Aggregation Framework and Patch-Based Image Representation for Optical Flow

Denis Fortun

ANNÉE 2014

UNIVERSITÉ DE RENNES 1

THÈSE / UNIVERSITÉ DE RENNES 1
sous le sceau de l'Université Européenne de Bretagne

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Traitement du Signal et Télécommunications

Ecole doctorale Matisse

présentée par

Denis Fortun

préparée à l'unité de recherche Inria
Centre Inria Rennes - Bretagne Atlantique
Université Rennes 1

Aggregation framework and Patch-Based Image Representation for Optical Flow

Thèse à soutenir à Rennes
le 10 juillet 2014

devant le jury composé de :

Nikos Paragios
Professeur, Ecole Centrale de Paris / rapporteur

Fabrice Heitz
Professeur, Université de Strasbourg / rapporteur

Julie Delon
Professeur, Université Paris Descartes / examinateur

David Tschumperlé
Chargé de Recherche, Université de Caen Basse-Normandie / examinateur

Etienne Mémin
Directeur de Recherche, Inria Rennes / examinateur

Patrick Bouthemy
Directeur de Recherche, Inria Rennes / Co-directeur de thèse

Charles Kervrann
Directeur de Recherche, Inria Rennes / Directeur de thèse

Contents

List of Figures

Résumé en français

1 Introduction au flot optique

La mise en correspondance d'images est une composante essentielle de la vision par ordinateur. Elle permet d'enrichir l'interprétation de l'information parfois ambigüe contenue dans une image unique. Par exemple, la mise en correspondance de plusieurs points de vue sur une même scène permet d'accéder à une mesure de profondeur et à la structure tri-dimensionnelle de l'espace observé. L'étude de la variabilité entre images de même type acquises dans des conditions différentes est à la base d'un grand nombre de tâches d'interprétation ou de classification, par exemple dans un contexte de diagnostique médical ou de classification de couverts végétaux en imagerie satellitaire. Quand l'ensemble d'images étudié correspond à séquence temporelle, c'est l'information de mouvement qui peut être extraite. Le principe commun à tous ces cas est donc la recherche de transformations spatiales permettant de caractériser les différences d'une image à une autre. Tous ces domaines applicatifs ont en commun un grand nombre de problématiques méthodologiques. Cette thèse traite du problème de la mise en correspondance d'images sous l'angle de l'estimation du mouvement.

Le mouvement dans une séquence d'images peut être abordé de plusieurs façons. Il est implicitement présent dans des opérations de détection de changements photométriques ou de suppression de flou de bougé. Quand il est explicitement estimé, le contenu dynamique d'une séquence peut être soit résumé par un ensemble de trajectoires d'objets d'intérêt pour une application donnée (véhicules, personnes, cellules...), soit défini localement par le déplacement de chaque pixel. Cette dernière représentation de mouvement est dénommée *flot optique*.

Comme beaucoup de problèmes de traitement d'image, l'estimation du flot optique peut être formulée comme un problème inverse mal posé, dont la mesure d'adéquation aux données produit un système d'équations sous contraintes insuffisant pour garantir l'unicité de la solution. Formellement, soit deux images successives $I_1, I_2 : \Omega \to \mathbb{R}$, où Ω désigne le domaine de l'image, et $\mathbf{w} : \Omega \to \mathbb{R}^2$ le champ de déplacement recherché, l'équation de conservation de l'intensité au cours du temps est de la forme suivante:

$$I_2(x + \mathbf{w}(x)) - I_1(x) = 0, \tag{0.1}$$

où $x \in \Omega$ désigne un point de l'image. Sous l'hypothèse de petits déplacements, sa forme linéarisée peut être obtenue:

$$\frac{\partial I}{\partial t}(x) + \nabla I(x)^\top \mathbf{w} = 0. \tag{0.2}$$

Cette équation seule ne donne néanmoins accès qu'à la composante parallèle au gradient de l'image de $\mathbf{w}$; le gradient ∇I doit par ailleurs être non nul. Cette sous-détermination peut être surmontée par une contrainte supplémentaire reflétant une hypothèse *a priori* sur la forme du champ de déplacement. Les deux grandes familles d'approches se distinguent par l'etendue *locale* ou *globale* de cette contrainte.

L'équation de conservation linéarisée 0.2 étant obtenue par développement de Taylor, son domaine de validité se limite à des déplacements de faible amplitude. Nous mentionnons d'emblée que les schémas d'estimation multi-résolution sont devenus l'approche standard pour surmonter cette difficulté. Les estimations à des résolutions grossières sont propagées incrémentalement aux résolutions les plus fines. Le principal effet indésirable de cette technique est le lissage des petits objets à grands déplacements. Résoudre ce problème fut un sujet de recherche très actif ces dernières années. Une solution efficace consiste à intégrer des techniques de mise en correspondance de plusieurs descripteurs dans des schémas d'estimation plus vastes [Brox and Malik, 2011; Weinzaepfel et al., 2013; Xu et al., 2012b; Chen et al., 2013]. Nous préconisons également une mise en correspondance en Section 3.

2 Approches locale et globale pour le calcul du flot optique

Nous résumons dans cette section les principaux axes méthodologiques structurant les méthodes existantes. Nous portons notre attention sur les limites de ces techniques et les points sur lesquels portent les contributions présentées dans les sections suivantes. Nous considérons trois éléments essentiels qui permettent de caractériser les méthodes de calcul du flot optique: le modèle de données, la paramétrisation locale et la régularisation globale. Nous présentons égalment une première contribution sous forme d'étude préliminaire sur la combinaison des approches locale et globale. Nous nous basons sur la méthode de Bruhn et al. [2005] et proposons une adaptation spatialle du filtrage du terme de données.

2.1 Modèle de données

L'équation de conservation de l'intensité permet de construire un potentiel de données ρ_{data} de la forme:

$$\rho_{data}(x, I_1, I_2, \mathbf{w}) = \phi(I_2(x + \mathbf{w}(x)) - I_1(x)), \tag{0.3}$$

où $\phi(\cdot)$ est une fonction de pénalisation. Le choix de $\phi(\cdot)$ est déterminé par la nature de la distribution des erreurs liées la contrainte de conservation. Dans certains cas, l'hypothèse statistique sous-jacente peut se révéler insuffisante et c'est la contrainte elle-même qui doit être adaptée.

Dans ce cas, une première approche consiste à élaborer des descripteurs aux propriétés de conservation plus générales que la simple intensité. L'invariance à différents types de changements d'illumination peut ainsi être obtenue en considérant le gradient de l'image [Uras et al., 1988; Brox et al., 2004], la composante texturée de l'image [Wedel et al., 2009b], différentes combinaisons d'espaces de couleurs [Mileva et al., 2007], des mesures de corrélation [Werlberger et al., 2010; Drulea and Nedevschi, 2013] ou encore la transformée de Census [Ranftl et al., 2012; Hafner et al., 2013]. L'efficacité de ces descripteurs est néanmoins souvent limitée à un nombre restreint de situations. Dans ce contexte, la sélection ou la combinaison locale optimale de plusieurs descripteurs a fait l'objet de recherches récentes [Xu et al., 2012b; Kim et al., 2013].

L'autre approche consiste à modifier explicitement l'équation de conservation de l'intensité via un terme d'erreur $e(x, I_1, I_2, \xi)$ paramétré par un vecteur inconnu ξ:

$$\rho_{data}(x, I_1, I_2, \mathbf{w}) = \phi(I_2(x + \mathbf{w}(x)) - I_1(x) - e(x, I_1, I_2, \xi)). \tag{0.4}$$

Le modèle général de Negahdaripour [1998] prend en compte les erreurs additive et multiplicative englobant un grand nombre de phénomènes de changement d'illumination possible. Ce modèle a été adopté avec quelques adaptations dans [Odobez and Bouthemy, 1995; Chambolle and Pock, 2011; Papadakis et al., 2013; Zach et al., 2008; Kim et al., 2005], sans pourtant jamais atteindre de résultats globaux complètement satisfaisants. Ceci s'explique en partie par la difficulté d'optimisation posée par l'estimation jointe, souvent menée de manière alternée, des deux variables $\mathbf{w}$ et ξ. La fonction $g(\cdot)$ peut aussi être spécifiée pour répondre à des besoins applicatifs précis, sur la base de considérations physiques [Haussecker and Fleet, 2001].

Nous revisitons cette approche dans la Section 4.1 en nous affranchissant de ce problème de minimisation alternée.

2.2 Paramétrisation locale

Comme évoqué en introduction, le modèle de données doit être intégré dans un schéma d'estimation qui tire profit du contexte spatial. Une classe de méthodes dites *locales*, s'appuie sur une modélisation paramétrique simple du champ de déplacement dans un certain voisinage $\mathcal{R} \in \Omega$. Le champ $\mathbf{w}_{\boldsymbol{\theta}} : \mathcal{R} \to \mathbb{R}^2$ est alors entièrement déterminé par le vecteur de paramètres $\boldsymbol{\theta}$, et l'énergie à minimiser est la somme des potentiels de données

sur le support $\mathcal{R}$, éventuellement pondérée par une fonction $g(x)$:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{x \in \mathcal{R}} g(x)\ \rho_{data}(x, I_1, I_2, \mathbf{w}_\theta). \tag{0.5}$$

Les modèles polynomiaux (e.g. constant, affine ou quadratique) sont une bonne approximation de la projection dans le plan 2D de mouvements simples d'objets individuels dans l'espace 3D. Ils sont les plus utilisés en pratique. Leur domaine de validité est donc restreint aux zones de mouvement cohérent, sans discontinuité de mouvement. La principale difficulté des approches locales est de déterminer les domaines $\mathcal{R}$ d'estimation appropriés.

L'approche la plus simple est celle de [Lucas and Kanade, 1981]. On considère des régions de forme et de taille fixes centrés en chaque pixel; l'estimation résultante ne concerne que le pixel central de chaque région. Les régions ainsi définies sont sous-optimales car susceptibles de se positionner soit au niveau d'une discontinuité de mouvement, invalidant le modèle polynomial, soit dans une zone homogène sans gradient de l'image, rendant impossible toute estimation fiable utilisant (0.2). Les tentatives d'estimation de la taille des voisinages [Maurizot et al., 1995; Senst et al., 2012] ou de leurs positions [Jodoin and Mignotte, 2009] n'ont jamais produit de résultats compétitifs sur les bases de vidéos d'évaluation les plus récentes.

Les régions optimales correspondraient davantage à une partition de l'image au sens du mouvement. Les approches conçues pour atteindre cet objectif sont nombreuses et peuvent être classées en deux catégories. Une première catégorie de méthodes repose sur une segmentation préalable de l'image et essaie d'ajuster un mouvement paramétrique sur chaque région [Xu et al., 2008; Black and Jepson, 1996; Bleyer et al., 2006], avec l'aide éventuelle d'une estimation variationnelle globale indépendante. L'inconvénient de cette approche vient de la sur-segmentation du mouvement induite par la segmentation au sens de l'intensité de l'image. La seconde catégorie de méthodes estime conjointement les supports des régions et leur modèles de mouvement associés, [Bouthemy and François, 1993; Cremers and Soatto, 2005; Memin and Perez, 2002; Odobez and Bouthemy, 1998; Sun et al., 2012; Unger et al., 2012]. L'énergie globale à minimiser est cependant sévèrement non convexe et l'optimisation doit faire face à un grand nombre de minima locaux. Pour un nombre de régions élevé, l'estimation ne peut en pratique pas s'écarter drastiquement de son initialisation, en général obtenue à l'aide des méthodes globales déjà très performantes [Sun et al., 2012; Unger et al., 2012].

Nous proposons dans la Section 3 un schéma d'agrégation sélectionnant implicitement le meilleur voisinage pour une estimation polynomiale sans étape de segmentation.

2.3 Régularisation globale

L'approche *globale* repose sur le choix d'un potentiel $\rho_{reg}(x, \mathbf{w})$ de régularisation explicite. Formellement, il s'agit de minimiser une fonctionnelle d'énergie de la forme:

$$E_{\text{global}}(\mathbf{w}) = \int_{\Omega} \rho_{data}(x, I_1, I_2, \mathbf{w}) + \lambda\, \rho_{reg}(x, \mathbf{w}) dx \tag{0.6}$$

où $\lambda \in \mathbb{R}$ est un coefficient de pondération qui équilibre les deux potentiels ρ_{data} et ρ_{reg}.

Le terme de régularisation impose un lissage du champ de déplacement en pénalisant ses variations spatiales et parfois temporelles. La difficulté est d'obtenir un lissage par morceaux propageant le flot dans les zones cohérentes tout en préservant les discontinuités de mouvement. La pénalisation quadratique du gradient de $\mathbf{w}$ dans la formulation initiale de [Horn and Schunck, 1981] conduit à une énergie convexe et dérivable pour laquelle un minimum global peut être atteint. Cependant, cette modélisation quadratique induit un sur-lissage et ne respecte pas les discontinuités de mouvement. Un grand nombre de fonctions robustes ont été proposées par la suite pour contourner cette difficulté. Parmi celles-ci, la Variation Totale (TV) présente l'avantage d'être convexe et de préserver les discontinuités spatiales et temporelles dans un grand nombre de cas. De nombreuses variantes de ce modèle de base ont été conçues par la suite [Trobin et al., 2008a; Bredies et al., 2010; Wedel et al., 2009b; Xu et al., 2012b; Zimmer et al., 2011; Werlberger et al., 2010; Sun et al., 2010a; Nagel, 1990; Volz et al., 2011].

La grande flexibilité des méthodes de régularisation se heurte néanmoins à des difficultés d'optimisation dès que les modèles d'énergie deviennent trop complexes. Les méthodes de minimisation continue sont le plus souvent basées sur la résolution des équations d'Euler-Lagrange [Brox et al., 2004], des approches de division proximales [Chambolle and Pock, 2011] ou une discrétisation anticipée de l'énergie [Sun et al., 2010a]. Dans tous les cas, dérivabilité et convexité de l'énergie sont requises pour assurer la convergence globale. Les modèles les plus performants pour l'estimation du flot optique ne remplissent cependant pas toutes ces conditions. Des compromis doivent être trouvés et des stratégies doivent être élaborées pour garantir un bon minimum local à l'aide de techniques itératives comme la non-convexité graduelle [Sun et al., 2014] ou des stratégies de points fixes [Brox et al., 2004], ou en procédant par relaxation convexe de l'énergie [Werlberger et al., 2010; Unger et al., 2012].

Une alternative aux méthodes variationnelles est le recours aux méthodes d'optimisation discrète.Un des avantages de l'optimisation discrète est qu'elle ne requiert pas de calcul de dérivation de l'énergie et autorise donc une plus grande variété de termes de données et de régularisation. En contrepartie, un équilibre doit être trouvé entre bonne précision de la discrétisation de l'espace des vecteurs de mouvement et un coût de calcul raisonnable. Une discrétisation de l'espace 2D des vecteurs de déplacement étant incapable de résoudre

ce dilemme, des stratégies alternatives ont été développées: contraintes sur les espaces de recherche [Mozerov, 2013], fusion des estimations globales continues [Lempitsky et al., 2010], raffinement itératif de l'espace de recherche [Glocker et al., 2008]. Dans cette thèse, nous proposons d'adopter une nouvelle approche générant un ensemble fini d'estimations locales continues paramétriques.

2.4 Filtrage adaptatif du terme de données

Nous présentons dans cette section une étude préliminaire conçue comme une première tentative de combinaison des approches locales et globales. Notre contribution s'appuie sur le travail de [Bruhn et al., 2005], qui intègre l'énergie des méthodes locales (0.5) dans un schéma de régularisation global, par filtrage Gaussien du potentiel de données. Le filtrage Gaussien est approprié pour réduire l'influence du bruit dans l'estimation. Cependant, il a pour effet indésirable de lisser exagérément les discontinuités de mouvement si le support de filtrage coïncide avec une composition de mouvements multiples.

Nous proposons de considérer un champ de déviations standard $\sigma : \Omega \to \mathbb{R}$ plutôt qu'une valeur unique pour toute l'image. Notre première contribution est un modèle qui s'inspire de l'énergie de [Bruhn et al., 2005] en incluant un terme de pénalisation portant sur le module du gradient de σ. L'optimisation de l'énergie globale suivante est menée conjointement par rapport au champ $\mathbf{w}$ et à la déviation standard σ du filtre Gaussien:

$$E(\mathbf{w},\sigma) = \int_\Omega \phi\left(k_\sigma * \left(\frac{\partial I}{\partial t}(x) + \nabla I(x)^\top \mathbf{w}(x)\right)^2\right) + \lambda \int_\Omega \phi(\|\nabla \mathbf{w}(x)\|^2)\,dx + \beta \int_\Omega \phi(\|\nabla \sigma(x)\|^2)\,dx. \quad (0.7)$$

On note k_σ le filtre Gaussien, $*$ l'opération de convolution, et λ et β sont des coefficients pondérant les différents termes. Les premiers résultats mettent en évidence une amélioration significative apportée par notre adaptation de σ vis à vis de la méthode de [Bruhn et al., 2005]. Les résultats restent cependant légèrement inférieurs à ceux obtenus en fixant $\sigma = 0$. En effet nous considérons dans cette étude un lissage isotrope, qui ne permet pas de respecter complètement les discontinuités. Une extension anisotrope de notre modèle est possible et fera l'objet de travaux futurs pour améliorer les premiers résultats.

Nous avons exploré dans un second temps un terme de données différent de celui de (0.7). Nous intégrons dans (0.7) une équation de conservation faisant intervenir une mesure d'incertitude de l'estimation sur le support de filtrage, dérivée des travaux de

[Corpetti and Mémin, 2012]. Nous obtenons l'énergie

$$E(\mathbf{w},\sigma(x)) = \int_\Omega \phi\left(k_{\sigma(x)} * \left(I_t + I_{x_1}u(x) + I_{x_2}v(x) + \frac{\sigma^2\Delta I}{2}\right)^2\right)dx + \lambda\int_\Omega \phi(|\nabla\mathbf{w}(x)|^2)dx + \beta\int_\Omega \phi(|\sigma(x)|^2), \quad (0.8)$$

qui diffère de (0.7) par le terme supplémentaire $\sigma^2\Delta I/2$ dans le terme de données, dérivé de la modélisation de l'incertitude locale du mouvement. Le support de filtrage estimé tend alors à minimiser cette incertitude. Cette formulation pose des problèmes d'implémentation qui ont empêché pour le moment son évaluation quantitative poussée.

3 Agrégation de candidats paramétriques locaux et gestion des occultations par recherche d'exemples

Dans cette section, nous proposons une méthode originale d'estimation du flot optique visant à répondre aux problèmes exposés précédemment. Notre approche comprend deux étapes. La première opère au niveau local, en réalisant des estimations paramétriques du mouvement sur une distribution de patches (fenêtres carrées de pixels) afin d'établir une liste de candidats de mouvement pour chaque pixel. Les candidats sont agrégés dans une deuxième étape en sélectionnant le meilleur candidat par optimisation d'une énergie globale.

Le problème de la détection des occultations et de l'estimation du mouvement dans les zones occultées est traité coopérativement au niveau des deux étapes de la méthode. La détection des zones occultées à l'étape d'agrégation exploite une carte de confiance estimée au niveau local dans la première étape. Un modèle d'estimation conjointe du mouvement et des occultations est proposé dans l'étape d'agrégation,. Un terme de parcimonie guidé par la mesure de confiance est notamment proposé. Le problème de l'estimation de mouvement dans les zones d'occultation est résolue en mettant en oeuvre une stratégie de recherche d'exemples générique, potentiellement applicable à d'autres approches d'estimation du flot optique.

3.1 Candidats locaux et mesure de confiance dans les zones d'occultation

3.1.1 Candidats de mouvement paramétrique

Notre approche locale se base sur une décomposition de l'image I_1 en patches de différentes tailles se recouvrant. Cet ensemble de patches $\mathcal{P}_{\mathcal{S},\alpha}$ est paramétré par l'ensemble des tailles de patch $\mathcal{S}$ et par le taux de recouvrement $\alpha \in [0,1]$ indiquant la proportion de

surface partagée par des patches voisins. Le mouvement est estimé indépendamment sur chaque patch selon la procédure en deux étapes suivante:

1. **Mise en correspondance de patches:**
   A chaque patch $P_1 \in \mathcal{P}_{\mathcal{S},\alpha}$, nous associons un ensemble $\mathcal{M}_N(P_1)$ de N patches dans l'image I_2 les plus similaires à P_1. Pour chaque paire de patches $P_{1,2} = (P_1, P_2)$ avec $P_2 \in \mathcal{M}_N(P_1)$, nous obtenons un vecteur de translation $\mathbf{w}_{P_{1,2}} \in \mathbb{Z}^2$ déplaçant P_1 en P_2.

2. **Raffinement par estimation affine:**
   Les déplacements estimés par mise en correspondance de patches correspondent à des translations en valeurs entières. Pour atteindre une précision sous-pixellique et autoriser des déformations plus complexes nous raffinons le vecteur $\mathbf{w}_{P_{1,2}}$ par une estimation affine $\delta\mathbf{w}_{P_{1,2}}$ régie par le vecteur de paramètres $\boldsymbol{\theta}_{P_{1,2}}$ [Odobez and Bouthemy, 1995].

Le recouvrement des patches et le nombre de tailles de patches $|\mathcal{S}| > 1$ implique qu'un pixel donné appartient à plusieurs patches. L'estimation en deux étapes décrite ci-dessus, appliquée à chaque patch de $\mathcal{P}_{\mathcal{S},\alpha}$, génère donc un ensemble $\mathcal{C}(x)$ de candidats de vecteurs de déplacement pour chaque pixel $x \in \Omega$:

$$\mathcal{C}(x) = \{\mathbf{w}_{P_{1,2}}(x) + \delta\mathbf{w}_{P_{1,2}}(x) : P_1 \in \mathcal{P}_{\mathcal{S},\alpha}(x), P_2 \in \mathcal{M}_N(P_1)\}, \tag{0.9}$$

où $\mathcal{P}_{\mathcal{S},\alpha}(x) = \{P \in \mathcal{P}_{\mathcal{S},\alpha} : x \in P\}$. Cette approche comporte plusieurs avantages:

- La distribution de patches $\mathcal{P}_{\mathcal{S},\alpha}$ permet de généraliser l'approche originale de Lucas and Kanade [1981], qui revient à choisir $\mathcal{P}_{s_0,1-\frac{1}{s_0}}$ où s_0 est une taille de patch unique. Tout comme l'approche locale de [Lucas and Kanade, 1981], notre procédure est simple sur le plan calculatoire et se prête à une parallélisation massive immédiate. Contrairement à l'idée classique [Lucas and Kanade, 1981], nous conservons tous les vecteurs de mouvement au sein de du patch, indexés par leurs positions dans le patch et l'image, en plus de l'estimation niveau du pixel central de chaque patch. La sélection du patch le plus approprié pour chaque pixel est réalisée au cours de l'étape d'agrégation via la sélection du candidat correspondant.

- La sélection d'un candidat s'apparente donc à la sélection de la "meilleure" région d'estimation pour chaque pixel, sans recours à une étape de segmentation coûteuse et une minimisation délicate de l'énergie associée.

- Les mises en correspondance de patches servent d'initialisations grossières, comme c'est également le cas dans quelques travaux récents [Chen et al., 2013; Leordeanu

et al., 2013; Mozerov, 2013]. Cependant, dans tous ces travaux, le raffinement est effectué dans un cadre de minimisation globale d'énergie, alors que notre procédure exploite uniquement des estimations paramétriques locales.

- A la différence des autres méthodes mettant en correspondance des descripteurs d'image [Brox and Malik, 2011; Chen et al., 2013; Weinzaepfel et al., 2013], nous ne conservons pas que la solution qui minimise l'erreur de mise en correspondance, mais mais les N meilleures correspondances. Cela nous permet d'assurer une meilleure robustesse aux bruit et aux erreurs d'assignement.

- En considérant plusieurs tailles de patch, nous accédons à plusieurs échelles de mouvement, notamment les grands déplacements des petits objets qui ne sont pas bien gérés par les schémas de multi-résolution classiques.

3.1.2 Extension de candidats dans les zones occultées

Par pixel occulté, nous entendons des pixels qui disparaissent entre l'image I_1 et l'image I_2, et qui n'ont donc pas de correspondant par définition. L'estimation des candidats $\mathcal{C}(x)$ ne différencie pas les pixels occultés des pixels non-occultés. Notre schéma d'estimation étant purement local, si tous les patches de $\mathcal{P}_{\mathcal{S},\alpha}(x)$ contiennent principalement des pixels occultés, les estimations dans ces patches seront systématiquement erronées. Il nous faut donc calculer les candidats de façon spécifique dans les zones occultées.

La carte d'occultation $o : \Omega \to \{0, 1\}$ est supposée dans un premier temps connue et est définie comme suit:

$$o(x) = \begin{cases} 1 & \text{si } x \text{ est occulté,} \\ 0 & \text{sinon.} \end{cases} \tag{0.10}$$

Soit $\mathcal{O}$ l'ensemble des pixels occultés de Ω. L'ensemble des candidats est alors étendu de deux manières:

- **Extension par recherche d'exemples:**
  Nous couplons à notre méthode de calcul du flot optique une méthode par recherche d'exemples basée sur le même principe que celle adoptée pour résoudre des problèmes d'*inpainting* en édition d'image. Pour tout pixel occulté $x \in \mathcal{O}$, soit $m(x)$ le pixel non occulté le plus similaire à x, et donc (sous réserve de la qualité de la métrique de similarité choisie) supposé appartenir au même objet en mouvement que x. Les nouveaux candidats de mouvement relatifs à $m(x)$ sont alors ajoutés à la liste qui concerne le pixel x. On obtient un ensemble étendu $\mathcal{C}_+(x)$:

$$\mathcal{C}_+(x) = \mathcal{C}(x) \cup \mathcal{C}(m(x)),\ \forall x \in \mathcal{O}. \tag{0.11}$$

	MPI Sintel	Middlebury
BCF avec extension de candidats	**0.792**	**0.0710**
BCF sans extension de candidats	1.851	0.0833
DeepFlow [Weinzaepfel et al., 2013]	4.691	0.386
MDP-Flow2 [Xu et al., 2012b]	4.006	0.223

Table 0.1: Erreurs obtenues sur les bases de vidéos MPI Sintel et Middlebury (distance euclidienne à la vérité-terrain) avec les deux versions de BCF, avec et sans extension de candidats dans les zones occultées, comparées avec les erreurs obtenues avec [Weinzaepfel et al., 2013] et [Xu et al., 2012b].

- **Extension aux mouvements de caméra:**
  Un type particulier d'occultation due au mouvement de caméra est pris en compte via un schéma d'estimation robuste du mouvement dominant $\mathbf{w}_{cam}$ dans l'image. L'ensemble final de candidats est donc:

$$\mathcal{C}_f(x) = \mathcal{C}_+(x) \cup \{\mathbf{w}_{cam}(x)\},\ \forall x \in \Omega. \tag{0.12}$$

3.1.3 Validation expérimentale

Pour valider notre méthode d'estimation de candidats, nous avons traité les séquences avec vérité-terrain extraites des bases de données MPI Sintel [Butler et al., 2012] et Middlebury [Baker et al., 2011]. Soit *Best Candidate Flow* (BCF) le champ de déplacements construit en sélectionnant pour chaque pixel x, le candidat appartenant à $\mathcal{C}_f(x)$ qui s'avère être le plus proche du vecteur de déplacement correspondant à la vérité-terrain. L'erreur globale moyenne pour le champ BCF ainsi reconstruit pour l'ensemble des séquences avec vérité-terrain issues des deux bases de vidéos est reportée dans le tableau 0.1. Les résultats de BCF sont également comparés aux résultats obtenus avec deux méthodes très performantes de l'état de l'art [Xu et al., 2012b; Weinzaepfel et al., 2013]. BCF surclasse nettement ces deux méthodes sur les deux bases de données, de manière plus significative encore sur la base de vidéos MPI Sintel qui comporte des déplacements de grande amplitude et des zones d'occlusion très importantes. L'intérêt d'étendre la liste des candidats dans les régions d'occultation est également illustré sur la Figure 0.1.

Ces résultats mettent clairement en évidence que de simples estimations paramétriques opérées sur des patches bien choisis suffiraient pour dépasser l'état de l'art de manière significative. Tout l'enjeu de l'étape d'agrégation qui fait l'objet de la Section 3.2, est de sélectionner les candidats de manière optimale en l'absence de vérité-terrain.

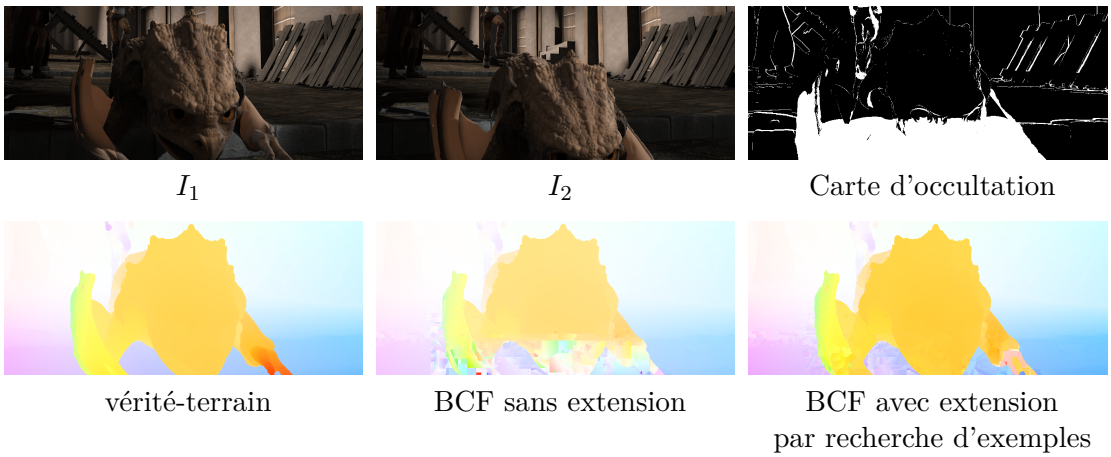


Figure 0.1: Extension par recherche d'exemples de candidats. Première ligne: deux images successives de la séquence *market_5* de la base de vidéo MPI SINTEL, et vérité-terrain de la carte d'occultations. Deuxième ligne: champs de déplacement correspondant à la vérité-terrain, BCF sans extension de candidats, et BCF après ajout de candidats par recherche d'exemples.

3.1.4 Carte de confiance d'occultation

Notre extension de candidats par recherche d'exemples suppose que la carte d'occultation o soit connue. En pratique, cette carte n'est pas disponible et doit donc être estimée conjointement avec le champ de mouvement lors de l'étape d'agrégation. Au premier niveau local, nous estimons une carte de confiance grossière $\omega_o : \Omega \to [0,1]$ qui nous informe sur la présence potentielle d'une occultation. Cette information sera exploitée dans l'étape d'agrégation pour guider l'estimation. Nous construisons ω_o à partir d'une première estimation grossière des occultations sur la base d'un critère de cohérence *forward/backward* appliqué à la distribution de patches $\mathcal{P}_{s_1,\alpha}$, avec s_1 la plus petite taille de patch. Une estimation de densité de Parzen fournit ensuite une distribution de probabilité de présence d'occultation en chaque pixel.

3.2 Agrégation globale discrète

Comme l'a démontré l'analyse BCF, la sélection du meilleur candidat en chaque pixel est potentiellement capable de produire d'excellents résultats. Nous formulons donc l'étape d'agrégation comme un problème d'optimisation discrète, où l'espace des labels est défini par l'ensemble des candidats $\mathcal{C}_f(x)$ en chaque point x. La carte d'occultation est estimée conjointement avec le champ de déplacement en exploitant la carte de confiance ω_o issue de la première étape. Soit le problème d'optimisation:

$$\{\widehat{\mathbf{w}}, \widehat{o}\} = \arg\min_{\{\mathbf{w},o\}} E(\mathbf{w}, o) \text{ sous la contrainte } \mathbf{w}(x) \in \mathcal{C}_f(x) \text{ et } o(x) \in \{0,1\}, \quad (0.13)$$

où l'énergie $E(\mathbf{w}, o)$ est composée de quatre termes,

$$E(\mathbf{w}, o) \;=\; E_{data}(\mathbf{w}, o, I_1, I_2) \;+\; E_{occ}(o, \omega_o) \;+\; E^{\mathbf{w}}_{reg}(\mathbf{w}) \;+\; E^{o}_{reg}(o) \qquad (0.14)$$

décrits brièvement par la suite.

3.2.1 Terme de données E_{data}

Le terme de données met en relation les variables de mouvement, d'occultation et d'intensité dans les images. Il est composé de deux potentiels ρ_{vis} et ρ_{occ}, dédiés respectivement aux pixels non occultés (ou visibles) et occultés:

$$E_{data}(\mathbf{w}, o, I_1, I_2) = \sum_{x \in \Omega} (1 - o(x))\, \rho_{vis}(x, \mathbf{w}) \;+\; \lambda_1\, o(x)\, \rho_{occ}(x, \mathbf{w}, m). \qquad (0.15)$$

Le terme ρ_{vis} est un potentiel classique similaire à ceux discutés en Section 2.1. Le potentiel ρ_{occ} correspond à une mesure d'adéquation aux données valide dans les zones occultées. Par analogie, le problème de l'estimation du mouvement sur des zones occultées est posé comme un problème d'*inpainting*. L'objectif est de synthétiser un contenu dans des zones où aucune mesure (de mouvement dans le cas du flot optique) n'est disponible. Les approches par diffusion sont systématiquement retenues en flot optique, alors que les méthodes par recherche d'exemples sont plus efficaces pour combler des "trous" en édition d'image. Nous proposons une approche par recherche d'exemples pour le traitement des occultations en flot optique en recourant au potentiel ρ_{occ} suivant:

$$\rho_{occ}(x, \mathbf{w}, m) = \left\| \mathbf{w}(x) - \mathbf{w}(m(x)) \right\|^2, \qquad (0.16)$$

où le champ $m : \mathcal{O} \to \Omega \backslash \mathcal{O}$ met en correspondance chaque pixel occulté avec un pixel non occulté en minimisant une distance appropriée. Le vecteur de déplacement d'un pixel occulté est donc contraint à être similaire au vecteur de déplacement du pixel correspondant non-occulté.

Si on s'intéresse à la détection des occultations (minimisation par rapport à o), ce terme de données favorise l'émergence de zones d'occultation dans les zones de violation de l'hypothèse de conservation de l'intensité. Cette modélisation peut être interprétée comme une version discrète de [Ayvaci et al., 2012].

3.2.2 Contrainte d'occultation E_{occ}

Le terme de données (0.15) favorise l'apparition de pixels occultés et doit être équilibré par un terme de pénalisation qui prend en compte le nombre de pixels occultés:

$$E_{occ}(o, \omega_o) = \lambda_2 \sum_x \omega_o(x) o(x). \tag{0.17}$$

Nous exploitons ici la carte de confiance ω_o issue de la première étape pour contrôler ce terme de pénalisation. Ce terme est en fait analogue à une contrainte de parcimonie dans un cadre continu comme cela a été exposé dans [Ayvaci et al., 2012]. La pondération ω_o est essentielle aussi pour éviter un couplage trop fort entre o et $\mathbf{w}$ qui conduit à la détection de minima locaux dès la première itération d'un schéma de minimisation alternée.

3.2.3 Termes de régularisation E^1_{reg} et E^2_{reg}

Un lissage spatial du champ de déplacement $\mathbf{w}$ et de la carte d'occultation o est obtenu via les termes E^1_{reg} et E^2_{reg}.

3.3 Résultats

La méthode proposée – *AggregFlow* – est évaluée sur les bases de vidéos MPI Sintel [Butler et al., 2012] et Middlebury [Baker et al., 2011], couvrant un large spectre de situations dynamiques en analyse de vidéos. Les évaluations quantitatives se basent sur la moyenne de la distance euclidienne entre la vérité-terrain et le vecteur de déplacement estimé en chaque pixel (EPE).

La base de vidéos MPI Sintel contient un grand nombre de séquences comportant des déplacements de grande amplitude (plusieurs dizaines de pixels) qui induisent de larges zones occultées. Les résultats reportés dans le tableau 0.2 mettent en évidence les performances de notre méthode, notamment dans les zones d'occultation ("EPE unmatched"). Notre modélisation s'avère donc pertinente pour traiter ces zones. Notre algorithme AggregFlow est également classé $2^{\text{ème}}$ si on examine l'erreur au voisinage de discontinuités de mouvement ("d0-10") et se classe 1^{er} dans les zones qui correspondent à des déplacements d'amplitudes supérieures à 40 pixels (s40+).

Les séquences de Middlebury comportent des déplacements de plus faibles amplitudes. L'enjeu principal est d'être capable de retrouver les déformations lisses, les discontinuités de mouvement et les petits détails. Les différences entre les méthodes sont donc moins importantes que sur MPI Sintel, comme l'atteste le tableau 0.3. L'algorithme AggregFlow demeure très compétitif si on le compare aux méthodes les mieux classées (mesure EPE).

	EPE all	EPE matched	EPE unmatched	d0-10	s40+
AggregFlow	**4.754**	**1.694**	**29.685**	3.705	**31.184**
DeepFlow [Weinzaepfel et al., 2013]	5.377	1.771	34.751	4.519	33.701
MDP-Flow2 [Xu et al., 2012b]	5.837	1.869	38.158	**3.210**	39.459
EPPM [Bao et al., 2014]	6.494	2.675	37.632	4.997	39.152
S2D-Matching [Leordeanu et al., 2013]	6.510	2.792	36.785	5.523	44.187
Classic+NLP [Sun et al., 2014]	6.731	2.949	37.545	5.573	45.290
FC-2Layers-FF [Sun et al., 2012]	6.781	3.053	37.144	5.841	45.962
MLDP-OF [Mohamed et al., 2014]	7.297	3.260	40.183	5.581	51.146

Table 0.2: Résultats sur la base de vidéos MPI Sintel (version "clean")

	EPE all	Avg. rank
MDP-Flow2 [Xu et al., 2012b]	**0.245**	**7.8**
FC-2Layers-FF [Sun et al., 2012]	0.283	19.3
Classic+NL [Sun et al., 2014]	0.319	27.1
EPPM [Bao et al., 2014]	0.329	32.6
AggregFlow	0.339	35.9
MLDP-OF [Mohamed et al., 2014]	0.349	32.6
S2D-Matching [Leordeanu et al., 2013]	0.347	34.6
DeepFlow [Weinzaepfel et al., 2013]	0.416	48.8

Table 0.3: Résultats sur la base de données Middlebury

3.4 Une autre approche: agrégation dans un cadre continu

Nous proposons une autre approche d'agrégation dans un cadre continu, qui peut être considérée comme une alternative à l'agrégation discrète décrite en Section 3.2. L'ensemble des candidats est alors considéré comme un dictionnaire de mouvement à partir duquel le champ global est reconstruit. La sélection d'un candidat se fait au travers d'une contrainte de parcimonie appliquée à un vecteur de poids $\boldsymbol{\alpha}(x)$ en chaque pixel. L'énergie considérée est de la forme:

$$E(\mathbf{w}, \boldsymbol{\alpha}) = \int_{\Omega} \left\| \mathbf{w}(x) - \boldsymbol{\alpha}(x)^{\top} \mathbf{W}_{\mathbf{c}}(x) \right\|_1 + \lambda_2 \|\boldsymbol{\alpha}(x)\|_{1,\boldsymbol{\beta}(x)} + \lambda_1 \|\nabla \mathbf{w}(x)\|_1 dx. \quad (0.18)$$

où $\|\cdot\|_{1,\boldsymbol{\beta}(x)}$ une norme L_1 pondéré par des mesures de confiance $\boldsymbol{\beta}(x)$, $\mathbf{W}_{\mathbf{c}}(\mathbf{x})$ est la forme vectorielle de l'ensemble des candidats en x, et les paramètres λ_1 et λ_2 pondèrent les différents termes. La minimisation dans l'espace continu tolère une déviation par rapport au candidat sélectionné. Cette formulation est convexe, ce qui rend possible une minimisation efficace.

les résultats quantitatifs globaux de l'agrégation continue sont légèrement inférieurs à

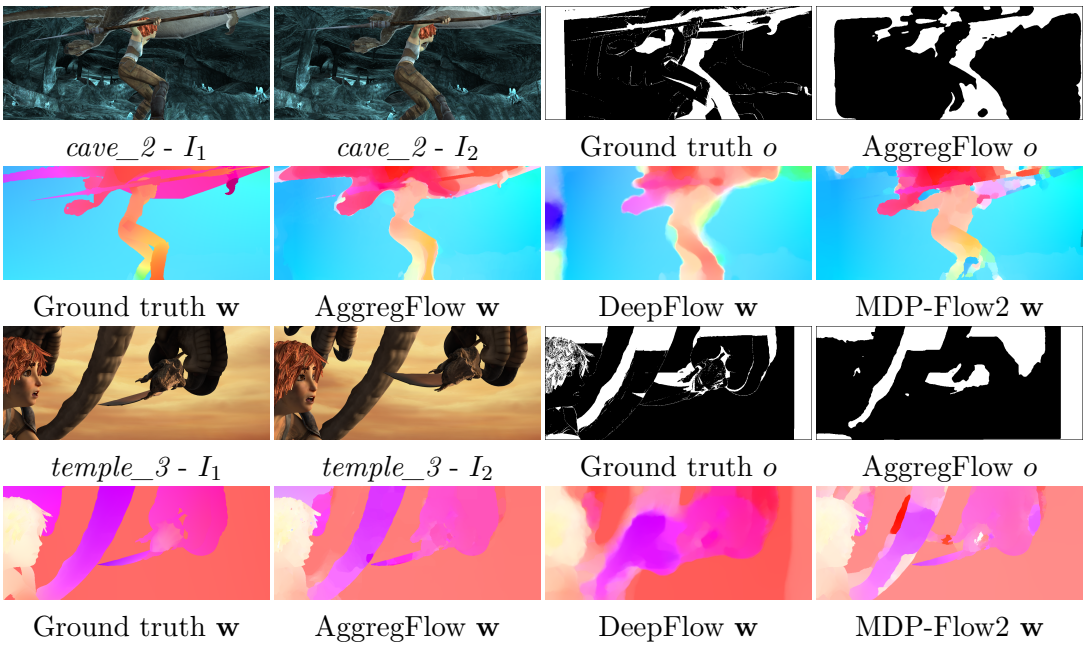


Figure 0.2: Résultats visuels d'estimation de mouvement et d'occultations sur quelques séquences de la base de vidéos MPI Sintel, comparés avec ceux obtenus par les méthodes décrites dans [Weinzaepfel et al., 2013] et [Xu et al., 2012b].

ceux de l'agrégation discrète. L'approche continue présente cependant l'avantage d'un temps de calcul plus faible. De plus la capacité de s'écarter des vecteurs mouvement définis par les candidats permet d'obtenir des résultats satisfaisant même avec des candidats moins pertinents, ce qui permet également de réduire le temps de calcul lié à la génération des candidats.

4 Estimation du mouvement et de la diffusion en imagerie biologique

Dans ce chapitre, nous nous intéressons à des problématiques d'estimation de mouvement spécifiques rencontrées en imagerie biologique, et plus particulièrement en microscopie optique de fluorescence. En particulier, nous proposons des solutions à deux problèmes distincts: i) les grands changements d'intensité causés par des fluctuations de fluorescence; ii) l'analyse de situations de diffusion de particules caractérisées par leur coefficient de diffusion.

4.1 Changements d'intensité en imagerie de fluorescence

Le changement d'intensité des objets en mouvement est un problème général à prendre en considération dans l'estimation du flot optique. Comme évoqué en Section 2.1, une manière simple est de choisir des descripteurs invariants à certains changements d'intensité de fluorescence. Ces descripteurs sont cependant imparfaits et s'avèrent efficaces seulement dans des cas particuliers. En imagerie de fluorescence, nous observons à la fois des zones où l'intensité est préservée au cours du temps et des fortes variations temporelles du signal de fluorescence. Ces phénomènes sont liés au photo-blanchiment (extinction progressive de la fluorescence) et aux mouvements des entités sous-résolues.

Pour traiter ces situations difficiles, nous reprenons le schéma d'agrégation défini au chapitre précédent, en y intégrant l'idée d'une modélisation explicite des changements d'intensité décrite en Section 2.1. L'objectif est d'estimer conjointement au champ de déplacements, une carte de changement local d'intensité de fluorescence. Nous considérons le potentiel de correction d'intensité additive $\xi_0 : \Omega \to \mathbb{R}$ suivant:

$$\rho_{data}(x, I_1, I_2, \mathbf{w}, \xi_0) = \phi(I_2(x + \mathbf{w}(x)) - I_1(x) - \xi_0(x)). \tag{0.19}$$

Des méthodes similaires ont déjà été proposées dans la littérature mais les performances demeurent limitées. Notre énergie d'agrégation globale est similaire à la formulation décrite dans [Chambolle and Pock, 2011; Kim et al., 2005]:

$$\begin{aligned} E(\mathbf{w}, \xi_0) \;=\; \sum_{x \in \Omega} \rho_{data}(x, I_1, I_2, \mathbf{w}, \xi_0) \;+\; \lambda_1 \sum_{<x,y>} \psi(\|\mathbf{w}(x) - \mathbf{w}(y)\|) \\ + \lambda_2 \sum_{<x,y>} \psi(|\xi_0(x) - \xi_0(y)|). \end{aligned} \tag{0.20}$$

La difficulté réside dans l'optimisation conjointe des deux variables $\mathbf{w}$ et ξ_0. L'originalité de notre approche consiste à éviter ce problème en proposant des candidats pour ξ_0 conjointement aux candidats servant à estimer $\mathbf{w}$.

L'étape de génération des candidats est menée de la façon suivante:

1. **Mise en correspondance de patches:**
   Nous adoptons le coefficient de corrélation normalisé, invariant aux changements additifs d'intensité, comme métrique de similarité. $\forall P_1 \in \mathcal{P}_{\mathcal{S},\alpha}(x), P_2 \in \mathcal{M}_N(P_1)$, $\xi_{P_{1,2}}$ est une estimation grossière du changement d'intensité définie comme la différence des moyennes des patches P_1 et P_2. Notons que $\xi_{P_{1,2}}$ est indissociable de l'estimation de mouvement $\mathbf{w}_{P_{1,2}}$ définie précédemment.

2. **Raffinement affine:**
   L'estimation de mouvement affine $\delta\mathbf{w}_{P_{1,2}}$ nécessite également l'estimation d'un paramètre supplémentaire $\delta\xi_{P_{1,2}}$ (potentiel de données (0.19)).

L'ensemble de candidats est maintenant défini en fonction des paires mouvement-changement d'intensité:

$$\mathcal{C}(x) = \left\{ \left(\mathbf{w}^c_{P_{1,2}}(x), \xi^c_{P_{1,2}} \right) : P_1 \in \mathcal{P}_{\mathcal{S},\alpha}(x), P_2 \in \mathcal{M}_N(P_1) \right\}. \tag{0.21}$$

L'espace discret considéré pour chercher la solution minimisante (0.20) contient donc des labels à la fois associés à $\mathbf{w}$ et à ξ_0, ce qui évite le recours à une minimisation alternée.

Les résultats obtenus sur des séquences d'imagerie de fluorescence démontrent les avantages de notre approche par rapport à des méthodes qui s'appuient sur des descripteurs invariants à certains changements d'intensité tel que le gradient de l'image [Brox and Malik, 2011] ou sa composante texturée [Sun et al., 2010a]. La Figure 0.3 illustre le type de séquences traitées et les améliorations apportées par notre méthode.

4.2 Une approche variationnelle pour l'estimation de la diffusion

Un comportement dynamique commun à un grand nombre de phénomènes observés en imagerie biologique est le mouvement de diffusion de particules. Le mouvement est davantage caractérisé par le coefficient de diffusion que par le champ de déplacement.

L'approche la plus généralement adoptée pour estimer le coefficient de diffusion est la méthode ICS (*Image Correlation Spectroscopy*) basée sur des mesures de corrélation de patches de grande dimension. Les deux inconvénients majeurs de cette approche sont d'une part le temps de calcul, et d'autre part l'incapacité à estimer des cartes denses de diffusion, et donc de détecter des discontinuités de diffusion. Nous proposons une approche variationnelle apportant une solution à ces problèmes.

Nous nous plaçons dans le cas du mouvement brownien des particules dont la taille est inférieure à la résolution optique utilisée. A chaque pixel sont alors associées potentiellement plusieurs particules. On peut donc considérer que l'intensité observée est proportionnelle à une mesure de concentration. Dans ces conditions, le modèle de diffusion suivant est exploitable:

$$\frac{\partial I}{\partial t} = D_0 \, \Delta I, \tag{0.22}$$

où D_0 représente le coefficient de diffusion isotropique recherché et Δ désigne l'opérateur laplacien de l'image.

Contrairement à la technique ICS estimant un scalaire D_0 constant pour toute l'image, nous proposons un schéma d'estimation variationnelle pour estimer une carte dense $D : \Omega \to \mathbb{R}$. Soit le problème de minimisation:

$$\widehat{D} = \arg\min_D \{E_{data}(D, I) + \lambda_D E_{reg}(D)\} \text{ sous la contrainte } D(x) \geq 0 \tag{0.23}$$

où le terme de données $E_{data}(D, I)$ est dérivé de l'équation de diffusion. Une mesure

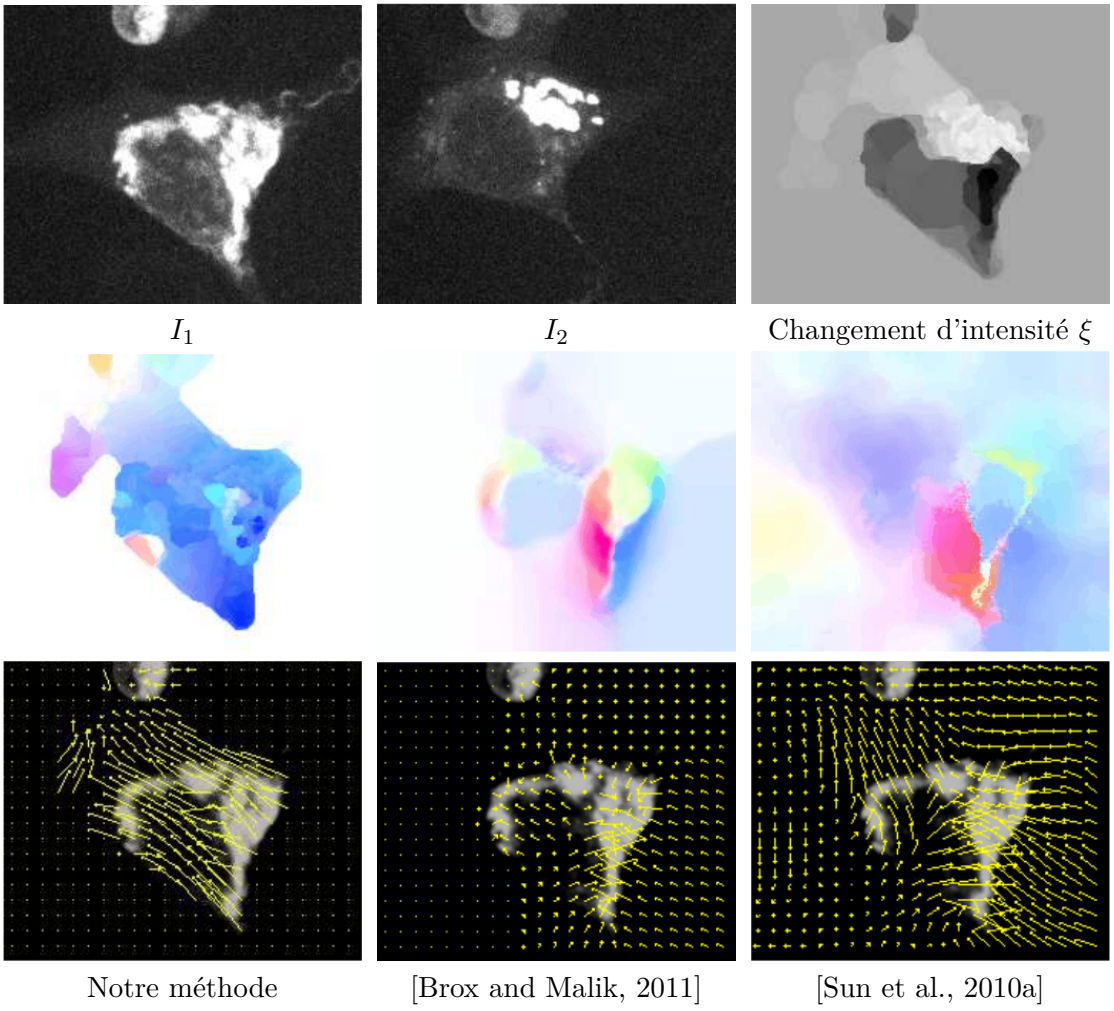
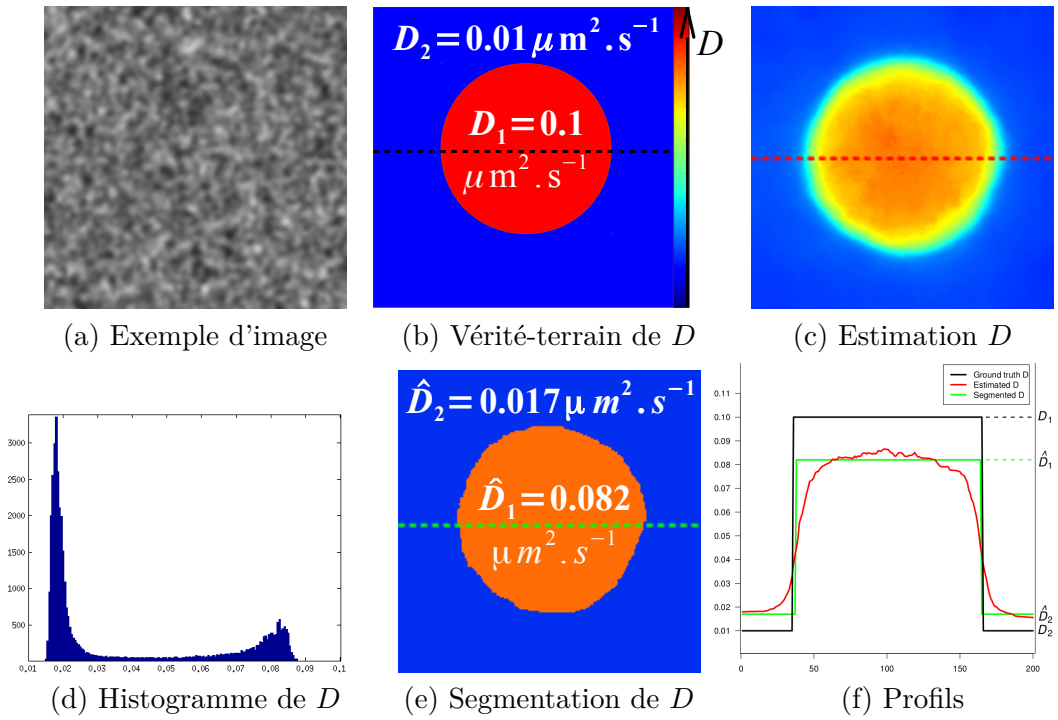


Figure 0.3: Résultats sur une séquence de déplacement de cellules "HeLa" (acquisition réalisée par le groupe de F. Perez, UMR 144 Institut Curie, PICT-IBiSA). Ligne du haut: les deux images successives et la carte de changement d'intensité estimée par notre méthode. Lignes du milieu et du bas: champs de déplacement esstimés respectivement par notre méthdode, [Brox and Malik, 2011] and [Sun et al., 2010a]. Les champs de déplacement sont visualisé par vecteurs et code couleur.

d'écart ponctuel à (0.22) est insuffisante à cause de la nature aléatoire du phénomène de diffusion. Nous adoptons donc une pénalisation non-ponctuelle en supposant que le coefficient de diffusion $D(x)$ est constant dans un voisinage de x:

$$E_{data}(D, I) = \int_{\Omega} \phi(\mathbf{D}^{\top} \mathbf{J}_{\rho} \mathbf{D}) dx \tag{0.24}$$



(a) Exemple d'image (b) Vérité-terrain de D (c) Estimation D

(d) Histogramme de D (e) Segmentation de D (f) Profils

Figure 0.4: Estimation variationnelle du coefficient de diffusion sur une séquence simulée de diffusion spatialement inhomogène. Les courbes de (f) sont les profils des lignes pointillées de (b),(c) et (e)

où $\mathbf{D} = \begin{pmatrix} D \\ 1 \end{pmatrix}$ et $\mathbf{J}_\rho = k_\rho * \begin{pmatrix} \Delta I^2 & -I_t\,\Delta I \\ -I_t \Delta I & I_t^2 \end{pmatrix}$.
Le terme de régularisation conduit à un lissage robuste du champ D. La contrainte $D(x) \geq 0$ est imposée par l'ajout d'une barrière logarithmique à l'énergie initiale.

En cas de diffusion constante sur toute l'image notre approche variationnelle produit des résultats similaires à la méthode ICS et ses variantes. Quand le coefficient de diffusion est inhomogène, comme c'est le cas sur l'exemple simulé de la Figure 0.4, notre méthode permet d'estimer de manière satisfaisante les discontinuités de diffusion. La méthode ICS est incapable de produire de telles estimations denses compte-tenu des temps de calcul importants sur chaque patch.

5 Conclusion

Nous avons traité au cours de cette thèse plusieurs problèmes liés à l'estimation du flot optique, que nous avons identifiés comme les facteurs limitant des méthodes actuelles.

Notre approche générale est la combinaison de modèles locaux et globaux.

Nous avons dans un premier temps proposé une méthode s'appuyant sur le travail de [Bruhn et al., 2005], qui intègre l'énergie des méthodes locales dans le terme de données d'un modèle global. Notre méthode adapte spatialement le noyau gaussien utilisé pour filtrer le potentiel de données standard. Les résultats préliminaires ont démontré la pertinence de cette approche. Nous avons également exploré une version modifiée intégrant le modèle d'incertitude stochastique de [Corpetti and Mémin, 2012].

L'axe principal de ce travail à été consacré à la conception d'un schéma d'estimation original d'agrégation, également basé sur la combinaison des approches locales et globales. L'idée principale est de revisiter les estimations locales paramétriques, et plus particulièrement la sélection des tailles et positions de patches servant à l'estimation. La procédure d'agrégation comprend deux grandes étapes. Une première étape réalise des estimations locales paramétriques de mouvement dans une distribution régulière de patchs, et une étape d'agrégation effectue une sélection parmi les candidats générée par la première étape locale. Nous avons construit deux approches pour le problème de l'agrégation, basées sur des méthodes d'optimisation discrète et continue. L'agrégation discrète fournit de meilleurs résultats. L'agrégation continue reste cependant intéressante d'un point du vue pratique pour son faible coût de calcul et sa robustesse à une moindre qualité des candidats.

Un processus de gestion des occultations est conçu coopérativement entre les deux étapes. Une mesure de confiance au niveau local est utilisé pour guider la détection des occultations au niveau global. L'estimation du mouvement dans les zones occultées ainsi détectées est réalisée par une approche par recherche d'exemples, qui surmonte les difficultés des approches usuelles de diffusion. Cette approche est générale et pourrait être intégrée dans d'autres schéma d'estimation. Une intégration originale de mise en correspondance de descripteur est également permise par le schéma d'agrégation Pour gérer les grands changement d'intensité, nous proposons une modélisation jointe du mouvement et des changement d'intensité aux deux étapes de la méthode.

L'évaluation expérimentale a démontré la supériorité de notre méthode AggregFlow sur les méthodes existantes dans des cas difficile de grands déplacement, en particulier sur la base de vidéos récente MPI Sintel. Sur des séquences de faibles déplacements comme celles de Middlebury, nos résultats restent compétitif avec la plupart des meilleures méthodes. La plus grosse amélioration est observée au niveau des zones occultées. Des exemples de grands changements d'intensité en imagerie de fluorescence montrent aussi les avantages de notre approche.

Nous avons finalement proposé des solutions à des problèmes spécifiques rencontrés en imagerie biologique. L'imagerie de fluorescence produit de grands changements d'intensité, pris en compte dans le schéma d'agrégation. Une estimation variationnelle du coefficient de diffusion permet également de caractériser des mouvement browniens de particules,

fréquents en imagerie de fluorescence. Notre approche détecte les discontinuités de diffusion plus fidèlement que les méthodes courante de corrélation.

General introduction

A large range of computer vision tasks cannot be limited to the analysis of the content of a single image, but requires to find correspondences between images. Indeed, the information contained in one image, that is the spatial relations between intensities, can be intrinsically ambiguous and insufficient to access essential interpretative elements about the observed scene. Example of uncertainties yielded by analyzing single images are numerous:

- optical illusions playing with uncertainties about depth ordering of objects in a scene are well known and reveal the impossibility to access depth information from a single point of view;
- objects with similar appearance can be indistinguishable in a static scene, and reveal their individual shapes through the perception of their relative motions;
- the interpretation and classification of specific measures obtained from images, for pathology diagnosis on anatomical structures in medical imaging, or classification of ground types in remote sensing, is often impossible without comparison with known reference data.

In all these cases, the information concerning spatial relations between pixels needs to be extended to the *transformation* of these relations when passing from one image to the other. Determining these transformations is one of the major field of research in computer vision.

Methodological approaches for the correspondence problem described above can be influenced by the type of transformation expected, and the type of images to deal with. These characteristics can be assessed from the knowledge on the set of images to be analyzed, depending on the targeted application. In medical imaging, the image set can be composed of acquisitions of similar anatomical structures for several patients, the same patient at different time steps, or with different imaging modalities. For depth estimation or 3D reconstruction, a single scene is observed under several points of view. When the objective is to estimate the motion of a scene, the images are successive frames of a temporal sequence. While these applications need specific investigations in modeling and adaptations, they share common methodologies. In this thesis, we are interested in motion estimation, where the searched image correspondences account for the projection

of the real 3D motion in the scene, creating a 2D dense motion field, the so-called *optical flow*.

Optical flow can be considered as the most low-level dynamical characterization of an image sequence since it is not concerned with any object detection or motion interpretation task. As such, it is an essential component for numerous applications using it as a fundamental module for a more semantic interpretation. Among the application fields, we can mention robotics where the knowledge of the visual environment is essential to achieve autonomous navigation and adaptation to unexpected situations in real conditions. Augmented reality also requires a complete control of motion to ensure consistency between augmented and original scenes, with applications in multimedia or assisted surgery. Handling motion in biomedical imaging is also an important field of research involving optical flow analysis to get rid of disturbances caused by heart beats or patient undesirable displacements. Video compression also makes use of optical flow to restrict the quantity of transmitted information exploiting motion redundancy. A large and active application field also concerns the development of assisted driving systems adapting responses and actions of a vehicle to the perceived motion. The quantification of fluid motion is particularly important in meteorology to provide predictions based on the motion of clouds. Tracking of individual objects can use optical flow as a key information to analyze trajectories in contexts of biological imaging or video surveillance. The success of these applicative systems, and of many others, is often based on the accuracy of the optical flow estimation step. The large diversity of situations described above gives rise to as many specific problems for motion estimation, and despite great progress over years, existing methods still fail to handle all these issues together.

Like many other computer vision problems, optical flow estimation can be formulated as an ill-posed inverse problem. As such, it relies on a data fitting measure leading to an under-constrained system unable to guarantee unicity of the solution. In the case of optical flow this measure is based on the assumption of constancy of a given image descriptor (usually image intensity) during motion. To cope with the single constraint insufficiency, modeling assumptions have then to be made to regularize the problem by adding a priori constraints on the expected form of the motion field. Existing methods can be broadly classified regarding the *local* or *global* extent of the spatial constraint.

Global models have always outperformed local approaches in terms of accuracy. In a broad sense, the problem of the choice of appropriate regions for local estimation has never been efficiently solved, whereas the global regularization framework offers appropriate modeling capabilities and efficient optimization techniques. However, the complexity of resulting global energy to cope with large displacements, occlusions or illumination changes induces optimization issues like the use of coarse-to-fine scheme or optimization in high dimensional spaces, which constitutes the main intrinsic limitations of global methods.

Based on the analysis of optical flow state-of-the-art and current limitations, we have investigated new approaches. In particular, we have explored several ways to revisit local motion estimation combined with global regularization models.

Our first contribution builds on the work of [Bruhn et al., 2005], which takes advantage of both local and global models by considering an extended local energy in the data term of the global energy model. We address the oversmoothing of the discontinuities implied by this method by adapting spatially the standard deviation of the Gaussian filtering.

Secondly, the central part of this thesis is dedicated to the design of an original aggregation framework for optical flow estimation. It is composed of two successive steps, combining sequentially local and global estimations. We first demonstrate the potential of local parametric methods to outperform existing global approaches. The framework allows us to address several important issues of optical flow. The problem of selection of the local estimation supports is solved by selecting a candidate in the aggregation step. Multiple patch correspondences are integrated in an original and efficient way. We solve the aggregation problem with discrete and continuous global optimization techniques. We propose also an occlusion handling framework in our aggregation scheme. The exemplar-based principle of our motion estimation in occluded regions is generic and adaptable to other methods. Finally we also deal with large intensity changes cooperatively in the two steps of the method.

In the last part of the thesis, we address specific issues for motion estimation occurring in biological imaging. There is a growing interest for the use of live cell imaging for cancer research and analysis of the dynamical behavior of biological structures. A lot of related questions requires a quantification of motion of biological structures. We proposed solutions on the one hand to cope with large intensity changes in fluorescence microscopy, and on the other hand, to estimate spatially varying diffusion.

Organization of the thesis

Part I

This part is conceived as an introductory tutorial on optical flow. We focus on recent developments and limitations of existing methods, which are presented as the context of the contributions developed in subsequent parts.

Chapter 1 We introduce basic definitions and concepts of optical flow and present standard evaluation procedures and benchmarks. These elements constitute the basis for the understanding of the rest of the thesis.

Chapter 2 We classify existing data conservation assumptions used to design data terms.

Chapter 8.1 The issues involved in the local parametric approaches are detailed and classified regarding the choice of the motion model, the optimization strategy, and more importantly the choice of an appropriate local region, which is directly linked with our aggregation approach described in Part II.

Chapter 4 Principles and practices in global regularized optical flow estimation are presented. The advantages and limitations of variational and discrete optimization techniques are discussed. They play an essential role in the modeling and the strategy involved in the contribution of the following parts.

Chapter 5 We analyze the recent trends of optical flow methods consisting in the integration of feature matching for dense motion estimation. We classify current practices and point out the main limitations.

Chapter 6 We make a first step towards combining local and global models based on the work of [Bruhn et al., 2005]. We propose an original spatial adaptation of the Gaussian filtering involved in the data term [Bruhn et al., 2005] to prevent oversmoothing of the discontinuities while keeping the advantage of integration of local information. We discuss this idea with the recent work of [Corpetti and Mémin, 2012].

Chapter 7 A summary of the main methodological directions is given, together with the respective limitations that emerged from the analysis of each class of methods.

Part II

This part is dedicated to the description and evaluation of a new optical flow estimation method based on a general aggregation framework.

Chapter 8 We describe our generation process for local parametric motion candidates. An original feature matching integration and occlusion handling combined with simple affine estimation are performed in a generalized patch distribution. We experimentally demonstrate that the computed set of candidates is able to potentially outperform state-of-the-art methods if the best candidate would be correctly selected at every image point.

Chapter 9 Our aggregation method is presented as a discrete optimization problem in the discrete label space composed of the candidates generated in the previous step. A generic joint modeling of motion and occlusion is proposed. An original exemplar-based occlusion filling strategy is integrated in the global energy. The *move-making* optimization method is also detailed.

Chapter 10 The method is evaluated on the most known computer vision benchmarks. The main improvements are reached for occluded regions and large displacements, which validates the models presented in previous chapters.

Chapter 11 We propose an alternative aggregation method to the discrete approach presented in Chapter 9, in the continuous optimization setting. The motion field is reconstructed from a sparse combination of motion candidates. Quantitative results are less accurate than those obtained with the discrete aggregation approach, but improvements are achieved regarding computational time and robustness to the quality of the finite set of candidates.

Chapter 12 We conclude and present the perspectives of the proposed method and modelling framework.

Part III

This part presents contributions to motion estimation in the context of biological imaging.

Chapter 13 An adaptation of variational methods to several biological imaging situations is presented. Results are compared with correlation approaches which are mostly used in biological imaging.

Chapter 14 One major problem to take into account in fluorescence microscopy imaging is the intensity variation caused by fluorescence fluctuations. We propose an adaptation of the aggregation framework for joint estimation of intensity changes and motion. Our method outperforms other variational approaches in the case of large intensity changes.

Chapter 15 We address the characterization of diffusive dynamics, which is a current situation in fluorescence imaging. We propose a variational approach for diffusion coefficient estimation, which overcomes limitations of usual correlation approaches in terms of discontinuity recovering and computational time.

Chapter 16 We conclude and present the main applicative achievements and potential impact of our contributions in biological imaging. We propose several perspectives of improvements of the proposed methods and general research directions in biological image analysis.

Part I

Local and global approaches for optical flow

Motion in image sequences can be characterized through different forms. It can be implicitly contained in operations like change detection methods or motion blur estimation. The motion of interest can be the tracking of individual objects like persons in video surveillance, cellular structures in biological imaging or vehicles for driving assistance purposes. The dynamical content is then determined by a sparse set of trajectories. In other cases, optical flow estimation is necessary to retrieve dense deformation fields. Complex fluid, organ or cell deformations, or global crowd displacement are examples of complex deformations requiring to produce a dense motion field. Moreover, as a low-level representation of motion, optical flow is often used as an input information for other motion analysis tasks.

The design of a reliable generic optical flow estimation method is a difficult task. A tremendous quantity of works have been carried out for thirty years, trying to overcome new issues emerging with the increasing use of computer vision in a large number of areas. This chapter is conceived as a tutorial aiming at organizing the main approaches and practices developed for optical flow estimation. We will not try to tell the whole story of the evolutions of optical flow, neither we give an exhaustive list of existing methods. We rather propose a synthetic classification of the main methodological principles at the basis of current methods, with a particular concern given to recent developments. We will insist on the modeling aspects, practical interests and limitations of each introduced methodological element. We will adopt a classifying approach, decomposing the optical flow estimation problem in independent parts. This viewpoint has the advantage of being didactic and giving a global view of existing method. However It should not hide that individual methods are usually conceived as homogeneous and coherent approaches, not restricted to a mere assembly of these elements.

Chapter 6 of this part is dedicated to a preliminary study. We investigate a first way to combine the two main classes of methods for optical flow, namely local and global approaches, through an adaptation of the work of [Bruhn et al., 2005] and [Corpetti and Mémin, 2012].

1 Basics of optical flow

1.1 Definition of optical flow

The concept of motion usually refers to the physical displacement of objects in a given referential. Extended to images, which are composed of intensity patterns on a two dimensional space, it can be understood as the shift of photometric features of the image, usually designated as *image motion*. It is associated with the constancy of the intensity patterns under motion. The difficulty in estimating *image motion* is to cope with the uncertainty due to the ill-posedness of the problem.

The motion of interest to be extracted from images is usually related to the displacements of objects in the physical scene. In video analysis, the *motion field* is defined as the projection on the image plane of the 3D motion in the scene. The problem arising is then that the temporal changes of image intensity are not necessarily caused by the displacements of objects of the scene, but can also be due to other disturbing phenomena like lighting changes, reflection effects or modification of the internal properties of the objects affecting their light emission or reflectance. These two sources of intensity changes (and consequently of *image motion*) have to be distinguished when computing the motion field. The term of *optical flow* is most commonly used to designate the *motion field*, and we will also consider this definition in the following.

1.2 Evaluation of optical flow

The accuracy of an optical flow estimation can be qualitatively evaluated through the visualization of the motion field. The two main visualization methods are illustrated in Figure 1.1. First, the arrow visualization directly represents motion vector and offers a good intuitive perception of physical motion. On the counterpart, a clean display requires to under-sample the motion field to prevent from overlapping arrows. Secondly, the color code associates a color hue to a direction and a saturation to the magnitude of the vector. It allows for a dense visualization of the flow field and a better visual perception of subtle differences between neighbor motion vectors. In the manuscript, we will use quasi-exclusively the color visualization. The visualization tools are useful to understand behaviours of estimation methods, especially when ground truth is not available.

Objective evaluation based on error metrics measured from ground truth motion fields

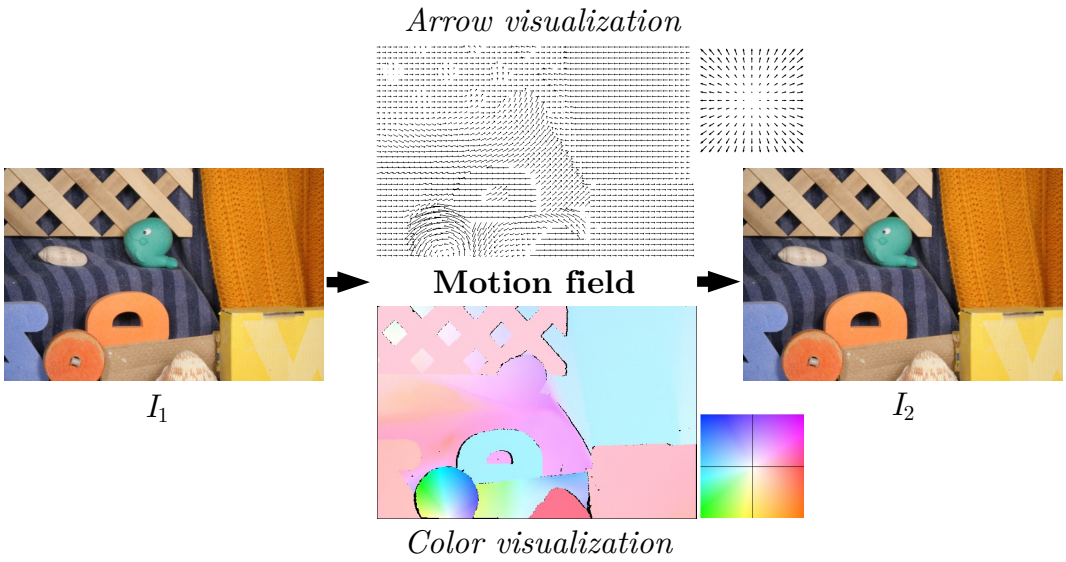


Figure 1.1: Arrow and color visualizations of optical flow

is necessary for an accurate comparison of methods performances When ground truth is available, two error measures are commonly used, namely the Angular Error (AE) and the Endpoint Error (EPE). The AE of an estimated motion vector $\mathbf{w} = (u, v)^\top$ w.r.t. the reference vector $\mathbf{w}_{ref} = (u_{ref}, v_{ref})^\top$ is defined by the 3D angle created by the extended vectors $(u, v, 1)^\top$ and $(u_{ref}, v_{ref}, 1)^\top$:

$$AE = \cos^{-1}\left(\frac{u \cdot u_{ref} + v \cdot v_{ref} + 1}{\sqrt{u^2 + v^2 + 1}\sqrt{u_{ref}^2 + v_{ref}^2 + 1}}\right). \tag{1.1}$$

The EPE is defined as the euclidean distance between the two vectors:

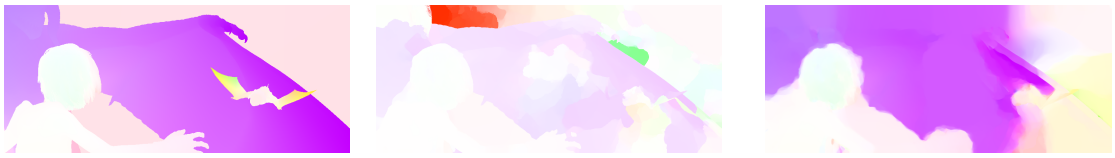
$$EPE = \sqrt{(u - u_{ref})^2 + (v - v_{ref})^2}. \tag{1.2}$$

The AE is more appropriate to accurately quantify small errors, occurring when the displacement magnitudes are low. In the example of Fig. 1.2, results of [Xu et al., 2012b] and [Chen et al., 2013] achieve both small errors, and the EPE is very small and cannot differentiate the performance of the two considered methods, while the AE allows for a more significant ranking. On the other hand, when errors are large, as in Fig. 1.3, AE tends to under-estimate large errors. The EPE is more impacted than the AE by large errors in the result of [Xu et al., 2012b] compared to the result of [Brox and Malik, 2011].

The design of challenging benchmarks with ground truth for evaluating optical

ground truth | Result of [Xu et al., 2012b] AE = 3.23 EPE = 0.08 | Result of [Chen et al., 2013] AE = 2.89 EPE = 0.08

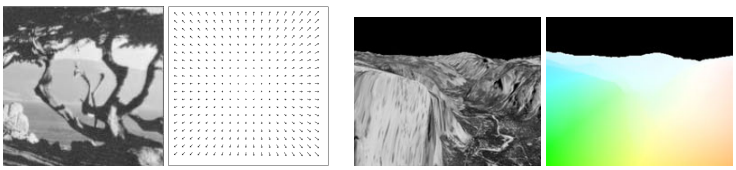
Figure 1.2: AE and EPE for small displacements.



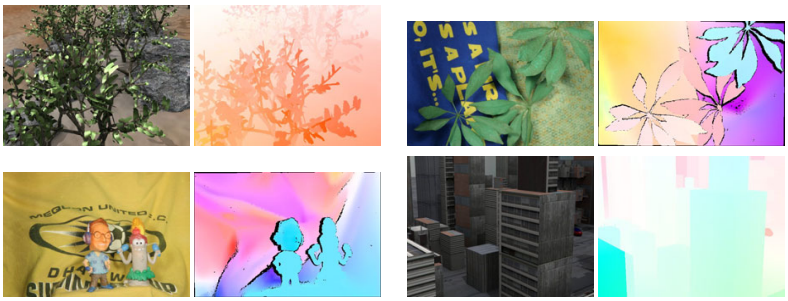
ground truth | Result of [Xu et al., 2012b] AE = 35.261 EPE = 48.034 | Result of [Brox and Malik, 2011] AE = 29.0133 EPE = 29.709

Figure 1.3: AE and EPE for large displacements.

flow methods has motivated a substantial amount of work. The first optical flow benchmark with ground truth was established by Barron et al. [1994]. The procedure was either to apply simple parametric transformations to real images, like translation or rotation, or to generate synthetic sequences for which the true motion is available by construction. The resulting motion fields were characterized by small displacements and absence of discontinuities. More recent benchmarks have been proposed with new challenges or applicative issues. The main successive benchmarks are illustrated in Fig. 1.4. The MIDDLEBURY benchmark [Baker et al., 2007, 2011] is composed of more challenging sequences, partly made of smooth deformations similar to the sequences described in [Barron et al., 1994], but also involving motion discontinuities and motion details. While some sequences are synthetic, several others were acquired in a strictly controlled environment allowing to produce ground truth for real scenes. Issues raised by MIDDLEBURY being almost solved by modern methods, the MPI SINTEL benchmark [Butler et al., 2012] has been recently proposed. It is extracted from a synthetic movie opening new issues mostly related to very large displacements, occlusions, illumination changes, and effects like blur or defocus. In parallel to this effort, dedicated datasets have been designed to solve specific problems related to applicative contexts, the most successful example being the KITTY benchmark [Geiger et al., 2012] devoted to assisted



[Barron et al., 1994]



Middlebury [Baker et al., 2007]

Kitty [Geiger et al., 2012]

MPI Sintel [Butler et al., 2012]

Figure 1.4: Main optical flow benchmarks.

driving applications.

1.3 Estimation principles

Let us denote an image sequence by $I : \Omega \times T \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^2$ is the image domain and $T \subset \mathbb{N}$ is the number of frames of the sequence. Every optical flow estimation method is based on an assumption on the relationship between the searched motion field $\mathbf{w} : \Omega \to \mathbb{R}^2$ at time t and the image I. The most natural and widely used assumption is that pixel intensity remains constant during displacement. The brightness constancy constraint equation (BCCE) is then defined by:

$$\frac{dI}{dt} = 0. \tag{1.3}$$

Other feature conservations can be chosen, each encoding specific image properties, which will be discussed in Chapter 2. The discrete approximation of (1.3) at a given pixel $x \in \Omega$ and time t yields:

$$I(x + \mathbf{w}(x), t + 1) - I(x, t) = 0. \tag{1.4}$$

However, the constraint (1.4) generates a particularly difficult optimization problem. It can be much more tractable to consider the expanded version of (1.3) with partial derivatives, resulting in a linear version of (1.4):

$$\frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \frac{\partial I}{\partial t} = 0, \tag{1.5}$$

where $\mathbf{w} = (u, v)^\top$.

The linearized brightness conservation constraint (1.5) provides only one equation to recover the two unknown components of $\mathbf{w}$. From this single constraint, the component of the motion vector $\mathbf{w}$ in the direction of the image gradient can be computed, but the two-dimensional problem remains under-constrained. This is known as the *aperture problem*, stating that motion of linear structures, as it is assumed by (1.5), is by nature ambiguous if the neighboring context is not taken into account.

To make the problem well-posed, it is necessary to introduce an additional constraint encoding *a priori* information on $\mathbf{w}$. The *a priori* will take the form of spatial coherency imposed by either *local* or *global* constraints. The classification of the usual practices and recent advances will be presented in Chapters 3 and 4.

It is important to mention that while (1.4) holds for motion of arbitrary magnitude, the continuous motion constraint (1.5) restricts its validity domain to the linear region of I, which usually corresponds to small displacements or very smooth images. The linearization is nevertheless necessary for methods relying on differential computations.

The standard technique to cope with large displacements is to embed the estimation in a coarse-to-fine scheme [Enkelmann, 1988; Black and Anandan, 1996; Mémin and Pérez, 1998]. The idea is to create a pyramid of coarse-to-fine downsampled versions of the original image. On coarse levels, the linearity domain of the image encompasses larger displacements and the estimation can be based on (1.5). The estimations at coarser levels serves to warp the image at subsequent finer levels, where the estimation then reduces to search for small motion increments. The solution is iteratively refined at each level until reaching the full image resolution. The solution at each level of the multi-resolution pyramid can be interpreted as a fixed point in the direct optimization of the non-linearized constancy equation (1.4) [Brox et al., 2004]. Almost all differential methods described in this part and concerned by large displacements resort to the multiscale approach, possibly with some additional strategies to avoid its drawbacks.

The main undesirable effect produced by smoothing at coarse levels is the loss of small and rapidly moving objects in the final estimated flow field. If the object extent is smaller than its displacement, it is likely to be smoothed out at coarse levels and then "forgotten". Avoiding this drawback has been a very active topic in recent years. It has been accomplished mainly by integrating feature matching in the estimation process, as will be discussed in Chapter 5, or by resorting to discrete optimization methods, as detailed in Section 4.3.2.

2 Data constancy

The data term of optical flow methods penalizes deviations from the constancy assumption, e.g. brightness constancy (1.4), via the sum of potentials $\rho_{data}(x, I_1, I_2, \mathbf{w})$ defined at each pixel $x \in \Omega$ as

$$E_{data}(\mathbf{w}, I_1, I_2) = \int_\Omega \rho_{data}(x, I_1, I_2, \mathbf{w}) dx \tag{2.1}$$

where $I_1 = I(\cdot, t)$ and $I_2 = I(\cdot, t+1)$ denote two successive frames. In the case of brightness constancy, the data potential writes

$$\rho_{data}(x, I_1, I_2, \mathbf{w}) = \phi(I_2(x + \mathbf{w}(x)) - I_1(x)) \tag{2.2}$$

where $\phi(\cdot)$ is the penalty function.

The brightness constancy assumption is in practice an imperfect photometric expression of the real physical motion in the scene. The typical counter-example consists in moving the light source of an immobile scene, producing brightness variations without motion of any objects. In general, while it is possible to create synthetic sequences for which the constraint strictly holds Butler et al. [2012], it is often violated in practice in case of changes in the illumination source of the scene, shadows, noise in the acquisition process, specular reflections or even complex motion.

Choosing a quadratic penalty function $\phi(z) = z^2$, as in early works [Horn and Schunck, 1981; Lucas and Kanade, 1981], makes optimization much easier, but assumes that the residual of the brightness constancy constraint equation (1.3) is normally distributed and thus gives a strong influence to large localized violations mentioned above. It is then common to resort to robust statistics [Huber, 1981] to reduce the impact of local errors considered as outliers [Odobez and Bouthemy, 1995; Black and Anandan, 1996; Mémin and Pérez, 1998]. Adapted optimization schemes must then be adopted to cope with non-linearity or non-convexity induced by the robust terms, as will be discussed in Sections 3.2 and 4.3. *A priori* smoothness assumption based on parametric constraint (Chapter 3) or explicit regularization (Chapter 4) also counterbalances local invalidity of data constancy.

Robust statistics and regularization treat the problem of violations of the constancy assumption by considering it as noise, with underlying distribution assumptions [Simoncelli et al., 1991; Krajsek and Mester, 1996]. The considered distributions may not suitably

model the possibly large localized violations implied by the above listed causes. Therefore, a large number of alternatives to brightness constancy have been proposed, aiming at more stable invariance properties. A few experimental studies have compared performances of different data costs given fixed optimization and regularization contexts [Steinbrucker et al., 2009; Vogel et al., 2013].

Let us notice as a preamble to this section that it is difficult in practice to design a data term independently from the spatial coherence constraint and the optimization strategy to which it will be associated. For example, sophisticated feature conservation usually involves specific optimization difficulties, and is thus closely intricate with the choice of the optimization solution. We will dedicate this chapter to a review of the main classes of data terms for optical flow estimation, emphasizing their validity domains and their limitations, independently from the estimation context in which they were elaborated.

We will not address in this chapter the problem of characterizing motion in occluded regions; we will focus only on the behaviour of data terms when correspondences exist. The specific treatment of occlusions will be addressed in Part II.

2.1 Beyond brightness constancy

We explore several matching costs aiming at compensating the drawback of the brightness constancy, in particular its sensitivity to noise and illumination changes.

2.1.1 Image filtering

The first class of data potentials has the same pixel-wise form as (2.2), but operates on a filtered version $f(I)$ of the original image sequence:

$$\rho_{data}(x, I_1, I_2, \mathbf{w}) = \phi_{data}(f(I_2)(x + \mathbf{w}(x)) - f(I_1)(x)) \tag{2.3}$$

Image smoothing We can first notice that Gaussian smoothing is applied as a pre-processing step by most methods [Brox et al., 2004; Zimmer et al., 2011], in order to reduce the influence of noise. It can be viewed as a modified version of brightness constancy, setting f as a Gaussian filtering operator.

High-order constancy Image derivatives possess illumination invariance properties that are well suited for motion estimation. The constancy of spatial image gradient, defined by $f(I) = \nabla I$, has been introduced in [Uras et al., 1988] for its ability to overcome the *aperture problem* when the determinant of the Hessian is non-zero. However, when applied on the directional derivative vectors, the gradient conservation only holds for translational or divergence motions. To achieve rotational invariance, the penalty should

rather be applied on the magnitude of the derivatives, that is $f(I) = \|\nabla I\|$ [Brox et al., 2004]. It was subsequently used in the context of the *local* approach (see Chapter 3) in [Tistarelli, 1996] and integrated in *global* variational methods (see Chapter 4) in [Brox et al., 2004].

Despite a demonstrated performance gain in the case of additive illumination changes compared to brightness constancy, gradient conservation is also much more sensitive to noise, and looses information in poorly textured regions. Therefore, it is always used in complementarity with brightness constancy. A large number of methods rely on this combination and achieve good results [Brox et al., 2004; Xu et al., 2012b; Mozerov, 2013]. Finally Papenberg et al. [2006] investigated higher-order constancy like Laplacian $f(I) = \Delta I$, or Hessian $f(I) = \mathcal{H}\, I$ conservation.

Texture Another way to obtain robustness against illumination changes is to work with the structure and texture components of the image, as proposed in [Wedel et al., 2009b]. The decomposition proposed in [Aujol et al., 2006] consists in first obtaining the structure part I_S by discontinuity-preserving smoothing (using the ROF model [Rudin et al., 1992] in [Wedel et al., 2009b]), and then deriving texture part I_T by subtracting I_S to the original image. Illumination changes only affect the structure image while the texture image is less impacted. However, similarly to the image gradient constraint, I_T misses a lot of other image information and is more sensitive to noise. To limit this drawback, the texture image used to compute optical flow is blended with the structure part by a parameter α: $f(I) = I - \alpha I_S$. A number of methods adopted this constraint in [Wedel et al., 2009a; Sun et al., 2010a; Krähenbühl and Koltun, 2012]. From our experiments, this constraint globally produces more erroneous measures than the combination of brightness and gradient constancy.

Colour spaces When dealing with colour images, several photometric invariant colour spaces can be exploited. In particular, multiplicative illumination invariance is essential for realistic illumination models [van de Weijer and Gevers, 2004] and is achieved in the HSV space by the hue channel (local and global changes) and the saturation channel (only global changes) [Mileva et al., 2007]. As for previously mentioned image transformations, the benefit in illumination change regions coincide with a loss of information in other parts, and the colour channels are in practice combined with the intensity valued channel [Zimmer et al., 2011]. Other colour spaces like normalized RGB [Golland and Bruckstein, 1997] or spherical space [van de Weijer and Gevers, 2004] have been proposed.

Combination of linear filters As mentioned before, it is often necessary to combine several constancy assumptions. In [Sun et al., 2008], a learning approach is proposed for finding the best combination of linearly filtered brightness (that is taking f as a linear

filter). Each data potential induced by a given linear filter J_k is penalized by a Gaussian Mixture Model (GSM) ϕ_{GSM} of L models. The weight accorded to each filter is encoded in the parameters of ϕ_{GSM}, $\Xi_k = \{\xi_1^k, \ldots, \xi_L^k\}$. We obtain

$$\rho_{data}(x, I_1, I_2, \mathbf{w}) = \sum_k \phi_{GSM}(J_k * I_2(x + \mathbf{w}(x)) - J_k * I_1(x), \Xi_k), \tag{2.4}$$

where $*$ denotes the convolution operator, and

$$\phi_{GSM}(z, \Xi_k) = \sum_{l=1}^{L} \xi_l^k \, \mathcal{N}(z, \sigma^2/s_l), \tag{2.5}$$

where s_l are the scales of the elements of the mixture. The parameters Ξ_k associated to each filter are learned from a set of ground truth sequences. Spatially adaptive combination is also of upmost importance and will be addressed in Section 2.2.1.

2.1.2 Patch-based measures

Rather than pre-filtering images, neighborhood information can be integrated directly in the data term by patch-based similarity measures. We can already stress that a major issue with patch-based measure is the determination of the size or shape of the patch. The methods we have developped address this issue. They are presented in Chapter 6 and Part II.

Filtering the data term In addition to pre-smoothing the images (2.3), Bruhn et al. [2005] proposed to filter the data potential as follows:

$$\rho(x, I_1, I_2, \mathbf{w}) = f\left(\phi(I_2(x + \mathbf{w}(x)) - I_1(x))\right). \tag{2.6}$$

Bruhn et al. [2005] chose f to be a Gaussian filter. While it was demonstrated beneficial for very noisy sequences, it also significantly blurs motion edges and degrades the overall performance for low amount of noise, compared to pixel-wise data term, as emphasized in [Zimmer et al., 2011]. This limitation is addressed in [Drulea and Nedevschi, 2011; Rashwan et al., 2013] by replacing the Gaussian filtering with anisotropic discontinuity-preserving filtering (e.g. bilateral filtering in [Drulea and Nedevschi, 2011] and tensor voting in [Rashwan et al., 2013]). We suggest an additional variant in Chapter 6.

Correlation-based measures Similarity measures based on cross-correlation have been extensively used for various correspondence problems. Normalized Cross Correlation (NCC) is usually preferred for its invariance to linear illumination changes. The NCC for

a window $\mathcal{W}(x)$ centered at pixel x is defined as

$$NCC(x, I_1, I_2, \mathbf{w}) = \frac{\sum_{y \in \mathcal{W}(x)} (I_2(y + \mathbf{w}(x)) - \mu_2(x + \mathbf{w}(x)))(I_1(y) - \mu_1(x))}{\upsilon_1(x)\upsilon_2(x + \mathbf{w}(x))} \tag{2.7}$$

where for $i = \{1, 2\}$, $\mu_i(x)$ is the mean and $\upsilon_i(x)$ the standard deviation of I_i in the window $\mathcal{W}(x)$. The associated data potential is

$$\rho_{data}(x, I_1, I_2, \mathbf{w}) = 1 - NCC(x, I_1, I_2, \mathbf{w}). \tag{2.8}$$

NCC is actually discriminative enough to be used in a matching procedure without additional regularization, and produces coarse but reasonably robust motion fields. It is used in several applications like stereovision [Delon and Rougé, 2007], fluid flow analysis [Becker et al., 2012] or biological imaging [Kolin and Wiseman, 2007] where it also enables direct physical measures for diffusion processes, as it will be explained in Chapter 15.

The cost in the computation of (2.7) is a major limitation. Unlike simple cross-correlation which can be efficiently computed with Fast Fourier Transform (FFT), the computation of NCC for matching purpose cannot be easily performed in the frequency domain. Lewis [1995] computes only the numerator with FFT and proposed to rewrite the denominator as a product of sums independent of the position of the pixel, and thus efficiently computable with integral images [Facciolo et al., 2013]. Luo and Konofagou [2010] generalized this idea and compute also the numerator with integral images, dramatically reducing the computation time and making it invariant to the patch size.

Integrating NCC in a variational optimization scheme is challenging because it requires differentiating it. Indeed, Taylor expansion on the terms containing $\mathbf{w}$ in (2.7) still yields a highly non linear potential. The approach of [Molnár et al., 2010; Werlberger et al., 2012] has been applied to NCC but is able to handle arbitrary data terms as well. The authors directly linearize the data term and compute its spatial derivatives with finite differences. Werlberger et al. [2012] keep a second order approximation to ensure the convexity of the energy, necessary in the primal-dual scheme used. Another recent technique allowing very fast computation of NCC relies on the fact that NCC is actually equivalent to the Sum of Square Differences (SSD) when the images are filtered with the cheaply computed *correlation transform* [Drulea and Nedevschi, 2013].

Census Census Transform [Zabih and Woodfill, 1994] recently regained interest and was promoted by Stein [2004] for optical flow estimation [Müller et al., 2011; Muller et al., 2011; Mohamed and Mertsching, 2012; Ranftl et al., 2012; Vogel et al., 2013; Hermann and Klette, 2013; Hafner et al., 2013]. The Census signature is a bit string reflecting relative value of pixels of a patch with the center pixel. By discarding the absolute

intensity values, only the structure of the neighborhood is encoded in the signature, which makes it robust to illumination changes. It has shown robust behaviour in outdoor scenes and vehicle driving scenarios [Vogel et al., 2013; Ranftl et al., 2012; Stein, 2004]. Integrating the Census transform in variational optical flow is not trivial since it cannot be easily linearized. Solutions to remedy this problem are convex approximation [Vogel et al., 2013], reformulation as a generalization of the gradient constancy conservation [Hafner et al., 2013] or linearization of the data term [Ranftl et al., 2012; Müller et al., 2011] as previously mentioned for NCC [Werlberger et al., 2012]

2.2 Spatially adaptive matching costs

The validity of each of matching costs is limited to a given range of visual situations. In a single frame regions can coexist satisfying a given constancy assumption, and violating others. One solution could be to linearly combine them to take advantage of their complementary invariance properties. Softly selecting the best constancy constraint at each pixel is usually devoted to the robust penalty function, limiting the influence of locally wrong assumptions. However, the data term should ideally be spatially adapted.

We distinguish two classes of methods achieving the spatial adaptivity: i) optimization of the weights of a linear combination of data potentials, and ii) estimation of the spatial distribution of the errors attached to a single data potential. The normalization of the data term used in [Simoncelli et al., 1991; Schoenemann and Cremers, 2006; Zimmer et al., 2011] could fall in this category since a spatially varying weight is applied to the data term. It is derived from the linearized brightness constancy constraint to prevent too strong data constraint in regions of high image gradient (see a detailed interpretation in [Zimmer et al., 2011]).

2.2.1 Spatially adaptive combination of constancy assumptions

The combination of P data constraints can be expressed as the weighted sum of their associated potentials $\rho_p(x, I_1, I_2, w)$:

$$\rho_{data}(x, I_1, I_2, \mathbf{w}) = \sum_{p=1}^{P} \lambda_p(x)\, \rho_p(x, I_1, I_2, \mathbf{w}). \tag{2.9}$$

The weights $\lambda_p(x)$ are spatially variant and have to be optimized to locally favor different data terms.

The idea of combining several data constraints has already been explored twenty years ago in [Heitz and Bouthemy, 1993]. In addition to the classical brightness constancy, the

authors exploit a complementary sparse edge-based constraint. The weights $\lambda_p(x)$ are binary confidence measures derived from hypothesis testing providing evidence on each constraint.

Xu et al. [2012b] combine intensity and gradient conservation, experimentally showing their complementarity. The weights are defined to operate a binary selection between the two constraints and are obtained by considering a mean field approximation of (2.9), which intuitively amounts to selecting the constraint having the lower (normalized) potential. This idea has been used subsequently in [Mozerov, 2013].

The work of [Kim et al., 2013] addresses the problem in its most general form (2.9), allowing the combination of an arbitrary number and type of data conservation assumptions. A confidence measure for arbitrary data term is designed as an extension of the feature discriminability [Shi and Tomasi, 1994] to data discriminability. The confidence measures are used as local constraints on the weights $\lambda_p(x)$ in (2.9), and a regularization on $\lambda_p(x)$ is also imposed. The weights are then optimized jointly with the motion field.

2.2.2 Modeling of data constancy violations

Another way to handle errors related to the constancy constraint is to explicitly model them as an additional variable of the problem. Considering the brightness constancy, errors can be modeled by a parametrized function $e(x, I_1, I_2, \xi)$:

$$\rho_{data}(x, I_1, I_2, \mathbf{w}, \xi) = \phi(I_2(x + \mathbf{w}(x)) - I_1(x) - e(x, I_1, I_2, \xi)). \tag{2.10}$$

The model proposed by [Negahdaripour, 1998] is composed of an offset change $\xi_o : \Omega \to \mathbb{R}$, accounting, e.g., for moving shadows or highlights, and a multiplicative change $\xi_m : \Omega \to \mathbb{R}$ encoding linear illumination variations. The error function can then be expressed as

$$e(x, I_1, I_2, \xi) = \xi_m I_1(x) + \xi_o. \tag{2.11}$$

This general formulation has been exploited in a number of works, considering either the offset parameter alone [Odobez and Bouthemy, 1995; Chambolle and Pock, 2011], the multiplier alone [Zach et al., 2008] or both parameters [Kim et al., 2005; Teng et al., 2005; Lai, 2000]. They may differ on their type of spatial coherency, the penalty function, or the optimization strategy. A smoothness constraint is assumed on ξ_o and ξ_m, either with a local parametric form [Odobez and Bouthemy, 1995; Negahdaripour, 1998] or a global regularization [Lai, 2000; Kim et al., 2005; Teng et al., 2005; Chambolle and Pock, 2011]. The offset formulation is also used in [Ayvaci et al., 2012], but it is constrained to be sparse, rather than smooth, with the aim of retrieving violations only due to occlusions.

The model (2.11) is based on general assumptions about physics of imaging. If specific

knowledge about the observed physical process is available, dedicated models can be designed. Haussecker and Fleet [2001] explored a number of physical constraint and design a generic local estimation framework based on a Taylor expansion of arbitrary data constraints similar to the subsequent methods [Molnár et al., 2010; Werlberger et al., 2012]. We will propose a variant of the offset approach in Chapter 14.

3 Parametric approach

As explained in Section 1.3, a spatial coherence constraint must be added to the previously described data terms. To this end, a class of methods impose the flow field to follow a parametric model in region $\mathcal{R} \subset \Omega$. The motion field $\mathbf{w}_{\boldsymbol{\theta}} : \mathcal{R} \to \mathbb{R}^2$ is then fully characterized by the associated parameter vector $\boldsymbol{\theta}$. When the region $\mathcal{R}$ is a small sub-domain of the image, these methods are referred to as *local* approaches. The objective energy to be minimized is the weighted sum of the potentials provided by each pixel of $\mathcal{R}$:

$$\widehat{\boldsymbol{\theta}} = \arg\min_{\boldsymbol{\theta}} \sum_{x \in \mathcal{R}} g(x) \rho_{data}(x, I_1, I_2, \mathbf{w}_{\boldsymbol{\theta}}) \tag{3.1}$$

where $g(x)$ is a spatial weighting function controlling the influence of pixel x in the estimation.

It is crucial to determine local estimation domains where the parametric form of the motion model is a valid approximation of the true motion. Low-order polynomial motion models like translation or affine deformation can usually represent motion in small neighborhoods, whereas more complex models like deviations from affine constraint or combination of basis functions can deal with larger regions.

We first give an overview of the mostly used motion models and their associated optimization strategies. Secondly, we discuss about the different ways to define appropriate local estimation domains. The aggregation method presented in Part II will propose a new approach for the implicit selection of the region $\mathcal{R}$.

3.1 Motion models

The choice of the motion model is driven by a trade-off between efficiency and representativeness. Complex nonlinear and physical-based models can be exploited to model deformations for image registration [Sotiras et al., 2013]. These models are particularly well adapted to physically constrained situations as they can be encountered in medical imaging, and in particular to capture smooth deformations. In contrast, optical flow is dealing with temporal sequences of arbitrary type, usually involving motions of several objects with unrelated behaviours, generating discontinuities as well as smooth parts. As a result, it is difficult to capture the whole complexity of motion fields with a single unifying and computationally tractable parametric model. Therefore, attempts in

this direction are not frequent and not among the best performing methods in optical flow benchmarks. The approach of most parametric methods for optical flow is rather to rely on much simpler motion models, mostly polynomial models, but to restrict their application to local domains, where they can represent accurate approximations.

In this thesis, we restrict ourselves to linear models of the form

$$\mathbf{w}_{\boldsymbol{\theta}}(x) = \sum_{k=1}^{K} \mathbf{b}_k(x)\boldsymbol{\theta}(x), \tag{3.2}$$

where $\mathbf{b}_k(x)$ are basis functions and $\boldsymbol{\theta}(x)$ the weights to be optimized. Other parametric models than those described here can be found, like planar surfaces or rigid body [Bergen et al., 1992], or wavelet basis [Wu et al., 2000; Dérian et al., 2011, 2012; Shen and Wu, 2010]. It can also be noted that parametric models are sometimes completed with explicit regularization terms (see Chapter 4) imposed on the parameters themselves [Nir et al., 2008; Dérian et al., 2011; Ju et al., 1996; Memin and Perez, 2002].

3.1.1 Polynomial models

Polynomial models are among the most compact parametric representations of motion fields and are also remarkably well suited to retrieve local physical motion of individual objects. Apart from the exception of [Nir et al., 2008] where the parameters are spatially variant and regularized, the parameters are kept constant over the estimation domain, $\boldsymbol{\theta}(x) = \boldsymbol{\theta}$. The basis $\mathbf{b}_k(x)$ are functions of the coordinates of the domain, and the order of the polynomial determines the complexity of the motion field. Low order polynomials like translational and affine models are usually sufficient to model smooth motion fields, and their small number of parameters allows for efficient computation:

$$\text{Translational}: \quad \boldsymbol{\theta} = (a_1, a_2)^\top; \quad \mathbf{b}_k(x) = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \tag{3.3}$$

$$\text{Affine}: \quad \boldsymbol{\theta} = (a_1, a_2, a_3, a_4, a_5, a_6)^\top; \quad \mathbf{b}_k(x) = \begin{pmatrix} 1 & x_1 & x_2 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & x_1 & x_2 \end{pmatrix} \tag{3.4}$$

with $x = (x_1, x_2)^\top$.

The translation assumption is very restrictive and must be applied to very small regions [Lucas and Kanade, 1981]. The physical assumption underlying the affine model is a rigid motion of 3D objects projected orthogonally on the image plane, which is often a good approximation. Higher order polynomials can model more complex situations, but are still too smooth to allow for motion discontinuities. For example, the 8-parameter quadratic model represents rigid motion of a plane surface in perspective projection. The small number of parameters of the affine model and its realistic local assumption make it

often considered as the best trade-off between complexity and descriptiveness [Odobez and Bouthemy, 1995; Black and Anandan, 1996; Memin and Perez, 2002].

3.1.2 Learned basis

The basis functions can also be learned from a set of training flow fields. As polynomial models, resulting motion fields cannot describe motion of complex scenes, but they are able to retrieve a larger diversity of local motion patterns, including discontinuities.

The design of the training set reflects the assumption on the form of the flow. In a generic point of view, Black et al. [1997] used synthetic motion fields representing simple motion patterns. Nieuwenhuis et al. [2010] relied on a large number of patches of ground truth motion fields. The training set can be dedicated to a specific application, as in [Fleet et al., 2000] where the aim is to estimate mouth motion. To avoid resorting to external ground truth, Garg et al. [2011] define the training set on the processed sequence itself. The basis set is composed of trajectories constructed by feature tracking on large temporal scales, in regions ensuring reliable tracks. In all these works, an orthogonal basis of flow fields is generated by PCA decomposition, conserving only the first K components containing most of the variance of the training set.

3.1.3 Free-form deformations

The free-form deformation model (FFD) [Rueckert et al., 1999] has been originally introduced for image registration and has demonstrated great robustness to retrieve smooth deformations. The displacements are defined on a coarse regular subgrid of the image and is interpolated on the final resolution with B-splines. The motion basis $\mathbf{b}_k(x)$ is thus the displacements of the k^{th} control point and the coefficient $\boldsymbol{\theta}_k(x)$ is a B-spline influence function. The dimensionality reduction induced by the subsampling of the image grid makes the computation much easier, and the spatial coherence of the deformation is ensured by the B-spline interpolation. On the counterpart, the framework cannot retrieve sharp motion discontinuities, while it is necessary for optical flow applications. Image-adaptive non-regular control points distribution [Schnabel et al., 2001] or coarse-to-fine spacing strategies [Rueckert et al., 1999] are possibilities to address this problem.

Szeliski and Shum [1996] were the first ones to apply this idea to optical flow, with non-uniform control points defined on image-driven quadtrees. Glocker et al. [2008] turned the problem in a discrete setting. They iteratively adapt the range of motion labels by estimating a local uncertainty covariance. They obtained good results but their method is still limited by over-smoothing. Shi et al. [2012] addressed the discontinuity problem with a sparsity constraint on the B-spline coefficients, allowing to modulate the influence of the control points. Recent approaches [Glocker et al., 2007, 2008; Shi

et al., 2012] also add an explicit regularization on the motion field, overlapping with the methodology described in Chapter 4.

3.2 Optimization

Parametric models are usually associated with the penalty of a pixel-wise data constancy constraint (2.2). In case of intensity constancy (1.5), the energy (3.1) then writes:

$$E(\boldsymbol{\theta}) = \sum_{x \in \mathcal{R}} g(x) \phi \left(\nabla I^\top(x) \mathbf{w}_{\boldsymbol{\theta}}(x) + I_t(x) \right) \tag{3.5}$$

where ∇I is the image gradient and $I_t = \partial I / \partial t$. The special case of a quadratic function ϕ and a translational model $\mathbf{w} = \boldsymbol{\theta}$ as in [Lucas and Kanade, 1981] leads to a very simple optimization problem, since the cancelling of the derivatives of (3.5) amounts to solving the following linear system:

$$\mathbf{M}\boldsymbol{\theta} = \mathbf{b}, \tag{3.6}$$

with

$$\mathbf{M} = \begin{pmatrix} \sum_x g(x)\, I_{x_1}^2(x) & \sum_x g(x)\, I_{x_1}(x) I_{x_2}(x) \\ \sum_x g(x)\, I_{x_1}(x) I_{x_2}(x) & \sum_x g(x)\, I_{x_2}^2(x) \end{pmatrix} \text{ and } \mathbf{b} = \begin{pmatrix} \sum_x g(x)\, I_{x_1}(x) I_t(x) \\ \sum_x g(x)\, I_{x_2}(x) I_t(x) \end{pmatrix} \tag{3.7}$$

where I_{x_1} and I_{x_2} are the partial derivatives of I respectively along the horizontal axis x_1 and the vertical axis x_2. The rank of $\mathbf{M}$ allows one to decide if a unique solution of the linear system (3.6) exists, and can be used to adapt the size of the local domain $\mathcal{R}$ (see Section 3.3.2). Despite the limitations of the quadratic penalty, this approach has become very popular for its implementation simplicity, low computational cost and available implementation in the OpenCV library [Bradski, 2000].

However, robust estimation is often advocated [Odobez and Bouthemy, 1995; Black and Anandan, 1996; Black et al., 1997; Dérian et al., 2012; Senst et al., 2012] as mentioned in Chapter 2, especially for polynomial models, to deal with the frequent case of multiple motions in the estimation domain. Among the variety of optimization methods used to optimize the robust penalty function, the Iterative Reweighted Least Squares (IRLS) [Holland and Welsch, 1977] and gradient descent approaches have mostly been used. IRLS proceeds by successive optimizations of quadratic problems weighted by a function of the current estimate [Odobez and Bouthemy, 1995; Senst et al., 2012]. Gradient descent approaches are often coupled with Graduated Non-Convexity (GNC) [Blake and Zisserman, 1987; Black and Anandan, 1996] to cope with non-convexity of (3.5). Regarding the slow convergence of steepest descent, it is preferable to use second order approximations and Newton methods, or quasi Newton methods like L-BFGS or Levenberg-Maquardt, approximating the Hessian for large dimension problems.

3.3 Neighborhood selection

As previously mentioned, the spatial adaptation of the parameters $\boldsymbol{\theta}(x)$ is a way to cope with complex and discontinuous flow fields [Nir et al., 2008; Garg et al., 2011; Shi et al., 2012]. This approach often involves *a priori* constraints on the spatial distribution of $\boldsymbol{\theta}(x)$ and is thus strongly related to the methods that will be presented in Chapter 4. Such a dense parameter map moves away from the compactness of parametric models.

On the other hand, when the parameters are constant over the estimation domain $\mathcal{R}$, the resulting motion field is imposed to be smooth and is a valid approximation in regions of coherent motion, free from motion discontinuities. The choice of the region $\mathcal{R}$ is then crucial, $\mathcal{R}$ must be large enough to enable motion estimation, while small enough to keep valid the parametric approximation. We will describe strategies for defining $\mathcal{R}$ in the case of constant parameters $\boldsymbol{\theta}(x) = \boldsymbol{\theta}_{const}$ over $\mathcal{R}$, for polynomial models.

3.3.1 Entire domain

Despite their inability to retrieve realistic motion in arbitrary scenes, polynomial models combined with robust estimation are well adapted to capture dominant motion. Applied in the whole image domain, they become particularly useful to estimate the camera motion [Odobez and Bouthemy, 1995].

3.3.2 Square patches

The approach initiated by [Lucas and Kanade, 1981] performs independent estimations in small square or circular patches. Most of the related methods use fixed patch size, and conserve the velocity vector deduced from the estimated model at the square center [Baker and Matthews, 2004; Bigun et al., 1991; Zach et al., 2008; Kim et al., 2004]. This choice is very popular for its simplicity of implementation and can be naturally parallelized nature [Zach et al., 2008; Sinha et al., 2011]. It is still extensively used for numerous applications. However, following the *generalized aperture problem*, patches centered at each pixel with a fixed size are likely to contain either multiple motions or no image gradient, regarding the position of the pixel.

Multiple motions in a single patch can be partially handled with robust estimation by rejecting secondary motions, considered as outliers [Odobez and Bouthemy, 1995; Black and Anandan, 1996; Gelgon et al., 1999; Senst et al., 2012].

The second option is to adapt the size or the position of the patch so that it contains an unimodal motion distribution. The size of the patch can be adapted with a bias-variance criterion Maurizot et al. [1995], or based on a confidence measure on the reliability of the local domain for parametric estimation. Starting from a small patch size, it is thus possible to increase the size of the patch until it fulfils the condition for a reliable domain

[Senst et al., 2012]. In the case of [Lucas and Kanade, 1981], such a condition can be found by analyzing the singularity of the matrix $\mathbf{M}$ of the linear system (3.7) and, for example, imposing a minimum threshold to the maximum eigenvalue of $\mathbf{M}$ [Barron et al., 1994]. Rather than adapting the size, Jodoin and Mignotte [2009] adapt the position of the patches. Patches corrupted by strong intensity edges are displaced by a mean-shift procedure to reach homogeneous regions.

However, all these variants have never been competitive with state-of-the-art approaches based on global regularization (Chapter 4) or segmentation (Section 3.3.3). Nevertheless, we will show in Chapter 8 that if the sizes and positions are appropriately chosen, square patches are sufficient approximations to yield better flow estimation than state-of-the-art methods.

3.3.3 Segmented regions

The optimal regions $\mathcal{R}$ to perform polynomial estimations ideally correspond to a segmentation of the image in coherently moving regions. We briefly describe two types of approaches: independent image segmentation and joint estimation of motion and region supports or frontiers.

Image segmentation While the ultimate goal is to segment the unknown motion field, color-based image segmentation is a much simpler alternative which can help motion estimation. It can be reasonably assumed that motion discontinuities coincide with image discontinuities (but the inverse is far from being true). It implies that an image segmentation is a motion field over-segmentation, and obtained regions are thus guaranteed to contain no motion discontinuity. However, merely estimating motion in the resulting regions is problematic for two reasons.

The first limitation is that the segmented regions, may not contain enough information for motion estimation. Parametric estimations in these regions must be performed by circumvented ways. The very fine over-segmentation of Zitnick et al. [2005] imposes for instance to perform region matching. Generally, an independent coarse and cheap motion estimation is fused with the color image segmentation to overcome the lack of information [Xu et al., 2008; Black and Jepson, 1996; Bleyer et al., 2006]. Xu et al. [2008] find hybrid regions by applying mean-shift segmentation in the extended space of color and motion. Differently, [Black and Jepson, 1996; Bleyer et al., 2006] fit a parametric flow field on the coarse initial motion field, obtained with a global regularized method [Black and Anandan, 1993] for [Black and Jepson, 1996], and with the sparse KLT tracker [Shi and Tomasi, 1994] for [Bleyer et al., 2006].

The second problem is that spatial coherence between estimated motion in neighboring segments is not ensured. Global regularization (see Chapter 4) can here be imposed,

either on the motion parameters associated to each region [Xu et al., 2008] (similarly to [Ju et al., 1996], not resorting to image segmentation), on the coarse motion field completing color information [Black and Jepson, 1996] or, in a layered approach (see below), on the layer assignment function [Bleyer et al., 2006]

Joint estimation and segmentation Color segmentation is usually too dependent on the image content to make it the basis of a robust motion estimation method. Rather than considering segmentation and estimation as two independent tasks, most methods have a coupled approach of the problem. Motion parameters and region supports are jointly estimated by minimizing a global energy imposing a coupling between them. This approach has first been addressed as a labelling problem [Bouthemy and François, 1993; Odobez and Bouthemy, 1998] where the label field $l : \Omega \to \{l_1, \ldots, l_N\}$ associated to the N regions is estimated jointly with the motion parameters in each region $\boldsymbol{\theta} = \{\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_N\}$, in a discrete Markov Random Field framework:

$$E(\boldsymbol{\theta}, l) = \sum_{x \in \Omega_d} \rho_{data}\left(x, I_1, I_2, \mathbf{w}_{\boldsymbol{\theta}_{l(x)}}\right) + \sum_{<x,y>} \rho_{reg}^{\text{MRF}}(l(x), l(y)) \tag{3.8}$$

where $\rho_{reg}^{\text{MRF}}(l(x), l(y))$ is a regularization prior on the label field, typically chosen as $\rho_{reg}^{\text{MRF}}(l(x), l(y)) = 1 - \delta(l(x), l(y))$, with δ the Kronecker function. Another viewpoint in a variational framework extends the Mumford-Shah formulation of image segmentation [Mumford and Shah, 1989] to motion segmentation [Paragios and Deriche, 2005; Cremers and Soatto, 2005]. In addition to the data fitting potential (6.5) inside each region $\mathcal{R}_i$, a constraint restricting the length $\mathcal{L}(C)$ of the set of region boundaries C is imposed globally with the energy:

$$E(\mathbf{w}, C) = \sum_{i=1}^{N} \int_{\mathcal{R}_i} \rho_{data}(x, I_1, I_2, \mathbf{w}_{\boldsymbol{\theta}_i}) dx + \nu\, \mathcal{L}(C), \tag{3.9}$$

where Ω_D is the discrete image domain. The minimization is performed alternatively on the flow and the boundaries. Minimizing (3.9) with respect to C requires a differentiable approximation of the contour length $\mathcal{L}(C)$. It is common to implicitly represent the partitioning of the image with level sets, which allows to represent the interior of the regions by the sign of the function, as well as the total length of the boundaries by their level lines. One level set function can only represent two regions. For an arbitrary number of N regions, it is possible to define N corresponding levels sets, at the price of a high computational cost and a more complex energy to prevent vacuums in the partitioning, or other strategies can be employed, as the one of Chan et al. [2002] re-used in [Cremers and Soatto, 2005], for more sophisticated combinations between functions. The optimization is done alternatively between motion and regions. In [Paragios and Deriche, 2005], this

level set framework is augmented with an edge-driven tracking and background detection. A graph-cut optimization scheme has been proposed in [Dupont et al., 2005]. These two formulations ((3.8) and (3.9)) can be found in numerous other works [Kervrann et al., 2011; Memin and Perez, 2002; Unger et al., 2012; Sun et al., 2010a, 2012; Schnörr and Peckar, 1995].

Two main drawbacks affect this joint estimation and segmentation approach based on Mumford-Shah functionals. First, alternate minimization of regions and flow fields is computationally expensive. The related layered approach [Sun et al., 2010b, 2012] achieving state-of-the-art results requires several hours to process a pair of 640×480 pixels, and even GPU-based implementation [Unger et al., 2012] can need up to an hour.

Second, the evolution of the contours throughout the minimization procedure is very dependent on the initialization of the segmentation. Therefore, the best performing methods have to initialize the minimization of (3.9) with optical flow estimation methods that are able to reach among the best results in optical flow benchmarks. For example, [Sun et al., 2010b] is initialized with [Sun et al., 2010b], and [Unger et al., 2012] is initialized with Werlberger et al. [2009].

Some works exploit more complex flow field representations than polynomial approximation, by authorizing deformations from an initial affine model [Black and Jepson, 1996; Sun et al., 2010b; Memin and Perez, 2002] or an explicit regularized model [Brox et al., 2006; Amiaz and Kiryati, 2006]. Allowing such complex and discontinuous motion fields in segmented regions actually tends too allow for larger regions, and ultimately leads to *global* approaches detailed in Section 4, where motion discontinuities are handled by the model itself. Consequently regions may not represent coherent motion, but rather a delimitation adapted to the specific estimation method.

4 Explicit regularization

Alternatively to the parametric representation of spatial coherence, mostly adapted to smooth deformations, the global form of the motion field can be imposed by an explicit regularization potential. Motion discontinuities are then no more represented by the boundaries of the regions delimiting parametric motion fields, but they are involved the global model, often considered as outliers w.r.t. smoothness assumptions. The variational approach has been initially proposed by Horn and Schunck [1981] and is usually referred to as the *global* approach, since the regularization term interconnects all the pixels of the image and thus requires the optimization of the objective energy to be performed globally. In this section, we review current versions of the regularization model and optimization strategies.

4.1 General approach an principle

In its most general form, the energy minimized by globally regularized methods can be written as:

$$E_{\text{global}}(\mathbf{w}) = \int_{\Omega} \rho_{data}(x, I_1, I_2, \mathbf{w}) + \lambda\, \rho_{reg}(x, \mathbf{w}) dx \tag{4.1}$$

where $\rho_{data}(x, I_1, I_2, \mathbf{w})$ is the data potential, as discussed in Chapter 2, $\rho_{reg}(x, \mathbf{w})$ is the regularization potential encoding an *a priori* assumption on the the field $\mathbf{w}$, and λ is a parameter tuning the balance between the two terms. Broadly speaking, the regularization potential aims at smoothing the motion field in regions of coherent motion while preserving motion discontinuities at the boundaries of moving objects. Finding the trade-off can also be partially addressed in the adaptation of the balance parameter λ [Ng and Solo, 1998; Zimmer et al., 2011; Krajsek and Mester, 2007; Héas et al., 2012].

A major interest of the global variational framework is its versatility, allowing one to model different forms of flow fields by combining different data and regularization terms. One must nevertheless keep in mind that minimizing (4.1) is often a tricky task. The potentially unlimited combinations of data terms and regularization terms is restricted in practice to those compatible with efficient minimization. Besides, advances in optical flow have often been correlated with new possibilities offered by optimization techniques. For example, efficient Primal-Dual minimization for Total Variation regularization [Chambolle and Pock, 2011] have motivated a number of optical flow models [Zach et al., 2007;

Werlberger et al., 2010; Unger et al., 2012]. The development of efficient discrete optimization techniques based on graph cuts [Boykov et al., 2001] or message passing [Kolmogorov, 2006] also inspired various works [Lempitsky et al., 2010; Chen et al., 2013; Mozerov, 2013]. Another consequence of the close intricacy between energy model and optimization method is the difficulty to compare performances of different models, as global optimum is in general not guaranteed for sophisticated energies and the quality of the local optimum depends on the type of optimization method.

We will detail in Section 4.2 existing regularization models independently from optimization techniques, for the sake of clarity. Section 4.3 focuses on the dependency between specific energy models and optimization methods.

4.2 Regularization models

4.2.1 Spatial gradient constraint

The most natural and widely used way to impose smoothness of the motion field is to penalize the magnitude of the flow gradient:

$$\rho_{reg}(x, \mathbf{w}) = f(x, I)\, \phi(\|\nabla \mathbf{w}\|^2) \tag{4.2}$$

where $\psi(\cdot)$ is the penalty function and $f(x, I)$ is a weighting function.

A taxonomy of optical flow regularizers has been proposed in [Weickert and Schnorr, 2001]. The authors focus on convex and rotational invariance properties, and prove uniqueness of the solution in each case. For each regularization of type (4.2), they show the equivalence between the resolution of the Euler-Lagrange equations associated with energy (4.1) and diffusion filtering. In addition, a diffusion tensor is derived for each particular variation of (4.2). We will give a more succinct overview, taking only some elements from this classification and integrating more recent approaches.

Flow-driven regularization In flow-driven approaches, no relation between the form of the flow field and the structure of the image is assumed. The weighting function is thus $\forall x \in \Omega,\ f(x, I_1) = 1$. The seminal formulation of [Horn and Schunck, 1981] adopts a quadratic penalization function:

$$\rho_{reg}(x, \mathbf{w}) = u(x)^2 + v(x)^2, \tag{4.3}$$

with $\mathbf{w} = (u, v)^\top$. The quadratic penalization is unable to capture motion discontinuities. Robust sub-quadratic penalties have soon been employed to overcome the problem [Mémin and Pérez, 1998; Deriche et al., 1996; Black and Anandan, 1996]. Among the wide panel of robust functions, the popular parameter-free Total Variation (TV)

prior, has interesting and useful properties [Brox et al., 2004; Zach et al., 2007; Xu et al., 2012b]. Contrary to most other robust norms, the TV yields a convex constraint facilitating optimization. The non-differentiability in 0 is generally circumvented by using the regularized version $\phi(z) = \sqrt{z^2 + \epsilon^2}$, where ϵ is a small constant. Associated to proximal splitting minimization, TV involves solving several ROF (Rudin-Osher-Fatemi) models [Rudin et al., 1992], for which very efficient algorithms exist [Chambolle, 2004]. A series of accurate optical flow estimation methods have exploited this idea for fast and accurate minimization [Zach et al., 2007; Wedel et al., 2009b; Chambolle and Pock, 2011; Werlberger et al., 2012].

TV regularization actually favors piecewise constant flow fields. This framework is known to transform smoothly varying motion to a succession of small discontinuous constant steps (*staircasing* artifacts). This undesirable effect can be reduced by replacing the L_1 penalization by a quadratic one for small gradient magnitude, which is the behaviour of the Huber norm [Shulman and Herve, 1989; Werlberger et al., 2009]. Another possibility is to penalize higher order derivatives of the flow, as done in [Trobin et al., 2008a] for the second derivative, to favor piecewise affine flow fields. The Total Generalized Variation (TGV) [Bredies et al., 2010], generalizes L_1 penalization to arbitrarily high order derivatives. The performance gain of the second order TGV has been experimentally shown for smooth deformation conditions, for which the *staircasing* effect is prominent with TV regularization [Ranftl et al., 2012; Vogel et al., 2013; Braux-Zin et al., 2013].

Despite the demonstrated performance of TV due to its algorithmic attractiveness, the real distribution of optical flow derivatives has been shown to follow a more heavy-tailed and concave distribution [Roth and Black, 2007]. Finding good approximate solutions for non-convex priors has motivated a number of works and will be discussed in Section 4.3. When appropriate minimization strategy is available, like graph cuts or the recent work of [Ochs et al., 2013], this kind of penalties has proven to yield improvements compared to the TV model.

Integrate image gradient information It is natural to assume a link between the motion field and its source image I_1. As already stated in Section 3.3.3 about the relationship between motion and image segmentation, it is reasonable to consider that motion discontinuities coincide with image discontinuities delineating moving objects. This information can be incorporated in the regularization through the weighting function $g(x, I)$ taken as a smooth decreasing function of $\|\nabla I\|^2$ [Alvarez et al., 1999; Wedel et al., 2009b; Xu et al., 2012b; Ayvaci et al., 2012; Mozerov, 2013], often defined as

$$g(x, I) = e^{-\left(\frac{\|\nabla I(x)\|^2}{\varsigma^2}\right)}. \tag{4.4}$$

where ς is a parameter setting the influence of the image gradient on the regularization.

Despite the risk of over-segmentation, this simple weighting strategy usually improves significantly the results in practice.

The previous approach is isotropic since the smoothing is modulated by the same value in all the directions. This is suboptimal since we would ideally like to prevent the smoothing only *across* the boundaries, and allow it *along* them. This can be achieved by defining the regularization axes differently from the horizontal and vertical axes. The eigenvectors s_1 and s_2 of the structure tensor $R_\rho = K_\rho * [\nabla I \nabla I^\top]$ are well adapted since s_1 is oriented across local image edges and s_2 is orthogonal to s_1. This idea has been first exploited by Nagel and Enkelmann [1986], which regularizer has been rewritten in [Zimmer et al., 2011] as:

$$\rho_{reg}(x, \mathbf{w}) = \frac{1}{\|\nabla I\|^2 + 2\kappa^2} \left(\kappa^2 \left(u_{s_1}^2 + v_{s_1}^2 \right) + (\|\nabla I\|^2 + \kappa^2) \left(u_{s_2}^2 + v_{s_2}^2 \right) \right) \tag{4.5}$$

where the eigenvectors s_1 and s_2 are obtained with a radius $\rho = 0$ for the Gaussian filtering of the structure tensor R_ρ, u_{s_i} are the derivatives of u along the s_i axis and κ is a regularization parameter. When κ is small, the regularization is reduced in the direction of the image gradient s_1 and strengthened along image edges s_2 depending on the image gradient magnitude $\|\nabla I\|^2$.

The classical artifact produced by purely image-driven regularizations is an over-fitting of the flow field on image boundaries, creating artificial motion discontinuities. To reduce the impact of image gradient in the regularization, Sun et al. [2008] proposed to keep the s_1 and s_2 directions, while suppressing the weighting on $\|\nabla I\|^2$ in (4.5) and to employing robust penalization:

$$\rho_{reg}(x, \mathbf{w}) = \phi(u_{s_1}^2) + \phi(v_{s_1}^2) + \phi(u_{s_2}^2) + \phi(v_{s_2}^2). \tag{4.6}$$

Zimmer et al. [2011] proposed a generalized computation of the regularization axes, oriented to follow the data constraint rather than the image edges. In analogy with the previous approach defining s_1, s_2 from the structure tensor, they compute the eigenvectors of a so-called *regularisation tensor*, designed to be complementary with the data term. The approach of Zimmer et al. [2011] can be generalized to data potentials built from the combination of L linear constancy constraints, that is,

$$\rho_{data}(x, I_1, I_2, \mathbf{w}) = \sum_{\ell=1}^{L} \phi(\mathbf{A}_\ell \, \mathbf{w} + B_\ell). \tag{4.7}$$

For this kind of data potential, the regularisation tensor is defined by

$$\mathbf{R}_\rho = \sum_{\ell=1}^{L} k_\rho * \mathbf{A}_\ell \mathbf{A}_\ell^\top, \tag{4.8}$$

where k_ρ is a Gaussian kernel and s_1,s_2 are the eigenvectors of $\mathbf{R}_\rho$. Taking $L = 1$, $\mathbf{A}_1 = \nabla I$ and $B_1 = I_t$ yields the brightness constancy constraint, and the regularization tensor reduces to the structure tensor. For more elaborated data terms, as the combination of normalized brightness and gradient constancy used in [Zimmer et al., 2011], the resulting axes are more consistent with the data constraints.

4.2.2 Non-local regularization

The gradient of the flow can only provide a local constraint on the interaction between pixels. Assuming longer range interactions can model more precisely the form of the motion field. Such non-local regularization has been recently investigated in [Sun et al., 2010a; Werlberger et al., 2010; Krähenbühl and Koltun, 2012; Drulea and Nedevschi, 2013] by describing the structure of the flow in an extended neighborhood $\mathcal{N}(x)$ in a discrete setting as:

$$\rho_{reg}(x, \mathbf{w}) = \sum_{y \in \mathcal{N}(x)} g(x, y, I_1) \phi \left(\|\mathbf{w}(x) - \mathbf{w}(y)\|^2 \right). \tag{4.9}$$

The weights $g(x, y, I_1)$ indicate which pixel $y \in \mathcal{N}(x)$ should share a similar motion with pixel x. They are derived from the bilateral filter, favoring small distances in the spatial and color spaces [Yoon and Kweon, 2006]:

$$g(x, y, I_1) = \exp\left(-\frac{\|x - y\|^2}{\sigma_s^2} - \frac{\|I_1(x) - I_1(y)\|^2}{\sigma_c^2} \right) \tag{4.10}$$

where σ_s and σ_c control the influence of spatial distance and color similarity. This approach is image-driven in a similar way to local weighting (4.4), in the sense that the smoothness is weighted by the image edges. Nevertheless, the integration on a large neighborhood reduces the influence of local gradients and exhibits more globally the structure of the objects. It is implemented as an alternate weighted median filtering in [Sun et al., 2010a] and interpreted as a low-level soft segmentation in [Werlberger et al., 2012].

The high order regularization causes severe optimization difficulties discussed in Section 4.3, in particular in terms of computational cost, increasing with the size of the neighborhood $\mathcal{N}(x)$.

4.2.3 Temporal coherence

A natural idea is to extend the spatial regularization described above to the temporal dimension, assuming that motion varies smoothly across consecutive frames. Similarly with the spatial case, smoothness on the time axis can be achieved either locally, based on

the temporal gradient, or taking into account a longer interval by working on trajectories.

Constraint on the temporal gradient The most straightforward way to model temporal smoothness is to penalize the temporal flow gradient, analogously with the spatial flow gradient in Section 4.2.1. As for spatial dimension, constant or quadratic smoothness assumption [Murray and Buxton, 1987] is unrealistic since it cannot account for the temporal discontinuities frequently occurring when objects change direction. Robust temporal consistency is achieved in [Nagel, 1990; Chin et al., 1994; Weickert and Schnörr, 2001; Zimmer et al., 2011] by simple extensions of the spatial gradient penalties described in Section 4.2.1 to the temporal dimension. However, the performances of the local temporal regularization are deceiving in most cases.

Constraint on the trajectory In case of large displacements, the temporal gradient at a given pixel can be high even if the motion remains constant across frames. Temporal coherence is then more adequately modeled by constraining the trajectories of objects. It was done by [Black, 1994], who does not model explicitly the trajectories but estimates temporal changes on warped images to alleviate the problem of large displacements. However, the warping is done sequentially in the forward direction and is thus prone to propagate errors. In the same spirit, [Volz et al., 2011] considers the same coordinate system for groups of five frames, which implies an implicit natural registration. The estimation is done jointly in all frames of the sequence, which overcomes lack of feedback with previous frames of [Black, 1994]. The trajectory constraint of Garg et al. [2011] is explicitly imposed by modeling the flow field as linear combination of long-term trajectory bases, obtained from reliable sparse tracks.

4.3 Optimization

As mentioned in Section 4.1, the optimization strategy employed to minimize (4.1) has a decisive influence on the final result. We give an overview of the main continuous and discrete optimization methods and point out their adaptability to specific energy terms.

4.3.1 Continuous methods

Resolution of the Euler-Lagrange equations The Euler-Lagrange equations give necessary conditions for minimizing energy of the form

$$E(\mathbf{w}) = \int_{\Omega} F(x, \mathbf{w}, \nabla \mathbf{w})\, dx, \tag{4.11}$$

which is the case of (4.1) when the regularization is a function of the flow gradient (4.2). With simplified notations, they provide the following system of partial differential equations:

$$\begin{cases} \frac{\partial \rho_{data}}{\partial u} + \operatorname{div}\left(\frac{\partial \rho_{reg}}{\partial \nabla u}\right) = 0, \\ \frac{\partial \rho_{data}}{\partial v} + \operatorname{div}\left(\frac{\partial \rho_{reg}}{\partial \nabla v}\right) = 0, \end{cases} \tag{4.12}$$

which can be rewritten by introducing the diffusion tensor $\mathbf{D}$ accounting for their relation with diffusion equations [Weickert and Schnorr, 2001]:

$$\begin{cases} \frac{\partial \rho_{data}}{\partial u} + \operatorname{div}\left(\mathbf{D}\,\nabla u\right) = 0, \\ \frac{\partial \rho_{data}}{\partial v} + \operatorname{div}\left(\mathbf{D}\,\nabla v\right) = 0. \end{cases} \tag{4.13}$$

The analogy with diffusion equations makes explicit the direction and magnitude of the smoothing, which correspond respectively to the eigenvectors and eigenvalues of $\mathbf{D}$.

The discretization of the gradient and divergence operators yields a large system of equation to be solved. If the system is linear, its sparsity makes it well suited for iterative solvers like Gauss-Seidel or successive over-relaxation (SOR) [Brox et al., 2004]. These methods are guaranteed to converge for strictly diagonally dominant systems and also converge in practice in case of small deviations from diagonal dominance, which occurs in practice for most optical flow models. Nevertheless, the linear case is in practice only encountered in the model of Horn and Schunck [1981] using quadratic penalties and linearized data constraint. To cope with non-linearity, the typical approach [Weickert et al., 2001; Brox et al., 2004] is to resort to time-lagged schemes [Ciarlet, 1978] by handling each source of non-linearity by fixed point iterations, turning the problem into a succession of linear systems, and updating iteratively the non-linear parts. Convergence of the scheme is ensured if the linear systems are solved exactly, but approximations and frequent updates often yield in practice good results and much faster convergence.

Fast computational schemes have been employed for achieving near real-time performance. A multigrid framework has been proposed in [Bruhn and Weickert, 2006]. The scheme is very efficient, but it is problem-specific and requires a substantial implementation effort. Another approach is to consider the solution of the Euler-Lagrange equation as the steady state of the corresponding diffusion process (4.13), and use the Fast Explicit Diffusion (FED) principle [Grewenig et al., 2010] to accelerate convergence by adapting the time steps. Gwosdek et al. [2012] exploited the natural parallelization of explicit schemes to implement a quasi real-time version of the variational method of [Zimmer et al., 2011] on GPU, based on FED.

This approach has become standard because of its simplicity, the wide range of models it can handle, and its good experimental performances even for non-convex energies [Brox et al., 2004; Li et al., 2013; Volz et al., 2011].

Early discretization The computation of the Euler-Lagrange equations can be complex or impossible for some forms of the energy (4.1). Moreover, one can argue that the discretization of the Euler-Lagrange equations can generate numerical errors with respect to the original energy [Pock et al., 2007]. A way to overcome these shortcomings is to avoid computing the Euler-Lagrange equations and directly discretize the energy (4.1). The equation to solve is then simply the cancellation of the differentiated discretized energy:

$$\frac{\partial E(\mathbf{w})}{\partial \mathbf{w}} = 0. \tag{4.14}$$

Solving (4.14) amounts to inversing a large linear system, similar to the one obtained by Euler-Lagrange equation discretization for energies of the form (4.11). Employing a fixed-point iterations scheme to cope with non-linearity amounts to an IRLS approach, which is shown by Liu et al. [2009] to be equivalent to the above described resolution of the Euler-Lagrange equations with fixed-point.

Contrary to the Euler-Lagrange scheme, the regularization is not imposed to be a derivative of $\mathbf{w}$, and non-local regularization terms (4.9) can be handled. However, such an approach yields a dense linear system, not solvable with standard iterative methods. Krähenbühl and Koltun [2012] proposes a linear-time method to compute matrix product by successive Gaussian filtering, allowing to efficiently inverse the dense linear system with a conjugate gradient solver. Sun et al. [2008] also exploited the ability of handling more general energies to optimize learned data and regularization terms. The general experimental study of [Sun et al., 2010a] and the complex method of [Sun et al., 2010b] also exploits this approach, with Graduated Non Convexity (GNC) to cope with multimodality of the energy [Blake and Zisserman, 1987; Black and Anandan, 1996].

Instead of solving (4.14), a gradient descent method can be applied to minimize the discretized energy. Due to the large scale of the problem, Newton methods requiring the inversion of the Hessian of the energy are not applicable, and only quasi-Newton methods are computationally tractable. A few works have explored this direction, with Truncated Newton [Kalmoun et al., 2011; Kalmoun and Garrido, 2013] or L-BFGS [Dérian et al., 2011].

Half-quadratic minimization Instead of solving directly the energy (4.1) a number of optimization approaches proceed to the addition of an auxiliary variable splitting the original problem into easier sub-problems. The half-quadratic minimization falls in this class. It can be shown that under large conditions [Geman and Reynolds, 1992; Charbonnier et al., 1997], a function ϕ can be rewritten as the following function Φ

introducing the additional variable $\boldsymbol{\gamma} \in]0,1]$:

$$\phi(\mathbf{z}) = \inf_{\boldsymbol{\gamma}} \Phi(\mathbf{z}, \boldsymbol{\gamma}) = \inf_{\boldsymbol{\gamma}} \left(\boldsymbol{\gamma}\, \mathbf{z}^2 + \psi(\boldsymbol{\gamma}) \right), \tag{4.15}$$

where the function ψ can be explicitly derived from ϕ. The robust non-convex penalty $\phi(\mathbf{z})$ in data and regularization potentials can be replaced by $\Phi(\mathbf{z}, \boldsymbol{\gamma})$ in the data and regularization terms, so that the optimization in $\mathbf{z}$ becomes easy since it involves only quadratic terms. Moreover, the minimization w.r.t. $\boldsymbol{\gamma}$ has a closed-form solution:

$$\widehat{\boldsymbol{\gamma}} = \arg\min_{\boldsymbol{\gamma}} \Phi(\mathbf{z}, \boldsymbol{\gamma}) = \frac{\phi'(\mathbf{z})}{2\mathbf{z}}. \tag{4.16}$$

The original minimization problem in $\mathbf{z}$ is thus turned into alternate simple optimizations on $\mathbf{z}$ and $\boldsymbol{\gamma}$. This approach also leads to the IRLS algorithm described in Section 3.2.

The introduction of this approach for optical flow estimation coincided with the use of robust penalties for discontinuity preserving regularization [Deriche et al., 1996; Black et al., 1997; Mémin and Pérez, 1998; Aubert et al., 1999] and was more recently exploited in [Corpetti et al., 2002; Héas et al., 2012].

Proximal splitting Another successful optimization method based on alternate minimization of simple sub-problems has been proposed by Aujol et al. [2006] and used for optical flow in [Zach et al., 2007]. The data and regularization terms are splitted and associated to separate variables, which are quadratically coupled by a third term:

$$E_{\text{split}}(\mathbf{w}, \mathbf{v}) = \underbrace{\int_\Omega \rho_{data}(x, I_1, I_2, \mathbf{w}) dx}_{E_{data}} + \frac{1}{2\varepsilon} \int_\Omega \|\mathbf{w}(x) - \mathbf{v}(x)\|^2 dx + \lambda \underbrace{\int_\Omega \rho_{reg}(x, \mathbf{v}) dx}_{E_{reg}}, \tag{4.17}$$

where $\mathbf{v}$ is an auxiliary variable. The parameter ε sets the intensity of the coupling. If ε is small, (4.17) tends to the original univariate problem (4.1). The minimization on each variable is the computation of the proximal operator of E_{data} and E_{reg}:

$$\begin{cases} \arg\min_{\mathbf{w}} E_{\text{split}}(\mathbf{w}, \widehat{\mathbf{v}}) = \text{prox}_{E_{data}}(\widehat{\mathbf{v}}) \\ \arg\min_{\mathbf{b}} E_{\text{split}}(\widehat{\mathbf{w}}, \mathbf{v}) = \text{prox}_{E_{reg}}(\widehat{\mathbf{w}}) \end{cases} \tag{4.18}$$

where the proximity operator of a function f is defined by

$$\text{prox}_f(\mathbf{z}) = \arg\min_{\mathbf{u}} \left(f(\mathbf{u}) + \frac{1}{2}\|\mathbf{u} - \mathbf{z}\|^2 \right). \tag{4.19}$$

The minimization problems (4.18) can be viewed as alternating coarse pixel-wise data matching and denoising of the flow field. A generalization of this approach and its

formalization in a primal-dual framework is described in [Chambolle and Pock, 2011].

This approach has initially been designed for L_1 penalty on the data and regularization terms, a.k.a. TV-L_1 model [Aujol et al., 2006; Zach et al., 2007], for the simplicity of the resulting optimization sub-problems. The optimization of the data term with $\mathbf{v}$ fixed is efficiently performed by a thresholding scheme, and fixing $\mathbf{w}$ yields the ROF model, optimized with the duality based algorithm of [Chambolle, 2004]. An advantage is that a differentiable approximation of the L_1 norm is not required, as in [Brox et al., 2004], and one can solve for the exact TV-L_1 model.

In a general point of view, the independence of the optimization of data and regularization parts allows one to design dedicated minimization schemes in a variety of cases. The restriction is to be able to compute the proximal operators, and the convergence is ensured only for convex energies. For the data part, the thresholding scheme of [Zach et al., 2007] is applicable in some cases for the L_1 norm, and efficient solutions have been found to handle an additional fundamental matrix constraint [Wedel et al., 2008], a truncated L_1 norm of normalized cross correlation [Werlberger et al., 2010], or mutual information [Panin, 2012]. Another advantage of the pixel-wise nature of the data part minimization is the naturally parallel implementation strategies which can dramatically speed up the algorithm and reach real-time [Zach et al., 2007]. Based on the pixel-wise observation, [Steinbrucker et al., 2009] proposes a discrete exhaustive matching for optimization in $\mathbf{w}$, which opens the usage of arbitrary data terms, only limited by the computational cost of the matching. Patch-based similarity measures have also recently been implemented with the fast PatchMatch algorithm [Barnes et al., 2009] by [Heise et al., 2013].

Concerning the regularization part, it is possible to minimize non-local regularization terms [Werlberger et al., 2012; Drulea and Nedevschi, 2013]. The decoupling also allows for a fair comparison of data or regularization models due to the absence of interference between the minimization of the two parts [Vogel et al., 2013].

4.3.2 Discrete methods

In a discrete setting, the solution of the minimization of (4.1) is searched in a set of discrete labels $\mathcal{L}(x)$ corresponding to a quantization of the continuous motion vector space or the selection of a finite subset of motion vectors. We give a fast overview of basic principles of discrete optimization methods. For a more complete analysis, see the recent review of [Wang et al., 2013]. The spatial discretization of (4.1) yields an energy in the Markov Random Field (MRF) framework:

$$E_D(\mathbf{w}) = \sum_{x \in \Omega_D} \rho_{data}(x, I_1, I_2, \mathbf{w}) + \sum_{y \in \mathcal{N}(x)} \rho_{MRF}(x, y, \mathbf{w}) \tag{4.20}$$

where Ω_D is the discrete image domain, $\mathcal{N}(x)$ denotes the pixels interacting with x in the model and ρ_{MRF} is the discrete version of $\rho_{reg}(x, y, I_1, I_2, \mathbf{w})$ which explicitly takes into account the interaction between two neighboring pixels x and y.

An important advantage of discrete optimization over the continuous approach is that it does not require differentiation of the energy and can thus handle a wider variety of data and regularization terms. On the counterpart, a trade-off has to be found between the accuracy of the motion labelling and the size of the search space $\mathcal{L}(x)$. Indeed, discrete optimization methods are usually severely limited in terms of accuracy and speed by the number of labels, particularly for optical flow where subpixel accuracy is necessary, and where the two-dimensional motion space becomes more rapidly intractable than the one-dimensional stereo case for instance. Therefore, the design of the label space $\mathcal{L}(x)$ is a crucial component of discrete optimization methods for optical flow.

Among early methods for minimizing (4.20), simulated annealing [Geman and Geman, 1984] offers proved convergence towards the global minimum based on stochastic relaxation, which can be viewed as the stochastic counterpart of the GNC scheme discussed above in Section 4.3.1. However, the optimal convergence is guaranteed only under prohibitive computational cost. An approximate solution can be obtained much faster with the Iterated Conditional Modes [Besag, 1986], operating by iterative local minimizations of the energy, but this local optimum often yields poor results compared to modern methods [Szeliski et al., 2008], especially for non-convex functions.

Graph cut The work of [Boykov et al., 1998] gave rise to rapidly growing research interest and advances on graph cut approaches for MRF minimization. The basis of graph cuts is the max-flow/min-cut problem consisting in finding the optimal path between to nodes in a directed graph, solvable by many algorithm in polynomial time [Fulkerson, 1962; Goldberg and Tarjan, 1988]. It is possible to model an undirected MRF structure as a directed graph by introducing two additional *source* and *sink* nodes, and then interpret the min-cut partition as a binary label segmentation of the MRF energy. The global minimum of the binary optimization can be guaranteed for pairwise interactions and submodular functions. In summary, it is possible to find the global optimum for the energy (4.20) under the following conditions:

- submodularity of $\rho_{MRF}(x, y, I_1, I_2, \mathbf{w})$,
- pairwise interactions,
- binary labels.

Research on graph cuts has attempted to overcome these three constraints, based on the original max-flow/min-cut algorithm.

Submodularity of the pairwise term is a required property for convergence of the algorithm. Finding good approximate solution can be achieved with the Quadratic

Pseudo Boolean Optimization (QPBO) [Boros et al., 1991; Kolmogorov and Rother, 2007; Rother et al., 2007], leading to an optimal but partial labelling. Dealing with higher-order local interactions between pixels has also been addressed in several recent works [Kohli et al., 2009; Ishikawa, 2009; Komodakis and Paragios, 2009; Fix et al., 2011].

Submodularity and pairwise interactions are not too restrictive constraints for a large range of computer vision problems and let room for a number of applications. On the contrary, the cardinality of the label space verifies almost always $|\mathcal{L}(x)| > 2$, so the binary label requirement is a much harder limitation. The extension of binary graph cuts to multi-label have mostly been realized through iterative *move-making* techniques. The idea is to create at each iteration a binary-labeled space composed of the current solution and a new proposal label. The label space $\mathcal{L}(x)$ is thus explored progressively by each new proposal. If each binary minimization performed with the max-flow/min-cut method is ensured to be optimal, a decreasing of the energy is guaranteed at each iteration, and the method converges to a local minimum. It can be noticed that the class of *move-making* methods is not restricted to graph cuts and the *moves* can be achieved by other techniques such as the variational approach of [Trobin et al., 2008b], optimizing in the continuous space by relaxing the binary variable.

The elements of a *move-making* method are the *move-space*, specifying the set of new labelling proposals at each iteration, and the way the *moves* themselves are performed. The most popular *move-space* is the *expansion-move*, where the proposal labelling is defined as a constant label map. Another common alternative is the $\alpha\beta$-swap move based on a label exchange at pixels having labels α or β. The *range-move* [Veksler, 2007] extends binary proposal choice to a larger range of labels, in the case of ordered label spaces. Pre-computed labellings computed with independent and possibly continuous methods, can also serve as proposals [Lempitsky et al., 2010].

Additionally, computational efficiency of graph cut approaches have been addressed by several works. The most representative ones are [Komodakis et al., 2008], operating in a Primal-Dual framework and speeding up convergence by minimizing the Primal-Dual gap, and on the other hand, [Kohli and Torr, 2007] working on dynamic MRF and exploiting previous iterations as initializations. For more details about existing *move-spaces*, see [Veksler, 1999; Wang et al., 2013]. Applications for optical flow have increased fastly in recent years [Lempitsky et al., 2010; Chen et al., 2013; Cooke, 2008; Glocker et al., 2008, 2010; Sun et al., 2012; Xu et al., 2012b; Li et al., 2013].

Message passing Belief propagation is based on the *max-product* algorithm, which is able to find the MAP of a probability distribution expressed as a product of factors, and thus representable in a factor graph (see [Kschischang et al., 2001] for a detailed introduction). Taking the negative logarithmic version of such distribution amounts to work on MRFs of the form (4.20), which can motivate to rename the algorithm *min-sum*

in this case. In a nutshell, the minimization is done by iteratively updating local *messages* reflecting influence of local label configurations on the energy. After convergence of the *messages*, they can be used to define the probability of assigning a given label to a node, and the label with the maximum probability is chosen. The *max-product* has originally been designed for tree structures and is guaranteed to find the global optimum in this case. Nevertheless, it can also be used for MRFs exhibiting cycles (it is referred to as loopy belief propagation in this case [Pearl, 1988]), without convergence guarantees but showing good experimental results in a large number computer vision problems [Szeliski and Shum, 1996; Kappes et al., 2013]. The Tree Reweighted message passing approach [Wainwright et al., 2005] deals with similar concepts, but with a particular message passing strategy based on tree representations. The sequential approach of [Kolmogorov, 2006] has proven to yield good results and computational performance compared to other discrete methods in [Szeliski et al., 2008]. It has been applied for optical flow estimation in [Mozerov, 2013; Lee et al., 2010; Grauer-Gray and Kambhamettu, 2009].

5 Combining feature matching and optical flow

An increasingly addressed challenge for optical flow estimation is to handle very large displacements and deformations, as it is reflected in the recent MPI Sintel benchmarks [Butler et al., 2012], where is it not rare to find displacements of more than 100 pixels. As a consequence, the gap between feature matching and optical flow tends to vanish, and several methods have tried to combine density and accuracy of optical flow with the ability to capture large displacements of feature matching.

Finding correspondences by matching image features can be considered as a *local* parametric approach since a given neighborhood of the pixel to match is assumed to translate or undergo another parametric transformation towards its correspondence location. The similarity measures can be patch-based distances (see Section 2.1.2), or more complex and sparse feature descriptions often based on histogram of oriented gradients [Dalal and Triggs, 2005; Lowe, 2004; Bay et al., 2006; Yu and Morel, 2009], or segment matching [Wang et al., 2009]. The fundamental difference lies in the optimization process, since the parametric formulation has a differential optimization process imposing linearization, whereas feature matching explores a discrete space of admissible correspondences. Although being integer displacements and prone to errors, feature matching can thus handle large displacements without coarse-to-fine-schemes, with arbitrarily complex similarity measures. Regarding their complementarity of advantages, the combination of feature matching with regularized approaches is therefore of upmost interest.

We consider three approaches to obtain dense and accurate motion fields with feature matching: local filtering of correspondences, integration in a variational framework, and generation of coarse initialization for variational refinement.

5.1 Correspondence filtering

Pure feature matching has long been considered unable to produce dense flow fields with competitive accuracy with the previously described *local* and *global* approaches. There are three reasons for this:

1. the optimization of the similarity measure is often performed with exhaustive search,

which induces prohibitive computational cost,

2. repetitive textures or uninformative regions are sources of ambiguities for the matching process and generate large errors,

3. the correspondence process usually limits the accuracy to integer displacements, contrary to *local* and *global* approaches working in the continuous $\mathbb{R}^2$ space.

Several recent advances attempted to overcome these three seemingly inherent limitations.

Research on speeding up block matching include multi-scale search strategies [Tzovaras et al., 1994], integral images [Facciolo et al., 2013] or search in trees [Kumar et al., 2008], but the recent most spectacular contribution was achieved in [Barnes et al., 2009, 2010] with the PatchMatch algorithm. The method scans the image in the lexicographic and inverse order, and alternates two simple steps at each pixel: the *propagation* step minimizes locally the data cost in the space composed by the current pixel and its two predecessors in the scanning order, the second step proposes a small set of new candidates randomly chosen in the neighborhood of the current correspondence. It is easy to extend the matching to more complex transformations than translation, like rotation or scale factor [Barnes et al., 2010] by increasing the degree of the search space. The method was originally designed for image editing and was applied to several other applications [Barnes et al., 2010], with impressive results regarding the low computation time. For motion estimation, the interesting property is that the computational cost is not affected by the spatial extension of the search space, so that no trade-off has to be found between speed and displacement range.

The problem of matching ambiguities comes from the lack of discriminative power of the data cost. Without resorting to explicit regularization (Chapter 4), the coherency induced by simple local filtering of patch correspondences [Hosni et al., 2013] has been shown to be sufficient in practice to reduce most ambiguities and provide excellent dense results. The filtering is achieved in [Hosni et al., 2013] by guided filtering [He et al., 2010] and in [Ma et al., 2013] by weighted median.

Subpixel accuracy is usually reached by upscaling image resolution. The induced computational cost is reduced in [Hosni et al., 2013; Steinbrucker et al., 2009] by GPU implementation, and can also be handled by iterative refinement [Lee et al., 2010].

The combination of the three ingredients has led to the development of competitive optical flow estimation methods based on pure feature matching locally filtered [Hosni et al., 2013; Ma et al., 2013; Tao et al., 2012]

5.2 Feature matching in global regularized model

As discussed in Section 1.3, a major limitation of the global variational framework is the loss of small and fast objects due to the use of a coarse-to-fine estimation scheme. One recently investigated approach to overcome this problem is to integrate an information from feature matching into the variational framework, thus combining advantages of both methods.

Brox and Malik [2011], inspired by [Héas et al., 2007], made the first step in this way by adding to the classical data potential a new constraint taking into account an off-line computation of sparse feature correspondences. Let us denote $\mathbf{w}_c$ the displacement field obtained with a possibly sparse feature matching process. The new combined data potential is then

$$\rho_{data}^{Extended}(x, I_1, I_2, \mathbf{w}, \mathbf{w}_c) = \rho_{data}(x, I_1, I_2, \mathbf{w}) + \beta \rho_{data}^{Corresp}(x, \mathbf{w}, \mathbf{w}_c) \tag{5.1}$$

where β is a trade-off parameter and the matching potential is defined as:

$$\rho_{data}^{Corresp}(x, \mathbf{w}, \mathbf{w}_c) = \delta(x, \mathbf{w}_c) c(x, \mathbf{w}_c) \phi(\|\mathbf{w}(x) - \mathbf{w}_c(x)\|^2). \tag{5.2}$$

The binary function $\delta(x, \mathbf{w}_c)$ returns 1 if $\mathbf{w}_c$ is defined at x and 0 otherwise, and the weights $c(x, \mathbf{w}_c)$ correspond to the matching cost. The matching term (5.2) imposes the motion field to be close to the motion vectors obtained by feature matching.

The advantage of this approach is that the term $\rho_{data}^{Corresp}(x, \mathbf{w}, \mathbf{w}_c)$ is both differentiable and valid for large displacements. Problems related to the use of coarse-to-fine schemes are thus avoided. The main drawback is that the importance of the matching term $\rho_{data}^{Corresp}(x, \mathbf{w}_c, \mathbf{w})$ relatively to the data term $\rho_{data}(x, I_1, I_2, \mathbf{w})$ is mostly determined by the value of the matching cost $c(x, \mathbf{w}_c)$. Consequently the final variational estimation is extremely dependent on the reliability of the confidence measure, which is very difficult to guarantee, as emphasized in [Braux-Zin et al., 2013]. Matching errors are thus easily driven by $c(x, \mathbf{w}_c)$ and are likely to have a high impact on the final result. Reducing the impact of local feature matching errors by the regularization and the robust penalization $\phi()$ is insufficient in practice in a lot of cases.

Recently, [Weinzaepfel et al., 2013] based their method on [Brox and Malik, 2011] by taking the model (5.2) and improving the feature matching stage. They demonstrated a significant performance improvement due to the increased reliability of the matching. [Braux-Zin et al., 2013] also built their method upon [Brox and Malik, 2011] and modified the matching component by using segment matching [Wang et al., 2009]. The matching term is generalized to handle weakly localized line features. It is done through a point-to-line distance in addition to the point to point distance of (5.2), combined with a confidence measure for segment matching based on the assumption of a linear

mapping between segment matches. Moreover, considering the unreliability of confidence measure, Braux-Zin et al. [2013] gets rid of the matching cost $c(x, \mathbf{w}_c)$ and relies only on the regularization for robustness to matching errors. To sum up, the tendency is to take into account the great dependency of the estimation on the quality of the feature matching, and thus to concentrate most efforts on the design of robust matching algorithms.

5.3 Coarse initialization

Another recently investigated approach taking advantage of feature matching consists in exploiting the integer displacements and possibly sparse result to provide a coarse but relevant initialization for a variational refinement [Xu et al., 2012b; Mozerov, 2013; Chen et al., 2013]. These methods are composed of three steps:

1. a feature matching step provides a limited number of candidates at each pixel;
2. these candidates serve as labels for the discrete optimization (see Section 4.3) of a global regularized model;
3. the resulting coarse optical flow field is refined with one of the variational optimization techniques described in Section 4.3.

The idea is that steps 1 and 2 handle indifferently small and large displacements of objects at any scale, and the initialization is assumed to be good enough to avoid the coarse-to-fine scheme in step 3. The specificity of [Xu et al., 2012b] is that the three steps are repeated at each level of a multiresolution pyramid. Since steps 2 and 3 have already been discussed in previous sections, we focus on the specificity of step 1, that is, the production of candidates from feature matching.

In [Xu et al., 2012b], the feature matching in only performed at a restricted number of keypoints. After pruning of similar vectors, only a few displacement vectors are retained. Each of these vectors is expanded to produce a global constant flow field, used as candidates for step 2.

Mozerov [2013] considers a integer discretization of the two-dimensional motion field. Based on the observation that correlation-based patch matching is able to reproduce coarsely the motion distribution pattern of ground truth motion fields, the discretized motion space is delineated by the matching vectors.

The approach of [Chen et al., 2013] is close to [Mozerov, 2013] since the idea is to rely on the dominant motion patterns of dense patch matching. In [Chen et al., 2013], the candidates to feed discrete optimization are obtained by explicitly clustering PatchMatch motion fields [Barnes et al., 2010] to keep only dominant patterns.

6 Adaptive filtering of data term

The previous analysis of optical flow literature methods exhibited local and global approaches as the two main classes of optical flow methods. We present in this chapter a preliminary study conceived as a first attempt to combine these two classes. We base our contribution on the work of [Bruhn et al., 2005], integrating the constant flow assumption of local methods in a global regularized model through Gaussian filtering of the data term. We propose a spatial adaptation of this filtering to prevent its induced over-smoothing effect.

6.1 Combined Local-Global method of [Bruhn et al., 2005]

The idea of combining advantages of local and global approaches in a single model has been investigated by Bruhn et al. [2005]. A similar approach, generalized to other image processing problems can be found in [Tschumperlé and Brun, 2011]. The so-called "Combined Local-Global" method (CLG) was motivated by the robustness of local methods to the presence of noise in input images. It is made possible by the local filtering of the data term, implying a neighborhood-wise data constancy constraint, less sensistive to noise than pixel-wise measures. In contrast, global methods tackle noise by increasing the regularization term, which also tends to over-smooth the motion field. Indeed, the role of motion regularization is to model the *a priori* on the motion field, regardless of the nature of data. Noise in the data should therefore be taken into account in the data term.

As already discussed in Chapter 2, the usual way to cope with noise in the optical flow estimation process is to apply a pre-filtering operation to the input images. Simple Gaussian filtering is the standard choice. We experimented that more advanced edge-preserving image smoothing globally has less beneficial impact on motion estimation than Gaussian smoothing. This is because fine texture information is often smoothed out by denoising filters, and the resulting homogeneous regions are then not informative enough for optical flow estimation. However, Gaussian smoothing also blurs image discontinuities and thus affects the recovery of discontinuities of the motion field, as analyzed in [Bruhn et al., 2005].

The idea of [Bruhn et al., 2005] is then to more deeply account for noise in the data term of the global energy, by considering no more pixel-wise data constraint, but patch-based

constraint, assuming locally constant motion. Let us consider the usual global energy

$$E_{\text{global}}(\mathbf{w}) = \int_{\Omega} \rho_{data}(x, I_1, I_2, \mathbf{w}) + \lambda\, \rho_{reg}(x, \mathbf{w}) dx. \tag{6.1}$$

The pixel-wise data fidelity potential obtained from the brightness constancy constraint equation (1.5) with quadratic penalization is given by

$$\rho^0_{data}(x, I_1, I_2, \mathbf{w}) = \left(I_t + I_{x_1} u + I_{x_2} v\right)^2, \tag{6.2}$$

where x_1 and x_2 are the image axis, $\mathbf{w} = (u, v)^\top$, $I_t = \partial I/\partial t$, $I_{x_1} = \partial I/\partial x_1$ and $I_{x_2} = \partial I/\partial x_2$. It is modified in [Bruhn et al., 2005] by Gaussian filtering:

$$\rho^\sigma_{data}(x, I_1, I_2, \mathbf{w}) = k_\sigma * \left(I_t + I_{x_1} u + I_{x_2} v\right)^2, \tag{6.3}$$

where k_σ is a Gaussian filter of standard deviation σ and $*$ is the convolution operator. In (6.3), the Gaussian filtering is applied only on the image variables I_t, I_{x_1}, I_{x_2} and not on $\mathbf{w}$ which is assumed to be locally constant. It can be written in a tensor-based representation taking into account the constancy of $\mathbf{w}$:

$$\rho^\sigma_{data}(x, I_1, I_2, \mathbf{w}) \;=\; \boldsymbol{\alpha}_{\mathbf{w}}^\top \mathbf{J}_\sigma \boldsymbol{\alpha}_{\mathbf{w}} \tag{6.4}$$

with $\boldsymbol{\alpha}_{\mathbf{w}} = \begin{pmatrix} u \\ v \\ 1 \end{pmatrix}$, $\mathbf{J}_\sigma = k_\sigma(x) * \begin{pmatrix} I_{x_1}^2 & I_{x_1} I_{x_2} & I_{x_1} I_t \\ I_{x_1} I_{x_2} & I_{x_2}^2 & I_{x_2} I_t \\ I_{x_1} I_t & I_{x_2} I_t & I_t^2 \end{pmatrix}$.

Besides, the energy minimized by local methods (see Chapter 3), is of the following general form:

$$E_{local}(\boldsymbol{\theta}) = \int_{x \in \mathcal{R}} g(x) \left(I_t + I_{x_1} u_{\boldsymbol{\theta}} + I_{x_2} v_{\boldsymbol{\theta}}\right)^2 dx, \tag{6.5}$$

where $\boldsymbol{\theta}$ is the parameter vector for the motion model over region $\mathcal{R}$. It is equivalent to the data term (6.3) if $g(x)$ is a Gaussian kernel. The global minimization of (6.1) with data term (6.3) can thus be interpreted as an integration of a local model in a global regularized framework, in order to take advantage of robustness of local approaches to noise.

Experiments in [Bruhn et al., 2005] demonstrate the increased robustness to noise of the results obtained by Gaussian filtering of the data term. The method has been applied to several applicative domains where acquisition conditions induce noise in the image [Dawood et al., 2008; Delpiano et al., 2012].

Nevertheless, isotropic Gaussian smoothing also tends to over-smooth the resulting flow field. For large signal-to-noise ratio, the motion field estimated with the Gaussian filtered data potential (6.3) is usually less accurate than the one obtained with the pixel-wise

potential (6.2), as experimented in [Zimmer et al., 2011]. One can thus be interested in improving the behaviour at motion discontinuities while keeping noise robustness.

It should also be noted that in addition to be more robust to noise, local filtering on the data term can also provide a richer description of the local structure of the image. Consequently, motion estimation could also be improved in the absence of noise if relevant local image information is integrated in the data term.

6.2 Adaptive filtering of data term

Over-smoothing occurs with the method [Bruhn et al., 2005], as for local methods, when the support of the local filtering contains multiple motions, that is, at motion discontinuities. The aim is then to restrict the spatial support to coherently moving regions by replacing Gaussian filtering with fixed standard deviation by an adaptive filtering. A few works, already mentionned in Section 2.1.2, have been done in that direction. Drulea and Nedevschi [2011] replaces Gaussian filtering by bilateral filtering, and Rashwan et al. [2013] exploit tensor voting. These latter approaches rely on image measurements to specify filters. Similarly to image-based regularization, they are still not robust enough to noise and produce over-segmentation of the motion field.

In this chapter, we propose a spatially adaptive filtering approach estimating the filter parameters jointly with motion. Considering the simplest case of Gaussian smoothing, the parameter to optimize jointly with $\mathbf{w}$ is the standard deviation of the Gaussian filter, now defined as a dense field $\sigma : \Omega \to \mathbb{R}^+$. The optimization problem is then:

$$(\widehat{\mathbf{w}}, \widehat{\sigma}) = \arg\min_{\mathbf{w},\sigma} E(\mathbf{w}, \sigma) \tag{6.6}$$

where we define $E(\mathbf{w}, \sigma)$, the energy coupling $\mathbf{w}$ and σ, as follows:

$$E(\mathbf{w}, \sigma) = \int_\Omega \phi_d \underbrace{\left(\boldsymbol{\alpha}_\mathbf{w}^\top \mathbf{J}_{\sigma(x)}\, \boldsymbol{\alpha}_\mathbf{w}\right)}_{\rho^\sigma_{data}(x, I_1, I_2, \mathbf{w})\ (6.3)} dx + \lambda \int_\Omega \phi_r(\|\nabla \mathbf{w}(x)\|^2)\, dx + \beta \int_\Omega \phi_r(|\nabla \sigma(x)|^2)\, dx \tag{6.7}$$

where λ and β are parameters that balance the contributions of the data and regularization terms, and ϕ_d, ϕ_r are penalization functions. The first term is identical to the data potential (6.3), and the two other terms are regularizations on $\mathbf{w}$ and σ. For the sake of simplicity, we consider a unique penalty function for the two regularization terms, but they could be different. The minimization of (6.9) w.r.t. $\mathbf{w}$ amounts to the CLG method [Bruhn et al., 2005]. Minimizing also w.r.t. σ adapts spatially the standard deviation of the convolution $\sigma(x)$ at each point x. The aim is to reduce σ at motion discontinuities, where Gaussian smoothing tends to blur the estimated motion field. If a

discontinuity is contained in the Gaussian support defined by σ, the locally constant motion assumption will be violated and lower values σ will be favoured. Rather than being adapted to the image content as in [Drulea and Nedevschi, 2011; Rashwan et al., 2013], σ is guided now by the data term. Its variations follow motion discontinuities rather than image discontinuities. The smoothness assumption on σ is also relevant since it is expected to vary linearly with the distance from a motion discontinuity. We considered a regularized L_1 penalty of the gradient of σ in our experiments, but a quadratic penalty could also be used under the assumption that the variation of σ is smooth in the whole image. We mention that a similar approach has been developed for image denoising in [Azzabou et al., 2007], where a spatially variable bandwidth of an integration kernel is also estimated jointly with the recovered image. The filtering was applied to the regularization term.

Optimization The optimization is performed alternatively on $\mathbf{w}$ and σ. In the following, we consider regularized L_1 norm in the smoothness terms, $\phi_r(z^2) = \sqrt{z^2 + \epsilon^2}$. Minimization w.r.t. $\mathbf{w}$ with σ fixed amounts to a classical optimization problem in optical flow. We derive the associated non-linear Euler-Lagrange equation and we consider a fixed point scheme, as detailed in [Brox, 2005].

The energy to minimize w.r.t. σ given an estimate of $\mathbf{w}$ is of the following form:

$$J(\sigma) = \int_\Omega \phi_d \left(\boldsymbol{\alpha}_\mathbf{w}^\top \mathbf{J}_{\sigma(x)} \boldsymbol{\alpha}_\mathbf{w} \right) dx + \beta \int_\Omega \phi_r(|\nabla \sigma(x)|^2)\, dx \tag{6.8}$$

We adopt a gradient-based minimization approach using the quasi-Newton method L-BFGS [Nocedal, 1980], approximating the Hessian for faster computation. The approximation requires the computation of the first derivative of the energy $dJ(\sigma)/d\sigma$. In what follows, we detail the analytical computation of the derivative in the case of a quadratic penalization $\phi_d(z^2) = z^2$. We work on a spatially discretized version of $J(\sigma)$:

$$J(\sigma) = \sum_{x \in \Omega} \phi_d \left(\boldsymbol{\alpha}_\mathbf{w}^\top \mathbf{J}_{\sigma(x)} \boldsymbol{\alpha}_\mathbf{w} \right) + \sum_{z \in \mathcal{N}(x)} \phi_r((\sigma(x) - \sigma(z))^2) \tag{6.9}$$

Where $\mathcal{N}(x)$ is the set of four neighbours of x. We first explicitly develop the convolution operation. As explained before on the equivalence between (6.3) and (6.4), we can write:

$$\begin{aligned} J(\sigma) &= \sum_{x \in \Omega} k_{\sigma(x)} * (I_t + I_{x_1} u(x) + I_{x_2} v(x))^2 + \beta \sum_{z \in \mathcal{N}(x)} \phi_r((\sigma(x) - \sigma(z))^2) \\ &= \sum_{x \in \Omega} \left(\sum_{y \in \Omega} k_{\sigma(x)}(x, y) \, (I_t + I_{x_1} u(y) + I_{x_2} v(y))^2 \right) + \beta \sum_{z \in \mathcal{N}(x)} \phi_r((\sigma(x) - \sigma(z))^2) \end{aligned}$$

Then, differentiation gives:

$$\frac{\partial J(\sigma)}{\partial \sigma(x)} = \left(\sum_{y \in \Omega} \frac{\partial k_{\sigma(x)}(x,y)}{\partial \sigma(x)} \left(I_t + I_{x_1} u(x) + I_{x_2} v(x)\right)^2 \right) \\ -2\,\beta \sum_{z \in \mathcal{N}(x)} (\sigma(x) - \sigma(z)) \phi_r^{'}((\sigma(x) - \sigma(z))^2). \tag{6.10}$$

The differentiation of weights are obtained from definition:

$$k_{\sigma(x)}(x,y) = \frac{1}{\sqrt{2\pi}\sigma(x)} e^{-\frac{\|x-y\|^2}{2\sigma^2(x)}},$$

leading to

$$\frac{\partial k_{\sigma(x)}(x,y)}{\partial \sigma(x)} = \frac{k_{\sigma(x)}(x,y)}{\sigma(x)} \left(\frac{\|x-y\|^2}{\sigma^2(x)} - 1 \right). \tag{6.11}$$

By substituting (6.11) in (6.10), we have:

$$\frac{\partial J(\sigma)}{\partial \sigma(x)} = \sum_{y \in \Omega} \frac{k_{\sigma(x)}(x,y)}{\sigma(x)} \left(\frac{\|x-y\|^2}{\sigma^2(x)} - 1 \right) \left(I_t + I_{x_1} u(y) + I_{x_2} v(y)\right)^2 \\ -2\,\beta \sum_{z \in \mathcal{N}(x)} (\sigma(x) - \sigma(z)) \phi_r^{'}((\sigma(x) - \sigma(z))^2).$$

We introduce the filter $h_{\sigma(x)}$ defined by:

$$h_{\sigma(x)} * I = \sum_{y \in \Omega} \frac{k_{\sigma(x)}(x,y)}{\sigma(x)} \left(\frac{\|x-y\|^2}{\sigma^2(x)} - 1 \right) I(y).$$

The derivative of $J(\sigma)$ w.r.t. σ can then be rewritten in a simpler form, with an analogy between $h_{\sigma(x)}$ and the Gaussian filter $k_{\sigma(x)}$ of the original data term (6.3):

$$\boxed{\frac{\partial J(\sigma)}{\partial \sigma(x)} = h_{\sigma(x)} * \left(I_t + I_{x_1} u(y) + I_{x_2} v(y)\right)^2 - 2\,\beta \sum_{z \in \mathcal{N}(x)} (\sigma(x) - \sigma(z)) \phi_r^{'}((\sigma(x) - \sigma(z))^2)}$$

Following the same steps, we can write its expression for an arbitrary ϕ:

$$\boxed{\frac{\partial J(\sigma)}{\partial \sigma(x)} = h_{\sigma(x)} * \left(I_t + I_{x_1} u(y) + I_{x_2} v(y)\right)^2 \phi' \left(k_{\sigma(x)} * \left(I_t + I_{x_1} u(y) + I_{x_2} v(y)\right)^2 \right) \\ -2\,\beta \sum_{z \in \mathcal{N}(x)} (\sigma(x) - \sigma(z)) \phi_r^{'}((\sigma(x) - \sigma(z))^2)}$$

Extended model As explained in Chapter 2, it is important to combine brightness constancy with other illumination-invariant descriptors. The combination with image gradient constancy has demonstrated good performance [Brox et al., 2004; Xu et al., 2012b]. Therefore, we adopt it in our experiments. We also add a normalization of the data term [Zimmer et al., 2011] that we found important in practice to improve results. The resulting energy is similar to (6.9):

$$E(\mathbf{w},\sigma) = \int_\Omega \phi\left(\boldsymbol{\alpha}_\mathbf{w}^\top J^0_{\sigma(x)} \boldsymbol{\alpha}_\mathbf{w}\right) + \phi\left(\boldsymbol{\alpha}_\mathbf{w}^\top(x) J^{x_1x_2}_{\sigma(x)} \boldsymbol{\alpha}_\mathbf{w}(x)\right) dx \\ + \lambda \int_\Omega \phi(\|\nabla \mathbf{w}(x)\|^2)\, dx + \beta \int_\Omega \phi(|\sigma(x)|^2)\, dx$$

where we denote x_1, x_2 the vertical and horizontal axes, and

$$\boldsymbol{\alpha}_\mathbf{w}(x) = \begin{pmatrix} u(x) \\ v(x) \\ 1 \end{pmatrix},\ \mathbf{w} = \begin{pmatrix} u \\ v \end{pmatrix},\ J^0_{\sigma(x)} = k_{\sigma(x)} * \begin{pmatrix} \eta_0 I_{x_1}^2 & \eta_0 I_{x_1} I_{x_2} & \eta_0 I_{x_1} I_t \\ \eta_0 I_{x_1} I_{x_2} & \eta_0 I_{x_2}^2 & \eta_0 I_{x_2} I_t \\ \eta_0 I_{x_1} I_t & \eta_0 I_{x_2} I_t & \eta_0 I_t^2 \end{pmatrix},$$

$$J^{x_1x_2}_{\sigma(x)} = k_{\sigma(x)} * \begin{pmatrix} \eta_{x_1} I_{x_1x_1}^2 + \eta_{x_2} I_{x_1x_2}^2 & \eta_{x_1} I_{x_1x_1} I_{x_1x_2} + \eta_{x_2} I_{x_1x_2} I_{x_2x_2} & \eta_{x_1} I_{x_1x_1} I_{x_1t} + \eta_{x_1} I_{x_1x_2} I_{x_2t} \\ \eta_{x_1} I_{x_1x_1} I_{x_1x_2} + \eta_{x_2} I_{x_1x_2} I_{x_2x_2} & \eta_{x_1} I_{x_1x_2}^2 + \eta_{x_2} I_{x_2x_2}^2 & \eta_{x_1} I_{x_1x_2} I_{x_1t} + \eta_{x_2} I_{x_2x_2} I_{x_2t} \\ \eta_{x_1} I_{x_1x_1} I_{x_1t} + \eta_{x_1} I_{x_1x_2} I_{x_2t} & \eta_{x_1} I_{x_1x_2} I_{x_1t} + \eta_{x_2} I_{x_2x_2} I_{x_2t} & \eta_{x_1} I_{x_1t}^2 + \eta_{x_2} I_{x_2t}^2 \end{pmatrix}.$$

$$\eta_0 = \frac{1}{I_{x_1} + I_{x_2} + a},\quad \eta_{x_1} = \frac{1}{I_{x_1x_1} + I_{x_1x_2} + a},\quad \eta_{x_2} = \frac{1}{I_{x_2x_1} + I_{x_2x_2} + a}$$

where $a = 0.1$ avoids division by 0.

The differentiation of E is similar to the one of J and is detailed in Appendix B.

6.3 Preliminary results

In our experiments, we considered the model (6.9). Without optimization on σ, it amounts to the method of [Bruhn et al., 2005], denoted CLG_0 when $\sigma = 0$ (pixel-wise data term) and CLG when $\sigma = 1.5$. Our method with optimization on σ is denoted CLG-adaptive. In our implementation, we embedded the estimation in a coarse-to-fine scheme to cope with large displacements. The value of σ is adapted at each level of the pyramid and the alternate optimization of $\mathbf{w}$ and σ is performed at each level. All the estimations have been carried out with constant regularization parameter $\lambda = 2.5$.

The method is designed to enhance discontinuities compared to the baseline method of [Bruhn et al., 2005]. Therefore, we evaluate the improvements yielded by CLG-adaptive on the Middlebury benchmark, which main challenge is the recovering of discontinuities. Figure 6.1 shows visual results obtained by the three versions CLG_0, CLG and CLG-adaptive, with corresponding endpoint errors (EPE). The map of σ estimated with CLG-adaptive is also displayed, encoding σ values by the image intensity (dark regions correspond to small value of σ and bright regions to large value of σ). Minimum

	RubberWhale	Venus	Urban3	Grove2
CLG_0	0.124	0.399	0.473	0.159
CLG	0.143	0.420	0.585	0.193
CLG-adaptive	0.126	0.410	0.486	0.176

Table 6.1: Results on sequences of the MIDDLEBURY benchmark. The method of [Bruhn et al., 2005] is refered as CLG_0 when $\sigma = 0$ and CLG when $\sigma = 1.5$. Our method is named CLG-adaptive.

and maximum values σ_{min} and σ_{max} of σ are also reported. In Table 6.1, we give more quantitative results on the MIDDLEBURY benchmark.

In terms of EPE, we can first notice from Table 6.1 that CLG_0 performs significantly better than CLG, which confirms the experiments in [Zimmer et al., 2011] stating that in the absence of noise CLG degrades the results. CLG-adaptive allows to significantly decrease the error compared to CLG. The visualization of σ in Fig. 6.1 shows that low values of σ are mostly localized at motion discontinuities. Consequently, the blurring of motion discontinuities observed in CLG and due to the Gaussian convolution is reduced by CLG-adaptive. The range of values given by $\sigma_{min} = 0.1$ and $\sigma_{max} = 1.7$ show that the algorithm does not converge to $\sigma = 0$, but retains large values of σ in appropriate regions.

However, the results obtained with the adaptive convolution of CLG-adaptive are still slightly less accurate than those obtained without convolution, by CLG_0. This could be due on one hand to the convergence of σ to a bad local minimum. On the other hand, considering isotropic Gaussian filtering is not the best choice for retrieving sharp discontinuities. The extension of our model and related optimization scheme for anisotropic Gaussian filtering is a line of work which should help to improve the results.

6.4 Relation with stochastic uncertainty models

In the energy (6.9), $\sigma(x)$ is determined according to the local validity of the data constancy assumption for the current value $\mathbf{w}(x)$. It is therefore very dependent on the accuracy of $\mathbf{w}$. To relax this dependency, it could be beneficial to investigate other data terms guiding σ more independently to $\mathbf{w}$. In this section, we propose such a data term in the line of work of [Corpetti and Mémin, 2012].

Stochastic uncertainty models for the luminance consistency assumption [Corpetti and Mémin, 2012] We briefly summarize the basic principle of the stochastic uncertainty model proposed in [Corpetti and Mémin, 2012].

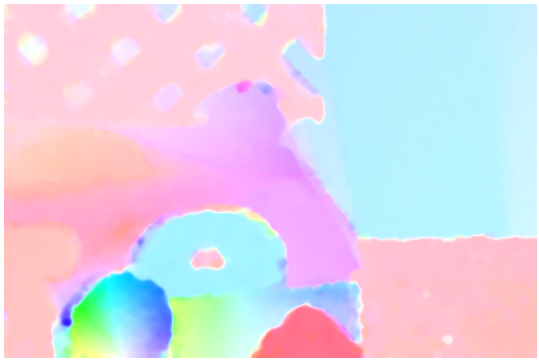
In the standard deterministic intensity differentiation, the image I is defined on a fixed grid of axes x_1 and x_2 such that we can write at pixel x:

$$dI(x,t) \quad = \quad I_t dt + I_{x_1} dx_1 + I_{x_2} dx_2,$$

I_1

I_2

CLG $\sigma = 0$, EPE = 0.124



CLG $\sigma = 1.5$, EPE = 0.143

CLG-adaptive, EPE = 0.126

Estimated σ
$\sigma_{min} = 0.1$, $\sigma_{max} = 1.7$

Figure 6.1: Comparison of CLG-adaptive with CLG and CLG_0. Top row: two input images. Middle row: motion field obtained with CLG_0 and CLG. Bottom row: motion field and σ field obtained with CLG-adaptive.

and $dx_1 = u\,dt,\ dx_2 = v\,dt$, where $\mathbf{w} = (u, v)^\top$. The classical motion constraint is then obtained by the chain rule:

$$\frac{dI(x,t)}{dt} \;=\; I_t + I_{x_1} du + I_{x_2} dv \;=\; 0. \tag{6.12}$$

The stochastic intensity constancy is analogously derived by considering a moving grid $\mathbf{X} = (\mathbf{X}^1, \ldots, \mathbf{X}^m)^\top$ composed of 2D points $\mathbf{X}^i \in \mathbb{R}^2$, and defined as a stochastic process. If the transport of the grid by the motion $\mathbf{w}$ is achieved up to a Brownian motion $\mathbf{B} = (\mathbf{B}^1, \ldots, B^m)^\top$, $d\mathbf{X}$ verifies:

$$d\mathbf{X} = \mathbf{w}\,dt + \mathbf{\Sigma}\,d\mathbf{B}, \tag{6.13}$$

where $\mathbf{\Sigma}$ is a covariance matrix. The image $I(\mathbf{X}, t)$ being a function of the stochastic process, its differentiation is given by the Itô formula:

$$dI(\mathbf{X},t) = I_t dt + I_{x_1} dX_1 + \frac{1}{2} I_{x_2} dX_2 + \sum_{(i,j)=(1,2)\times(1,2)} \frac{\partial^2 I}{\partial X_i \partial X_j} d < X_i, X_j > . \tag{6.14}$$

We consider only isotropic uncertainty of the form:

$$\mathbf{\Sigma}\,d\mathbf{B} = \mathrm{diag}(\sigma) \otimes \mathbb{1}_2\,d\mathbf{B}, \quad \mathbb{1}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \tag{6.15}$$

where σ is an uncertainty variance map. Applying the expression (6.15) to (6.14) yields:

$$\frac{dI(\mathbf{X},t)}{dt} = \underbrace{I_t + I_{x_1} u + I_{x_2} v}_{\text{deterministic intensity variation}} + \underbrace{\frac{\sigma^2 \Delta I}{2}}_{\text{uncertainty / deterministic part}} + \underbrace{\sigma \nabla I \,\frac{d\mathbf{B}}{dt}}_{\text{uncertainty / stochastic part}}. \tag{6.16}$$

The image derivative is thus composed of the usual deterministic intensity variation (6.12), augmented by two uncertainty terms. The uncertainty is derived from the stochastic assumption of Brownian motion of the grid (6.13). Since (6.16) contains a stochastic part, the constancy constraint equation is defined through the expectation, which can be written as (see [Corpetti and Mémin, 2012] for details):

$$\mathbb{E}\left(\frac{dI(\mathbf{X},t)}{dt}\right) \;=\; k_{\sigma(x)} * \left(I_t + I_{x_1} u + I_{x_2} v + \frac{\sigma^2 \Delta I}{2}\right) \;=\; 0. \tag{6.17}$$

An intuitive interpretation of (6.17) is to consider the Gaussian convolution as an operator which captures the uncertainty yielded by the intensity constancy constraint. In homogeneous regions without image gradients, the uncertainty is high and the value of σ

should therefore be large to extend the neighborhood until it contains enough image information to cancel the uncertainty. At an image discontinuity, Δ is high and thus favors small values of σ.

Relation with adaptive filtering of the data term We recognize a close similarity between the stochastic conservation (6.17) and the filtered data term of (6.3). The first main crucial difference is that, as explained in Section 6.1, the convolution in (6.3) does not apply to the motion $\mathbf{w}$. In contrast, the convolution in (6.17) also concerns motion by construction. This difference reflects the diverging assumptions underlying the Gaussian filtering in the two cases. In (6.3), the Gaussian convolution represents an area of local *motion constancy*. Consequently, it has a regularization effect. In (6.17) the Gaussian convolution is rather the expression of *estimation uncertainty*. It represents the most appropriate regions to remove ambiguity of motion estimation (*aperture problem*).

The second difference is the presence of the uncertainty term in (6.17). Intuitively, the presence of the Laplacian adapts σ to image edges, and the presence of σ^2 inhibits too large Gaussian kernels.

In [Corpetti and Mémin, 2012], motion estimation based on the stochastic brightness constancy is performed with a local parametric approach [Lucas and Kanade, 1981]. The uncertainty map σ is estimated independently with local image-based measurements. In contrast, we integrate (6.17) in our global energy:

$$\begin{aligned} E(\mathbf{w},\sigma(x)) &= \int_\Omega \phi\left(k_{\sigma(x)} * \left(I_t + I_{x_1}u(x) + I_{x_2}v(x) + \frac{\sigma^2\Delta I}{2}\right)^2\right) dx \\ &\quad + \lambda\int_\Omega \phi(|\nabla\mathbf{w}(x)|^2)dx + \beta\int_\Omega \phi(|\sigma(x)|^2). \end{aligned} \quad (6.18)$$

Compared to our previous energy (6.9), the minimization on σ is less dependent on the value of $\mathbf{w}$ owing to the uncertainty term. The integration of the stochastic model in a global approach also allows for an alternative uncertainty estimation to [Corpetti and Mémin, 2012]. We also expect to increase the quantitative performance of motion estimation compared to the local framework based on [Lucas and Kanade, 1981] used in [Corpetti and Mémin, 2012].

As in [Corpetti and Mémin, 2012], we approximate the energy by assuming that the convolution does not apply to motion. A tensor notation similar to (6.9) can then be applied:

$$E(\mathbf{w},\sigma) = \int_\Omega \phi\left(\boldsymbol{\alpha}^\top_{\mathbf{w},\sigma(x)} J_{\sigma(x)} \boldsymbol{\alpha}_{\mathbf{w},\sigma(x)}\right) dx + \lambda\int_\Omega \phi(|\nabla\mathbf{w}|^2)\,dx + \beta\int_\Omega \phi(|\nabla\sigma|^2)\,dx \quad (6.19)$$

$$\text{with } \boldsymbol{\alpha}_{w,\sigma(x)} = \begin{pmatrix} u \\ v \\ \sigma(x)^2 \\ 1 \end{pmatrix},\ J_{\sigma(x)} = k_{\sigma(x)} * \begin{pmatrix} I_{x_1}^2 & I_x I_y & I_{x_1}\frac{\Delta I}{2} & I_{x_1} I_t \\ I_x I_y & I_{x_2}^2 & I_{x_2}\frac{\Delta I}{2} & I_{x_2} I_t \\ I_{x_1}\frac{\Delta I}{2} & I_{x_2}\frac{\Delta I}{2} & \left(\frac{\Delta I}{2}\right)^2 & \frac{\Delta I}{2} I_t \\ I_{x_1} I_t & I_{x_2} I_t & \frac{\Delta I}{2} I_t & I_t^2 \end{pmatrix}.$$

Optimization The optimization problem w.r.t. σ does not differ fundamentally from the one described in Section 6.2. We also resort to the L-BFGS technique, and the computation of the gradient of (6.19) is provided in Appendix C.

Optimization w.r.t. $\mathbf{w}$ is more difficult than in Section 6.2. Indeed, in Section 6.2 we adopted the approach briefly described in Section 4.3.1 and precisely detailed in [Brox, 2005] based on solving Euler-Lagrange equations. It amounts to solving a large and sparse linear system, which is usually very efficiently achieved with iterative methods like Gauss-Seidel methods or its SOR variant (Successive Over Relaxation). A sufficient condition for these iterative methods to converge is the diagonal dominance of the system. A matrix is diagonally dominant if for each row, the absolute value of its diagonal element is superior to the sum of the absolute values of the other elements of the row, which writes:

$$|a_{i,i}| \geq \sum_{j \neq i} |a_{i,j}| \tag{6.20}$$

where $a_{i,j}$ are the elements of the matrix. In practice, this condition rarely strictly holds for typical optical flow energies, but Gauss-Seidel methods still converge without this guarantee if the deviations from (6.20) are not too strong. For the energy (6.9), the diagonal elements correspond to

$$a_{i,i} = I_\ell^2 + \phi'(I_t + I_{x_1} u + I_{x_2} v) - 2\phi'(\|\mathbf{w}\|^2), \tag{6.21}$$

with ℓ being the x_1 or x_2 axis. It is most of the time largely greater than the non-diagonal terms. In the case of stochastic uncertainty model, the diagonal elements differ by the additional uncertainty term:

$$a_{i,i} = I_\ell^2 + \phi'\left(I_t + I_{x_1} u + I_{x_2} v + \frac{\sigma^2 \Delta I}{2}\right) - 2\phi'(\|\mathbf{w}\|^2). \tag{6.22}$$

The frequent large negative values of the Laplacian in (6.22) implies a large number of very low values of $a_{i,i}$, and consequently implies strong violations of diagonal dominance. In our experiments, the lack of diagonal dominance was prohibitive for the convergence of Gauss-Seidel and SOR solvers, mostly used for optical flow energies. It made also diverge all the more sophisticated solvers implemented in the reference PETSc library [Balay et al., 2014] for linear system solvers.

Quasi-Newton schemes and proximal splitting algorithms have been tested but all led to divergence. As a consequence, we cannot present results of numerical evaluation of the method. Solving this optimization problem will be the subject of future work.

6.5 Conclusion - Perspectives

In the preliminary study presented in this chapter, we have proposed two energy models for joint optimization of motion and Gaussian kernel, based on [Bruhn et al., 2005]. First results on the model presented in Section 6.2 demonstrate significant improvements compared to the baseline method [Bruhn et al., 2005]. A persisting oversmoothing due to the isotropy of the filtering still remains and prevents from outperforming the pixel-wise version of the method. The full potential of our framework could be exploited by considering anisotropic Gaussian filters to reduce smoothing across discontinuities.

The second model presented in Section 6.4, inspired by [Corpetti and Mémin, 2012], provides a new data potential modeling the estimation uncertainty in the support of the filtering. This approach sounds more appropriate to an adaption of the filtering. However, numerical issues prevented from a performance evaluation of this approach. The optimization problem caused by the uncertainty modeling will be addressed in the future.

Despite potential important improvements at motion discontinuities, this adaptive framework is not designed to handle major issues in optical flow like occlusions, intensity changes or limits of the coarse-to-fine scheme. Therefore, in Part II, we explore another approach for the combination of local and global models. We design it with the purpose of addressing the aforementioned limiting issues of optical flow.

7 Summary of main challenges

Through the analysis of optical flow literature, we have introduced principles of optical flow estimation and classified the main methodological aspects of existing methods. From this taxonomy, limiting points of each class have been put forward. We summarize the main conclusions of this study.

Since the intensity constancy assumption is not guaranteed in a lot of situations, more robust data terms have been investigated The most appealing direction is the explicit estimation of deviations from the intensity constancy constraint. However, it is limited by optimization issues. In Chapter 14, we will propose to revisit the latter category to overcome its difficulties.

Concerning local parametric methods, polynomial models have been proven to be appropriate for local motion representation. The main issue remains the appropriate delineation of estimation supports. The aggregation framework described in Part II can be viewed as a new way to select appropriate neighborhoods.

Global regularization offers very powerful modeling capacities, which have to be tempered by optimization tractability constraints. We will exploit variational and discrete optimization principles in different parts of our thesis work. In particular, we will tackle several issues (occlusions, intensity changes, Gaussian kernel adaptivity) by jointly minimizing motion and other variables, while taking care of the validity of the proposed optimization.

The multi-resolution scheme is necessary to recover large displacements when the estimation is based on a linearized data constraint. However, it also generates artifacts. The integration of feature matching in dense optical flow estimation frameworks is a fruitful alternative. We will describe in Part II an original feature matching integration in the aggregation framework.

Occlusion handling has not been addressed in this part but is another crucial issue for optical flow estimation. Joint occlusion detection and motion estimation will be investigated in Part II. We also propose an exemplar-based approach for motion estimation in occluded areas.

Part II

Aggregation of local parametric motion candidates with exemplar-based occlusion handling

In this part, we propose a new method for optical flow computation called *AggregFlow* which exhibits several distinctive and original features. First, we advocate the systematic computation of affine motion models over a set of size-varying square patches combined with patch-based pairings. Indeed, we experimentally demonstrate that the set of motion vectors computed that way comprises at least one accurate motion vector for each pixel. On this basis, we build an optical flow estimation method composed of a first step computing local parametric candidates followed by a second step aggregating these candidates to produce the global flow field. The motion vector candidates are independently estimated on local supports without segmentation step. The aggregation is performed by a discrete optimization algorithm which selects one candidate at each pixel while ensuring piecewise smoothing of the resulting flow field.

Secondly, we address the occlusion problem in an original way by blending it with the motion estimation issue through the two steps of AggregFlow. In particular, we properly deal with very large displacements producing large occluded regions. Motion candidates are extended in occlusion areas with an exemplar-based approach. The estimated parametric model of the dominant motion in the image also contributes to create supplementary motion candidates. We extract an occlusion confidence map in the first step of AggregFlow and exploit them to guide the joint estimation of the occlusion map and motion field in the aggregation step.

Our method can thus be viewed as a novel and efficient combination of local and global approaches for occlusion-aware optical flow computation. The main original features and contributions of our method AggregFlow are listed below:

- Motion candidates are locally estimated by a general parametric patch-based method which ensures relevant and accurate motion vectors at every point among all the computed candidates.
- Feature matching is integrated in an original and efficient way in the two-step aggregation framework.
- We define a generic exemplar-based method for occlusion filling with motion vectors.
- We propose a joint motion and occlusion estimation framework based on a sparse model guided by a local occlusion confidence map.
- AggregFlow outperforms existing methods on the MPI Sintel benchmark which involves large displacements and occlusions, and it is competitive in the Middlebury benchmark composed of videos depicting smaller movements.

We propose an alternative continuous apporach to the discrete optimization of the aggregation stage. A sparse dictionnary model is exploited to achieve the selection of a few candidates with global continuous optimization.

Another variant of AggregFlow with variational candidates estimation is proposed and analyzed in Appendix A.

Positioning from related work

Hereunder, we briefly emphasize on the relations of AggregFlow with the vast literature on optical flow computation, more detailed in Part I.

Feature correspondences and large displacements The integration of feature correspondences in dense motion estimation has been investigated in several recent works. A first class of methods integrates feature correspondences in a global energy model. Variational methods [Braux-Zin et al., 2013; Brox and Malik, 2011; Héas and Mémin, 2008; Weinzaepfel et al., 2013; Hellier and Barillot, 2003] include an additional term to a classical global energy to impose the flow to be close to pre-computed correspondences. Giving a fixed weight to the correspondences, this approach is sensitive to matching errors. To overcome this problem, [Braux-Zin et al., 2013; Weinzaepfel et al., 2013] focused on improving the matching step. Another class of methods use correspondences to reduce the search space for discrete optimization and provide a coarse initialization for subsequent refinement [Chen et al., 2013; Mozerov, 2013; Xu et al., 2012b]. The main motivation of the attempts based on feature matching is to get rid of the drawbacks of the coarse-to-fine scheme imposed by variational optimization, in particular the loss of large displacements of small objects.

Our patch correspondence is related to [Chen et al., 2013; Mozerov, 2013; Xu et al., 2012b] in the sense that it is used in the candidate generation process, but differently from these approaches. Our method does not produce coarse approximations to be refined in a continuous subsequent step and we do not adopt any global variational optimization.

Occlusions Occlusions play a crucial role for motion estimation [Stein and Hebert, 2009], especially under large displacements, since no motion measurements are available in occluded areas. Therefore, a proper occlusion handling must distinguish between *occlusion detection*, segmenting the image into occluded and non-occluded regions, and *occlusion filling*, applying a specific treatment to motion estimation in occluded regions. Occlusion detection has been mostly undertaken as a subsequent operation to motion computation, by thresholding a consistency measure issued from the estimated motion field, like geometric forward-backward motion mismatch [Ince and Konrad, 2008], mapping uniqueness [Xu et al., 2012b] or data constancy violation [Xiao et al., 2006]. The main limitation of this post-processing is that accuracy of occlusion detection is highly dependent on the quality of the initial motion estimation. Several flow and image criteria have been be combined in a learning framework [Humayun et al., 2011]. Retrieving

occlusion map without optical flow computation have been realized in [Kervrann et al., 2011] based on semi-local patch-based hypothesis testing. Other approaches estimate the occlusion map jointly with the motion field in an alternate optimization scheme [Ayvaci et al., 2012; Ince and Konrad, 2008; Papadakis et al., 2013]. Our occlusion detection falls in the latter category.

The problem of filling occlusion regions with estimated velocity vectors when the occlusion map is known is closely related to the inpainting problem. While inpainting aims at filling missing regions without image information, occlusion filling aims at estimating motion in regions without motion information. Inpainting methods can be coarsely divided into two classes : diffusion-based methods [Chan et al., 2002] and exemplar-based methods [Criminisi et al., 2004; Komodakis and Tziritas, 2007; Daisy et al., 2013; Forbin et al., 2005]. A synthesis of these two approaches has been investigated in [Bugeau et al., 2010; Arias et al., 2011] which provide a general variational framework for non-local image inpainting. Occlusion handling is usually achieved by diffusion-based (or geometry-oriented) methods, propagating motion from non-occluded regions to occluded regions via partial derivative equation (PDE) resolution [Ayvaci et al., 2012; Ince and Konrad, 2008; Papadakis et al., 2013; Xu et al., 2012b]. In exemplar-based inpainting, the missing part is filled by copying pixels of the observed images. The framework is non local in the sense that similar pixels can be sought wherever in the image. We adapt this strategy to occlusion filling. Finally, we mention the work of [Ballester et al., 2012] wich defines a data constraint for occluded pixels under the assumption of temporal motion constancy between three consecutive frames. At occluded pixel, the intensity constancy constraint is thus applied between the current frame and the previous frame.

Parametric motion estimation

The use of a parametric model has been widely investigated in motion estimation [Black and Anandan, 1996; Cremers and Soatto, 2005; Leordeanu et al., 2013; Mémin and Pérez, 1998; Odobez and Bouthemy, 1995; Sun et al., 2012]. Applied on the whole image domain, affine or quadratic models are adequate to estimate the dominant image motion induced by the camera motion [Odobez and Bouthemy, 1995]. For accurate dense motion estimation, parametric approximations are only valid locally. Local regions are usually defined as square patches centered on each pixel [Black and Anandan, 1996; Lucas and Kanade, 1981]. However, this choice is suboptimal in particular in two situations illustrated in Fig. 7.1: when the patch contain a motion discontinuity invalidating the motion model, and when it does not contain enough image gradient to estimate motion. Adaptation of the size of the patch [Maurizot et al., 1995; Senst et al., 2012], or its position [Jodoin and Mignotte, 2009] have been investigated. This approach has the merit

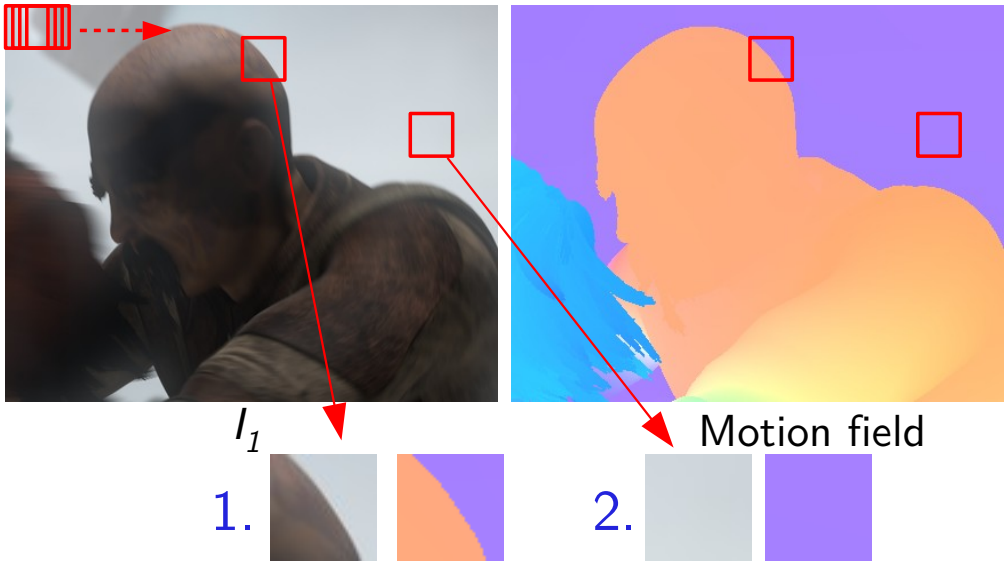


Figure 7.1: Illustration of the two typical issues of patch distribution in [Lucas and Kanade, 1981]. Case 1.: motion discontinuity invalidating polynomial motion model. Case 2.: not enough image gradient to estimate motion in the patch.

of being easy to implement with a low computational cost, but it is clearly outperformed by sophisticated extensions of [Horn and Schunck, 1981] introduced in modern global optical flow methods.

As aforementioned, more complex region shapes can be estimated by joint motion segmentation and estimation. Existing approaches can be divided in two classes. A first class of methods relies on an independent image color segmentation and tries to fit parametric motion in each region [Black and Jepson, 1996; Bleyer et al., 2006; Gelgon and Bouthemy, 2000; Xu et al., 2008; Zitnick et al., 2005], possibly with the help of an independent global variational estimation [Black and Jepson, 1996; Xu et al., 2008]. The drawback of this approach is that image color segmentation may lead to an over-segmentation of the motion field. The second class of methods jointly estimates supports of regions and parametric motion models in these regions [Bouthemy and François, 1993; Cremers and Soatto, 2005; Odobez and Bouthemy, 1998; Sun et al., 2012]. It is achieved by minimizing a global energy with respect to supports and motion parameters of the regions. However, this global energy is highly non-convex and consequently difficult to minimize and particularly sensitive to the initialization of the optimization procedure, as illustrated in Fig. 7.2.

The motion field produced by AggregFlow is composed of affine motion vectors estimated in square patches, without motion segmentation. Our aggregation strategy

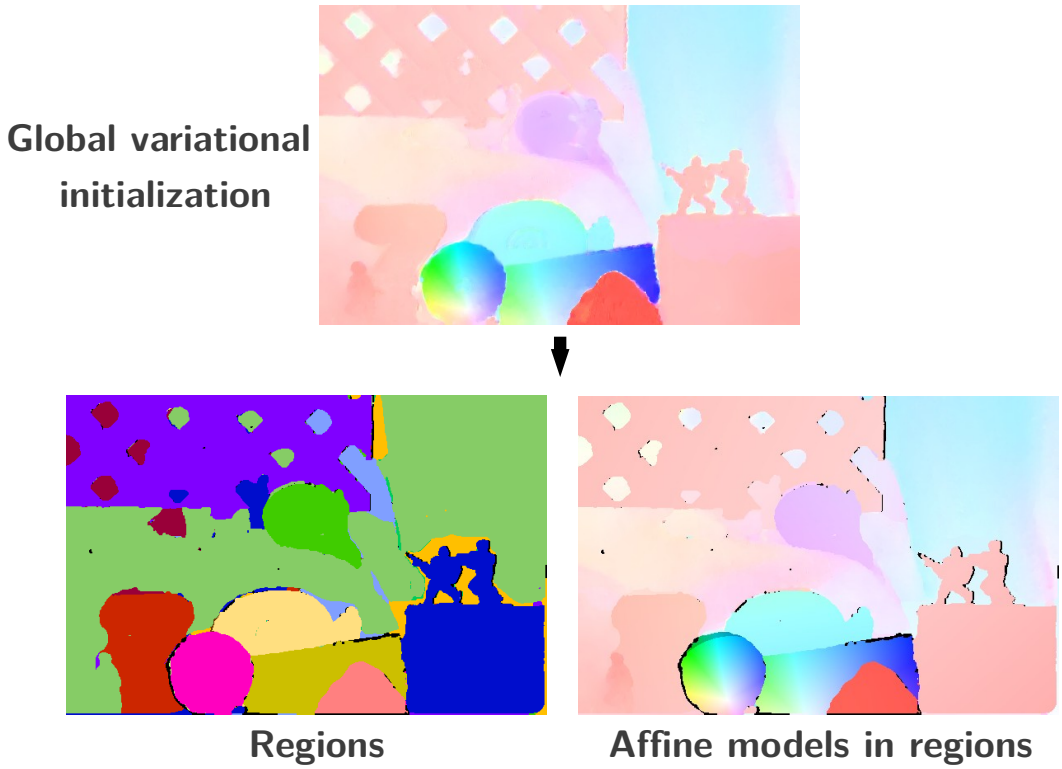


Figure 7.2: Example of the importance of initialization in joint motion estimation and segmentation methods, with the result of [Unger et al., 2012]. First row: intialization with [Werlberger et al., 2009]. Second row: Final segmented regions and motion estimation result.

allows us to select the best patch size and position for each pixel.

Motion discontinuities In the variational setting, the problem of preserving discontinuities has been addressed by modifying the regularization term. The seminal work of [Horn and Schunck, 1981] used a quadratic penalty function on the gradient magnitude of motion vectors. The first attempt to preserve discontinuities was investigated in [Heitz and Bouthemy, 1993] where a binary map of local motion discontinuities was introduced and estimated jointly with the motion field using two interwoven Markov Random Fields (MRF). The regularization is thus canceled on motion discontinuities. Subsequent improvement has then been reached with the use of robust penalty functions in the regularization term [Black and Anandan, 1996; Mémin and Pérez, 1998]. The robust L_1 norm has demonstrated its efficiency in a variational context thanks to its convexity [Brox et al., 2004; Werlberger et al., 2010]. These methods still suffer from local minima when minimizing non-convex functions which usually supply more accurate results, and are limited by the coarse-to-fine scheme.

Discrete optimization and aggregation paradigm Discrete optimization is an alternative to variational methods and is able to find good local minima for more general, non differentiable and non-convex energy functionals. To combine the subpixel accuracy of the continuous variational approach and the efficiency of discrete minimization, the authors of [Lempitsky et al., 2008] built a discrete motion space from motion fields delivered by several global variational estimations with different parameter settings. A classical energy model is then optimized by successive fusions of global proposals, efficiently performed by binary graph-cut methods. In [Alba et al., 2010], a set of candidate motion vectors is computed at each pixel using phase correlation in overlapping patches. The candidates are then linearly combined to create a global motion field. Recent works [Chen et al., 2013; Mozerov, 2013] also exploit discrete graph-cut optimization in a two-step paradigm. However, the philosophy of their method is different. Indeed, their motion candidate generation step only aims at finding dominant displacements and the aggregation provides a coarse initialization for a subsequent global variational estimation. Discrete optimization is also associated with a variational framework in [Xu et al., 2012b] as an intermediate stage between scales of a coarse-to-fine framework, in order to limit the loss of details of the flow.

8 Local motion candidates and occlusion cues

We describe in this chapter the first step of our method AggregFlow. It exploits local constraints to supply motion candidates and occlusion cues. A set of motion vector candidates is generated at every pixel by a combination of patch correspondences and local parametric motion model estimations. A specific treatment is applied to occluded regions by exemplar-based extension of the motion candidates set. We also exploit the dominant motion in the image due to camera motion. Motion candidates and occlusion cues form the input of the second stage of AggregFlow described in Chapter 9.

Local motion estimations are performed in overlapping square patches of different sizes, so that each pixel is contained in various patches. Our approach can be viewed as a new way to address the problem emphasized in the introduction of this part of the choice of the local neighborhood for parametric estimation. Rather than adapting the regions *a priori* or jointly with the motion field, we operate in two steps: 1) estimation of motion candidates on several supports at every pixel, 2) implicit selection of the best support through the selection of the optimal candidate at each pixel within the aggregation step (Chapter 9).

8.1 Local parametric motion candidates

In this section, we describe a combination of parametric estimation and patch correspondences for candidates computation. Nevertheless, the genericity of the framework allows for other types of local estimation. In particular, the reader can refer to Appendix A for an analysis of candidates computation with a regularized variational method.

8.1.1 Set of overlapping patches in I_1

The local supports for motion candidates computation are overlapping square patches of different sizes. Let us denote $\mathcal{P}_{s,\alpha}$ the patch set for a fixed patch size s and an overlapping ratio $\alpha \in [0, 1]$ indicating the proportion of surface shared by neighboring patches (see illustration of Fig. 8.1). Let $\mathcal{S} = \{s_1, \ldots, s_n\}$ be a set of n patch sizes, we then define $\mathcal{P}_{\mathcal{S},\alpha} = \bigcup_{s\in\mathcal{S}} \mathcal{P}_{s,\alpha}$. To capture different motion scales, the patch sizes must cover a large range of values. In all our experiments, we will use $\mathcal{S} = \{16, 44, 104\}$. Due

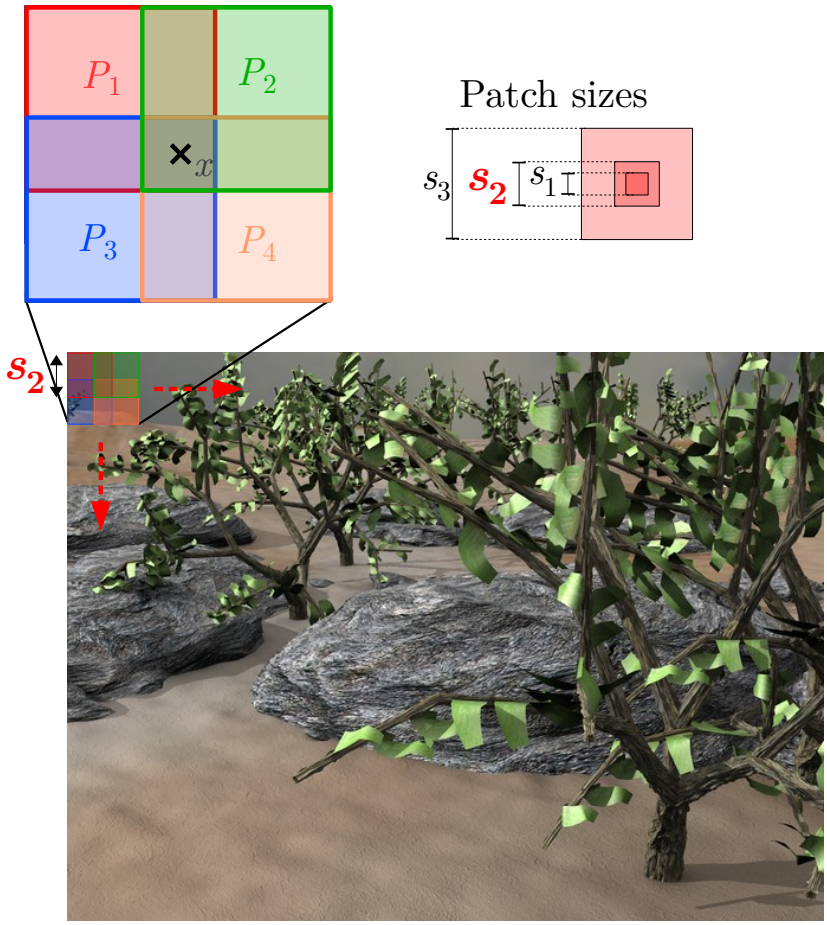


Figure 8.1: Four patches of set $\mathcal{P}_{s_2,\alpha}$ for a given size s_2 of the set $\mathcal{S} = \{s_1, s_2, s_3\}$, and overlapping ratio $\alpha = 0.3$. The pixel x is contained in the patches $P_1, \ldots, P_4$. Motion estimation in each of these patches provide motion candidates for x.

to the overlap and the number of patch sizes ($n > 1$), one given pixel $x \in \Omega$ belongs to several patches. The motion vectors are estimated independently in each patch in two sub-steps described below: patch correspondences and affine motion estimations.

8.1.2 Patch correspondences

For each patch $P_1 \in \mathcal{P}_{\mathcal{S},\alpha}$, we first determine the set $\mathcal{M}_N(P_1)$ of the N most similar patches to P_1 in I_2. Let us point forward that we do not aim at keeping at this stage the best correspondence only but at selecting N relevant correspondences to subsequently constitute motion candidates. The matching step is generic and could be achieved with

any arbitrary feature matching algorithm. We use a combination of the saturation and value channels of the HSV color space to gain partial robustness to illumination changes [Zimmer et al., 2011] and we use the Sum of Absolute Distances (SAD) to compare patches. To avoid that the set $\mathcal{M}_N(P_1)$ uselessly contains too close patches, we impose a minimal distance between two patches of $\mathcal{M}_N(P_1)$. Hence, for each established pair of corresponding patches $P_{1,2} = (P_1, P_2)$ with $P_2 \in \mathcal{M}_N(P_1)$, we get the translation vector $\mathbf{w}_{P_{1,2}} \in \mathbb{Z}^2$ shifting P_1 onto P_2.

8.1.3 Affine motion refinement

The displacements estimated by patch correspondences are integer-pixel translational approximations. To reach subpixel accuracy and to allow for more complex motion, we refine the first sub-step of coarse translation computation with the estimation of a local affine motion model in every pair $P_{1,2}$. Denoting Ω_{P_1} the pixel domain of P_1, the affine motion model $\delta\mathbf{w}_{P_{1,2}} : \Omega_{P_1} \to \mathbb{R}^2$ between P_1 and P_2 is defined at a pixel $x = (x_1, x_2)^\top$ as:

$$\delta\mathbf{w}_{P_{1,2}}(x) = (a_1 + a_2x_1 + a_3x_2, a_4 + a_5x_1 + a_6x_2)^\top. \tag{8.1}$$

The affine model parameter vector $\boldsymbol{\theta}_{P_{1,2}} = (a_1, a_2, a_3, a_4, a_5, a_6)^\top$ is estimated using the brightness constancy constraint:

$$\widehat{\boldsymbol{\theta}}_{P_{1,2}} = \arg\min_{\boldsymbol{\theta}_{P_{1,2}}} \int_{\Omega_{P_1}} \psi(P_2(x + \delta\mathbf{w}_{P_{1,2}}(x)) - P_1(x))dx \tag{8.2}$$

where the penalty function $\psi(\cdot)$ is chosen as the robust Tukey's function. The problem (14.3) is solved with the publicly available Motion2D software[1] [Odobez and Bouthemy, 1995], which implements a multi-resolution incremental minimization scheme involving an IRLS (Iteratively Reweighted Least Squares) technique for solving the successive linearizations of the penalty function in (14.3).

8.1.4 Final set of motion candidates

The above described two-step estimation is repeated for every patch of $\mathcal{P}_{\mathcal{S},\alpha}$ and generates a set of candidate motion vectors $\mathcal{C}(x)$ at each pixel $x \in \Omega$ defined as follows:

$$\mathcal{C}(x) = \{\mathbf{w}_{P_{1,2}}(x) + \delta\mathbf{w}_{P_{1,2}}(x) : P_1 \in \mathcal{P}_{\mathcal{S},\alpha}(x), P_2 \in \mathcal{M}_N(P_1)\}, \tag{8.3}$$

where $\mathcal{P}_{\mathcal{S},\alpha}(x) = \{P \in \mathcal{P}_{\mathcal{S},\alpha} : x \in P\}$.

Let us make a few comments on the estimation scheme for computing motion candidates. A coarse motion estimation followed by a refinement step has been investigated in several

[1] http://www.irisa.fr/vista/Motion2D/

previous works [Chen et al., 2013; Leordeanu et al., 2013; Mozerov, 2013], but it has always been dedicated to global motion fields. In our case, the refinement is local and adapted to each patch correspondence. Classical local motion estimation methods based on [Lucas and Kanade, 1981] also rely on square patches, but assign the computed motion vector only to the center point of each patch. On the opposite, parametric motion estimation in segmented regions as in [Cremers and Soatto, 2005] apply to regions of arbitrary shape. Our patch distribution can be considered as an intermediate level between these two extremes. Indeed, we use square patches as in [Lucas and Kanade, 1981] and thus avoid the complex segmentation step. However, we exploit the whole vector field issued from the affine model estimated in each patch. As a consequence, every pixel inherits several motion candidates from the affine motion estimations performed in patches of different positions and sizes which the given pixel belongs to. Finally, in contrast to several other methods using feature correspondences [Brox and Malik, 2011; Chen et al., 2013; Weinzaepfel et al., 2013], we do not select one single patch correspondence but we keep the N best ones.

The interest of the local set of motion candidates supplied by AggregFlow is three-fold. First, the correspondence sub-step enables to capture large displacements even for small patch sizes. Thus, it allows us to correctly deal with small structures undergoing large displacements in contrast to coarse-to-fine schemes. Second, by considering a large variety of patches, we get rid of the predefined choice of the local neighborhood encountered in parametric motion estimation. The selection of the proper patch via its corresponding motion candidate is transferred to the aggregation stage. Third, introducing patches of several sizes enables to tackle motion of different scales.

8.2 Motion candidates in occluded areas

The generation of motion candidates described in Section 8.1 does not differentiate between occluded and non-occluded pixels. For a given pixel x, if all the patches of $\mathcal{P}_{\mathcal{S},\alpha}(x)$ mainly contain occluded pixels, there is no chance to correctly estimate a relevant motion candidate at x in that way. Therefore, we compute motion candidates in occluded regions in a specific manner.

Let us define the occlusion map $o : \Omega \to \{0, 1\}$

$$o(x) = \begin{cases} 1 & \text{if } x \text{ is occluded,} \\ 0 & \text{otherwise.} \end{cases} \tag{8.4}$$

The occluded regions are denoted $\mathcal{O} = \{x \in \Omega : o(x) = 1\}$. The computation of map o will be addressed in Section 8.4 and Chapter 9, and we assume for now that o is known.

8.2.1 Occlusion filling

Let us first emphasize the relation between occlusion filling and inpainting. When occlusion regions are known, occlusion filling is conceptually closely related to image inpainting, since it recovers motion in regions where it is by definition *not observable*. Classical methods for occlusion filling operate in a variational framework by cancelling the data term and letting the diffusion process of the regularization propagate the optical flow in occluded regions [Ayvaci et al., 2012; Xu et al., 2012b]. This is also the approach of the diffusion-based class of inpainting methods [Bertalmio et al., 2000]. This class of inpainting methods performs well in case of thin missing areas or cartoon-like images, but they are usually outperformed by the class of exemplar-based inpainting methods [Criminisi et al., 2004] for large missing regions. The idea is to copy image pixels from the observed regions of the image to the region to be filled. In order to deal with large occlusions produced by large displacements, we follow the inpainting analogy and we overcome the problem of local motion candidates estimation in occluded areas by designing an exemplar-based scheme. In the first step of AggregFlow, the motion candidates set is thus augmented by copy-paste operations.

8.2.2 Exemplar-based candidates extension

We rely on the assumption that motion at an occluded pixel $x \in \mathcal{O}$ is similar to the motion of a close non-occluded pixel $m_o(x) \in \Omega \backslash \mathcal{O}$ belonging to the same object or the same background part. To provide relevant motion candidates at x, we copy motion candidates from $\mathcal{C}(m_o(x))$ to $\mathcal{C}(x)$. The search domain $\mathcal{V}_o(x) \subset \Omega \backslash \mathcal{O}$ for $m_o(x)$ is constrained to be close to the occlusion boundaries. Figure 8.2(e) and 8.3(e) represents the occluded regions $\mathcal{O}$ (in white) and the search domain $\mathcal{V}_o$ (in red), and Fig. 8.2(f) and 8.3(f) superimposes the two sets on I_1. Searching for the pixel $m_o(x)$ is actually easier for occlusion filling than for image inpainting. Indeed, occluded regions are not completely uninformative, while inpainted regions are, since we have access to the information supplied by image I_1 even in $\mathcal{O}$. Thus, as $m_o(x)$ is expected to belong to the same object as x, we use color similarity to find the match in I_1:

$$m_o(x) = \underset{y \in \mathcal{V}_o(x)}{\arg\min}\, D(I_1, x, y), \tag{8.5}$$

where $D(I_1, x, y)$ is the distance between patches centered respectively in x and y. As in Section 8.1, we resort to a SAD in the HSV space.

An extended candidate set $\mathcal{C}_+(x)$ is created for occluded pixels by adding to the initial

set $\mathcal{C}(x)$, which is usually empty, the motion candidates of their matched pixel $m_o(x)$:

$$\mathcal{C}_+(x) = \mathcal{C}(x) \cup \mathcal{C}(m_o(x)),\ \forall x \in \mathcal{O}. \tag{8.6}$$

By convention, $\forall x \in \Omega \backslash \mathcal{O}$, $\mathcal{C}_+(x) = \mathcal{C}(x)$.

8.2.3 Occlusions due to camera motion

A particular class of occluded (or disappearing) regions occurs at image borders in the case of large camera motion (Fig. 8.4). We cope with this issue by estimating the dominant image motion due to camera motion. To do so, we use again the robust parametric estimation described in Section 8.1, but now, we apply it to the whole image [Odobez and Bouthemy, 1995], to retrieve the dominant motion. We found in our experiments that the quadratic model was more adequate to accurately cope with large and sometimes complex camera motion. The resulting parametric motion field $\mathbf{w}_{cam} : \Omega \to \mathbb{R}^2$ is added to the motion candidates, and we end up with the final set of motion candidates $\mathcal{C}_f$:

$$\mathcal{C}_f(x) = \mathcal{C}_+(x) \cup \{w_{cam}(x)\},\ \forall x \in \Omega. \tag{8.7}$$

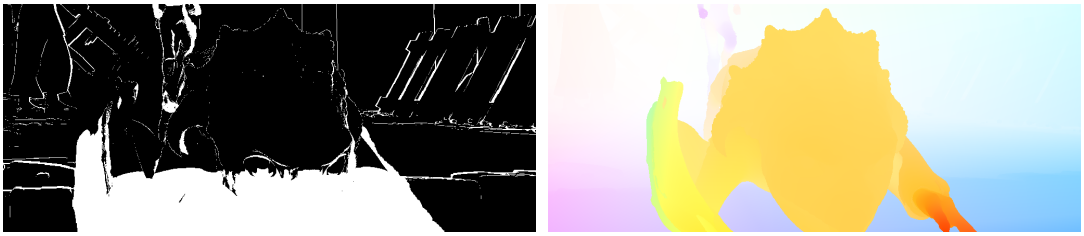
The camera motion candidates are mostly useful for occluded pixels, but it can sometimes provide relevant motion candidates in unoccluded regions of the background as well, so that we finally add it to all pixels in Ω.

8.3 Best candidate flow

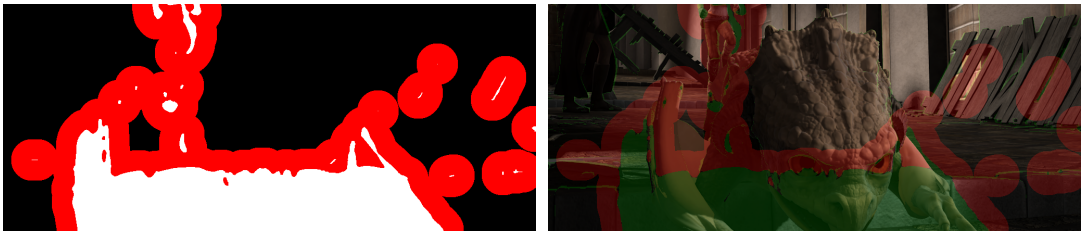
To validate our method for computing motion candidates, we have exploited sequences from MPI Sintel and Middlebury datasets [Baker et al., 2011; Butler et al., 2012] provided with ground truth. We create the *Best Candidate Flow* (BCF) by selecting at each pixel x the candidate motion vector of $\mathcal{C}_f(x)$ closest to the ground-truth vector. In order to evaluate our occlusion module, we distinguish between the BCF determined with the candidates extension described in the preceding section (or full BCF) and the BCF without it. Parameters involved in the local motion computation are set to $\mathcal{S} = \{16, 44, 104\}$, $\alpha = 0.75$, $N = 2$.

Illustrations of the accuracy of the BCF are provided in Fig. 8.2, Fig. 8.3 and Fig. 8.4 on sequences of the MPI Sintel benchmark with large occluded regions. Besides, we make a specific focus on the improvements obtained with the candidates extensions. The difference between BCF without candidates extension and the full BCF is clearly visible for occluded pixels and testify to the importance of the exemplar-based and camera motion candidates extensions. Overall, the full BCF is very close to the ground-truth motion

(a) I_1 (b) I_2



(c) Ground truth occlusion (d) Ground truth $\mathbf{w}$



(e) Occlusions (white) and search domain $\mathcal{V}_o$ (red)

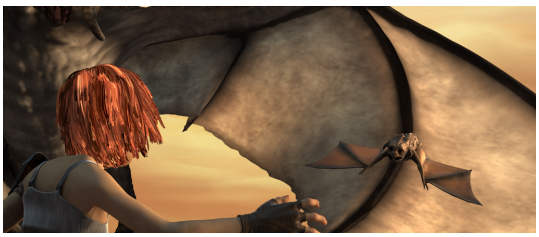
(f) I_1 with occlusions (green) and search domain $\mathcal{V}_o$ (red)

(g) BCF without exemplar-based candidates extension

(h) BCF with exemplar-based candidates extension

Figure 8.2: Illustration of the performance improvement with exemplar-based candidates extension. First row: two successive input images. Second row: ground-truth occlusion map and motion field. Third row: representation of the search domain $\mathcal{V}_o$ (displayed here after median filtering of the occlusion map for the sake of visibility only). Fourth row: Best Candidate Flow obtained respectively without and with the exemplar-based candidates extension.

(a) I_1

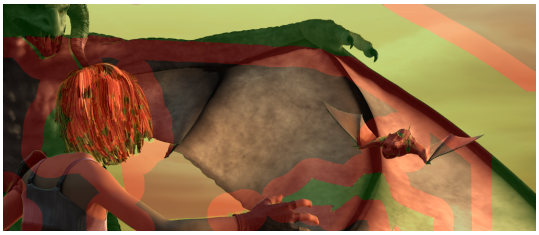


(b) I_2

(c) Ground truth occlusion

(d) Ground truth $\mathbf{w}$

(e) Occlusions (white) and search domain $\mathcal{V}_o$ (red)



(f) I_1 with occlusions (green) and search domain $\mathcal{V}_o$ (red)

(g) BCF without exemplar-based candidates extension

(h) BCF with exemplar-based candidates extension

Figure 8.3: Illustration of the performance improvement with exemplar-based candidates extension. First row: two successive input images. Second row: ground-truth occlusion map and motion field. Third row: representation of the search domain $\mathcal{V}_o$ (displayed here after median filtering of the occlusion map for the sake of visibility only). Fourth row: Best Candidate Flow obtained respectively without and with the exemplar-based candidates extension.

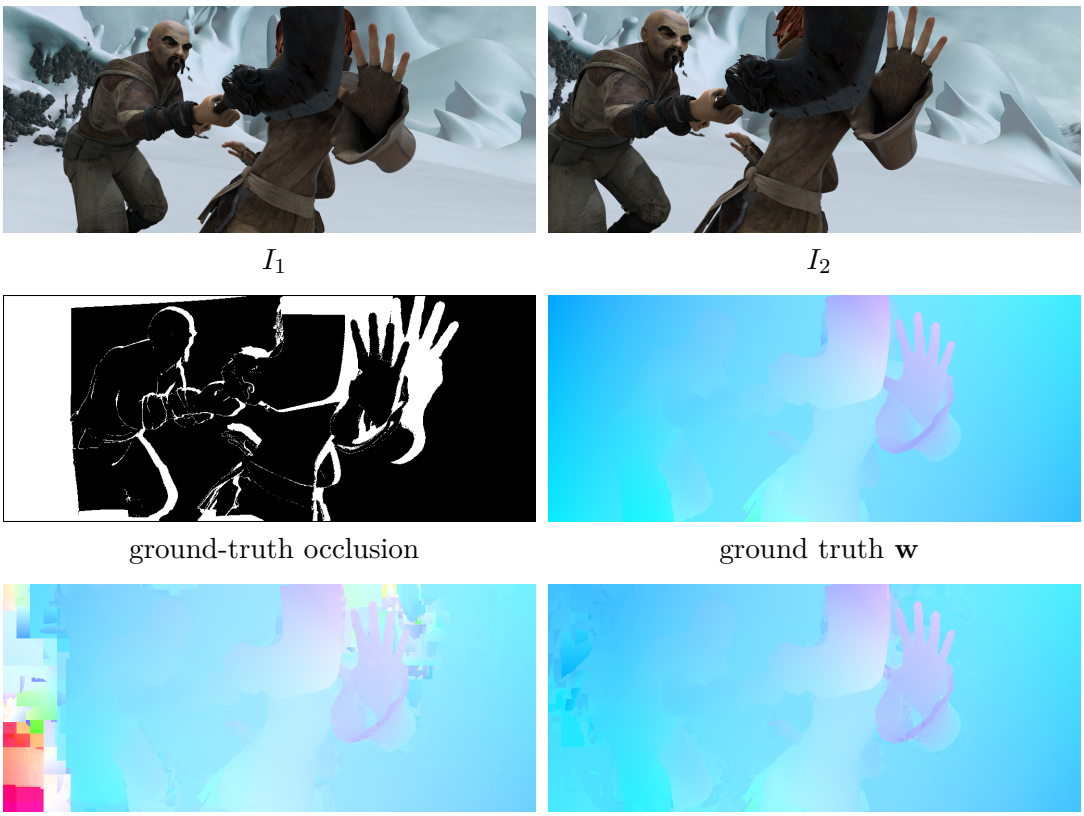


Figure 8.4: Illustration of the performance improvement with camera motion candidates extension. First row: two successive input images. Second row: ground-truth occlusion map and motion field. Third row: Best Candidate Flow obtained respectively without and with the camera motion candidates extension.

field which demonstrates the performance of the local parametric motion computation in the first step of AggregFlow. Indeed, we report in Table 8.1 the objective evaluation given by the Endpoint Error (EPE) scores for the full BCF and BCF without candidates extensions, on the sequences provided with ground-truth in the datasets MPI Sintel and Middlebury. We also compare them with those of motion fields supplied by [Weinzaepfel et al., 2013; Xu et al., 2012b], as obtained with publicly available code. Both BCFs outperform state-of-the-art methods [Weinzaepfel et al., 2013; Xu et al., 2012b] in the two benchmarks. Accuracy is further significantly improved with full BCF, especially for the MPI Sintel sequences where large displacements and wide occluded regions are present. It demonstrates that the combination of local affine estimations in square patches with patch correspondences as described in Section 8.1, is quite relevant and sufficient

Table 8.1: EPE-all scores of motion fields on sequences with ground-truth from MPI MPI SINTEL and MIDDLEBURY datasets

	MPI SINTEL	MIDDLEBURY
Full BCF	**0.792**	**0.0710**
BCF w/o candidates extension	1.851	0.0833
DeepFlow [Weinzaepfel et al., 2013]	4.691	0.386
MDP-Flow2 [Xu et al., 2012b]	4.006	0.223

to recover very accurate motion fields. The challenge now is to select the best velocity vector among the motion candidates at every pixel.

8.4 Occlusion confidence map

In Section 8.2, the occlusion map o was assumed to be known, and we addressed the occlusion filling problem by recovering motion candidates for occluded pixels from non-occluded areas. The occlusion detection task, that is the determination of o, will be performed through the two steps of AggregFlow. In the first step, we compute a coarse occlusion confidence map, which will be used in the aggregation to guide the estimation. Our procedure is simple and exploits the patch distribution and $\mathcal{P}_{\mathcal{S},\alpha}$ and the correspondences used for motion candidates estimation. Nevertheless, from a more general point of view, other coarse occlusion confidence map could be designed differently, for example in the framework of [Kervrann et al., 2011].

We first perform a coarse occlusion detection at the patch level. We consider the smallest patch size s_1 of the set $\mathcal{S}$ defined in Section 8.1 and detect the occluded patches of the set $\mathcal{P}_{s_1,\alpha}$. A common and simple occlusion detection consists in checking the consistency of forward and backward estimated motion vectors [Humayun et al., 2011; Ince and Konrad, 2008; Mozerov, 2013]. We apply the same principle to patches of $\mathcal{P}_{s_1,\alpha}$. Simplifying the notations of Section 8.1 for the sake of readability, let us denote T_P^f the forward displacement between a patch $P \subset I_1$ and its matched patch $M_P \subset I_2$, and T_P^b the backward displacement between M_P and its matched patch in I_1. The forward-backward consistency criterion states that the patch P is occluded if $\|T_P^f + T_P^b\| > \nu$, where ν is a threshold. We then infer a patch-based occlusion map o_P as follows:

$$o_P(x) = \begin{cases} 1 & \text{if } \exists P \in \mathcal{P}_{s_1,\alpha}(x) \text{ such that } P \text{ is occluded} \\ 0 & \text{otherwise.} \end{cases} \tag{8.8}$$

Let us now consider the point set $\mathcal{X}_{o_P}$ composed of the centers of each occluded patch $\mathcal{X}_{o_P} = \{x \in \Omega : \exists P \in \mathcal{P}_{s_1,\alpha},\ x \text{ is the center pixel of } P\}$. We use the density of this point

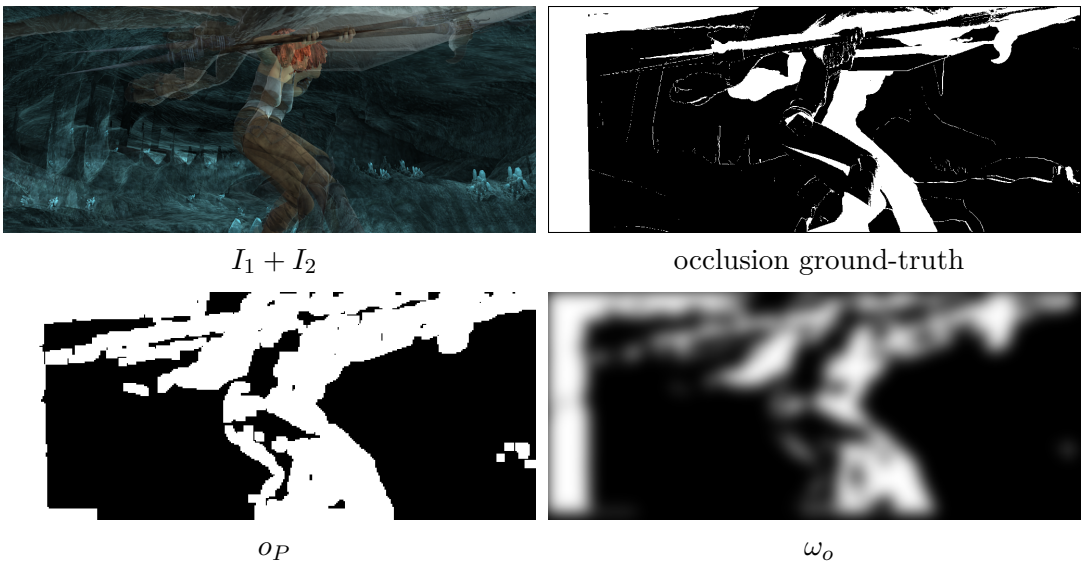


Figure 8.5: Illustration of patch-based occlusion detection. First row: Overlap of the two successive input images and occlusion ground-truth. Second row: Corresponding computed patch-based occlusion map o_P and occlusion confidence map ω_o.

set as an indicator of the presence of occlusions. We apply a Parzen density estimation on $\mathcal{X}_{o_P} = \{x_1, \ldots, x_{N_P}\}$, with N_P the number of occluded patches:

$$\omega_o(x) = \frac{1}{N_P} \sum_{i=1}^{N_P} \frac{1}{\sigma} K\left(\frac{x - x_i}{\sigma}\right), \tag{8.9}$$

where σ is the bandwidth parameter and we choose K to be a Gaussian kernel. We set $\sigma = s_1$. The occlusion confidence map ω_o is thus built as a probability density of the occlusion state. Figure 8.5 shows an example of o_P and ω_o. The map ω_o will be exploited in the aggregation stage to guide a sparsity-constrained occlusion detection.

The output variables of the whole first step are the motion candidates set $\mathcal{C}_f(x)$ and the occlusion confidence map ω_o. They will be exploited in the aggregation stage described in the following chapter, to generate final motion and occlusion fields.

9 Discrete aggregation

The first step of AggregFlow yields the set of motion candidates $\mathcal{C}_f(x)$ along with the occlusion confidence map ω_o. They are aggregated in the second step of AggregFlow to produce the global flow field $\mathbf{w} : \Omega \to \mathbb{R}^2$ and the final occlusion map $o : \Omega \to \{0, 1\}$. The analysis of the Best Candidate Flow in Section 8.3 has shown that the set of candidates at each pixel contains at least one motion vector very close to the ground truth. Therefore, we conceive the aggregation as the selection of the best candidate at every pixel. To this end, we formulate the aggregation as a discrete optimization problem, where the discrete finite motion vector space at each pixel x is composed of the motion candidates $\mathcal{C}_f(x)$. The occlusion map will be estimated jointly with the motion field while exploiting the occlusion confidence map ω_o. The aggregation step amounts to the minimization of the global energy function $E(\mathbf{w}, o)$ as follows:

$$\{\widehat{\mathbf{w}}, \widehat{o}\} = \underset{\{\mathbf{w},o\}}{\arg\min}\ E(\mathbf{w}, o) \ \text{ s.t. } \mathbf{w}(x) \in \mathcal{C}_f(x),\, o(x) \in \{0, 1\}. \tag{9.1}$$

In the following, we detail the design of $E(\mathbf{w}, o)$ and the optimization strategy we have adopted.

9.1 Global energy

The aggregation energy is composed of four terms:

$$E(\mathbf{w}, o) \ = \ E_{data}(\mathbf{w}, o, I_1, I_2) \ + \ E_{occ}(o, \omega_o) \ + \ E^{\mathbf{w}}_{reg}(\mathbf{w}) \ + \ E^{o}_{reg}(o). \tag{9.2}$$

In the following we describe the modeling assumption leading to each term.

9.1.1 Data term E_{data}

The data term accounts for the relations between motion, occlusion and input images. At non-occluded pixels, i.e., $o(x) = 0$, we rely on the usual constancy assumption of image intensity and of spatial image gradient, and we robustly penalize the deviation from the constraints. The potential ρ_{vis} associated to non-occluded (or visible) pixels is given by:

$$\rho_{vis}(x, \mathbf{w}) = \phi(I_2(x + \mathbf{w}(x)) - I_1(x)) \ + \ \gamma\phi(\nabla I_2(x + \mathbf{w}(x)) - \nabla I_1(x)). \tag{9.3}$$

where ϕ is the L_1 norm and γ balances intensity and gradient constancy constraints. Resorting to discrete optimization allows us to use the non-linearized brightness constancy constraint. Thus, coarse-to-fine scheme is not required to cope with large displacements, and we avoid drawbacks related to the loss of small objects with large displacements.

At occluded pixels, no correspondence can be established by definition, and consequently none image feature constancy constraint can be formulated. Therefore, coherently with the motion candidate extension of the first step, we define an exemplar-based constraint for occluded pixels, encoded in the potential ρ_{occ}:

$$\rho_{occ}(x, \mathbf{w}, m) \;=\; \|\mathbf{w}(x) - \mathbf{w}(m(x))\|^2 \tag{9.4}$$

where $m(x)$ is the visible pixel matched with x as obtained in (8.5). The motion vector of an occluded pixel is thus expected to be similar to the motion vector of its matched non-occluded pixel. The data term is finally formed by incorporating the selection of either the visible or the occlusion potential using the occlusion map:

$$E_{data}(\mathbf{w}, o, I_1, I_2) = \sum_{x \in \Omega} (1 - o(x))\, \rho_{vis}(x, \mathbf{w}) \;+\; \lambda_1\, o(x)\, \rho_{occ}(x, \mathbf{w}, m). \tag{9.5}$$

Contrary to other occlusion filling methods which only cancel the visibility term ρ_{vis} in occluded areas and fill the occlusions with motion vectors by diffusion [Ayvaci et al., 2012; Xu et al., 2012b; Papadakis et al., 2013], the potential ρ_{occ} acts as a valid data constraint at occluded pixels.

Concerning the occlusion detection (i.e., the optimization on o), the data term favors the selection of the occluded label at pixels where the data constancy constraint is strongly violated. The continuous approach of [Ayvaci et al., 2012] operates in a similar way. In [Ayvaci et al., 2012], the conservation constraint score is balanced by an estimated continuous residual intensity field, from which occluded points are retrieved by thresholding. In contrast, our occlusion map is binary by nature, and strongly prevents the influence of irrelevant data-constancy constraints on motion estimation in occluded areas, as in the recent work of [Papadakis et al., 2013].

9.1.2 Occlusion constraint E_{occ}

The constraint (13.4) favours detection of occluded pixels and must be counterbalanced by another constraint penalizing occlusion occurrence. It is defined as

$$E_{occ}(o, \omega_o) = \lambda_2 \sum_x \omega_o(x) o(x) \tag{9.6}$$

where ω_o is the occlusion confidence map computed in the first stage. The penalty of occlusion occurrence can be interpreted as a sparsity constraint on the binary occlusion

field o. A sparsity constraint for occlusion detection was also proposed in [Ayvaci et al., 2012] in a continuous setting, and in [Papadakis et al., 2013] for a binary occlusion variable, but without confidence map.

If we set $\forall x \in \Omega, \omega_o(x) = 1$, which would be similar to what is done in [Ayvaci et al., 2012; Papadakis et al., 2013], the data-driven occlusion detection would boil down to the data term (9.5) and (9.6) would be a pure sparse prior constraint. The detection of the occlusion map would be then too tightly coupled with the currently estimated motion field. We would face a chicken-and-egg problem, where o is determined by $\mathbf{w}$, which also depends on o. The consequence on the alternate optimization scheme would be a rapid trap into a local minimum.

Illustrations are given in Fig. 9.1. The results of two variational methods without occlusion handling [Brox and Malik, 2011; Weinzaepfel et al., 2013] are displayed in Fig. 9.1 (e,f). In both cases, the motion field in the occluded region, highlighted by the red bounding box, is wrongly estimated because no occlusion detection is performed. If the occlusion map is initialized to $o(x) = 0,\ \forall x \in \Omega$, the occlusion terms of our energy (9.2) are canceled in the very first iteration of the alternate optimization, which results in a similar behaviour to the methods [Brox and Malik, 2011; Weinzaepfel et al., 2013]. If $\forall x \in \Omega, \omega_o(x) = 1$, the convergence remains trapped in the initial local minimum, as displayed in Fig. 9.1 (g,h). The reason is that the occlusion map is determined by the motion field and cannot deviate from the output of the first iteration. The role of the confidence map ω_o is then to act as an additional evidence for occlusion detection, relaxing the coupling between $\mathbf{w}$ an o. The guidance of ω_o enables to deviate from the output of the first iteration and to converge to the result shown in Fig. 9.1 (i,j).

9.1.3 Regularization terms E^1_{reg} and E^2_{reg}

The regularization term $E^{\mathbf{w}}_{reg}(\mathbf{w})$ enforces piecewise smoothness of the motion field:

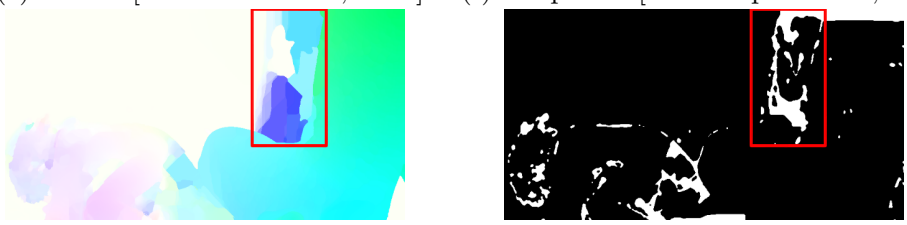
$$E^{\mathbf{w}}_{reg}(\mathbf{w}) = \lambda_3 \sum_{<x,y>} \beta(x)\phi(\|\mathbf{w}(x) - \mathbf{w}(y)\|^2) \tag{9.7}$$

where $< x, y >$ denotes the two-site clique issued from the 8-neighborhood system. The weights $\beta(x)$ are given by $\beta(x) = \exp\left(-\|\nabla I^0_1(x)\|^2/\tau^2\right)$ to modulate the regularization according to the intensity edge strength. To limit the influence of noise and textured regions on the weights, we consider a smoothed version I^0_1 of I_1 obtained with the L_0 smoothing of [Xu et al., 2011], favouring piecewise constant images and preserving only the abrupt edges.

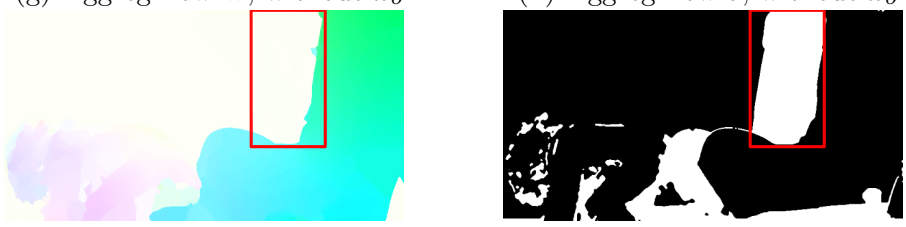
(a) I_1 (b) I_2

(c) ground truth $\mathbf{w}$ (d) Ground truth o

(e) LDOF [Brox and Malik, 2011] (f) DeepFlow [Weinzaepfel et al., 2013]



(g) AggregFlow $\mathbf{w}$, without ω_o (h) AggregFlow o, without ω_o



(i) AggregFlow $\mathbf{w}$, with ω_o (j) AggregFlow o, with ω_o

Figure 9.1: Influence of the occlusion confidence map ω_o on motion and occlusion estimation. (e),(f): variational methods [Brox and Malik, 2011; Weinzaepfel et al., 2013] without occlusion handling. (g),(h): similar behaviour of our method without occlusion confidence map and impact on the occlusion detection. (i),(j): output of AggregFlow when integrating the occlusion confidence map.

It is also important to impose smoothness of the occlusion map with the term E^o_{reg}:

$$E^o_{reg}(o) = \lambda_4 \sum_{<x,y>} (1 - \delta(o(x) = o(y))), \tag{9.8}$$

where δ designates the Kronecker function equal to 1 if its argument is true. The term $E^o_{reg}(o)$ completes the exemplar-based occlusion filling described in Section **??** with diffusion-based occlusion filling.

9.2 Optimization

The optimization problem (9.1) is addressed by alternating minimization of $\mathbf{w}$ and o. The initial value of o is given by the coarse patch-based occlusion detection o_P defined in (8.8). The matching variable m attached to the exemplar-based candidates extension is initialized with m_o defined in (8.5) and recomputed after each update of the occlusion map. Convergence was empirically observed after three iterations in most cases. Thus, to avoid unnecessary computational cost, we fix the number of iterations to 3 for all sequences. Table 9.1 gives an overview of the main steps of AggregFlow. Hereafter, we give details on the minimization procedure concerning $\mathbf{w}$ and o.

When $\widehat{\mathbf{w}}$ is fixed, the energy to optimize w.r.t. o amounts to:

$$\begin{aligned}\min_o \sum_{x\in\Omega} (1 - o(x))\, \rho_{vis}(x, \widehat{\mathbf{w}}) \;+\; \lambda_1\, o(x)\, \rho_{occ}(x, \widehat{\mathbf{w}}, m) \\ + \lambda_2 \sum_x \omega_o(x) o(x) \;+\; \lambda_4 \sum_{<x,y>} (1 - \delta(o(x) = o(y))).\end{aligned} \tag{9.9}$$

Since o takes binary values and the pairwise term is submodular, this problem can be solved exactly with standard graph cut method [Boykov et al., 2001].

The optimization w.r.t. $\mathbf{w}$ with $\widehat{o}$ fixed is more difficult. The reduced energy function writes:

$$\begin{aligned}\widehat{\mathbf{w}} = \min_{\mathbf{w}} \sum_{x\in\Omega} (1 - \widehat{o}(x))\, \rho_{vis}(x, \mathbf{w}) \;+\; \lambda_1\, \widehat{o}(x)\, \rho_{occ}(x, \mathbf{w}, m) \\ + \lambda_3 \sum_{<x,y>} \beta(x) \phi(\|\mathbf{w}(x) - \mathbf{w}(y)\|^2).\end{aligned} \tag{9.10}$$

Our overall label set is the union of the sets of motion candidates over the image points. As a consequence, our discrete optimization problem has several specific features:

1. *Large number of labels*: we have a large number of motion candidates (typically around 200) at each point.

2. *Local labels*: the motion candidates are generated by local (or semi-local) estimations. Therefore, proposal label field to be fused with the current labeling cannot be naturally derived, as in fusion-move.

3. *Redundancy of the labels*: the motion candidate set is redundant, that is, it usually contains groups of similar candidates (generated by estimation on similar patch-based supports).

The number of labels is an important limitation of existing discrete optimization algorithms. Message passing methods like belief propagation [Felzenszwalb and Huttenlocher, 2006] and TRW-S [Kolmogorov, 2006] can be applied to spatially varying label sets, as investigated in [Ulén and Olsson, 2013] for stereo, but we found these methods to be too slow for the minimization of (9.2). An alternative is to resort to graph-cut *move-making* methods [Boykov et al., 2001], generalized in [Lempitsky et al., 2008] to spatially varying label sets. In this setting, each *move* is a binary optimization problem defined on an auxiliary variable selecting between global proposals. The design and ordering of the proposals has a crucial impact on the result [Nieuwenhuis et al., 2013].

Move-making algorithms [Boykov et al., 2001; Komodakis et al., 2008; Lempitsky et al., 2008; Rother et al., 2007; Kohli and Torr, 2007] proceed by iterative modifications of the current labelling, called *moves*. A move is achieved by tackling the original optimization problem in a restricted label space making the solving easier. The restricted space is composed of the current labelling augmented by a *move-space*. In order to apply efficient $s-t$ mincut algorithms, the move-space is usually defined as a proposal labelling, turning the problem into a binary optimization : keep the current label or switch to the proposal label. Extension to larger move-spaces enabling to select between several labels at a time, is possible when a natural ordering can be defined on the range of involved labels [Veksler, 2007].

Building appropriate move-spaces at each iteration is important to ensure the quality of the minimum reached by the algorithm. The popular α-expansion [Boykov et al., 2001] defines each move-space as a constant labelling and passes in turn over all labels in an arbitrary order. Such move-spaces allow for optimal moves when the energy is submodular. In Fusion-move [Lempitsky et al., 2008], proposals are independent estimations performed with several methods. Such elaborated move-spaces come at the price of non-submodularity and only suboptimal moves can be achieved using quadratic pseudo-boolean optimization (QPBO) [Rother et al., 2007].

While several works have focused on improving or speeding up the moves themselves [Kohli and Torr, 2007; Komodakis et al., 2008], they all use fixed pre-defined move-spaces. However, the final result is highly dependent upon the succession order of the proposals as emphasized in [Nieuwenhuis et al., 2013]. Therefore, it is not sufficient to have the right label somewhere in the label space, but it should also be included in a move-space

at the wright iteration. The move-space must *i)* be adapted to the energy, *ii)* be updated at each iteration in accordance with the current labelling. α-expansion satisfies none of the two conditions, whereas fusion-move fulfills the first one. To the best of our knowledge, only the two works [Batra and Kohli, 2011] and [Ishikawa, 2009] actually take into account both requirements. In [Batra and Kohli, 2011], the move-space is chosen so as to maximize the primal-dual gap. In [Ishikawa, 2009], the move-space is constructed by gradient descent of the energy, which must be differentiable.

In our case the motion candidates are locally determined. In contrast, [Lempitsky et al., 2008] exploits global flow fields that can be directly used as proposals in the *move-making* process. Thus, we have to build global flow field proposals at each iteration from the local motion candidates computed in patches. The important point is to ensure spatial smoothness of the proposals, in accordance with the regularization term of the model (9.12). Therefore, we build a global flow field proposal by considering a tiling of non-overlapping patches of a given size and by selecting at every pixel in each patch the motion candidate precisely issued from that patch. This construction maintains the spatial coherency of the local affine estimations. We build as many global proposals as necessary to reasonably explore the motion candidate space. Designing more general and adaptive move-space generation over the *move-making* iterations is an important line of research to improve our method.

Another issue arises from the non-local interaction involved in the exemplar-based term $\rho_{occ}(x, \mathbf{w}, m)$. To make the optimization problem tractable, we transform $\rho_{occ}(x, \mathbf{w}, m)$ to a pixel-wise term at each *move-making* iteration by fixing the exemplar-based constraint $\mathbf{w}(m(x))$ to its value at the previous iteration. At a given *move-making* iteration i, denoting $\widehat{\mathbf{w}}^{(i-1)}$ the value of $\mathbf{w}$ at iteration $i-1$, the potential becomes:

$$\rho_{occ}(x, \mathbf{w}, m) = \left\| \mathbf{w}(x) - \widehat{\mathbf{w}}^{(i-1)}(m(x)) \right\|^2. \tag{9.11}$$

In the next Chapter, we analyse the performance of the overall method composed of the candidates computation step and aggregation, through experiments on challenging data.

1. Local step

1.1. Generate the motion candidates sets $\mathcal{C}(x)$ (8.3)

$$\mathcal{C}(x) = \{\mathbf{w}_{P_{1,2}}(x) + \delta\mathbf{w}_{P_{1,2}}(x) : P_1 \in \mathcal{P}_{\mathcal{S},\alpha}(x), P_2 \in \mathcal{M}_N(P_1)\}$$

1.2. Compute patch-based occlusion map o_P
Derive the occlusion confidence map ω_o from o_P

1.3. Compute the matching variables $m_o(x)$ (8.5)

$$m_o(x) = \arg\min_{y \in \mathcal{V}_o(x)} D(I_1, x, y),$$

Extend motion candidates in occluded regions to obtain $\mathcal{C}_f$

$$\mathcal{C}_f(x) = (\mathcal{C}(x) \cup \mathcal{C}(m_o(x))) \cup \{w_{cam}(x)\}$$

Output of the 1^{st} step: $\mathcal{C}_f$, ω_o

2. Global aggregation
Initialize $o = o_P$ and $m = m_o$
Iterate:

2.1. Estimate $\mathbf{w}$ (9.12)

$$\widehat{\mathbf{w}} = \min_{\mathbf{w}} \sum_{x\in\Omega} (1 - \widehat{o}(x))\, \rho_{vis}(\mathbf{w}) \;+\; \lambda_1\, \widehat{o}(x)\, \rho_{occ}(\mathbf{w}, m(x)) + \lambda_3 \sum_{<x,y>} \beta(x)\phi(\|\mathbf{w}(x) - \mathbf{w}(y)\|^2).$$

2.2. Estimate o (9.12)

$$\widehat{o} = \min_{o} \sum_{x\in\Omega} (1 - o(x))\, \rho_{vis}(\widehat{\mathbf{w}}) \;+\; \lambda_1\, o(x)\, \rho_{occ}(\widehat{\mathbf{w}}, m(x)) + \lambda_2 \sum_{x} g_o(x)o(x) + \lambda_4 \sum_{<x,y>} (1 - \delta(o(x) = o(y))).$$

2.3. Update m (8.5)

Output of the 2^{nd} step: $\mathbf{w}, o$

3. Post-processing : weighted median filtering on $\mathbf{w}$

Table 9.1: Overview of AggregFlow

10 Experimental results

We provide in this chapter an extensive evaluation of AggregFlow in reference computer vision benchmarks. We detail the influence of the different aspects of the method, with a particular concern on the impact of our occlusion handling process.

10.1 Implementation details

All the patch correspondences involved in AggregFlow are computed with the PatchMatch algorithm [Barnes et al., 2009] based on the minimal C++ code provided by the authors[1]. A weighted median filtering with bilateral weights [Xu et al., 2012a] is performed as a post-processing step to enhance motion edges as advocated in [Sun et al., 2014]. For the discrete minimization, we use available QPBO and max-flow code[2]. After extensive experimental tests, the aggregation parameters have been set to $\lambda_1 = 5$, $\lambda_2 = 50$, $\lambda_3 = 500$, $\lambda_4 = 20$ for the MPI SINTEL benchmark and to $\lambda_1 = 2$, $\lambda_2 = 10$, $\lambda_3 = 250$, $\lambda_4 = 4.5$ for the MIDDLEBURY dataset. As a representative example (the one used to compare methods), the computation time for the *Urban2* sequence of the MIDDLEBURY benchmark is 27 minutes on a Intel Xeon laptop with 2.20GHz clock speed and 64Gb RAM. Nevertheless, the first step of AggregFlow can be massively parallelized, which should lead to a far less computation cost with a GPU implementation for instance. Most of the computation time is consumed in the patch correspondence sub-step for the largest patch size (106×106 pixels). The determination of the matching variable m is performed with patches of size 11×11.

10.2 Quantitative results on computer vision benchmarks

We have evaluated AggregFlow on the two most representative benchmarks for optical flow: MPI SINTEL flow dataset[3] [Butler et al., 2012] and MIDDLEBURY flow dataset[4] [Baker et al., 2011], which offer different and complementary challenges. We have retained the Endpoint Error measure (EPE) for quantitative evaluation. Results of [Xu et al.,

[1] http://gfx.cs.princeton.edu/pubs/Barnes_2009_PAR/index.php
[2] http://pub.ist.ac.at/ vnk/software.html
[3] http://sintel.is.tue.mpg.de/
[4] http://vision.middlebury.edu/flow/

DeepFlow	Weinzaepfel et al. [2013]
MDP-Flow2	Xu et al. [2012b]
EPPM	Bao et al. [2014]
S2D-Matching	Leordeanu et al. [2013]
Classic+NL	Sun et al. [2014]
FC-2Layers-FF	Sun et al. [2012]
MLDP-OF	Mohamed et al. [2014]

Table 10.1: References of the method names.

2012b] and [Weinzaepfel et al., 2013] reported in Table 10.5 and from Fig. 10.1 to Fig. 10.7 have been obtained with the public codes provided by the authors[5],[6]. The references associated to the name of each method used for comparison are given in Table 10.1.

MPI Sintel flow dataset Sequences of the most recent MPI Sintel benchmark [Butler et al., 2012] are characterized by long-range motion, motion blur, non-rigid motion, and wide occluded areas. Methods are evaluated on two versions of the sequences named *Clean* and *Final*. The Final version adds motion and defocus blur along with atmospheric effects like fog on some sequences. We reproduce in Tables 10.2 and 10.3 public results of the top ranked methods at the submission date of this manuscript (May 22^{nd} 2014), which are available on the MPI Sintel website. Results are analyzed through several indicators: "EPE all" is the average EPE on all the sequences; "EPE matched" and "EPE unmatched" restrict the error measure respectively to regions that remain visible in adjacent frames (non-occluded pixels) and to regions that are visible only in one of two adjacent frames (occluded pixels); "d0-10" denotes EPE over regions closer than 10 pixels to the nearest occlusion boundary, and thus reveals the ability to recover motion discontinuities; "s40+" denotes EPE over regions with velocities larger than 40 pixels per frame. Methods are ranked regarding their EPE all. Visual comparison with results supplied by [Weinzaepfel et al., 2013] and [Xu et al., 2012b] on training sequences (i.e., MPI Sintel sequences provided with ground truth) is available from Fig. 10.1 to Fig. 10.4.

As for the *Clean* subset, our method AggregFlow ranks first over the published methods. The most significant improvement is obtained on the unmatched category, which emphasizes the efficiency of our occlusion framework. AggregFlow is ranked second for the d0-10 metric which demonstrate its capacity to recover motion discontinuities as confirmed by results displayed from Fig. 10.1 to Fig. 10.4. First, it is due to the robust affine estimation of the motion candidates able to capture locally dominant motion in case of two or even several motions present inside patches, which preserves motion discontinuities.

[5]http://www.cse.cuhk.edu.hk/ leojia/projects/flow/
[6]http://lear.inrialpes.fr/src/deepmatching/

	EPE all	EPE matched	EPE unmatched	d0-10	s40+
AggregFlow	**4.754**	**1.694**	**29.685**	3.705	**31.184**
DeepFlow	5.377	1.771	34.751	4.519	33.701
MDP-Flow2	5.837	1.869	38.158	**3.210**	39.459
EPPM	6.494	2.675	37.632	4.997	39.152
S2D-Matching	6.510	2.792	36.785	5.523	44.187
Classic+NLP	6.731	2.949	37.545	5.573	45.290
FC-2Layers-FF	6.781	3.053	37.144	5.841	45.962
MLDP-OF	7.297	3.260	40.183	5.581	51.146

Table 10.2: Results on the MPI Sintel Clean test subset

It is also made successful by the efficient occlusion module, which allows us to moderate the need for motion field regularization. Indeed, large errors in occluded regions are usually alleviated by imposing high regularization with the result of over-smoothing the rest of the motion field (see motion fields computed with DeepFlow [Weinzaepfel et al., 2013] from Fig. 10.1 to Fig. 10.4). In case of very large displacements (s40+ metric), all the first five methods (AggregFlow, [Weinzaepfel et al., 2013; Xu et al., 2012b; Bao et al., 2014; Leordeanu et al., 2013]) somehow integrate feature matching in their motion estimation process to capture the largest deformations. The top rank of AggregFlow demonstrates the efficiency of the aggregation framework for integrating feature matching.

As for the *Final* version AggregFlow is ranked second in terms of EPE all. The slight decreasing in performance compared to the Clean subset is mainly due to large errors caused by the added fog effect in the two *ambush* sequences. As emphasized in [Bao et al., 2014], local intensity-based displacement computation tends to capture the motion of the fog rather than the motion of objects appearing in transparency. As our candidates estimation is local, it is subject to this limitation. Global variational approaches are able to diffuse motion estimates in these regions and are consequently better suited for this kind of situations. Despite this shortcoming, our method still yields significant improvement in unmatched regions and on motion discontinuities. One solution to improve results in fog regions would be to incorporate a more sophisticated feature correspondence technique as the ones proposed in [Leordeanu et al., 2013; Weinzaepfel et al., 2013].

Middlebury **dataset** The Middlebury benchmark is composed of sequences with small displacements, where the main challenge is to be able to recover both complex smooth deformation, motion discontinuities and motion details. Table 10.4 reproduces public results at the submission date of this manuscript (May 22^{nd} 2014) for the same methods as those taken for comparison on the MPI Sintel benchmark. Visual comparative results

	EPE all	EPE matched	EPE unmatched	d0-10	s40+
DeepFlow	**7.212**	**3.336**	38.781	5.650	**44.118**
AggregFlow	7.329	3.696	**36.929**	**5.538**	44.858
S2D-Matching	7.872	3.918	40.093	5.975	48.782
FC-2Layers	8.137	4.261	39.723	6.537	51.349
MLDP-OF	8.287	4.165	41.905	6.345	50.540
Classic+NLP	8.291	4.287	40.925	6.520	51.162
EPPM	8.377	4.286	41.695	6.556	49.083
MDP-Flow2	8.445	4.130	43.430	5.703	50.507

Table 10.3: Results on the MPI Sintel Final test subset

	EPE all	Avg. rank
MDP-Flow2	**0.245**	**7.8**
FC-2Layers-FF	0.283	19.3
Classic+NL	0.319	27.1
EPPM	0.329	32.6
AggregFlow	0.339	35.9
MLDP-OF	0.349	32.6
S2D-Matching	0.347	34.6
DeepFlow	0.416	48.8

Table 10.4: Results on the Middlebury benchmark.

are displayed from Fig. 10.5 to Fig. 10.7. It can be observed that the absolute EPE values, together with the differences between methods, are much lower than on the MPI Sintel dataset. The average EPE score computed over the considered methods is equal to 6.22 for the MPI Sintel Clean subset and to 0.327 for the Middlebury dataset, with respective variance of 0.613 and 0.0025. We also provide the average rank over the 8 test sequences for each method which is the metric used for global ranking on the Middlebury website.

On the whole Middlebury benchmark, AggregFlow is ranked 38 over 97 submitted methods in terms of average rank on the results (evaluated with the average endpoint error on the sequence) obtained on the eight test sequence. Notwithstanding, it is still very close to the ranked two MDP-Flow2 method [Xu et al., 2012b] in terms of EPE metric, knowing that the top ranked published method OFLAF [Kim et al., 2013] has an average rank of 6.8 and an EPE all of 0.197 (OFLAF method was not tested on the MPI Sintel benchmark). Visual results reported from Fig. 10.5 to Fig. 10.7 confirm the tightness of performance gap. In particular, the preservation of motion discontinuities with AggregFlow is more satisfying than with the DeepFlow method [Weinzaepfel et al.,

	AggregFlow	**AggregFlow w/o occlusion**	DeepFlow	MDP-Flow2
ambush_2	**5.632**	9.456	14.743	12.083
ambush_4	**11.923**	16.515	14.647	15.570
ambush_5	**5.042**	5.4996	8.333	6.591
ambush_6	**5.854**	6.251	9.928	8.466
market_5	**9.957**	11.958	15.056	12.816
market_6	**3.626**	4.547	6.606	5.384
cave_2	**6.029**	8.228	10.082	8.347
cave_4	**3.706**	4.185	4.234	3.815
temple_3	**5.875**	8.314	11.895	9.011
Average	**6.002**	8.417	10.614	9.120

Table 10.5: Results on the MPI Sintel training subset. Scores correspond to the EPE all metric.

2013]. These results also show that AggregFlow is competitive for recovering motion details in addition to the large velocities of the MPI Sintel benchmark.

10.3 Occlusion handling

As aforementioned, the impact of our occlusion framework on optical flow estimation is demonstrated by the EPE unmatched metric scores obtained on the MPI Sintel benchmark (Tables 10.2 and 10.3). Results from Fig. 10.1 to Fig. 10.4 reveal the superiority of AggregFlow in coping with occluded regions. Since the occlusion framework is composed of several elements, we detail the influence of each one in the following. The efficiency of the motion candidates extension in occluded regions has already been highlighted in Section 8.3 and Table 8.1 through the analysis of the Best Candidate Flow.

To evaluate the occlusion model of the aggregation step, we report in Table 10.5 results obtained on a selection of training sequences of the MPI Sintel benchmark with the largest displacements. We distinguish between the full AggregFlow method, and AggregFlow without the occlusion-related terms in (9.2), that is, by setting $\lambda_1 = 0$, $\lambda_2 = 0$ and $\lambda_3 = 0$. The improvement due to the occlusion terms is clearly significant since the average EPE is 8.417 for AggregFlow without occlusions and 6.002 for full Aggregflow. It can also be noticed that even without handling occlusion AggregFlow still performs better than competing methods. The role of the occlusion confidence map involved in the sparsity constraint (0.17) has already been explained and illustrated in Section 9.1.2 and Fig. 9.1.

Recovered occlusion maps are displayed from Fig. 10.1 to Fig. 10.7. For the large occluded regions of Fig. 10.1 to Fig. 10.4 for which ground truth is available, the estimated occlusion map is correct in most cases. A specific behaviour is particularly prominent in

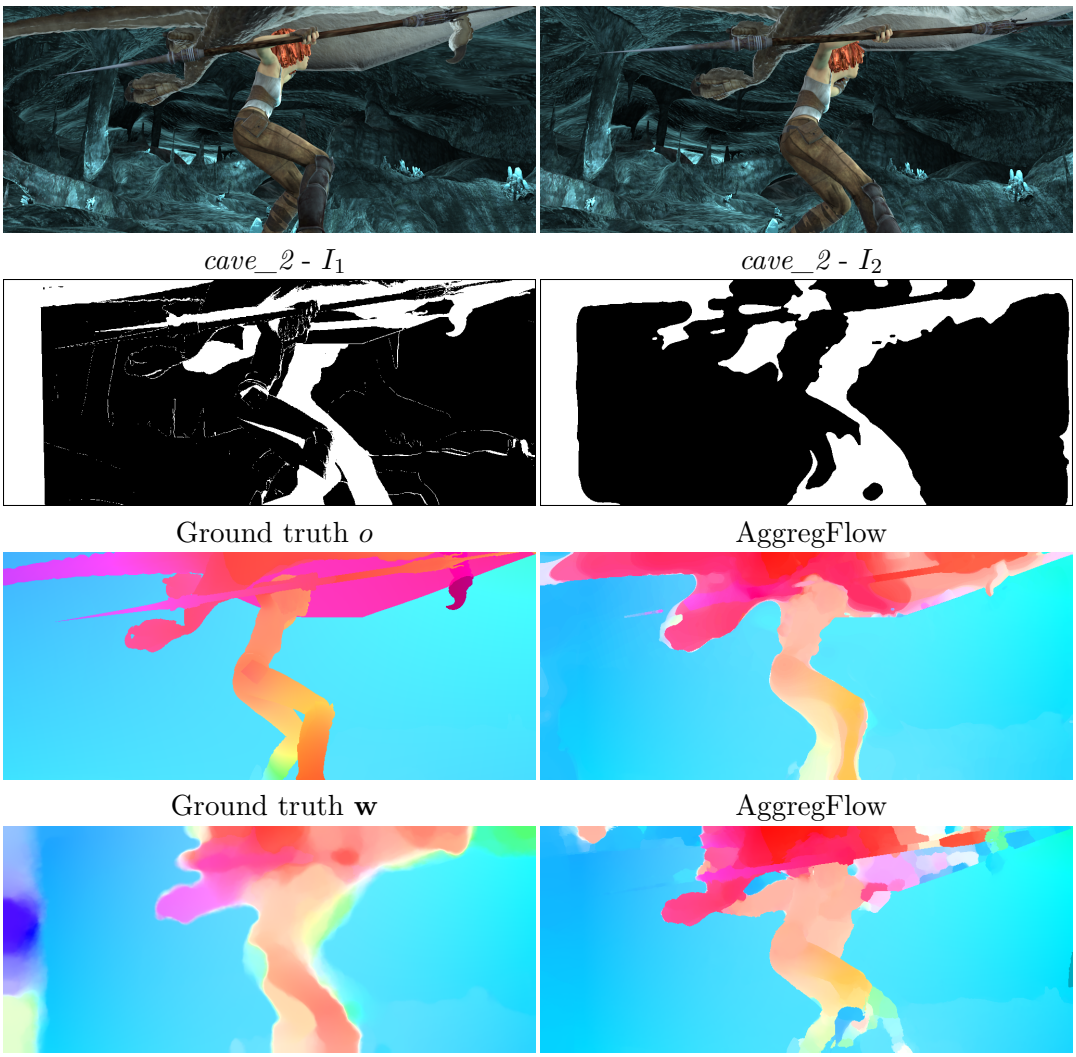


Figure 10.1: Comparative evaluation with [Weinzaepfel et al., 2013] and [Xu et al., 2012b] on the *cave_2* sequence of the MPI SINTEL dataset. First row: successive input images. Second row: ground truth occlusion and occlusion map computed with AggregFlow. Third row: ground truth motion field and motion field computed with AggregFlow. Fourth row: motion fields computed resp. with DeepFlow [Weinzaepfel et al., 2013] and MDP-Flow2 [Xu et al., 2012b].

10.3 example, where occlusions are over-detected. This is due to the modeling assumption stating that occluded regions correspond to large violations of the data conservation constraint. Large regions of illumination changes can thus be detected as occlusions. While it leads strictly speaking to wrong occlusion detection, it can still be beneficial to motion estimation by treating illumination changes.

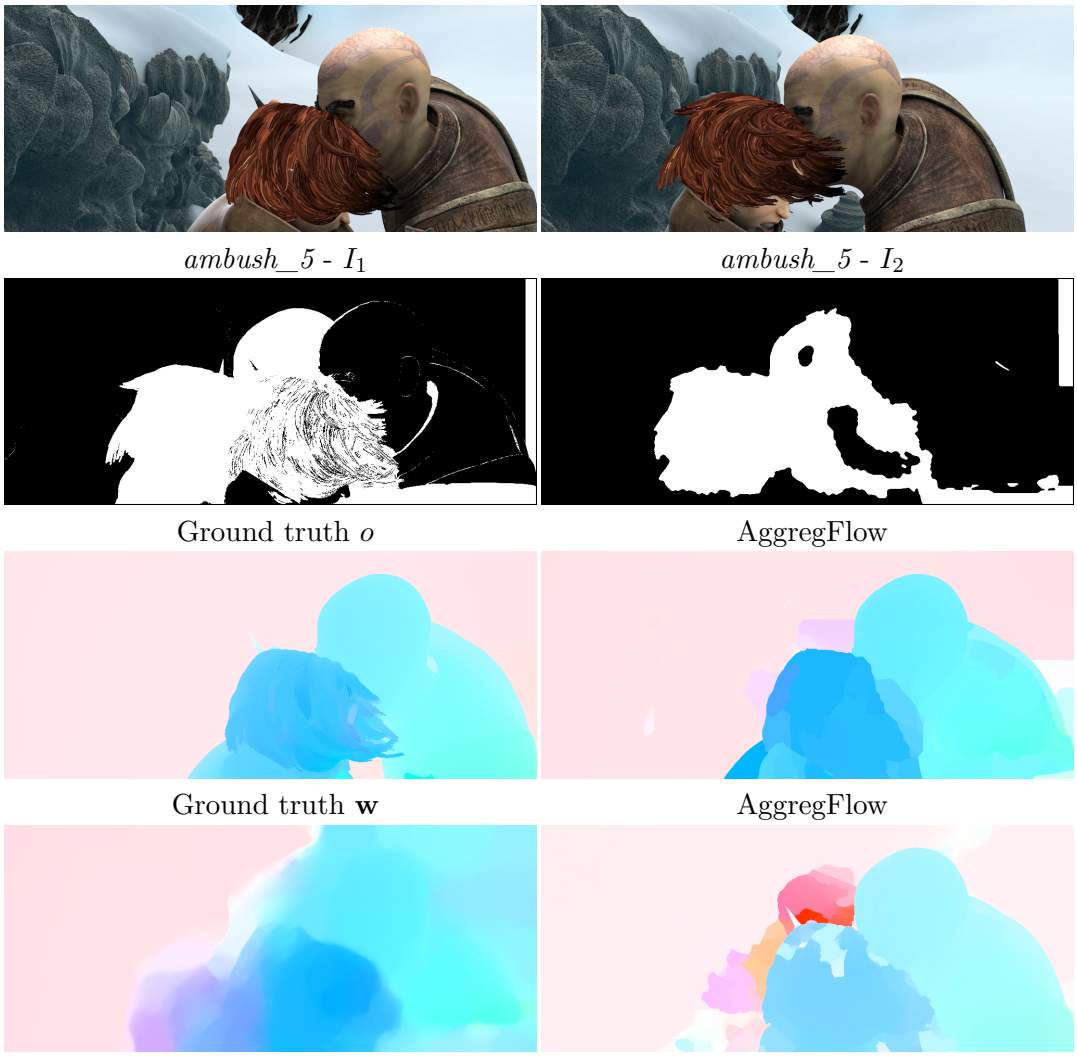


Figure 10.2: Comparative evaluation with [Weinzaepfel et al., 2013] and [Xu et al., 2012b] on the *ambush_5* sequence of the MPI SINTEL dataset. First row: successive input images. Second row: ground truth occlusion and occlusion map computed with AggregFlow. Third row: ground truth motion field and motion field computed with AggregFlow. Fourth row: motion fields computed resp. with DeepFlow [Weinzaepfel et al., 2013] and MDP-Flow2 [Xu et al., 2012b].

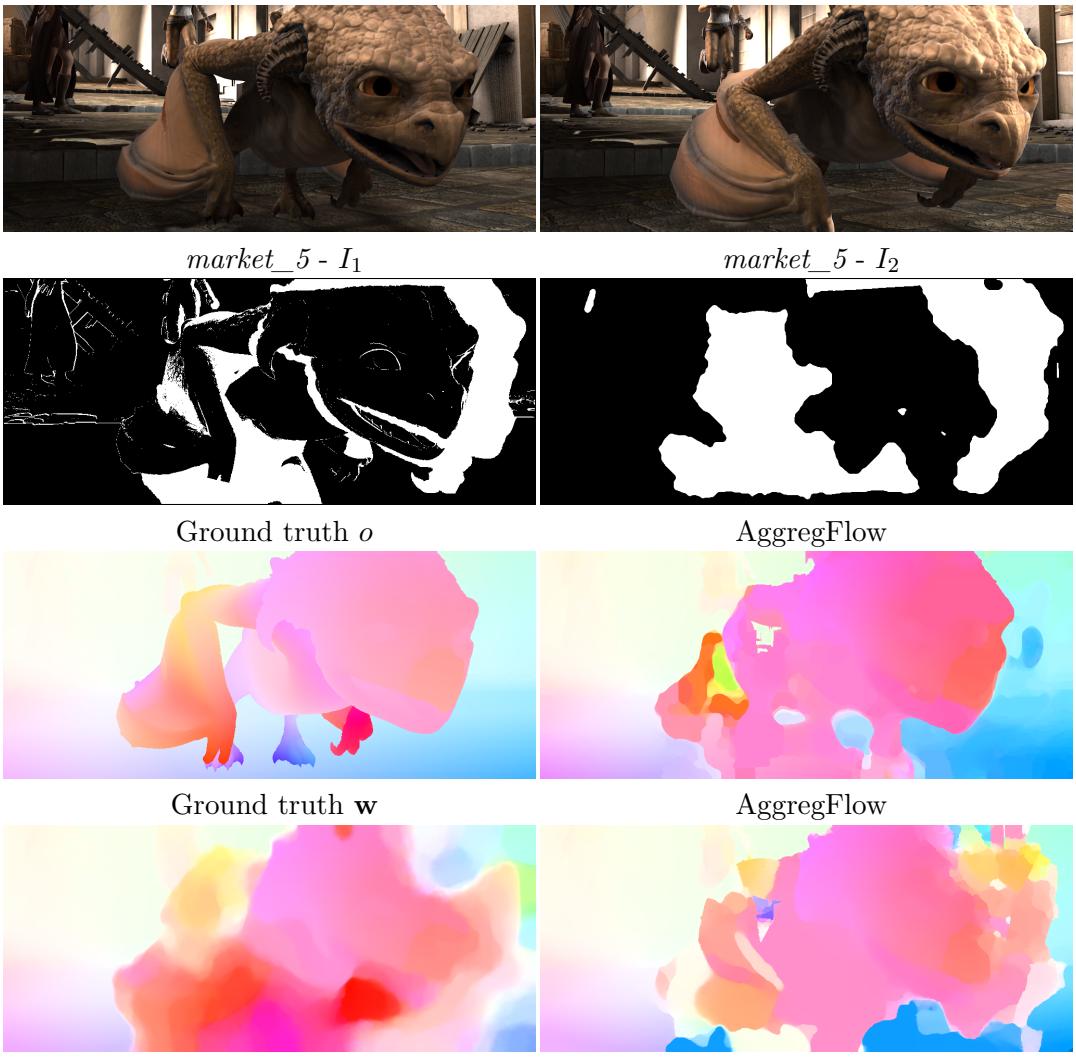


Figure 10.3: Comparative evaluation with [Weinzaepfel et al., 2013] and [Xu et al., 2012b] on the *market_5* sequence of the MPI Sintel dataset. First row: successive input images. Second row: ground truth occlusion and occlusion map computed with AggregFlow. Third row: ground truth motion field and motion field computed with AggregFlow. Fourth row: motion fields computed resp. with DeepFlow [Weinzaepfel et al., 2013] and MDP-Flow2 [Xu et al., 2012b].

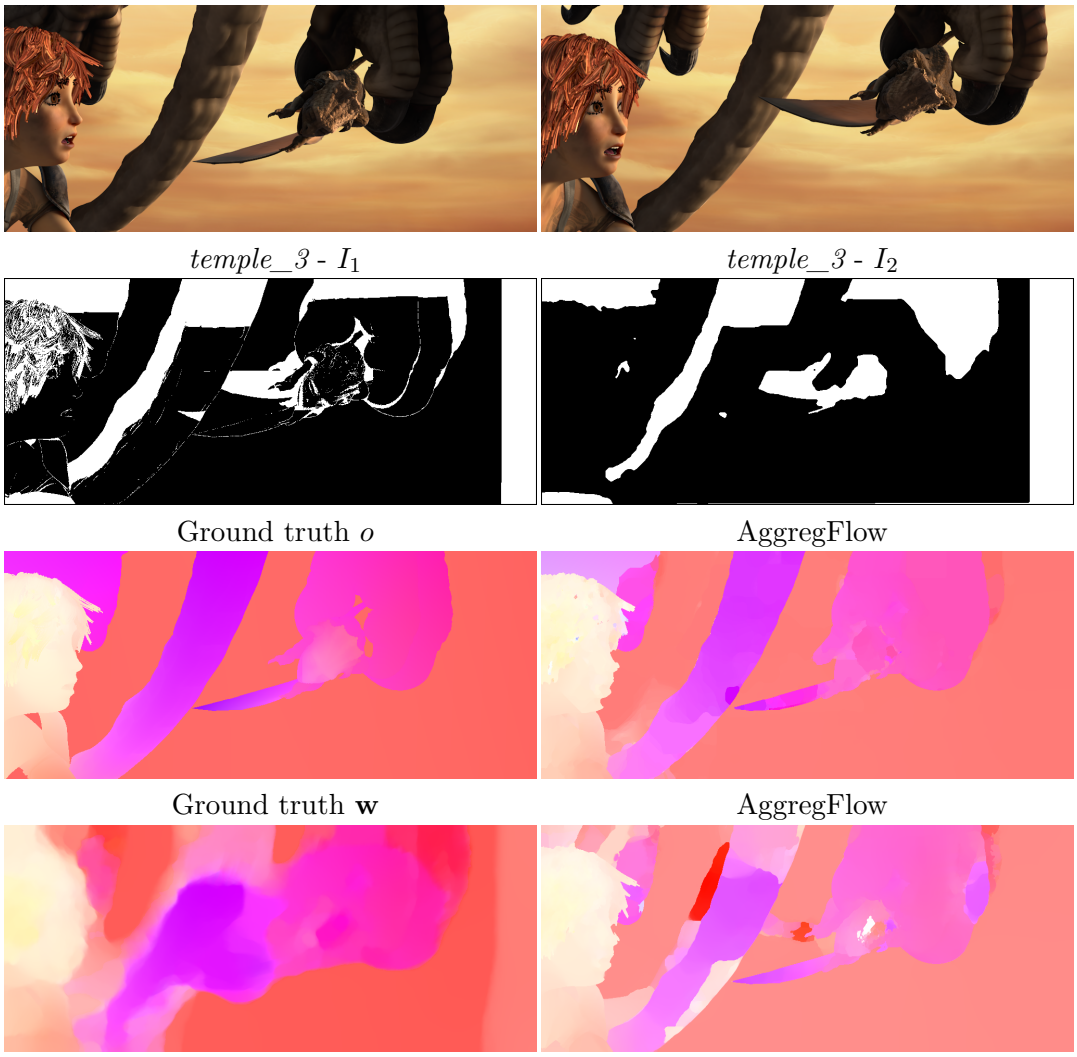


Figure 10.4: Comparative evaluation with [Weinzaepfel et al., 2013] and [Xu et al., 2012b] on the *temple_3* sequence of the MPI SINTEL dataset. First row: successive input images. Second row: ground truth occlusion and occlusion map computed with AggregFlow. Third row: ground truth motion field and motion field computed with AggregFlow. Fourth row: motion fields computed resp. with DeepFlow [Weinzaepfel et al., 2013] and MDP-Flow2 [Xu et al., 2012b].

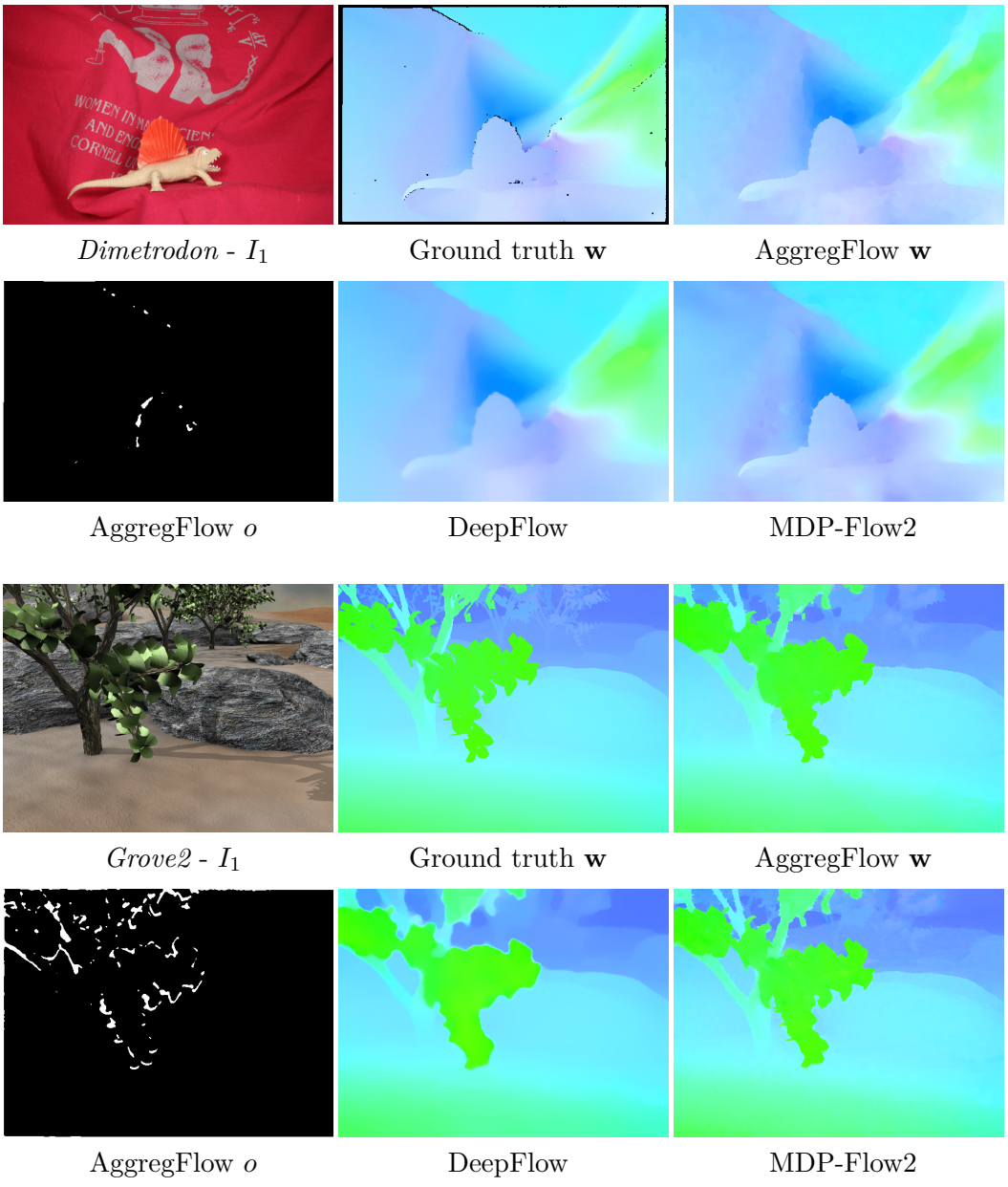


Figure 10.5: Comparative evaluation with [Weinzaepfel et al., 2013] and [Xu et al., 2012b] on the *Dimetrodon* and *Grove2* sequences of the MIDDLEBURY dataset. For each sequence, from left to right: in the first row, first input image, ground truth motion field, motion field computed with AggregFlow; in the second row, occlusion map computed with AggregFlow, motion field computed with DeepFlow [Weinzaepfel et al., 2013] and MDP-Flow2 [Xu et al., 2012b].

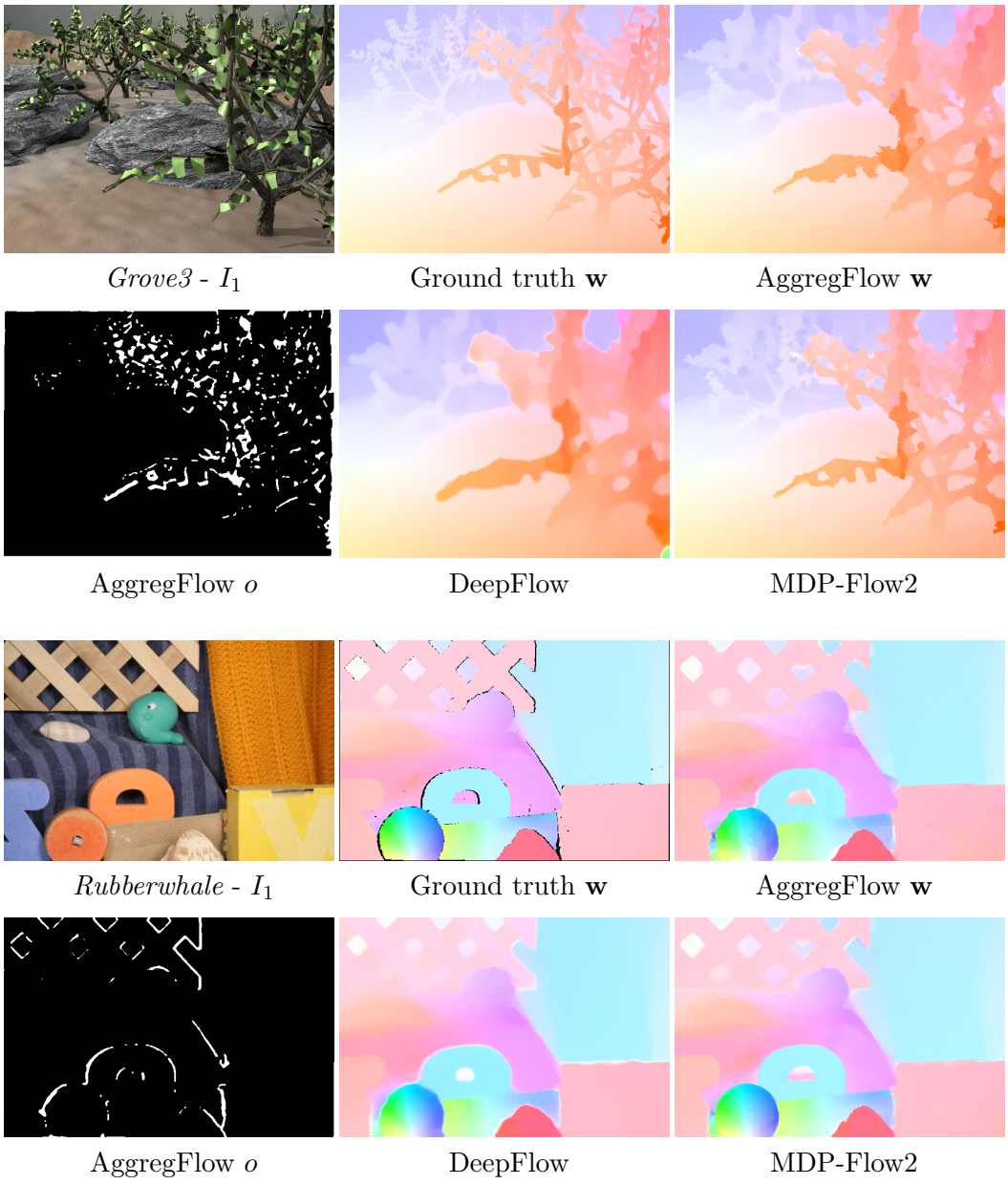


Figure 10.6: Comparative evaluation with [Weinzaepfel et al., 2013] and [Xu et al., 2012b] on the *Grove3* and *RubberWhale* sequences of the MIDDLEBURY dataset. For each sequence, from left to right: in the first row, first input image, ground truth motion field, motion field computed with AggregFlow; in the second row, occlusion map computed with AggregFlow, motion field computed with DeepFlow [Weinzaepfel et al., 2013] and MDP-Flow2 [Xu et al., 2012b].

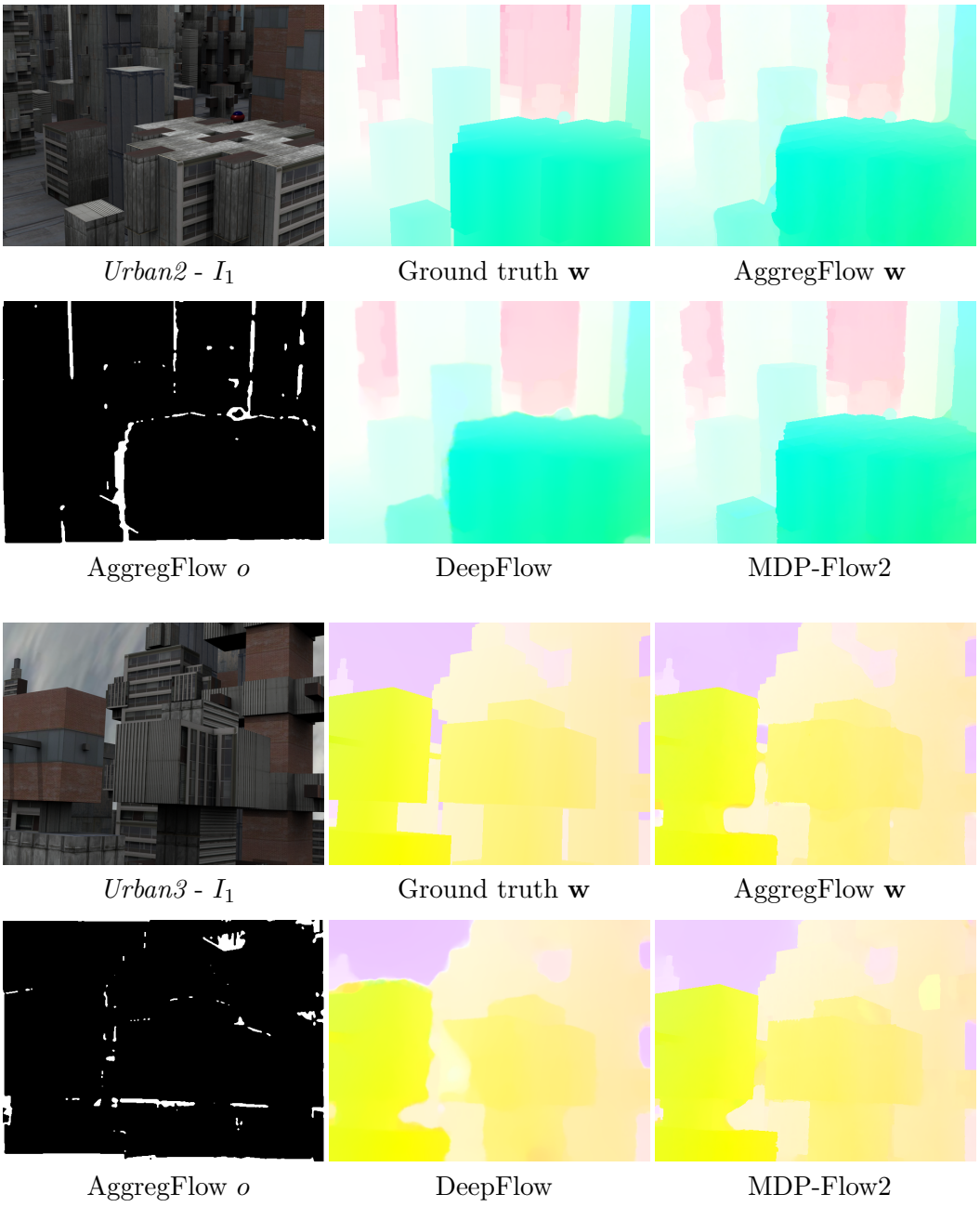


Figure 10.7: Comparative evaluation with [Weinzaepfel et al., 2013] and [Xu et al., 2012b] on the *Urban2* and *Urban3* sequences of the MIDDLEBURY dataset. For each sequence, from left to right: in the first row, first input image, ground truth motion field, motion field computed with AggregFlow; in the second row, occlusion map computed with AggregFlow, motion field computed with DeepFlow [Weinzaepfel et al., 2013] and MDP-Flow2 [Xu et al., 2012b].

11 Another strategy: aggregation in a continuous setting

In this section, we present an aggregation strategy in a continuous setting which can be considered as an alternative to the discrete aggregation method described in Chapter 9. Despite the convincing results obtained with the *move-making* approach described in Section 9.2, we have pointed out the issues raised by the large label set composed of the motion candidates. The difficulty to build global proposals from locally estimated candidates for the *move-making* optimization, and the important computational time were the two limitating issues. The continuous aggregation we propose in this section is not affected by the local nature of candidates estimation, and achieves lower computational time. While the quantitative results of this approach are globally less convincing than those obtained with the discrete aggregation, it still brings several local improvements and advantages.

In a continuous setting, we minimize an energy of the form

$$E(\mathbf{w}) = \int_{\Omega} \rho_{data}(x, \mathbf{w}, \mathcal{C}_f) + \lambda_1 \phi(\nabla \mathbf{w}(x)) dx, \tag{11.1}$$

where $\rho_{data}(x, \mathbf{w}, \mathcal{C}_f)$ is the data term and the second term imposes smoothness of $\mathbf{w}$, balanced by the parameter λ_1. In the following, we consider a Total Variation regularization: $\phi(\nabla \mathbf{w}(x)) = \|\nabla \mathbf{w}(x)\|_1$. Unlike usual approaches for optical flow discussed in Part I, the input images are not a parameter of the data potential $\rho_{data}(x, \mathbf{w}, \mathcal{C}_f)$, but are replaced by the motion candidates set $\mathcal{C}_f$. It means that in the continuous aggregation stage, the data is not the image sequence, but the motion candidates. Thus, the potential $\rho_{data}(x, \mathbf{w}, \mathcal{C}_f)$ does not encode constancy of image feature, but a relationship between $\mathbf{w}$ and the candidates set $\mathcal{C}_f$.

Minimizing in the continuous domain w.r.t. $\mathbf{w}$ implies that the estimated motion field is allowed to deviate from the motion vectors of $\mathcal{C}_f$. This could seem contradictory with the BCF analysis of Section 8.3 showing that the hard selection of one candidate is sufficient to outperform existing methods. However, from a practical point of view, one could be interested in achieving good results even when the set of candidates is less accurate. Critical parameters for the computational cost of the method, such as the overlapping ratio α, could then be adapted to speed up candidates computation.

An alternative approach could be to follow the class of methods presented in Section 5.2, considering both images and candidates as input data. In this case the motion candidates contribute to additional constraint to usual intensity constancy constraint (see (5.1)). Drawbacks of this approach are given in Section 5.2.

11.1 Candidate distribution

If the motion candidates set is considered as the input data for the aggregation stage, we have to study the distribution of the candidates. Figure 11.1 illustrates the 2D distribution of candidates $\mathcal{C}_f(x)$ at several locations in an image (blue points), while also plotting the ground truth motion vector among them (red triangle).

We can first observe that it is not always possible to identify modes of the distribution of motion candidates. While in regions of constant or smoothly varying motion most motion vectors are clustered around the same mode, the distribution at motion discontinuities is unpredictable. Secondly, when a mode exists, the ground truth motion vector does not always correspond to this mode. The best motion candidate is sometimes isolated from the rest of the candidates. As a matter of fact, the two cases (absence of modes and isolated best candidate) frequently occur in all types of sequences.

We can conclude that the candidates distribution is not a relevant information for modelling of the continuous aggregation. Options like linear combination of candidates, fitting of a statistical distribution or clustering are then excluded.

11.2 Continuous aggregation

We propose two versions for the data potential of (11.1). We impose $\rho_{data}(x, \mathbf{w}, \mathcal{C}_f)$ to be a measure of proximity of $\mathbf{w}$ to the set of candidates $\mathcal{C}_f$. However, it must not be a distance measure to a mode of $\mathcal{C}_f$ or a weighted average of candidates, as pointed out in the previous section. We rather define it as the distance to a single appropriately selected candidate from $\mathcal{C}_f$. Therefore, as for discrete aggregation, we still aim at selecting one candidate, but we will exploit it as a constraint in the data potential.

11.2.1 Minimum distance constraint

The selection of one candidate can be achieved by the following data potential:

$$\rho_{data}(x, \mathbf{w}, \mathcal{C}_f) = \min_{\mathbf{w}_c \in \mathcal{C}_f(x)} \|\mathbf{w}(x) - \mathbf{w}_c\|_1. \tag{11.2}$$

The min function naturally selects one candidate used for distance measure. However, the non-differentiability of the min function makes the minimization difficult. Methods

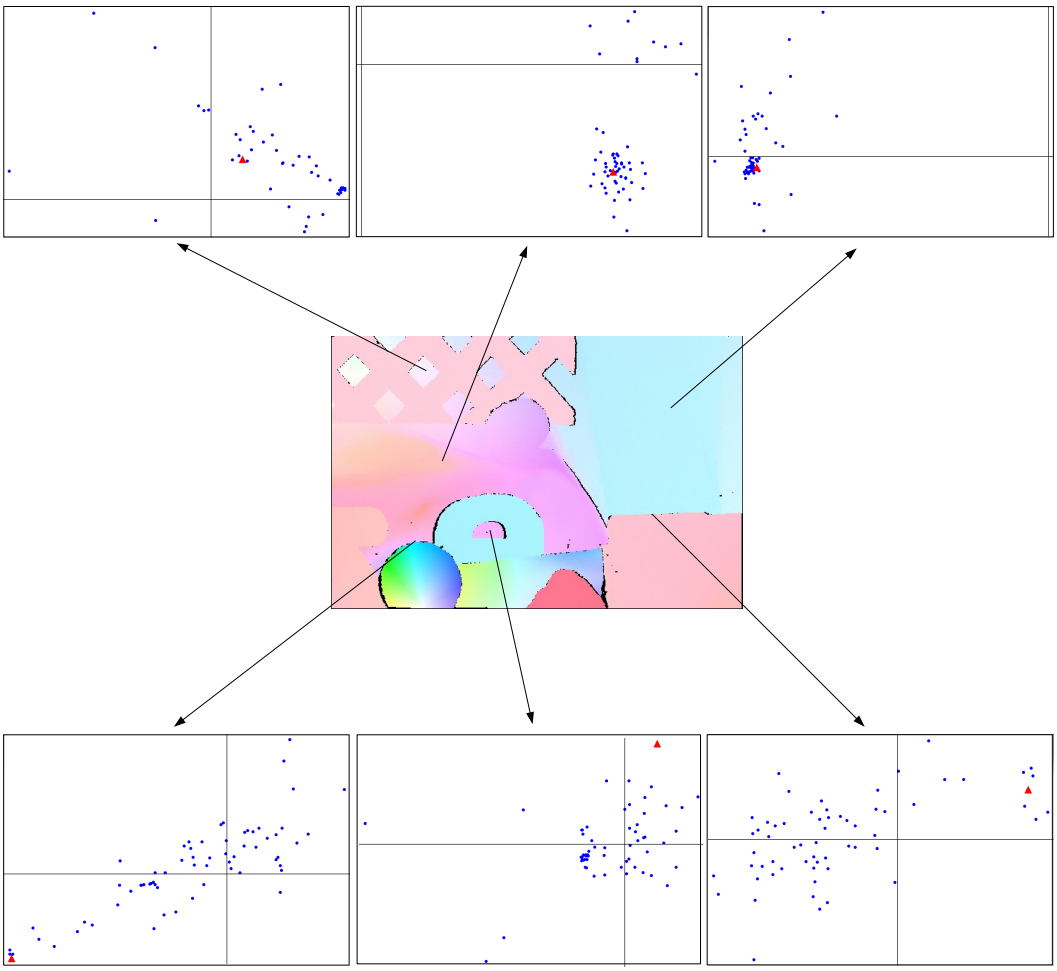


Figure 11.1: Visualization of the distribution of the motion candidates $\mathcal{C}_f(x)$ at several pixels x. The central image is the ground motion field of the *RubberWhale* sequence of the MIDDLEBURY benchmark. The six plots represent the motion vector candidates and the motion vector ground truth at each corresponding pixel. The horizontal and vertical axes are respectively the horizontal and vertical components of the motion vectors. Blue points are motion candidates and red triangle are ground truth motion vectors.

based on gradient descent or on resolution of Euler-Lagrange equations cannot be applied here. To minimize (11.1) with the data potential (11.2) we use the primal-dual approach of [Chambolle and Pock, 2011]. The main steps of the method are the computation of proximal operators:

$$\begin{cases} \arg\min_{\mathbf{u}} \int_\Omega \frac{1}{2\varepsilon}\|\mathbf{u}(x) - \widehat{\mathbf{v}}(x)\|_2^2 + \lambda_1 \|\nabla \mathbf{u}(x)\|_1 dx \\ \arg\min_{\mathbf{v}} \int_\Omega \frac{1}{2\varepsilon}\|\widehat{\mathbf{u}}(x) - \mathbf{v}(x)\|_2^2 + \min_{\mathbf{w}_c \in \mathcal{C}(x)} \|\mathbf{v}(x) - \mathbf{w}_c\|_1 dx. \end{cases} \tag{11.3}$$

The first problem corresponds to a ROF model [Rudin et al., 1992] and can be efficiently solved with the method of [Chambolle, 2004]. The second problem is pixel-wise and can be solved as follows. First, by inversing the min functions, we can write:

$$\min_{\mathbf{v}} \left\{ \|\widehat{\mathbf{u}}(x) - \mathbf{v}(x)\|_2^2 + \min_{\mathbf{w}_c \in \mathcal{C}_f(x)} \|\mathbf{v}(x) - \mathbf{w}_c\|_1 \right\} = \min_{\mathbf{w}_c \in \mathcal{C}_f(x)} \left\{ \min_{v} \left\{ \|\widehat{\mathbf{u}}(x) - \mathbf{v}(x)\|_2^2 + \|\mathbf{v}(x) - \mathbf{w}_c\|_1 \right\} \right\}. \quad (11.4)$$

The subproblem

$$\min_{\mathbf{v}} \left\{ \|\widehat{\mathbf{u}}(x) - \mathbf{v}(x)\|_2^2 + \|\mathbf{v}(x) - \mathbf{w}_c\|_1 \right\} \quad (11.5)$$

can be solved very efficiently by thresholding, as in [Zach et al., 2007]. The thresholding operation being very fast, it is computationally tractable to solve the minimization problem (11.5) for all the motion vectors $\mathbf{w}_c \in \mathcal{C}_f(x)$. Thus, the minimization problem (11.4) can be solved by exhaustive search over $\mathcal{C}_f(x)$.

The problem of the potential $\rho(x, \mathbf{w}, \mathcal{C}) = \min_{w_c \in \mathcal{C}(x)} \|\mathbf{w}(x) - \mathbf{w}_c\|_1$ is its high non convexity, leading inevitably to local minima. In practice, we experimentally observe a convergence of the algorithm, but it stays trapped in a local minimum very dependent on the initialization.

In the next section, we adopt a second modeling to relax the selection of a unique candidate and achieve more efficient minimization.

11.2.2 Sparse dictionary constraint

We want to keep the idea of selecting one candidate, but with a convex formulation. To this end, we define a relaxed data potential with an auxiliary variable $\boldsymbol{\alpha}$, imposing proximity to a sparse linear combination of candidates:

$$\rho_{data}(\mathbf{w}, \boldsymbol{\alpha}) = \left\| \mathbf{w}(x) - \boldsymbol{\alpha}(x)^\top \mathbf{W_c}(x) \right\|_1 + \lambda_2 \|\boldsymbol{\alpha}(x)\|_1 \quad (11.6)$$

where $\boldsymbol{\alpha}(x) = (\alpha_1(x), .., \alpha_{n(x)}(x))^\top$ is a sparse coefficient vector associated to the $n(x)$ candidates at pixel x, $\mathbf{W_c}(x) = (\mathbf{w}_1(x), .., \mathbf{w}_{n(x)}(x))^\top$ is the candidates set written as a vector, and λ_2 balances the influence of the two terms. The first term imposes a reconstruction by a linear model, considering the candidates as a motion dictionary. The second term imposes sparsity of the coefficients $\boldsymbol{\alpha}(x)$. If the balance coefficient λ_2 is high enough, only one or a few components of $\boldsymbol{\alpha}(x)$ will be non null, which will amount to the selection of almost one single candidate in (11.6). Besides, potential (11.6) is convex and thus more easily minimizable than (11.2), while having a similar behaviour.

Considering potential (11.6), we are facing a very similar problem to the one we encountered in Section 9.1.2 with the modelling of the occlusion map. Indeed, in an

alternate optimization scheme, the tight coupling between $\mathbf{w}$ and $\boldsymbol{\alpha}$ could imply that in practice $\boldsymbol{\alpha}(x)$ stays trapped in the local minimum of the first iteration. We overcome this problem similarly to the occlusion case of Section 9.1.2 by replacing the pure sparsity constraint of (11.6) by a weighted sparsity constraint defined by:

$$\|\boldsymbol{\alpha}(x)\|_{1,\boldsymbol{\beta}(x)} = \sum_{i=1}^{n(x)} \beta_i(x) \, |\alpha_i(x)| \tag{11.7}$$

where $\beta_i(x)$ is a confidence measure associated to the i^{th} candidate of pixel x. Apart from [Kondermann et al., 2008; Kybic and Nieuwenhuis, 2011], existing confidence measures are dedicated to specific motion estimation methods. For a variational approach, [Bruhn and Weickert, 2006] uses the inverse of the global energy. For local approaches like [Lucas and Kanade, 1981], eigenvalues of the structure tensor are usually exploited [Mota et al., 2001]. For parametric estimations in general, the variance of the estimate is also a possible confidence measure. To keep the generality and simplicity of our method, we consider the following general weights:

$$\beta_i(x) = \exp\left\{ -\frac{\sum_{y\in\Omega_d} g(x,y,I_1)\rho_0(y,w_i(x),I_1,I_2)}{\sigma^2} \right\} \tag{11.8}$$

where Ω_d is the discrete image grid and $g(x,y,I_1)$ are bilateral weights defined by:

$$g(x,y,I_1) = \exp\left(-\frac{\|x-y\|_2^2}{\sigma_s^2} + \frac{|I_1(x)-I_1(y)|}{\sigma_g^2} \right) \tag{11.9}$$

and $\rho_0(y,w_i(x),I_1,I_2)$ is a classical data potential penalizing deviations from brightness constancy

$$\rho_0(y,w_i(x),I_1,I_2) = |I_2(x+\mathbf{w}(x)) - I_1(x)|. \tag{11.10}$$

The weight $\beta_i(x)$ is then a measure of the local coherency of brightness constancy through bilateral filtering.

The final energy is:

$$E(\mathbf{w},\boldsymbol{\alpha}) = \int_\Omega \left\| \mathbf{w}(x) - \boldsymbol{\alpha}(x)^\top \mathbf{W}_\mathbf{c}(x) \right\|_1 \; + \; \lambda_2 \, \|\boldsymbol{\alpha}(x)\|_{1,\boldsymbol{\beta}(x)} + \lambda_1 \|\nabla \mathbf{w}(x)\|_1 dx. \tag{11.11}$$

We minimize $E(\mathbf{w},\boldsymbol{\alpha})$ alternatively on $\mathbf{w}$ and $\boldsymbol{\alpha}$. Minimization w.r.t. $\mathbf{w}$ is realized by solving the Euler-Lagrange equations with fixed point iterations [Brox, 2005]. To minimize w.r.t. $\boldsymbol{\alpha}$, we resort to a greedy algorithm. From an initial configuration of $\boldsymbol{\alpha}$, we search for possible configurations of $\boldsymbol{\alpha}$, and a configuration is kept if it leads to a decreasing of the energy. The search strategy consists in iteratively adding non null components ordered by decreasing value of the confidence measure.

Table 11.1: Angular errors obtained with AggregFlow-C, AggregFlow-D, [Brox and Malik, 2011] and [Chambolle and Pock, 2011] on sequences of the MIDDLEBURY benchmark.

	Grove2	Grove3	Hydrangea	Urban2	Urban3
SL-D	2.19	5.43	2.47	2.47	3.42
SL-C	2.43	5.92	2.29	2.53	4.12
[Brox and Malik, 2011]	2.38	5.97	2.10	2.50	3.91
[Chambolle and Pock, 2011]	2.92	6.72	2.29	43	6.10

11.3 Results

We have evaluated our method on sequences of the MIDDLEBURY benchmark [Baker et al., 2011]. We provide comparisons with our discrete aggregation method and the methods of [Brox and Malik, 2011; Chambolle and Pock, 2011]. Local improvements related to discontinuity preservation and large displacements are illustrated visually. The candidates computation and discrete aggregation are performed without occlusion handling, and no post-processing is applied on the flow fields. The candidates set was obtained with parameters $\mathcal{S} = \{15, 45, 115\}$, $\alpha = 0.8$, $N = 2$. We refer to the continuous aggregation version as AggregFlow-C, and the discrete version as AggregFlow-D. Other parameters are set to $\sigma = 0.1$, $\sigma_s = 5$ and $\sigma_c = 20$.

Table 1 contains Angular Errors, defined in (1.1), obtained with AggregFlow-C, AggregFlow-D and the variational methods [Brox and Malik, 2011; Chambolle and Pock, 2011] for sequences of the MIDDLEBURY benchmark. The results of AggregFlow-C are globally less accurate than those of AggregFlow-D. It is particularly obvious on sequences with small motion details or sharp motion discontinuities. Another drawback of AggregFlow-C is the impact of the confidence measures β_i on final results. Large errors of confidence measures can significantly decrease the accuracy of AggregFlow-C. Nevertheless AggregFlow-C yields better performance than [Chambolle and Pock, 2011] and is competitive with [Brox and Malik, 2011].

Figure 11.2 illustrates the ability of AggregFlow-C and AggregFlow-D to capture motion discontinuities and small details. Motion fields computed with AggregFlow-C are less sharp than with AggregFlow-D but, they are significantly better than [Brox and Malik, 2011; Chambolle and Pock, 2011].

In Fig. 11.3, the large displacement of the small ball is typically badly handled by variational methods using coarse-to-fine schemes, as [Chambolle and Pock, 2011]. In contrast, AggregFlow-C and AggregFlow-D satisfyingly retrieve the large displacement.

The method of [Brox and Malik, 2011] integrating feature matching in a variational framework also captures the motion of the ball, but the shape of the ball is less preserved. Also, it is more impacted by the associated occlusion region.

Figure 11.4 illustrates the ability of AggregFlow-C to deal with less accurate motion candidates. In this experiment, we set the overlap ratio, setting the proportion of area shared by two neighbor patches, to $\alpha = 0.5$. This parameter is essential to deliver good candidates. In Fig. 11.4, typical artifacts of AggregFlow-D can be observed. At motion discontinuities, the patches are not overlapping enough to produce accurate candidates, which implies block artifacts for AggregFlow-D, due to the hard selection of one candidate. In contrast, AggregFlow-C can deviate from the set of candidates to preserve clean discontinuities.

Discrete aggregation presented in Section 9.2 tends to produce block artifacts for complex smooth deformations, as illustrated in Fig. 11.5. The variational optimization of AggregFlow-C does not have this problem and estimate more accurately smooth flow fields.

Finally, the computational time of AggregFlow-C is around 5 minutes while Aggreg-Flow requires 20 minutes.

11.4 Conclusion

We have proposed an aggregation strategy minimizing a global energy in the continuous setting, as an alternative to the discrete aggregation presented in Chapter 9. A first version uses the min function and is minimized with a primal-dual scheme. It is however limited because of severe non-convexity of the energy. A more attractive convex formulation exploits a sparse dictionary model. Experiments show that the overall quantitative performance remains lower than with discrete aggregation. However results are still competitive with standard variational methods. Moreover, the ability to reconstruct motion vectors different from the candidates set makes the continuous aggregation more robust to smaller and suboptimal candidate sets. It is also better suited in case of complex and smooth motion fields. The computational cost of continuous aggregation is finally lower than for discrete aggregation

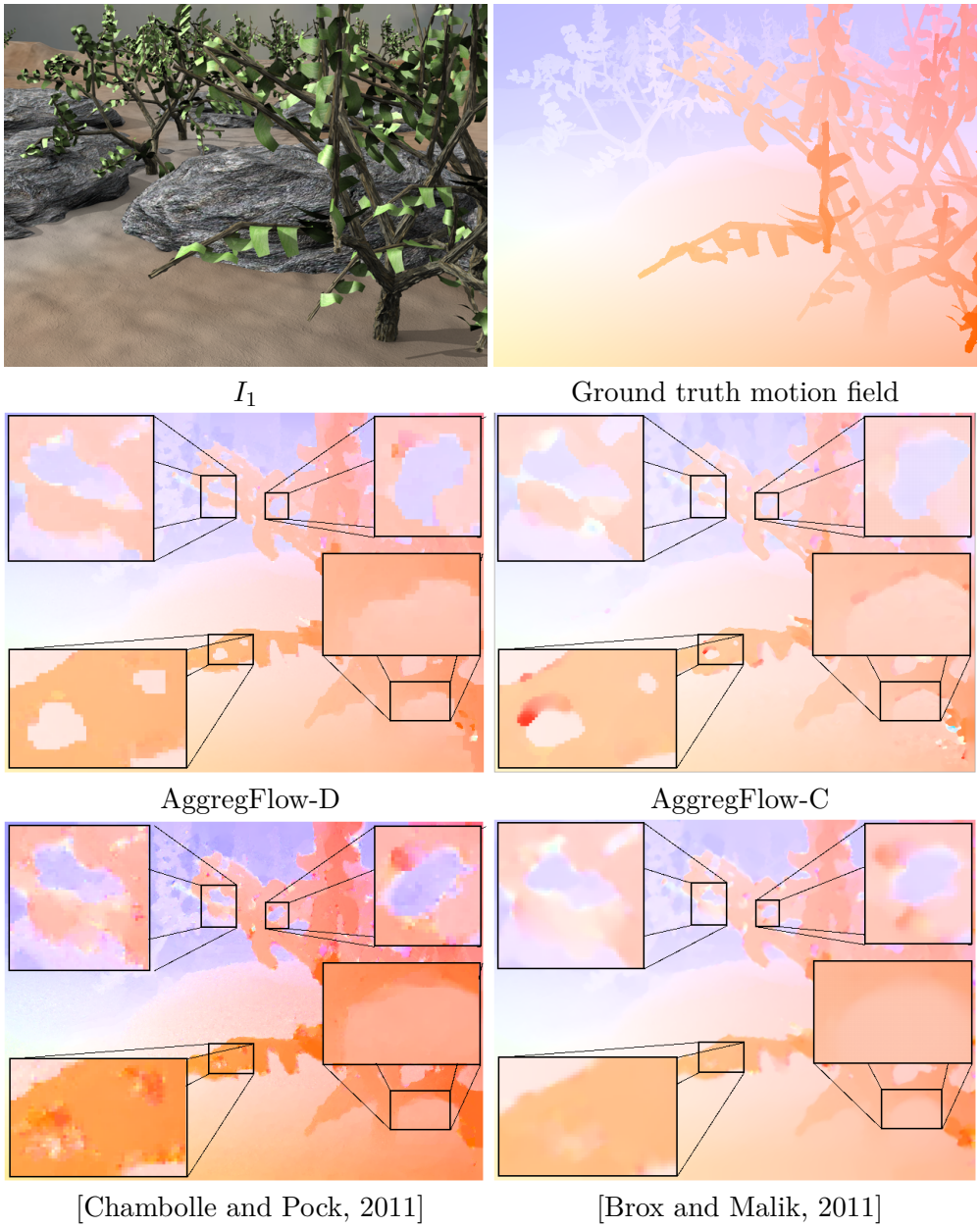


Figure 11.2: Ability of preserving small motion details and discontinuities on the *Grove3* sequence of the MIDDLEBURY benchmark. Top row: first frame and ground truth motion field. Middle row: motion field estimated with AggregFlow-D and AggregFlow-C. Bottom row: motion field estimated with [Chambolle and Pock, 2011] and [Brox and Malik, 2011]. Zooms on regions of interest overlay the images.

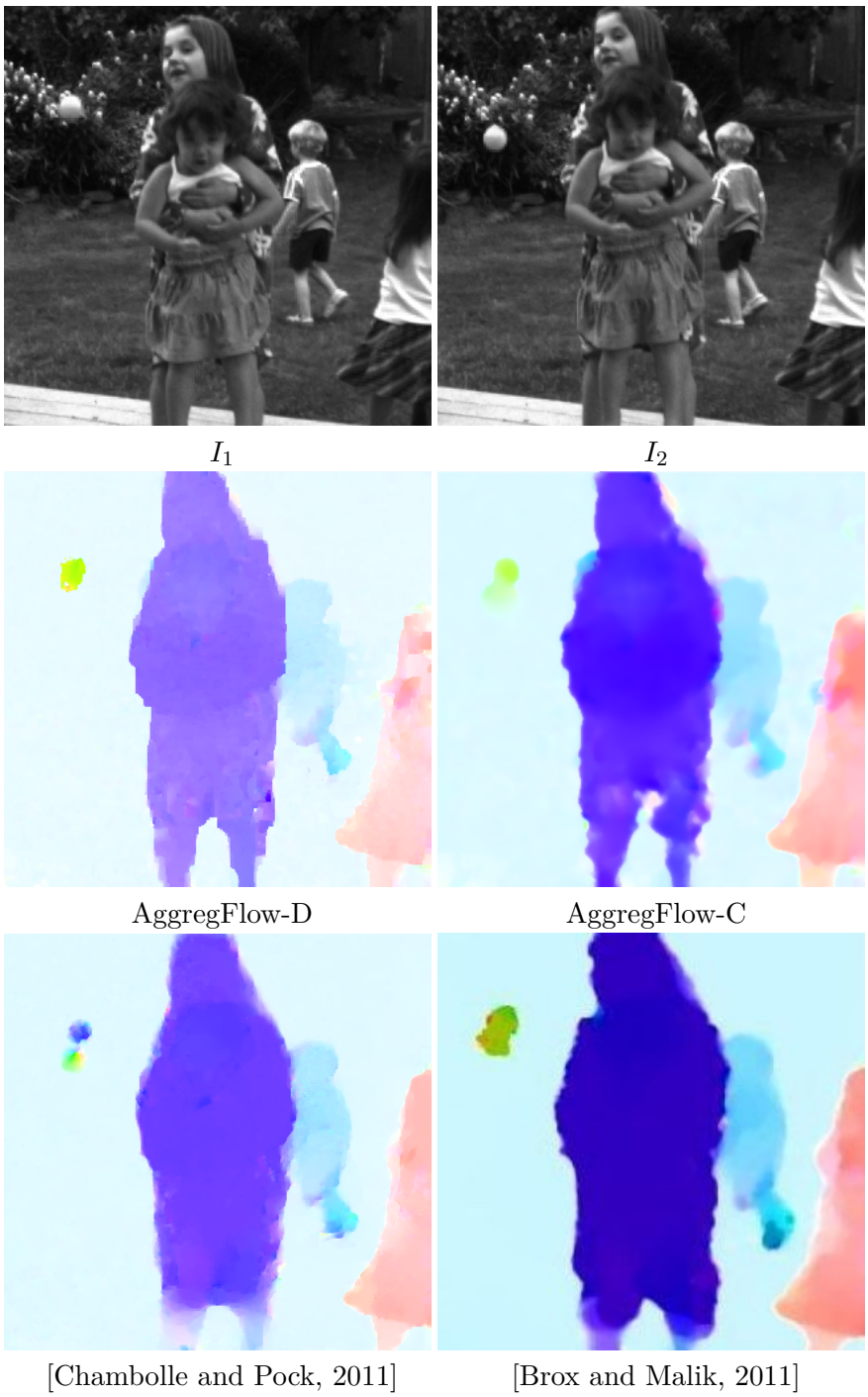
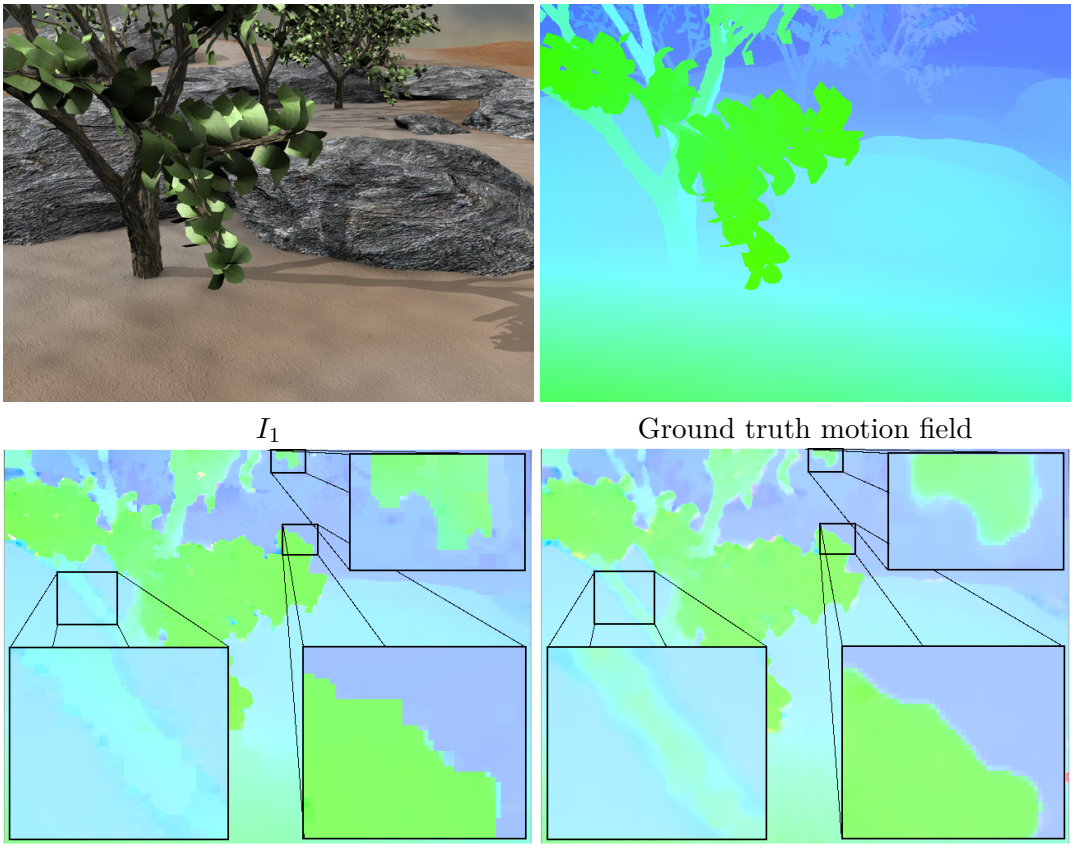
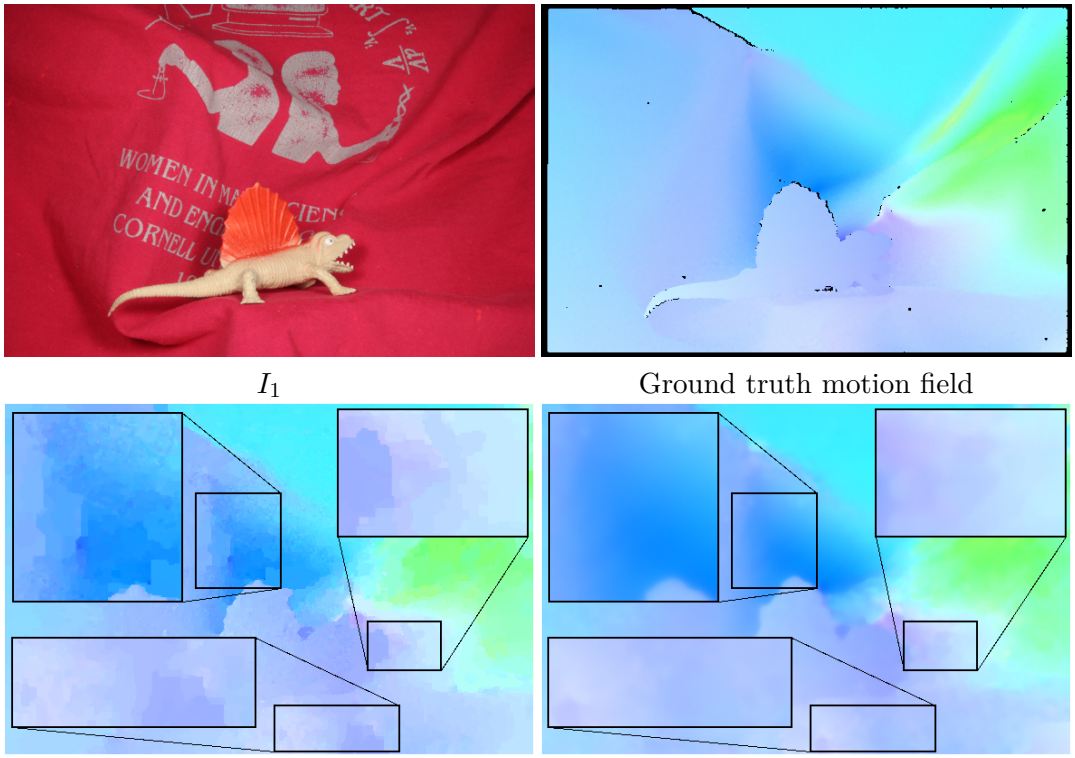


Figure 11.3: Results on the *Backyard* sequence of the MIDDLEBURY benchmark. Top row: first and second frames (ground truth is not available for this sequence). Middle row: motion field estimated with AggregFlow-D and AggregFlow-C. Bottom row: motion field estimated with [Chambolle and Pock, 2011] and [Brox and Malik, 2011].



(a) AggregFlow-D, $\alpha = 0.5$ (b) AggregFlow-C, $\alpha = 0.5$

Figure 11.4: Comparison of discrete and continuous aggregation for a small set of candidates on the *Grove2* sequence of the MIDDLEBURY benchmark. The candidates were computed with $\alpha = 0.5$. Top row: first frame and ground truth motion field. Middle row: motion field estimated with AggregFlow-D and AggregFlow-C. Zooms on regions of interest overlay the images.



I_1 Ground truth motion field

(c) AggregFlow-D (d) AggregFlow-C

Figure 11.5: Comparison of discrete and continuous aggregation for complex and smooth flow fields on the *Dimetrodon* sequence of the MIDDLEBURY benchmark. Top row: first frame and ground truth motion field. Middle row: motion field estimated with AggregFlow-D and AggregFlow-C. Zooms on regions of interest overlay the images.

12 Conclusion and perspectives

We have presented a new two-step optical flow estimation method called AggregFlow. This method combines the computation of local motion candidates and their global aggregation while jointly recovering occlusion maps. The framework is generic, and both the local and global steps could be adapted for specific purposes. We demonstrated the added value of combining patch correspondences and patch-based affine motion estimation to produce highly accurate motion candidates. It advocates the relevance of patch-based parametric motion estimation, provided size and position of the patches are appropriately defined. Candidates estimation with a variational regularized method is also envisaged in Appendix A. The integration of multiple patch correspondences in the candidate generation process allows us to deal with local matching ambiguities. We formulated the aggregation step as a discrete optimization problem, selecting the best motion candidate at every pixel while preserving motion discontinuities and achieving occlusion recovery. The occlusion scheme acts in both steps of AggregFlow. An exemplar-based occlusion term is incorporated in the global aggregation energy. Incidentally, it could be integrated in other estimation paradigms as well, e.g., in variational approaches. Occlusion cues derived from the computed motion candidates are exploited in the sparse modeling of occlusions. Overall, AggregFlow achieves state-of-the-art results on the MPI Sintel benchmark. The most significant improvements are reached in occluded regions and for large displacements. We proposed an alternative aggregation approach operating in a continuous setting. Despite lower global quantitative results, this continuous aggregation is more robust to suboptimal candidates set and computationally more efficient than the discrete method.

Extensions of the method could tackle remaining matching errors in the patch correspondence and in the exemplar search stages. A more elaborate and discriminative distance than the pixel-based L_1 distance could be envisioned for patch matching. Future work could also deal with the GPU implementation of AggregFlow to largely improve computation efficiency. The discrete optimization problem could be adapted to the specific label set constituted by the motion candidates.

Part III

Applications in fluorescence imaging and light microscopy

The methods described in the previous chapters have been designed to cope with the largest possible collection of problems involved in optical flow estimation. Our aim was to identify typical dynamic behaviours (large displacements, discontinuities, occlusions...) and image effects (illumination changes, motion blur...) affecting motion estimation, and to try to take into account all these issues together in a single methodological framework. While this unifying ambition may be appealing, handling too diverse situations and contradictory requirements is likely to end up in a compromise, averaging the performances to satisfy all the conditions. When a well defined problem in a clearly delineated environment is identified, it can be beneficial to focus the methodology on the restricted number of motion an image patterns defined by the application.

In this chapter, we explore such an applicative field and we propose to adapt our modeling approach to images acquired in biological imaging. Estimating motion in live cell imaging is a fundamental task to analyse dynamic properties of biological phenomena [Pinot et al., 2012; Boncompain et al., 2012]. The difficulties arise from the variety of situations encountered in microscopy images. Indeed, we have to deal with different imaging modalities (e.g. fluorescence microscopy, contrast-phase imaging) in order to observe interacting structures with different sizes and shapes such as cells, vesicles, microtubules, with different types of motion and deformation.

A majority of approaches for motion analysis in biological sequences is based on individual tracking of biological objects [Meijering et al., 2006]. However tracking methods are not adapted to answer to a set of biological questions, especially when the density and the lack of prominent features prevent the individual extraction of objects of interest undergoing complex motion (e.g. protrusions or membrane deformation in bio-mechanical studies). Accordingly, estimating the global deformation field can be more appropriate to capture complex dynamics observed in biological sequences [Lecomte et al., 2012; Kim et al., 2011; Delpiano et al., 2011]. In this part, we first compare features and performances of traditionally used correlation-based methods with variational approaches on a set of sequences representing typical dynamic patterns observed in biological imaging. Secondly, we show how the generic aggregation approach described in Part II is able to handle a variety of problems proposed in biological imaging. We adapt our method to the specific purpose of large intensity changes caused by fluorescence variations. Finally we are interested in analyzing the diffusion of particles, often occurring in intra-cellular processes. We propose a variational method for diffusion coefficient estimation and compare it with the standard approach based on correlation measures

Introduction to fluorescence microscopy

In this part, we will be mainly interested in sequences acquired by fluorescence imaging. We provide here a short description of fluorescence and Green Fluorescence Protein (GFP) tagging for application to microscopy and live cell imaging.

Basics in fluorescence In 1852, George G. Stokes observed for the first time the fluorescence phenomena corresponding to the light emitted by a mineral (fluospar) excited by ultra-violet lights. He noticed that that the emitted wavelength was longer than the incident wavelength.

Formally, the electron energy is known to be depend on its orbital. Due to the quantum nature of electron energy, the molecular energy is quantized in several discrete states. For each molecule, the lowest energy level is the so-called ground state. When a photon hits a molecule at its ground energy state, several electrons undergo an orbital leap. If the photon has sufficient energy, it can reach an excited quantum of higher energy (absorption phenomenon). In order to return to the stable ground state, an usual de-excitation pathway is the immediate emission of a photon. This rapid light emission that happens within nanoseconds (10^{-9} to 10^{-10} seconds) is the so-called fluorescence.

Fluorescent staining Fluorescence staining consists in the injection inside the cell of a fluorescent probe. This probe is introduced in the form of a fluorochrome chemically linked to a biological vector molecule that binds the protein targeted for visualization. The most popular tagging fluorescent protein is probably Green Fluorescence Protein (GFP) which has become a major tool for biologists to tag and to quantify dynamics of specific target proteins *in vivo*. The fluorescently labelled protein (or chimera) is a mutant protein resulting from the fusion of two protein genes (using genetic recombination technique) forming a recombinant DNA. The recombinant DNA is then transfected into the cell using a plasmid or a viral vector (transduction). Inside the cell, the recombinant DNA is translated by the ribosome to produce a specific amino-acid chain that will later folds into a fluorescent protein.

Photobleaching and photoxicity Two main limitations in the image acquisition process in fluorescence microscopy are photobleaching and phototoxicity.

Photobleaching (or fading) is a permanent loss of fluorescence of a fluorophore. A fluorescent molecule in the excited state presents a varying probability of interaction with an other molecule, a reaction that will cause irreversible covalent modifications. This probability depends on the fluorophore and on the molecular environment. Photobleaching is an issue that seriously hinder the acquisition of strong signal or time lapse acquisition in fluorescence microscopy. It is thus very important to manage the fluorescent capacity

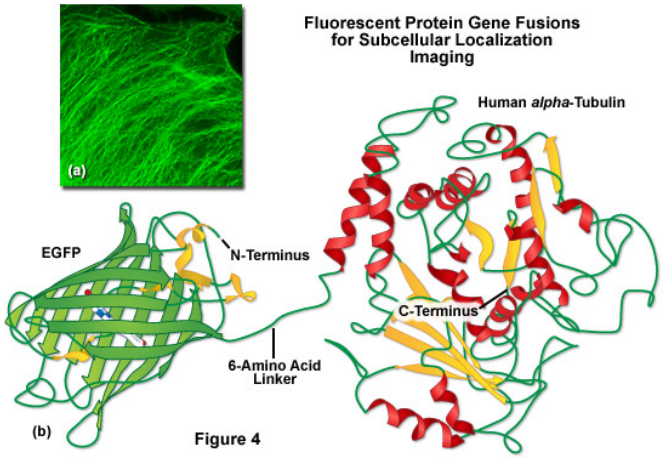


Figure 12.1: α-tubulin labeled with GFP. **a)** this chimera localize to microtubule **b)** the EGFP is tagged on the N-terminus of the α-tubulin (copyright Zeiss).

by controlling the sample excitation during acquisition. Photobleaching effect can also be exploited positively to measure biophysical quantities in fluorescence microscopy. For example, in fluorescence recovery after photobleaching (FRAP) experiments, the fluorophores are intentionally bleached in a selected area using excessive illumination. The dynamic of non-bleached fluorophore molecules diffusing into the bleached area can then be measured. Using this technique, the local dynamics of fluorescently labeled molecules can be assessed with a spatial resolution of 2 to 5 micrometers.

An important limitation in live cell imaging is phototoxicity which results from the interaction between the excited fluorophore and an other molecule, especially with oxygen. This reaction causes the release of free radical oxygens in the living cells that can damage chemically destroy other structures in the cell. Phototoxicity often occurs upon repeated laser exposures of fluorescently labeled cells.

Fluorescence microscopy for live cell imaging The fluorescence signal emitted by the fluorophores inside the cell can be imaged using a large range of modern microscopy techniques chosen depending on phenotypes under study and the preservation of the integrity of the cell. The quantum efficiency of the fluorescently labelled protein defines the illumination needed for signal detectability. Thickness of the sample will guide the decision to wide-field or optical sectioning. An other important factor is the amount of signal and acquisition speed actually needed for object detection or quantification of dynamics. The spatiotemporal scale of a given phenotype can limit the spatiotemporal resolution, *via* depth resolution in 3D or acquisition speed, of the acquisition to harness photodamaging effect. Finally, a microscopy setup is also chosen according to fluorescent tagging technologies, data processing and analysis methods.

The most popular microscopy approach in live cell imaging is wide-field microscopy which allows for vesicle localization with a resolution of 200 nm by illuminating the whole sample with a single light source. This technology provides fast imaging and flexibility at low cost. However, wide-field microscopy presents some limitations such as blurring effect induced by out-of-focus fluorescence signals at the focal plane. The plain illumination of the sample can also cause undesirable photodamages.

Generally, confocal microscopy is more recommended to reduce out-of-focus interferences. In this approach, a convergent laser beam is used to focus on a single point on the focal plane. A pinhole aperture is placed at the detection end to reduce out-of-focus fluorescence, thus providing an optical sectioning effect. By reducing the size of that pinhole below the size of the central airy disc pattern, the resolution of the confocal microscope can be enhanced by a factor 1.4 when compared with the wide-field microscope resolution. The precision of this technique is limited by the diffraction of laser focal point. Two confocal technologies are recommended depending on the application. First, laser-scanning confocal microscopy consists in a raster scan of the focal plane to image a two-dimensional slice of the sample. The sample can be thus excited with high precision. Secondly, a faster method is the spinning (a.k.a. Nipkow) disk setup which allows the acquisition of multiple points on the CCD sensor at the same time. The spinning disk setup is more affordable than the instrumentation required for laser-scanning and offers higher sensitivity due to CCD detector, thus reducing phototoxicity. However, controlled localization of the excitation is not possible and optical resolution is usually lower. Finally, three-dimensional images can be acquired with confocal microscopy by sequential movement of the objective or the sample. However, acquisition time and photodamage limit the axial resolution.

13 Analysis of correlation and variational approaches for motion estimation

The most popular techniques for motion estimation in live cell imaging are based on correlation measures. They have been well suited for a set of problems verifying temporal stationarity of fluorescence signals, and affected by illumination changes. In this chapter, we briefly review the Spatio-Temporal Image Correlation Spectroscopy technique (STICS), representative of correlation-based methods for motion and diffusion estimation. We compare this correlation approach to variational principles [Zimmer et al., 2011] described in Chapter 4. Global variational approaches have been recently investigated in biological imaging [Lecomte et al., 2012; Delpiano et al., 2011; Pizarro et al., 2011] but were applied to non-dense deformations or synthetic motion.

13.1 Correlation approach

Motion estimation in biological imaging is often performed as a block matching procedure, minimizing a patch-based distance measure to find corresponding pixels between the two consecutive images [Ji and Danuser, 2005; Rohr et al., 2010; Würflinger et al., 2004; Goobic et al., 2005; Baheerathan et al., 1998; Bornfleth et al., 1999]. Variations and improvements of the matching procedure have been added when assumptions adapted to the application can be made. Under temporal stationarity, temporal integration improves results [Ji and Danuser, 2005]. Multi-resolution techniques are also employed in [Rohr et al., 2010] to avoid local minima structures when similar shapes are present in the image.

Simple feature matching based on the Normalized Cross Correlation measure defined in (2.7) is often used in practice, and yields coarse but robust results. Its invariance under linear intensity changes is particularly important to cope with fluorescence variations. As already mentionned in Section 2.1.2, the computational complexity is a major limitation, which can be addressed by accelerating distance computation [Lewis, 1995; Luo and Konofagou, 2010] or correspondences space exploration [Barnes et al., 2009, 2010].

In a fluorescence scenario, a more theoretically justified exploitation of the correlation

ratio can be derived from physical fluorescence models. The Image Correlation Spectroscopy (ICS) approaches have then been developed specifically for fluorescence microscopy, taking advantage of physical properties of fluorescent particles and their relationship with correlation. These methods integrate the variations of fluorescence over space and/or time via correlation measures to access to information at the molecular level, such as diffusion coefficients or dominant flow speed and direction [Hebert et al., 2005]. The generalized spatial and temporal correlation expression is defined as follows

$$C(\mathbf{w}, I, \tau) = \frac{1}{N} \sum_{t=1}^{N-\tau} \frac{\langle \delta I(x,t)\ \delta I(x+\mathbf{w}, t+\tau) \rangle}{\langle I(x,t) \rangle \langle I(x,t+\tau) \rangle} \tag{13.1}$$

where $\langle \cdot \rangle$ denotes the spatial average over a patch. We define $\delta I(x) = I(x,t) - \langle I(x,t) \rangle$ as the intensity variation and N is the number of frames. The parameter τ is the temporal offset, that we set to $\tau = 1$ in this section, so we drop this notation in the following. The number N of frames is chosen such that an assumption of temporal stationarity of motion is valid in the considered subsequence. We point out that $C(\mathbf{w}, I)$ is not a normalized correlation criterion but enables to recover the biophysical parameters associated to density, motion of molecules, and diffusion coefficient [Hebert et al., 2005]. In Chapter 15, we use it to estimate the diffusion coefficient.

For motion estimation purpose, the goal is to estimate the translation vectors corresponding to the tracking of the correlation peak over time (Fig. 13.1). In our experiments, the static or immobile molecule population is filtered by local averaging and $C(\mathbf{w}, I)$ is computed by Fast Fourier Transform (FFT). We define

$$C(\mathbf{w}, I) = C(\mathbf{w}_p, I) e^{-\frac{(u-u_p)^2 + (v-v_p(\tau))^2}{\kappa_0^2}} + r_\infty \tag{13.2}$$

where $\mathbf{w}_p = (u_p, v_p)^\top$ is the motion vector at the previous frame, r_∞ is the spatial lag offset and κ_0 is the laser beam size which depends on the microscope. The correlation peak is tracked by linear regression to find the velocity vector from $\mathbf{w}_p$ (Fig. 13.1). We consider a 2D Gaussian function to estimate accurately the correlation peak over time [Hebert et al., 2005] using a Levendberg-Marquardt optimization scheme. In the experiments, the analysis is performed on image blocks. The size of the blocks determines the scale of moving objects retrieved and the maximum complexity of the estimated deformation (we take 64×64 pixels block size). The spatial lag between blocks is chosen to achieve an acceptable trade-off between spatial accuracy and computational time (we take 16 pixels spatial lag).

The requirement of temporal stationarity of motion limits the application of STICS to sequences without temporal motion discontinuities. This assumption is quite restrictive in practice. When it does not hold, we use in our experiments simple block matching

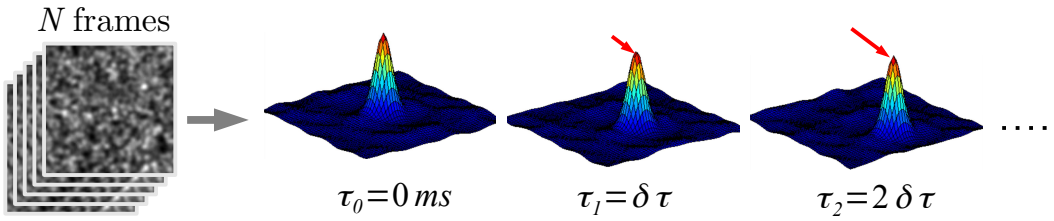


Figure 13.1: Motion estimation with STICS. Tracking of the correlation peak by Gaussian fitting on correlation maps.

with Normalized Cross Correlation (that will be referred as NCC) measures between two consecutive frames.

13.2 Variational approach

In this section, we describe the variational approach based on the principles stated in Chapter 4. The dense flow field is estimated as the minimizer of a global energy functional composed of two terms:

$$\widehat{\mathbf{w}} = \arg\min_{\mathbf{w}} E_{data}(\mathbf{w}, I) + \lambda E_{reg}(\mathbf{w}) \tag{13.3}$$

Let us recall that E_{data} is a data term penalizing deviations from a data conservation assumption over time, E_{reg} is a regularization term enforcing smoothness of the flow field and $\lambda > 0$ serves as regularization parameter to balance the contributions of E_{data} and E_{reg}. A high value of λ allows to retrieve only dominant motions of large structures by smoothing the flow field, while a small value of λ allows to distinguish between close spatial variations of small objects.

In our experiments we used the data term of [Zimmer et al., 2011] based on the assumption of constancy of intensity and spatial gradient of the image. The spatial gradient constraint is robust to additive illumination changes, which is necessary for several biological applications. The resulting data energy is:

$$\begin{aligned} E_{data}(\mathbf{w}, I) = \int_{\Omega} (1-\gamma)\phi(\eta_0 |\nabla I^T \mathbf{w} - I_t|^2) \\ +\gamma\phi(\eta_x |\nabla I_x^T \mathbf{w} - I_{xt}|^2) + \eta_y |\nabla I_y^T \mathbf{w} - I_{yt}|^2)) dx, \end{aligned} \tag{13.4}$$

where $\phi(z^2) = \sqrt{z^2 + \epsilon}$ is the regularized L_1 norm with $\epsilon = 0.001$, $\nabla\cdot$ denotes the spatial gradient operator, the subscripts $\cdot_x$, $\cdot_y$ and $\cdot_t$ are respectively the derivatives along the x, y and t axis and $\gamma \in [0, 1]$ balances the influence of intensity and gradient constancy terms. Normalization coefficients η_0, η_x, η_y prevent too strong data constraint in regions

of high image gradient [Zimmer et al., 2011], and are defined as

$$\eta_0 = \frac{1}{I_x + I_y + a}, \quad \eta_x = \frac{1}{I_{xx} + I_{xy} + a}, \quad \eta_y = \frac{1}{I_{yx} + I_{yy} + a}, \tag{13.5}$$

where $a = 0.1$ avoids division by 0.

The regularization term penalizes high gradients of the motion field $\mathbf{w} = (u, v)^\top$ with the convex and discontinuity-preserving $\phi(\cdot)$. We obtain the following energy term:

$$E_{reg}(\mathbf{w}) = \int_\Omega \phi(\|\nabla u\|^2 + \|\nabla v\|^2) dx. \tag{13.6}$$

We follow the minimization method of [Zimmer et al., 2011] by successively solving the Euler-Lagrange equations associated to the problem (13.3) at each level of a coarse-to-fine decomposition. The non-linearity due to the penalization function $\phi(\cdot)$ is removed by fixed point iterations and the remaining linear system is solved with SOR ("Successive Over Relaxation") (see [Brox, 2005] for details).

To cope with the largest intensity changes occurring in fluorescence sequences, it turns out that the gradient constancy constraint in (13.4) is insufficient. Accordingly, we found beneficial to apply the Midway image equalization method of [Delon, 2004] to improve the estimation.

13.3 Experimental comparison

In this section, we identify several classes of typical biological problems for which dense motion estimation can bring useful information, and we compare the results obtained with variational and correlation methods. We provide only qualitative visual results, due to the absence of ground truth. Our objective is to give an intuitive overview of the potential of the two approaches. Apart from the *Cell deformation* experiment, the results of NCC are given for three patch sizes $s = 15, 35, 75$, and three values of the regularization coefficient of the variational method are also compared, $\lambda = 3, 5, 8$.

Cell deformation

In this experiment, we evaluate the potential of the methods to accurately quantify cell deformation. Figure 13.2 shows a fluorescent protein attached to a membrane protein Clathrin in spinning-disk confocal microscopy. The sequence is acquired during chemical fixation, causing a contraction of the cell. We need to estimate motion in this case in order to compensate the cell deformation induced by the chemical fixation process. Furthermore, the biological sample is analyzed in electron microscopy to quantify structures and details but the bias due to the chemical fixation should be corrected if it can be estimated.

The fixation process is slow and the motion can be considered stationary, so we

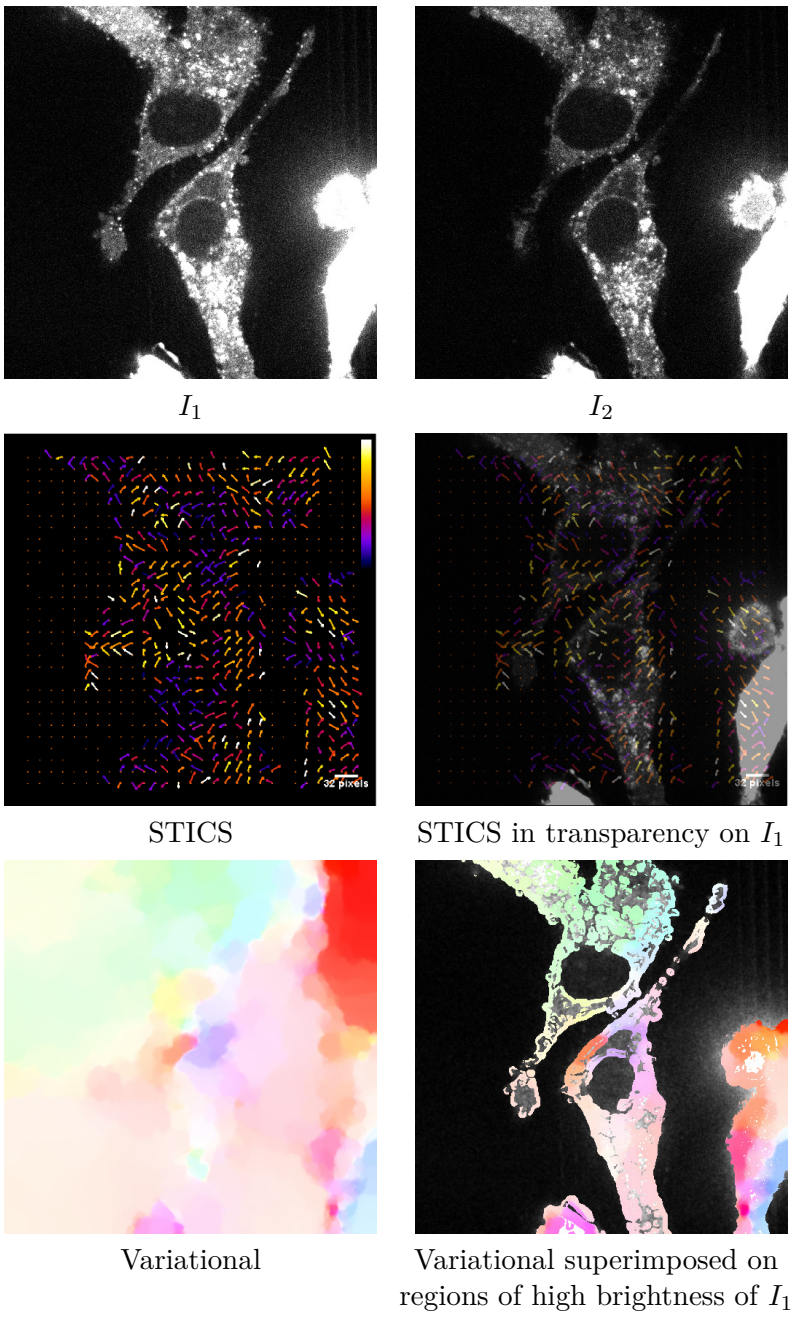


Figure 13.2: Consequence of shrinking due to fixation of proteins Clathrin GFP. Top row: two input images. Middle row: motion field obtained by STICS with arrow visualization, and transparency of the first image. Bottom row: motion field obtained by the variational method on the left, and restricted visualization in regions of high luminance on the right. Courtesy of V. Fraisier (J. Salamero's team), UMR 144 Institut Curie CNRS, PICT-IBiSA.

took $N = 100$ for the STICS analysis and selected the frames #0 and #100 for variational computation. Due to the large computational time required by STICS, we only compute motion in a sub-sampled grid, and we display the results with arrow visualization. Analysis of velocities obtained by STICS shows that the main cell deformations are satisfyingly estimated. Due to the time integration performed by the STICS method, we observe a regularization effect of the velocity map. The dense map provided by the variational method recovers more accurately the complex deformations in the membrane and nucleus regions. However, the motion estimated in these moving regions is propagated in static regions around the cell, whereas STICS is able to capture null motions. It is interesting to notice that the important illumination change in the sequence does not affect the variational estimation. We found in this particular case the Midway image equalization [Delon, 2004] applied in pre-processing to improve substantially the results. For a deformation compensation purpose, the accuracy of the motion estimation is of up-most importance, and is better achieved with the variational method. Also note that STICS requires 5 hours to produce the result of Fig. 13.2, whereas the variational method only requires 30 seconds.

Cell migration in phase contrast imaging

Cell migration is commonly studied in near *in situ* situation, such as collagen matrix. This dynamical mechanism is a highly integrated, multi-step process that plays an important role in the progression of various diseases including cancer and which is extensively studied in biology. One of the protocols used to study cell migration is to culture cells on a plate where some area (the empty space in Fig. 13.3) have been covered with gel preventing cell migration. The protocol consists then in removing this gel and retrieving the speed of migration by measuring the area progressively covered by the cells. But this area information is very global and will not give any information about the way cells migrate. Indeed, cells are very dense, which prevents individual tracking. A question that can be answered by flow estimation is to discriminate collective and individual motions of cells.

Figure 13.3 shows two phase-contrast images taken from a time lapse movie acquired in video-microscopy. The displacements and deformations of the cell are changing at each frame, so that temporal stationarity assumption is not valid. Therefore, the STICS method cannot be applied here and we use block matching with NCC measure. The same scale selection effect can be observed with the variation of the patch size s of NCC and the regularization coefficient λ of the variational method. For small values, motion of individual cells are detected, whereas for large values the dominant displacement of the migrating front is recovered. A more accurate analysis shows that for small λ and s, the variational method delineates very well every single cell, whereas NCC yields coarser results. Moreover, small patch sizes involves a number of large errors of NCC due to the

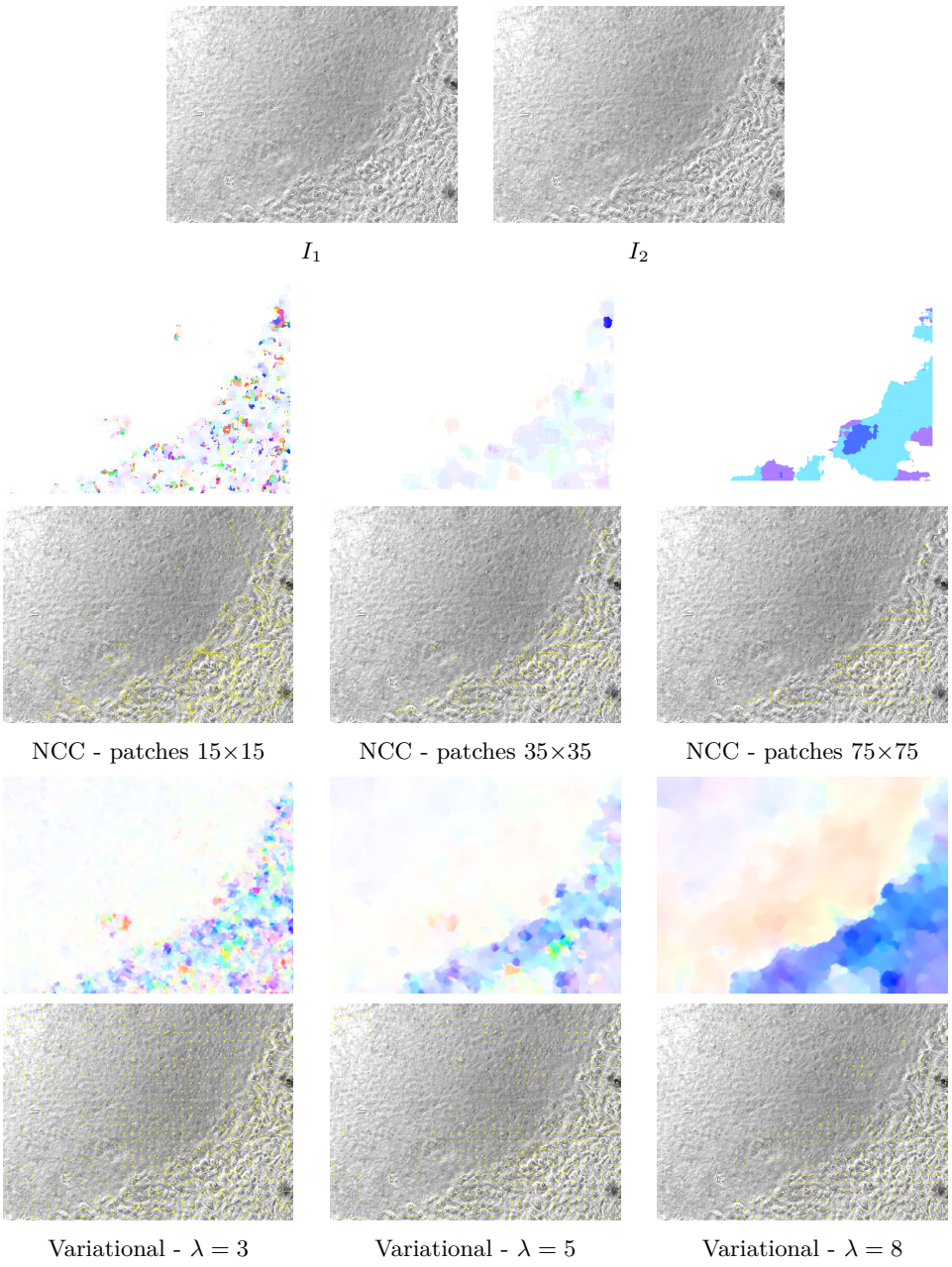


Figure 13.3: Results on the "Cell migration in phase contrast imaging" sequence (courtesy of P. Chavrier,Äôs team, UMR 144 Institut Curie CNRS, PICT-IBiSA). 1^{st} row: two input images. 2^{nd} and 3^{rd} rows: color and arrow visualizations of the results obtained with the NCC block matching, for several patch sizes. 4^{th} and 5^{th} rows: color and arrow visualizations of the results obtained with the variational method, for several regularization coefficients.

lack of local information for matching.

It is interesting to notice that the variational result could not be achieved with a sparse tracking method, due to the high density of cells and large deformations which makes impossible a robust detection of each cell. Exploiting this result for cell tracking in dense conditions could be the subject of a future work since a simple motion segmentation on a single frame already gives good individual cell detection. This idea has been exploited recently in [Liu et al., 2014].

Cell migration in fluorescence imaging Figure 13.4 also represents migrating cells, but is acquired in different conditions. Three differences can be observed with the previous cell migration example. First, the sequence is acquired in fluorescence imaging, which implies strong intensity variations occurring when cells undergo large deformations while keeping the same amount of fluorescence. Differently from phase contrast imaging of Fig. 13.3, the background is homogeneous and gives no information for motion estimation, which produces large errors for the local NCC measurement. Secondly, the cells are not dense, and most pixels of the image domain belong to the static background. Finally the time step between two frames is much larger than in the previous example, resulting in larger displacements and deformations.

The variational method is unable to retrieve the large displacements of small objects and creates typical colorwheel artifacts for small values of the regularization coefficient λ, but the smoothness constraint on the flow field allows for a good approximation of the null motion of the background. Thus, this example shows the insufficiency of the variational approach to capture large displacements of small objects. The NCC has an opposite behaviour since it captures most of the displacements of the cell, but also produces large errors in homogeneous regions, due to the lack of local information for matching. Moreover motion fields produced by are globally very noisy. It is noticeable that NCC is not affected by the strong local intensity changes. The tendency for the variational method to produce null motion field when the regularization increases can also be observed.

Actin network Dynamical behaviors of actine networks are involved in several fundamental biological processes [Pinot et al., 2012]. In Fig. 13.5, an F-actine network is evolving in vitro in a confined droplet. The motion field of the filaments network is dense and smooth, thus providing an opposite behaviour to the sparse and discontinuous motion field of the cell migration example (Fig. 13.4). This case is typically well suited for variational estimation, which retrieves well the complex deformation, if the regularization coefficient is well chosen. NCC produces coarser results, with large errors for small patch sizes.

HeLa cell The sequence of Fig. 13.6 is a real-time imaging of the synchronized trafficking

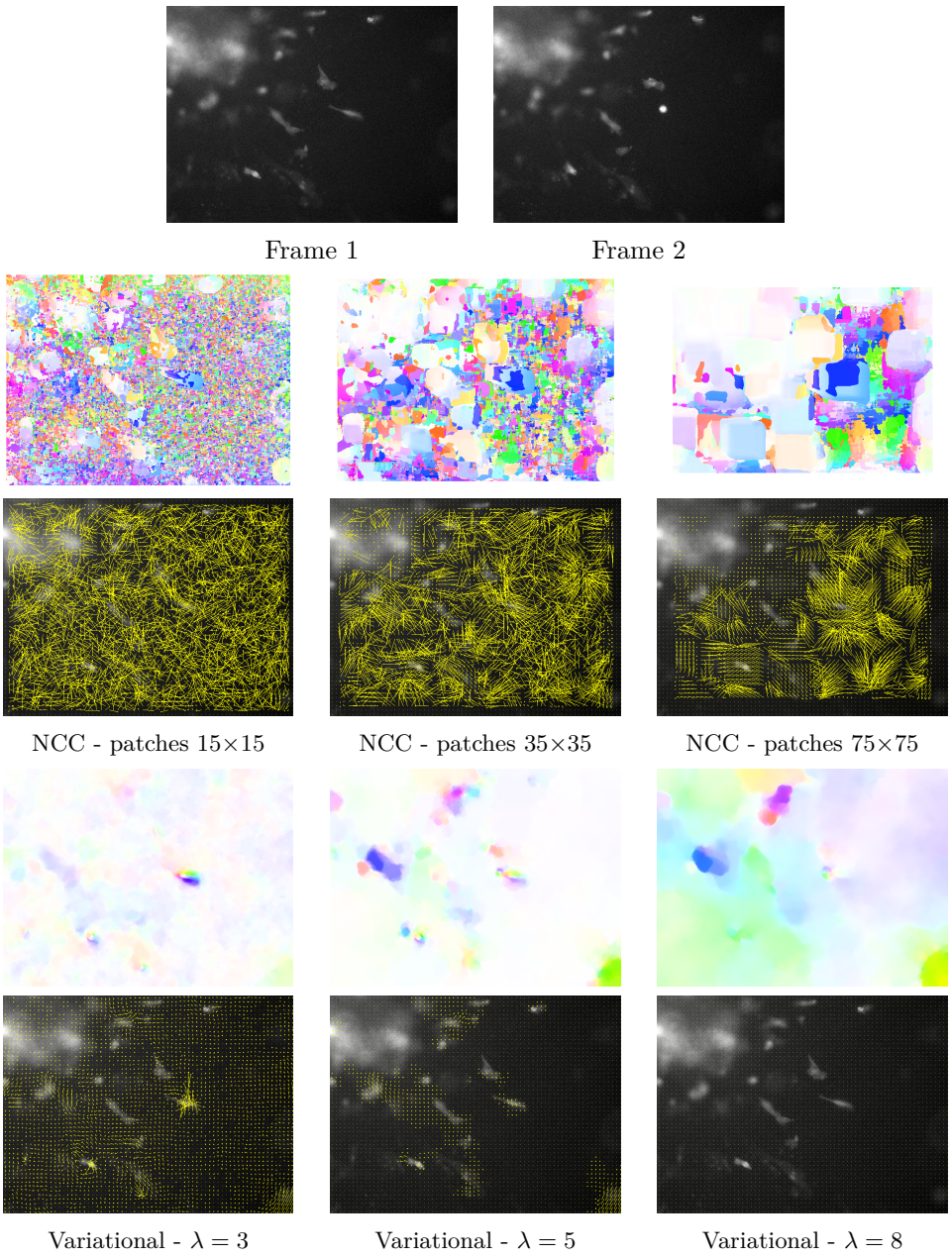


Figure 13.4: Results on the "Cell migration in fluorescence imaging" sequence (acquisition by P. Chavrier's group, UMR 144 Institut Curie, PICT-IBiSA). 1^{st} row: two input images. 2^{nd} and 3^{rd} rows: color and arrow visualizations of the results obtained with the NCC block matching, for several patch sizes. 4^{th} and 5^{th} rows: color and arrow visualizations of the results obtained with the variational method, for several regularization coefficients.

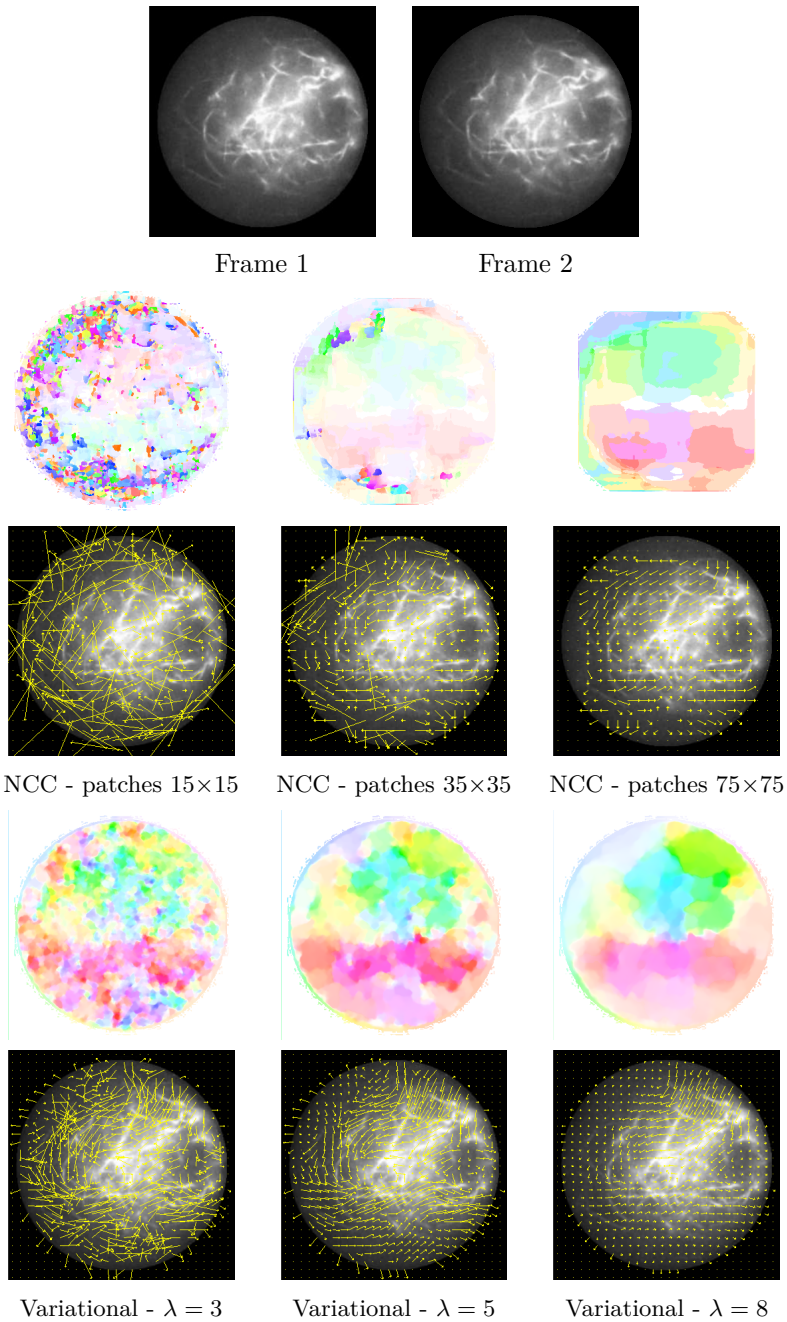


Figure 13.5: Results on the "Actin network" sequence Pinot et al. [2012]. 1^{st} row: two input images. 2^{nd} and 3^{rd} rows: color and arrow visualizations of the results obtained with the NCC block matching, for several patch sizes. 4^{th} and 5^{th} rows: color and arrow visualizations of the results obtained with the variational method, for several regularization coefficients.

of protein tagged with GFP in HeLa cells and has been exploited in [Boncompain et al., 2012]. The strong locals intensity changes occurring inside the cell correspond to the recruitment of fluorescent proteins to the Golgi. The very large intensity change affecting the cell, combined with the large displacement, cannot be handled by the variational method. On the contrary, NCC captures the displacement of the cell, but, as in Fig. 13.4 a large number of matching errors also occurs in non informative regions, and a lack of global smoothness of the motion field can be observed.

Collagen matrix The last example is issued from a protocol on plate trying to mimic a multi-cell environment such as in tissue, but on a 2D plate. The field is evolving toward placing the cells in much more realistic conditions, such as 3D collagen matrix, mimicking the conditions if a 3D collagen network which is the physiologic environment of cells. Then in addition to the cell movement it self, the interaction with the collagen matrix are of interest, in particular because they may be interpreted as forces applied by the cells to perform its locomotion. The sequence of Fig. 13.7 is taken from a 2D time lapse movie in video microscopy: images are taken simultaneously in one plane imaging fluorescence with two different filters: one would allow to see the cells stained (not shown), the other channel is showing the collagen fibers. Usually when biologists have to measure this kind of flow field, they will consider placing nano beads to be tracked or on which motion is estimated by correlation-based approaches. Showing that the fibers itself can be used, without adding any beads that may perturb the biological behavior, would be of great interest.

In terms of motion field, similarly to the actin filaments case of Fig. 13.5, the deformation field is dense, complex and smooth, without large intensity changes and large displacements of small structures. These conditions perfectly fit with the modeling purposes of variational methods. In practice the motion field obtained with the variational method is indeed able to retrieve very accurately the most complex deformations. The result of the NCC matching captures coarsely the main displacements, but delineates less accurately the structures. Moreover, as already observed in the cell migration example, large errors occur frequently in locally non-informative regions. The tendency of the variational approach to under-estimate motion when the regularization increases can also be observed.

13.4 Conclusion

From the previous experiments on biological sequences, we are able to identify a wide range of challenges for motion estimation in live cell imaging. The motion field can be sparse or dense dense. The lack of information for motion estimation in static regions occurs mostly in fluorescence imaging, producing homogeneous background. Complex smooth deformations often coexist with motion discontinuities. Large local intensity

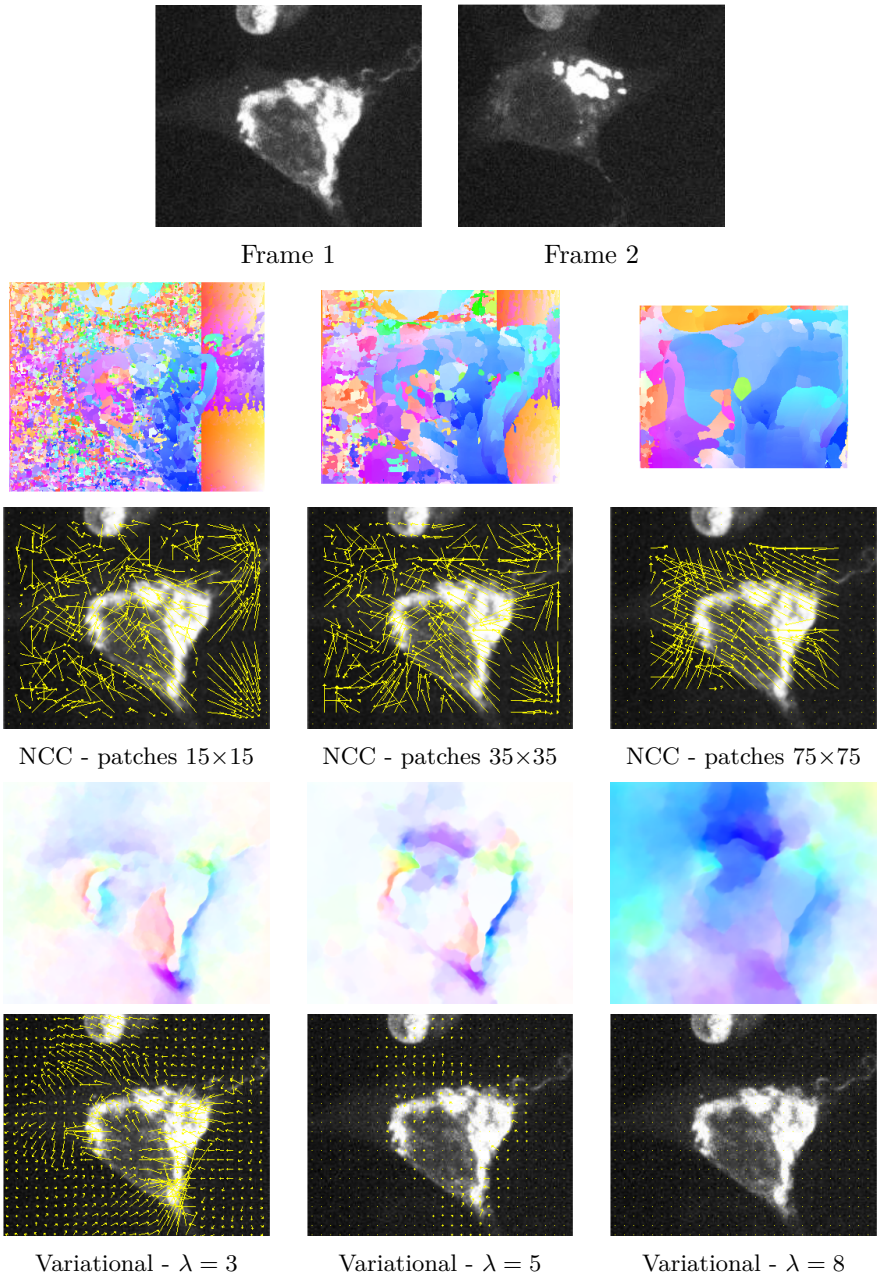


Figure 13.6: Results on the “HeLa cell" sequence (acquisition by Perez' group, UMR 144 Institut Curie, PICT-IBiSA). 1^{st} row: two input images. 2^{nd} and 3^{rd} rows: color and arrow visualizations of the results obtained with the NCC block matching, for several patch sizes. 4^{th} and 5^{th} rows: color and arrow visualizations of the results obtained with the variational method, for several regularization coefficients.

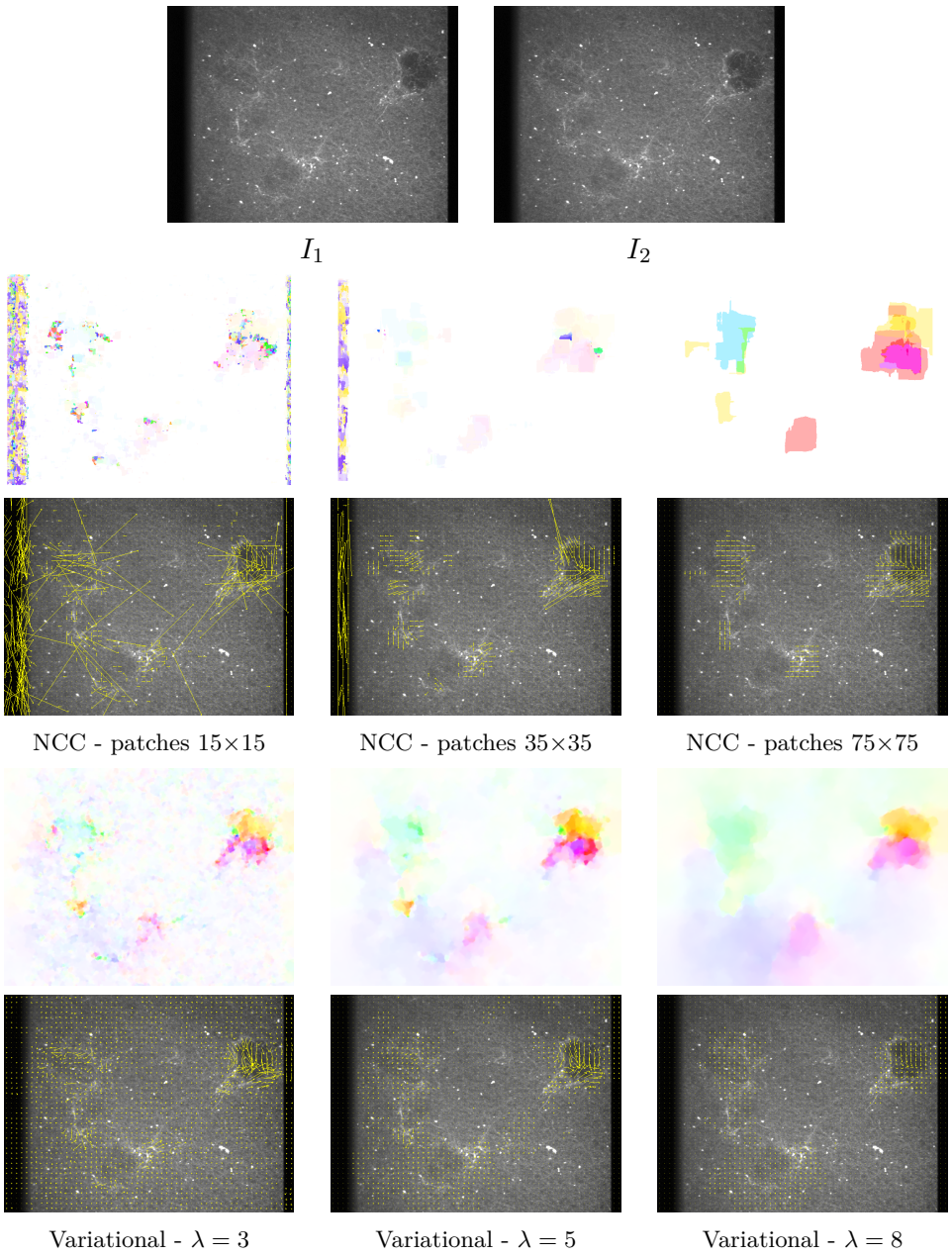


Figure 13.7: Results on the “Collagen" sequence (Courtesy of P. Chavrier’s team, UMR 144 Institut Curie CNRS, PICT-IBiSA). 1^{st} row: two input images. 2^{nd} and 3^{rd} rows: color and arrow visualizations of the results obtained with the NCC block matching, for several patch sizes. 4^{th} and 5^{th} rows: color and arrow visualizations of the results obtained with the variational method, for several regularization coefficients.

changes occur frequently in fluorescence imaging. Large time steps in the acquisition process can yield large displacements and large deformations.

Experimental comparison between representative variational and correlation approaches lead us to several conclusions. In terms of accuracy, our visual comparisons clearly shows the superiority of the variational approach. Indeed, the results of correlation approaches are affected by a block effect, due to the patch-based extent of the measurements, which is the only regularization constraint. Another consequence is the presence of frequent large localized errors in uninformative regions. In contrast, complex deformations and motion details are well recovered by the variational method in all our experiments. However, correlation methods are less affected by intensity changes due to fluorescence variations and are able to capture large displacements of small objects. We also emphasized the analogous behaviours of the regularization parameter and the patch size, allowing to capture several motion scales.

In the next chapter, we combine these two approaches with our aggregation framework to take advantage of their complementarity, and we address the problem of large intensity changes in fluorescence imaging.

14 Aggregation framework for fluorescence imaging

In this chapter we first evaluate the ability of the aggregation method presented in Part II to deal with issues encountered in fluorescence imaging. This method can be viewed as a combination of the correlation and global regularized approaches presented in Chapter 13 for biological imaging. Secondly, we adapt the method to handle a particularly prominent and challenging characteristic of fluorescenceng images, namely large intensity changes. We evaluate the limits of traditional intensity change handling, and we propose to integrate the estimation of the intensity change map into the aggregation framework, yielding significant improvements.

14.1 Intensity correction model and related works

As mentioned in Chapter 13, pre-processing images of a sequence to reduce contrast changes across frames [Delon, 2004; Delon and Desolneux, 2010] can be a way to reduce the impact of intensity changes on motion estimation. When integrated in the motion estimation process, handling intensity changes can be achieved in two ways (as already explained in Chapter 2). The first one avoids brightness constancy violations by looking for more robust descriptors, and the second one corrects the invalidity by explicitly estimating the deviations. Illumination invariant descriptors as those discussed in Chapter 2 are able to help the estimation up to a certain extent. However they also have their own restricted validity domain and have thus to be combined with intensity to achieve robustness to a reasonably large range of situations. The most appropriate combination of descriptors is often application specific, or even possibly region specific in a single image, so that a trade-off has to be made to find the best average solution. Some works addressed this problem by spatial adaptivity of the combination [Heitz and Bouthemy, 1993; Xu et al., 2012b; Kim et al., 2013].

The explicit estimation of intensity constancy deviations is conceptually more satisfying because it is not dependent on the list of descriptors arbitrarily chosen or to the difficulty to combine them. It rather relies on a single model of intensity correction which aims at estimating any intensity change. Several correction models have been briefly discussed in Section 2.2.2. We consider in this chapter the simple spatially varying offset model

leading to the data potential

$$\rho_d(x, I_1, I_2, \mathbf{w}, \xi) = \phi(I_2(x + \mathbf{w}(x)) - I_1(x) - \xi(x)), \tag{14.1}$$

where $\xi : \Omega \to \mathbb{R}$ compensates local additive intensity changes. This model has already been used in a few works [Chambolle and Pock, 2011; Kim et al., 2005; Teng et al., 2005; Lai, 2000], and for a spatially constant offset in [Odobez and Bouthemy, 1995]. However, the superiority of the explicit estimation over classical robust data terms has never really been demonstrated. We believe that it is due to the difficulty of optimizing jointly $\mathbf{w}$ and ξ. Indeed, most works perform alternate minimization steps between the two variables, which is prone to fall into local minima. Other alternatives to be explored could be high dimensional optimization methods like [Papadakis et al., 2013] or splitting schemes [Ayvaci et al., 2012].

We propose to overcome the optimization problem coming from the minimization w.r.t. to ξ owing to our aggregation framework. Local candidates for ξ are reliably estimated jointly with motion candidates, so that every motion candidate is associated with an intensity change candidate, and the minimization is performed on a single two-component label set.

14.2 Computation of local candidates

We describe the computation of augmented candidates, adding intensity change estimation to the motion estimation procedure described in Chapter 8. We work on the patch distribution $\mathcal{P}_{\mathcal{S},\alpha}$ described in Section 8.1.1, which is composed of overlapping square patches with several sizes of the set $\mathcal{S}$ and overlapping ratio α.

Patch correspondences For each patch $P_1 \in \mathcal{P}_{S,\alpha}$, the set $\mathcal{M}_N(P_1)$ was created by computing N patch correspondences to P_1 (Section 8.1.2). The distance used for the matching was the sum of point-to-point L_1 distances in the saturation and value channels of the HSV space. The saturation channel is invariant to local multiplicative intensity changes and the value channel has no invariance. To satisfy our additive model (14.1) we replace this measure by a Normalized Cross Correlation distance defined in Section 2.1.2, invariant to additive intensity changes. For each pair of corresponding patches $P_{1,2} = (P_1, P_2)$ with $P_2 \in \mathcal{M}_N(P_1)$, we can then estimate a coarse intensity change candidate $\xi_{P_{1,2}}$ defined as the difference between the intensity means computed over respectively P_2 and P_1. It is coupled with the coarse motion candidate $w_{P_{1,2}} \in \mathbb{Z}^2$ defined as in Section 8.1.2 as the location difference between the centers of P_2 and P_1.

Affine estimations For each pair of corresponding patches, denoted $P_{1,2}$, the intensity

change candidates are refined jointly with motion. We recall the notations Ω_{P_1} for the pixel domain of P_1, $\delta\mathbf{w}_{P_{1,2}} : \Omega_{P_1} \to \mathbb{R}^2$ for the affine motion field between P_1 and P_2, defined at a pixel $x = (x_1, x_2)^\top$ with $\delta\mathbf{w}_{P_{1,2}}(x) = (a_1 + a_2x_1 + a_3x_2, a_4 + a_5x_1 + a_6x_2)^\top$, and $\boldsymbol{\theta}_{P_{1,2}} = (a_1, a_2, a_3, a_4, a_5, a_6)^\top$ for the affine model parameter vector.

Parametric motion estimation usually relies on brightness constancy constraint. However, to refine the previous intensity change estimates $\xi_{P_{1,2}}$, we compute a constant intensity change increment $\delta\xi_{P_{1,2}}$ between P_1 and P_2, integrated in the modified brightness conservation constraint:

$$P_2(x + \delta\mathbf{w}_{P_{1,2}}(x)) - P_1(x) = \xi_{P_{1,2}} + \delta\xi_{P_{1,2}}. \tag{14.2}$$

Hence, we estimate the extended parameter vector $\Theta = (\boldsymbol{\theta}^\top, \delta\xi)^\top$ (we drop subscripts $\cdot_{P_{1,2}}$ for the sake of clarity) as follows:

$$\widehat{\Theta} = \arg\min_{\Theta} \int_{\Omega_{P_1}} \rho_{data}(x, \delta\mathbf{w}, \delta\xi, P_{1,2})dx, \tag{14.3}$$

where $\rho_{data}(\cdot)$ is the data potential penalizing deviations from the data constancy assumption (14.2) and is defined by:

$$\rho_{data}(x, \delta\mathbf{w}, \delta\xi, P_{1,2}) = \psi(P_2(x + \mathbf{w}(x) + \delta\mathbf{w}(x)) - P_1(x) - \xi - \delta\xi). \tag{14.4}$$

The penalty function $\phi(\cdot)$ is chosen as the robust Tukey's function. The minimization problem (14.3) is solved as in Section 8.1.3 with the Motion2D software[1] [Odobez and Bouthemy, 1995].

Final set of candidates The above described estimation is repeated for every pair of patches and generates a set of candidate motion vectors and intensity change parameters $\mathcal{C}(x)$ at each pixel $x \in \Omega$ defined as follows:

$$\mathcal{C}(x) = \left\{ \left(\mathbf{w}^c_{P_{1,2}}(x), \xi^c_{P_{1,2}} \right) : P_1 \in \mathcal{P}_{\mathcal{S},\alpha}(x), P_2 \in \mathcal{M}_N(P_1) \right\} \tag{14.5}$$

with $\mathbf{w}^c_{P_{1,2}}(x) = \mathbf{w}_{P_{1,2}}(x) + \delta\mathbf{w}_{P_{1,2}}(x)$, $\xi^c_{P_{1,2}} = \xi_{P_{1,2}} + \delta\xi_{P_{1,2}}$ and $\mathcal{P}_{\mathcal{S},\alpha}(x) = \{P \in \mathcal{P}_{\mathcal{S},\alpha} : x \in P\}$.

14.3 Global aggregation

We conceive the aggregation step similarly to Chapter 9 as a discrete optimization problem where the set $\mathcal{C}(x)$ of candidates forms the finite label space considered at pixel x. The

[1] http://www.irisa.fr/vista/Motion2D/

global flow field $\mathbf{w} : \Omega \to \mathbb{R}^2$ and the global intensity change field $\xi : \Omega \to \mathbb{R}$ are recovered by:

$$(\widehat{\mathbf{w}}, \widehat{\xi}) = \arg\min_{\mathbf{w},\xi} E(\mathbf{w}, \xi, I_1, I_2), \text{ s.t.}(\mathbf{w}(x), \xi(x)) \in \mathcal{C}(x), x \in \Omega.$$

The global energy is defined by:

$$\begin{aligned} E(\mathbf{w}, \xi, I_1, I_2) &= \sum_{x \in \Omega} \psi(I_2(x + \mathbf{w}(x)) - I_1(x) - \xi(x)) \\ &+ \lambda_1 \sum_{<x,y>} \phi(\|\mathbf{w}(x) - \mathbf{w}(y)\|) + \lambda_2 \sum_{<x,y>} \phi(|\xi(x) - \xi(y)|), \end{aligned} \tag{14.6}$$

where $\psi(\cdot)$ is the L_1 norm, $< x, y >$ is a two-site clique and λ_1, λ_2 are balance coefficients between data and regularization terms. The data term is the same as the one used for affine estimation in the first step. The second term is a classical regularization on the motion field. We expect intensity variations to vary piecewise smoothly as the flow field, so we impose regularization on ξ as expressed in the third term.

Energy (14.6) is minimized with the *fusion-move* algorithm described in Chapter 9. The important point is that we do not have two distinct label sets for $\mathbf{w}$ and ξ, but a single set of candidates $\mathcal{C}(x)$. Each motion candidate is coupled with an intensity change candidate under a common label, which avoids minimization on the two variables. Motion and intensity change candidates are estimated locally coherently with the additive model (14.1), so that the coupling is by construction appropriate and there is no need to look for other possible combinations of $\mathbf{w}$ and ξ.

It is interesting to notice the link between our global occlusion handling model (9.2) and our global intensity change handling model (14.6). The two problems are similar since they try to deal with the invalidity of data conservation. The first difference is the nature of the invalidity, which is a binary violation in occlusion areas, and a continuous deviation in the case of an intensity change. The second difference is in the *a priori* form of the two fields, which is sparse (and possibly smooth) for the occlusion field, and smooth for the intensity change field. These observations are reflected in the energy terms (9.2) and (14.6).

14.4 Results

We have evaluated our method on three sequences presented in Section 13, representative of the diversity of challenges previously identified. As explained in Section 13, in the absence of available ground truth, we provide visual results and we emphasize on visually clear differences. The design of sequences with ground truth is a major objective for future work. We compare our method with public implementations of state-of-the-art

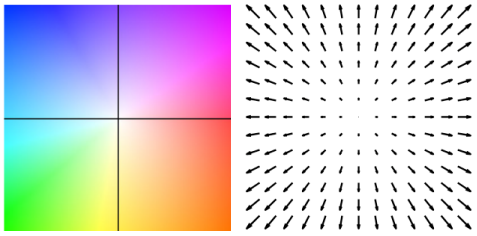


Figure 14.1: Equivalence between color and arrow visualizations.

optical flow methods [Brox and Malik, 2011; Sun et al., 2010a]. The method of [Sun et al., 2010a] is a typical efficient implementation and combination of best performing elements of variational methods. In [Brox and Malik, 2011], block matching is integrated in a variational regularized approach [Brox et al., 2004] as an additional constraint to deal with large displacements of small objects and get rid of the limitations of coarse-to-fine-schemes. We use the same color-code for dense and accurate visualization and arrow visualization for intuition of the physical displacement (Fig. 1.1). We also display the estimated intensity change map ξ. Low negative values of ξ correspond to a decreasing of intensity and are represented by dark regions, and high positive values correspond to increasing intensity and are represented by bright regions. The parameters are set to $\mathcal{S} = \{15, 35, 75\}$, $\alpha = 0.75$, $N = 3$ and for the aggregation, $\lambda_1 = 1$, $\lambda_2 = 0.5$. We point out that the three patch sizes of $\mathcal{S}$ correspond to the patch sizes used for our block matching experiments in Section 13.

Cell migration in fluorescence imaging The specificities of this sequence are the sparsity of the motion field, the large displacements and deformations of the cells, and the local large intensity change (Fig. 14.2). Variational methods [Brox and Malik, 2011; Sun et al., 2010a] fail to recover null displacement of the background and do not always retrieve correctly the motion of cells, as shown in Fig. 14.2. In contrast, our method successfully detects the static background and recovers the motion of the cells, even when they undergo large displacements, large deformations or intensity changes. In particular, large displacements of small cells are captured owing to NCC correspondences, and are selected in the aggregation stage thanks to the non linearization of the data term in (14.6). Our usage of patch correspondences is thus more efficient than the variational integration of [Brox and Malik, 2011]. The intensity change map is also visually coherent with the observed fluorescence variations between the two frames. In addition to providing accurate cell deformation, these results could be used to create cell correspondences for tracking applications. Unlike most cell tracking methods, we do not need any prior cell segmentation step.

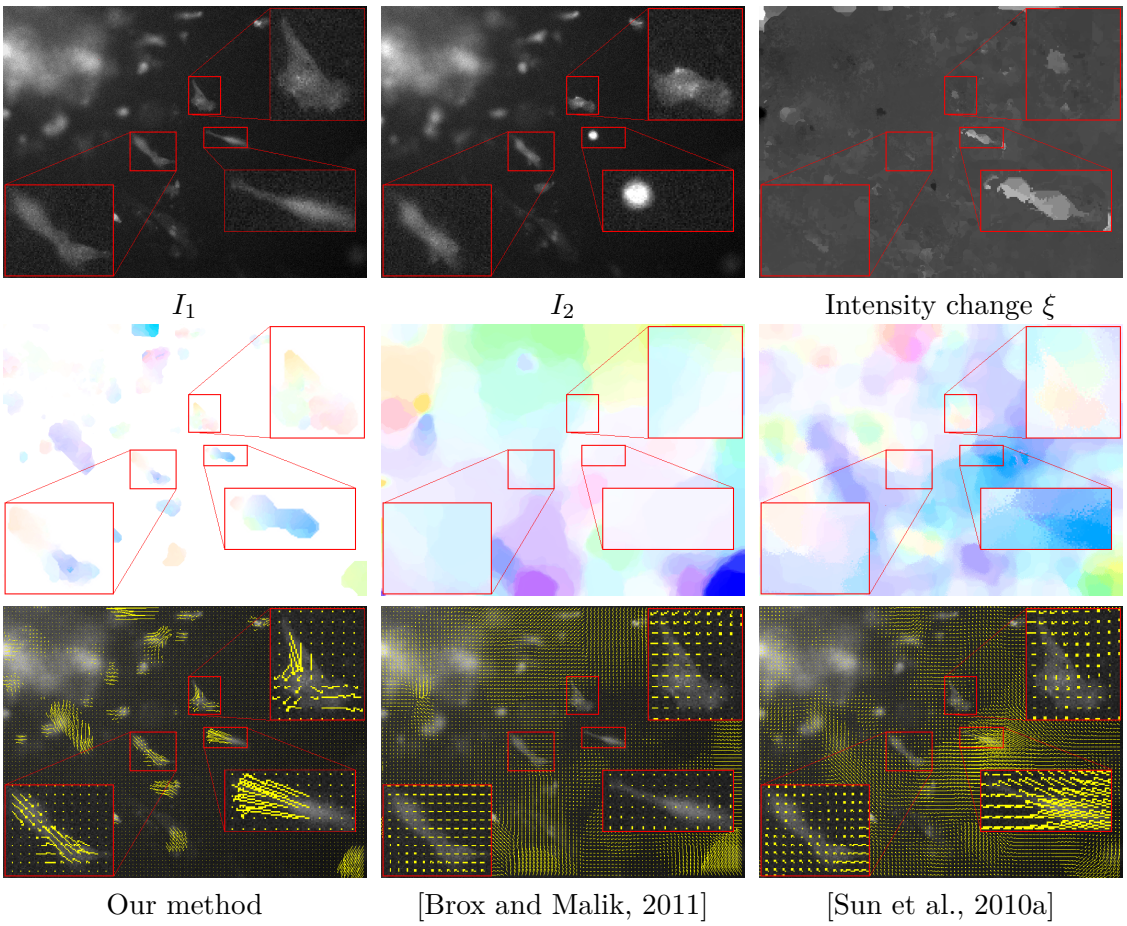


Figure 14.2: Results on the "Cell migration" sequence with zooms on regions of interest overlapping the images (acquisition by P. Chavrier's group, UMR 144 Institut Curie, PICT-IBiSA). Top row: the two input images and the reconstructed intensity change by our method. Middle and bottom rows: motion field respectively estimated by our method [Brox and Malik, 2011] and [Sun et al., 2010a].

Actin network This sequence exhibits complex, smooth and small deformations, which is a typical favourable case for variational estimation (Fig. 14.3). Our method successfully estimates this smooth deformation with the same parameters as for the case of cell migration. The method [Brox and Malik, 2011] gives similar results but propagates the flow to static regions. The method [Sun et al., 2010a] creates artificial motion clusters and high errors in sta tic regions. This example shows that our method is able to be competitive variational approaches in "simple" cases of small displacements and smooth motion field.

HeLa cell The main challenge of this sequence is the large intensity change due to the

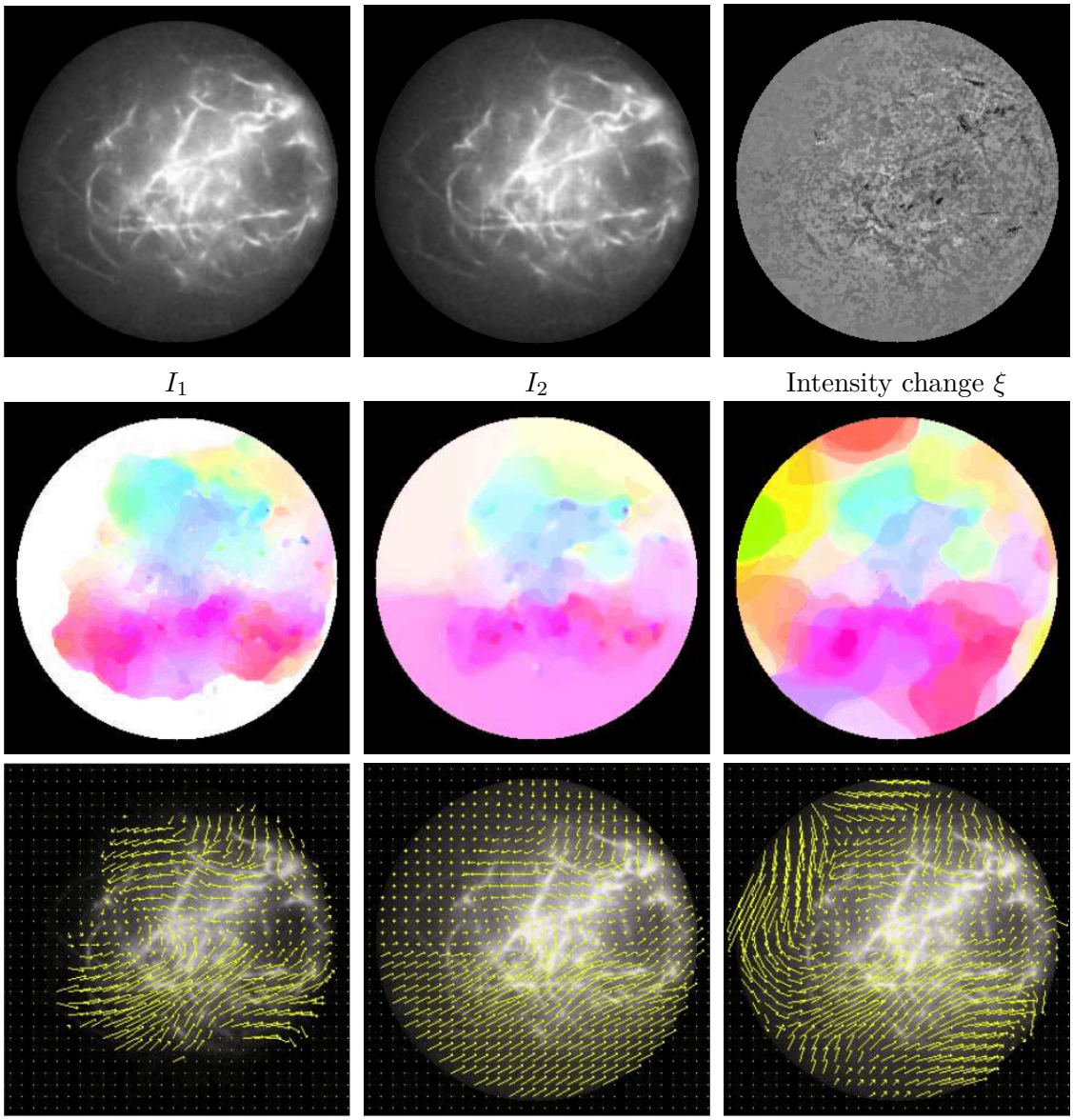


Figure 14.3: Results on the "Actin network" sequence [Pinot et al., 2012]. Top row: the two input images and the reconstructed intensity change by our method. Middle and bottom rows: motion field respectively estimated by our method [Brox and Malik, 2011] and [Sun et al., 2010a].

recruitment of fluorescent proteins to the Golgi, combined with the large displacement of the cell (Fig. 14.4). It is visually clear that the motion field estimated with our method captures more accurately the real cell deformation than the results of [Brox and Malik, 2011; Sun et al., 2010a]. The method [Brox and Malik, 2011] handles intensity

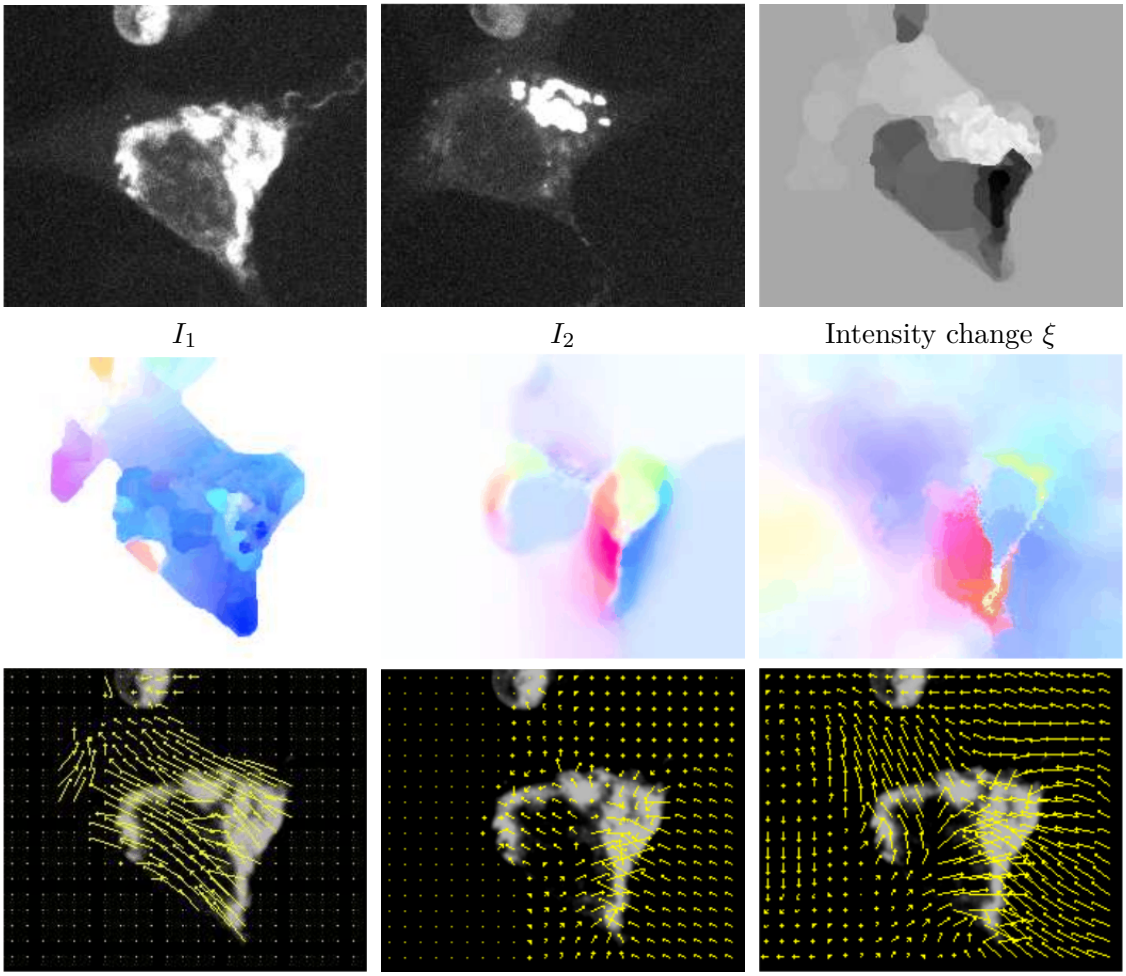


Figure 14.4: Results on the "HeLa cells" sequence (acquisition by Perez' group, UMR 144 Institut Curie, PICT-IBiSA). Top row: the two input images and the reconstructed intensity change by our method. Middle and bottom rows: motion field respectively estimated by our method [Brox and Malik, 2011] and [Sun et al., 2010a].

variations by the addition of an intensity gradient constancy constraint in the data term, and the method of Sun et al. [2010a] performs a structure/texture decomposition, which is insufficient to handle strong local intensity variations. As shown in Fig. 14.4, NCC is robust to intensity changes in the cell. Our method successfully exploits this invariance of NCC and the regularization effect of global models. The accuracy of the estimation is made possible by the efficiency of our explicit estimation of local intensity changes, as confirmed by the intensity change map shown in Fig. 14.4.

15 Variational approach for diffusion estimation

In the previous chapters, we focused on motion fields representing structure deformations or directed coherent motion of biological objects. At the protein level some motion behaviour also correspond to active transport, but the main mode of transport is often diffusive, i.e. proteins undergo Brownian motion. This type of motion is adequately modeled by its corresponding diffusion coefficient, representative of local change of the medium, or of the protein complex under study. In this section, we address the problem of diffusion coefficient estimation in time-lapse fluorescence microscopy, by first reviewing the standard Image Correlation Spectroscopy (ICS) method, and then proposing a new variational approach.

15.1 Image Correlation Spectroscopy

Image Correlation Spectroscopy techniques have been developed in the continuation of the signal-based Fluorescence Correlation Spectroscopy (FCS), as its extension to images. The principle of FCS is to measure the fluorescence fluctuations by photons counting in a small volume excited by a laser beam, as illustrated in Fig. 15.1. The recorded fluorescence variations can be caused diverse phenomena, which can be characterized by the analysis of its autocorrelation. The typical form of the autocorrelation function and the diverse physical process it expresses is shown in Fig. 15.1. Among these, we are interested in the characterization of diffusion (blue curve) occurring when fluorescent particles undergo Brownian motion. In this case, the density of particles can be retrieved from the value of the autocorrelation at time 0, and the diffusion coefficient D_0, which is the parameter of interest in this chapter, can be easily derived from the characteristic decay time τ_d of the correlation function with the following expression (see Schwille [2006] for a physical derivation of this expression):

$$D_0 = \frac{\omega_0^2}{4\tau_d} \tag{15.1}$$

where ω_0, is the laser beam size.

The idea of ICS is to extend this temporal analysis of a point-wise signal to the

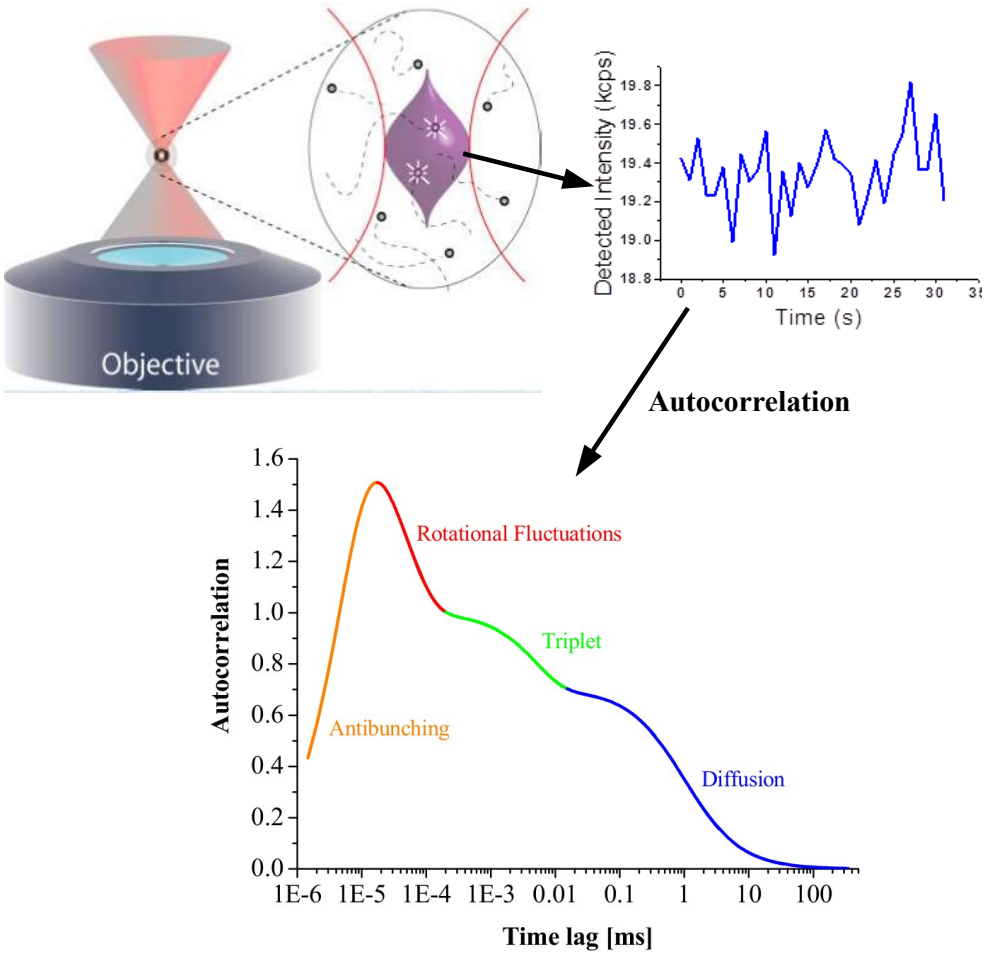


Figure 15.1: Principle of Fluorescence Correlation Spectroscopy (FCS).

spatial dimension. The basic ICS works with a a single image and replaces the temporal autocorrelation of the signal by the spatial autocorrelation of the image, which gives only access to the density of particles. The diffusion coefficient can be retrieved when considering image sequences, leading to the Temporal ICS (TICS). The computation of the spatio-temporal correlation function, illustrated in Fig. 15.2, yields a similar function to the FCS case of Fig. 15.1. More precisely, the theoretical function can be derived from the diffusion equation

$$I_t = D_0 \, \Delta I, \tag{15.2}$$

originally obtained for particles concentration in a diffusive scenario. This equation is also valid when applied to the images when image resolution is superior enough to size of the particles to be considered as a concentration measure. The closed-form solution of

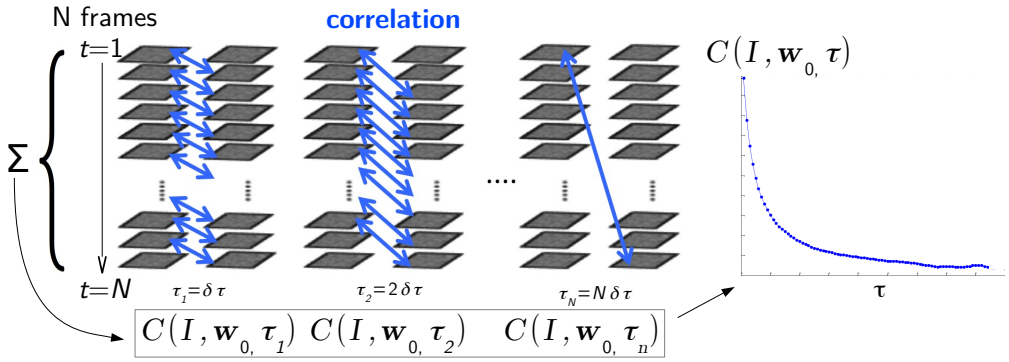


Figure 15.2: Illustration of the Temporal Image Correlation Spectroscopy (TICS) method.

(15.2) is inserted in the correlation expression (13.1) and leads to the following theoretical correlation function (see [Sergeev, 2004] for derivation):

$$C(I, \mathbf{w}_0, \tau) = C(I, \mathbf{w}_0, 0) \left(1 + \frac{\tau}{\tau_d}\right)^{-1} + C_\infty(\tau) \tag{15.3}$$

where $\mathbf{w}_0$ is the null motion field, since no motion field is estimated, $C_\infty(\tau)$ is an offset, and τ_d is the characteristic time decay, from which diffusion can be obtained from the same expression as for FCS (15.1). Each pixel can be considered as a FCS measurement, so that considering TICS as a the average of as many FCS as number of pixels in the image is expected to yield more robust estimation.

TICS is designed for spatially constant diffusion. To retrieve inhomogeneous diffusion maps, it must be applied locally in image patches. The local constancy assumption is however violated at diffusion discontinuities, which tends to over-smooth the estimated diffusion field. Moreover the patches must be large enough to validate the model, which implies, on one hand a loss of resolution, and on the other hand a large computational cost.

15.2 Variational diffusion coefficient estimation

The diffusion estimation problem has not been addressed yet in a variational framework. In the next section, we propose to fill this gap with a global regularized model of diffusion, optimized by variational techniques, and overcoming the above mentioned limitations of TICS.

Rather than estimating a constant diffusion coefficient over patches as performed with the TICS method, we consider a dense diffusion field $D : \Omega \to \mathbb{R}$. The estimation problem

is then formulated as

$$\widehat{D} = \arg\min_{D} \{E_{data}(I, D) + \lambda E_{reg}(D)\} \text{ s.t. } D(x) \geq 0. \tag{15.4}$$

The data fidelity is given by the diffusion equation (15.2). A data term penalizing pixel-wise deviations from this constraint is of the form

$$E_{data}(I, D) = \int_{\Omega} \phi((I_t - D\Delta I)^2) dx. \tag{15.5}$$

However, point-wise measurements are insufficient because of the random nature of the diffusion process. Therefore we design a neighborhood-wise data penalization by assuming a constant diffusion coefficient $D(x)$ over a neighborhood of a pixel x. This approach is similar to the CLG method for optical flow [Bruhn et al., 2005] (see Chapter 6). The data term is defined as

$$E_{data}(I, D) = \int_{\Omega} \phi(\mathbf{D}^{\top} \mathbf{J}_{\rho} \mathbf{D}) dx \tag{15.6}$$

$$\text{with } \mathbf{D} = \begin{pmatrix} D \\ 1 \end{pmatrix} \text{ and } \mathbf{J}_{\rho} = K_{\rho} * \begin{pmatrix} \Delta I^2 & -I_t \Delta I \\ -I_t \Delta I & I_t^2 \end{pmatrix}.$$

The regularization term imposes smoothness of D :

$$E_{reg}(D) = \int_{\Omega} \phi(\|\nabla D\|^2) dx. \tag{15.7}$$

The constraint $D(x) \geq 0$ is achieved by adding a logarithmic barrier to the energy of (15.4), such that the final energy writes

$$E(D) = \int_{\Omega} \phi(\mathbf{D}^{\mathbf{T}} J_{\rho} \mathbf{D}) dx + \lambda \int_{\Omega} \phi(\|\nabla D\|^2) dx - \mu \int_{\Omega} \log(D) dx. \tag{15.8}$$

The minimization procedure is the same as for motion estimation in Chapter 13. Up to our knowledge, it is the first time a diffusion field is estimated by variational minimization of a global regularized energy.

Contrary to the TICS method, this variational approach produces dense diffusion fields. Spatial variations of diffusion can thus be recovered more accurately, in particular at diffusion discontinuities usually occurring across membranes. Another difference is that the variational method exploits only two frames for the diffusion estimation whereas TICS uses extended sub-sequences. This choice sharpens the detection of temporal diffusion changes. However taking more frames into consideration can also improve the robustness of the estimation inside constant diffusion phases, therefore the estimations are averaged in practice over each phase.

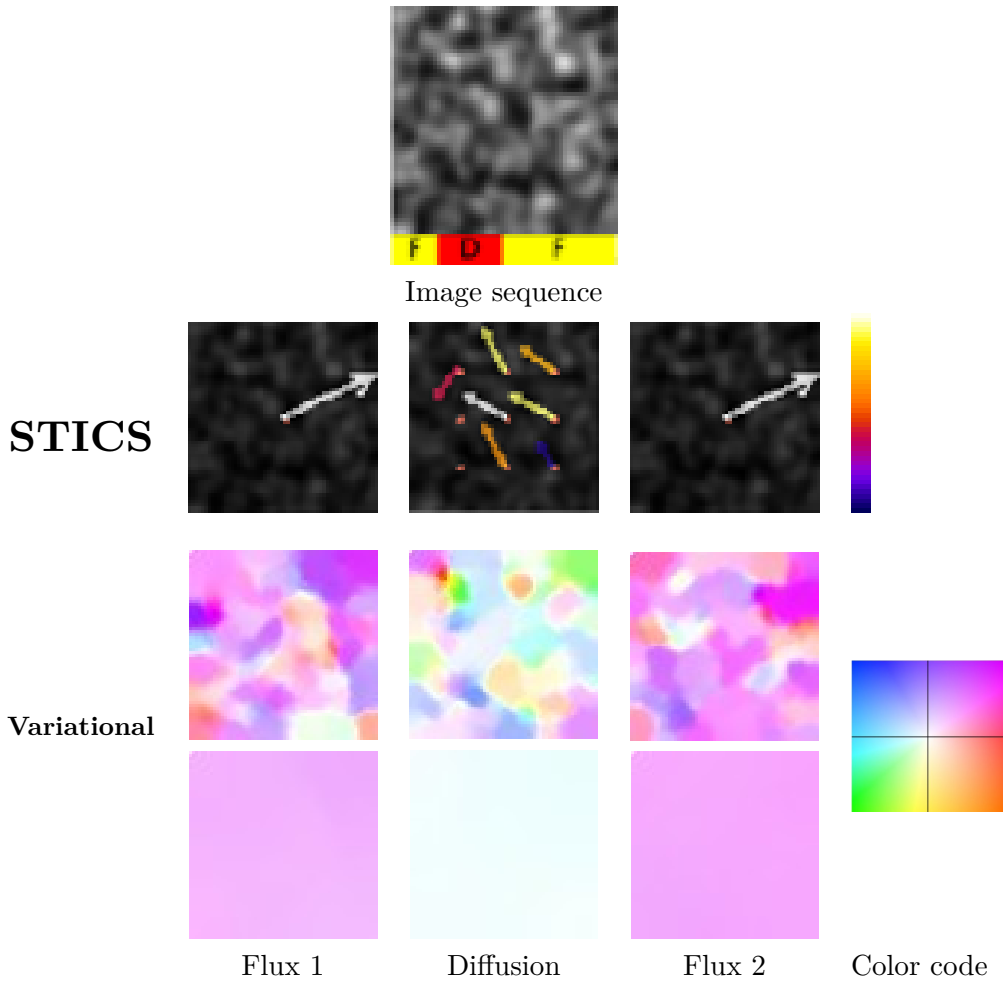


Figure 15.3: Analysis of TICS and variational method on synthetic image time series with three phases. First row: first frame of the sequence and temporal description of the 3 phases: F: Flux, D: Diffusion. Second row: TICS motion estimation for each phase. Third and fourth rows: Variational motion estimation for a selection of image pairs of each phase with $\lambda = 5$ (third row) and $\lambda = 11$ (fourth row) (we set $\gamma = 0.75$)

.

15.3 Experimental results

In this section, we present three different biological problems which necessitate motion or diffusion estimation. We compare the variational and TICS approaches and demonstrate their potential.

Temporally varying diffusion We simulated an image time sequence shown in Fig.15.3-a and composed of three phases: pure directional flow (North-East translation,

	Flux 1	Flux 2
TICS	0.744 / 0.0125	0.733 / 0.0126
Variational, $\lambda = 5$	0.728 / 0.0124	0.782 / 0.0133
Variational, $\lambda = 11$	2.20 / 0.0372	1.67 / 0.0281

Table 15.1: Error of estimated translational motion of the Flux phases obtained with TICS and variational method for two values of λ (each cell of the table reports "Angular Error / Endpoint Error").

	Flux1	Trans.	Diffusion	Trans.	Flux2
Ground truth	1/20	-	21/50	-	51/90
TICS	1/10	11/17	18/39	40/48	49/90
Variational	1/18	-	19/48	-	49/90

Table 15.2: Detection of the three phases of the simulated sequence with TICS and variational method (each cell of the table reports "begin frame / end frame" of the phase).

0.28μm/s), pure diffusion ($D = 0.01\mu\text{m}^2/\text{s}$), and pure directional flow (North-East translation, 0.28μm/s), respectively referred as Flux 1, Diffusion and Flux 2. This artificial example mimics a possible scenario observed in biological experiments with molecules in cells or beads in solutions. This simulation is used to demonstrate the ability of both methods to compute the dynamic parameters and to identify the three phases.

Figure 15.3 shows visual results obtained with TICS and variational methods. Two regularization parameters λ are compared in the variational case. We adopt two different motion visualizations : the motion vectors estimated every 16 pixels by TICS are represented by arrows, whereas the dense motion field of the variational method is more accurately visualized with the standard color-code presented in Fig. 15.3-e. The TICS exploits sub-sequences of 12 images.

The directional flow of Flux phases is well captured by the two methods. In the variational approach, small values for λ result in a detection of small moving structures. Higher values for λ allow us to recover the global flux, which is almost null during the diffusion phase and constant ("North-East" direction) during the two Flux phases.

In Table 15.1, we have reported the Angular Error and Endpoint Error of motion estimations in the two flux phases (see [Baker et al., 2011] for definitions). The results show the high accuracy of TICS for estimating the Flux translation. With the variational approach, the translation is obtained by averaging the estimations at each pixel and for each frame of the Flux phases. We have reported the results in Table 15.1 for two regularization parameters. We notice that, in contradiction with the visual impression of Fig. 15.3, the estimated motion for $\lambda = 11$ corresponds to a high error, whereas for $\lambda = 5$, the error is consistent with the TICS method. This is due to the fact that large regularization coefficient favors smoothness against accuracy and tends to under-estimate

the motion magnitude.

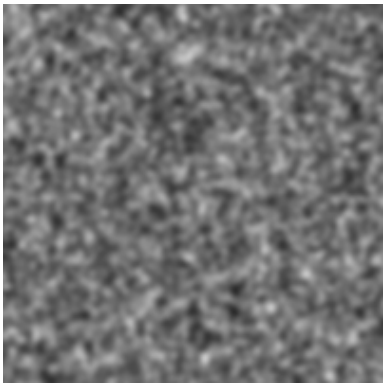
Based on this motion estimation, we identify the three phases. With the TICS method, we use backward and forward hidden Markov model on three states : Flux, Diffusion and Transition. With the variational method, a threshold is applied on the magnitude of the spatially averaged motion vectors of each frame to classify Flux and Diffusion phases. The classification results presented in Table 15.2 are consistent with the ground truth. Nevertheless, the TICS approach necessitates to consider a temporal analysis of a significant number of frames to produce satisfying diffusion estimation results. In our experiments, we processed 12 frames at each time point, which introduce a transition phase making the detection of transitions less accurate. The use of only two frames by the variational method allows to detect abrupt transitions.

Furthermore, the diffusion phase was analyzed separately to estimate the biophysical parameter related to diffusion rate. The diffusion coefficient of the diffusion phase was set to $D = 0.01\mu\text{m}^2/\text{s}$. The TICS estimates a coefficient $\widehat{D} = 0.013\mu\text{m}^2/\text{s}$, and the variational methods estimates in average $\widehat{D} = 0.016\mu\text{m}^2/\text{s}$. Thus, in this case of spatially constant diffusion, TICS achieves highest accuracy in the diffusion coefficient estimation.

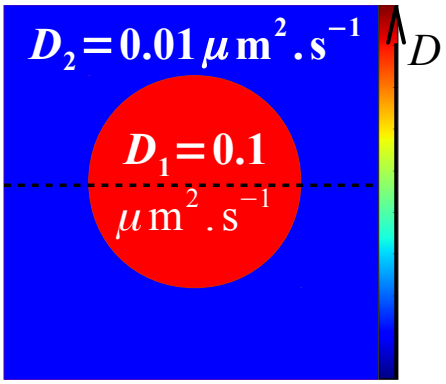
Spatially varying diffusion To evaluate the ability of the variational approach to produce dense diffusion field and thus to accurately recover spatial diffusion changes, we simulated a spatially varying diffusion sequence. The particles density is the same in the whole image domain, as shown in Fig. 15.4-a, but the diffusion coefficient is higher inside the circle represented in Fig. 15.4-b ($D_1 = 0.1\mu\text{m}^2/\text{s}$) than outside ($D_2 = 0.01\mu\text{m}^2/\text{s}$). The parameters of the variational method are set to $\rho = 5$ and $\lambda = 500$.

The estimated diffusion field shown in Fig. 15.4-c is visually very close the ground truth. The two regions are clearly delimited by the circle, despite an over-smoothing of the transition. The histogram of the diffusion field (Fig. 15.4-d) shows two modes corresponding to the two diffusion regions, but it also indicates that the diffusion inside the circle is under-estimated. Figure 15.4-e shows the result of the mean-shift algorithm [Comaniciu and Meer, 2002] applied to the diffusion field in order to delineate the two regions. Finally, 2D profiles of a line crossing the circle are represented in Fig. 15.4-d for ground truth, estimated and segmented diffusion fields. We observe that spatial discontinuities are accurately recovered, but $\widehat{D}_1 = 0.082\mu\text{m}^2/\text{s}$ is slightly under-estimated, while $\widehat{D}_2 = 0.017\mu\text{m}^2/\text{s}$ is slightly over-estimated.

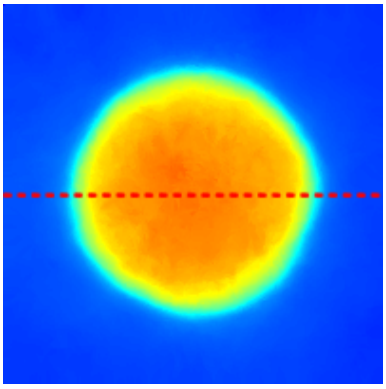
The TICS method is enable to produce such dense diffusion segmentation map because of the spatial lag between the patches used to estimate diffusion. Besides, the large patch size leads to erroneous estimations near diffusion boundaries.



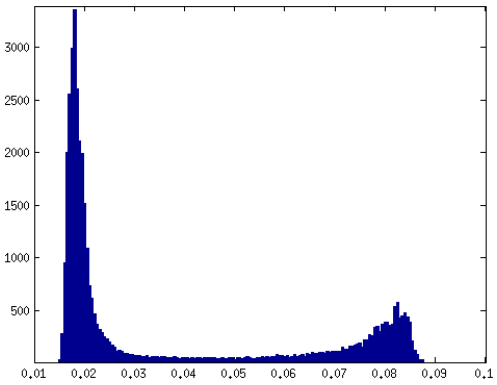
(a) Frame 1



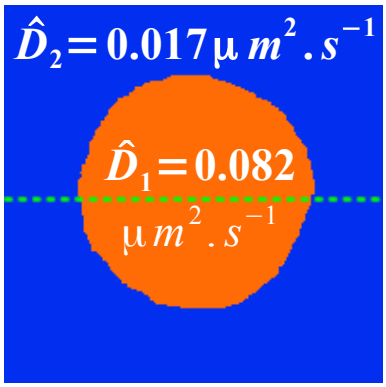
(b) Ground truth diffusion field



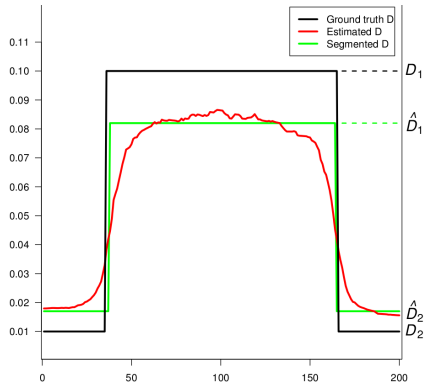
(c) Estimated diffusion field D



(d) Histogram of D



(e) Segmented diffusion field



(f) Profiles

Figure 15.4: Variational diffusion estimation on a simulated sequence with spatially variant diffusion. The curves of (f) are profiles of the dashed lines in (b),(c) and (e)

15.4 Conclusion

In this chapter, we have proposed a new variational diffusion estimation method able to recover spatially varying diffusion coefficients. The standard ICS approach is block-based and takes advantages of a fitting model to recover the physical parameters. Our method achieves similar results to ICS for the estimation of spatially constant diffusion coefficient. When diffusion is spatially varying, ICS is unable to recover discontinuities of the diffusion field, whereas our variational method yields accurate discontinuous diffusion maps.

An interesting future research direction would be to combine these two approaches, to take advantage of both the physical modeling of TICS and the regularization properties of the variational approach. It could be naturally carried out by using the aggregation framework described in Part II. Indeed, the patch-based TICS diffusion estimations could serve to generate candidates and the global energy model in Section 15.2 can be used as objective energy involved in the aggregation stage. Similarly to the case of optical flow, the aggregation scheme would implicitly select the most appropriate local support for the estimation of spatially constant diffusion produced by TICS at each pixel.

16 Conclusion

In this part we have first performed a comparative evaluation of correlation-based and variational approaches for motion estimation in live cell imaging. We identified a wide range of situations where dense motion estimation can bring important information to answer biological questions. The advantages and limitations of correlation and variational approaches have been pointed out from these experiments, and we emphasized on the complementarity of their performances, in particular for fluorescence imaging.

In a second time, we have adapted the the two-step aggregation framework introduced in Part II to overcome limitations of state-of-the-art variational approaches in fluorescence imaging. Our approach is designed as a combination of correlation and variational methods, and we address the problem of large local fluorescence variations by explicitely estimating the intensity change map. We have demonstrated the performance of our method in various biological contexts of dense and smooth motion fields as well as sparse and discontinuous ones. We successfully handle frequent and challenging situations arising from live cell imaging sequences like large displacements of small structures (e.g. cell migration) or local fluorescence intensity changes (e.g fluorescent protein recruitment).

Finally, we have proposed a new variational diffusion estimation method, as an alternative to traditional TICS method. In simple cases of spatially constant diffusion, our method is as accurate as TICS. For spatially discontinuous diffusion maps TICS is limited by its block-based approach preventing from retrieving discontinuities, whereas our method generates accurate discontinuous diffusion fields. An integration of the TICS and our global model in the aggregation framework is an appealing approach which would allow to combine the local physical modelling of TICS with global energy models.

Finally, more quantitative analysis are required to evaluate the performance of proposed methods in biological imaging. We plan to calibrate the proposed methodology from sequences showing molecules undergoing Brownian motion with a known diffusion coefficient.

General conclusion

Contributions

We have addressed throughout this thesis several issues related to optical flow estimation, which we identified as the main limiting points of modern methods. Our general approach is to combine local and global models to overcome these issues.

We first proposed a method based on the work of [Bruhn et al., 2005] integrating local energy as a data term of a global model. Our method spatially adapts the Gaussian kernel used to filter standard data potential. Preliminary results have demonstrated the relevance of the approach. We also investigated a modified version integrating the stochastic uncertainty model of [Corpetti and Mémin, 2012].

Secondly, we have designed an original aggregation framework exploring another way to combine local and global models. The main idea is to revisit local parametric estimations in patches, especially the selection of appropriate patch sizes and positions. Our aggregation framework is conceived as a two-step procedure, beginning with local parametric computations repeated in a general patch distribution, followed by an aggregation step selecting among the motion candidates generated by the first local step. We investigated two versions of the aggregation problem, using discrete and continuous optimization. The discrete aggregation method yields more accurate results. The continuous optimization alternative is still interesting from a practical point of view, for its reduced computational time and robustness to lower quality of the candidates. We introduce a generic exemplar-based approach for occlusion handling to overcome limitations of traditional diffusion-based techniques.

Our AggregFlow method has brought significant contributions to open problems related to optical flow. We list hereafter the main features investigated and problems solved:

- We have demonstrated that simple local parametric estimations can potentially outperform any existing methods when patches are appropriately chosen.

- The aggregation framework combines originally local and global models. Considering local estimations in various spatial supports to produce candidates, the selection of one candidate in the global aggregation step amounts to select the best spatial support for estimation at each pixel, without segmentation step.

- The aggregation allows an original and efficient integration of feature matching in the optical flow estimation process, as coarse initialization of local parametric estimation. The main advantages of our integration scheme are: 1) we keep the N best matches and not the single best one, which reduces the risk of matching errors; 2) We do not incorporate matches as additional constraints in a global energy, as done by most methods, which also reduces the sensitivity to matching errors; 3) We do not resort to global variational optimization in the refinement process, as other methods do.

- An occlusion detection process is designed cooperatively on both steps, under the form of local confidence measure in the first step and used to guide a sparsity constraint in the aggregation step. Estimation of motion in those detected occluded regions is performed in both steps with an exemplar-based approach, outperforming traditional diffusion-based methods. The occlusion filling model is generic and could be integrated in other global methods as well.

- To deal with large intensity changes, we propose a joint modeling of motion and intensity changes at the two steps of the method. The first estimates intensity change candidates along with motion candidates, producing a single label space for these two variables. The global optimization of the aggregation therefore does not need to alternate between optimizations w.r.t. the two variables, but directly operates in the two-component label space, which makes optimization easier.

We have experimentally evaluated the improvements yielded by our approach on reference computer vision benchmarks. On small displacement sequences, where occlusions and illumination changes are not essential and the main challenges lies in the fine structures and motion discontinuities, our method gives satisfying results, competitive with most current methods. When larger displacements occur, our occlusion handling framework and patch correspondences strategy allows us to outperform existing methods. The improvements in occluded regions are particularly important. Examples of very large intensity changes occurring in fluorescence imaging situations also attest for the superiority of our method compared to other approaches.

Motion in biological imaging sequences often significantly differs from standard challenges in other computer vision application domains. We have proposed solutions to deal with two kind of situations, namely large intensity changes due to fluorescence variations, and diffusion coefficient estimation in the frequent case of Brownian motion of particles. The first problem was treated in the aggregation framework and qualitative experiments confirmed the efficiency of our method. To overcome the limitations of existing correlation-based methods for diffusion coefficient estimation, we have designed a variational approach retrieving accurately space-varying diffusion discontinuities.

Perspectives

Several research directions remain opened at the end of this work:

- Our study on adaptation of Gaussian kernel for combined local-global methods (Chapter 6) opened the way to several extensions: anisotropic Gaussian filtering should improve current results; variations on the regularization terms are also possible; optimization difficulties for integrating stochastic models in the adaptive framework should be addressed.

- The label space composed by the candidates exhibits several unusual properties to be dealt with by the discrete optimization involved in the aggregation. Indeed, traditional graph-cut *move-making* methods are used to deal with either small label sets, or fusion of global proposals [Lempitsky et al., 2010]. In our case, the candidates locality, and their redundancy could be exploited to generate adapted proposals in the *move-making* optimization process.

- The computation time of the method could be dramatically reduced by GPU implementation. Indeed, the candidates computation is a naturally parallelizable process.

- An important cause of errors in AggregFlow is due to matching errors in the exemplar-based occlusion handling process. We used a straightforward sum of point-to-point L_1 distance as a similarity measures of patches. More efficient feature matching could be used. In particular, since we are not looking for exact matches but for pixels belonging to the same object, texture comparison could be more appropriate.

- The patch correspondences used to produce motion candidates could also be improved by recently proposed matching methods [Weinzaepfel et al., 2013; Leordeanu et al., 2013].

- While we have proposed a version of the method adapted to the wide variety of challenges proposed by computer vision benchmarks, other more dedicated applications could adapt each feature of the framework to its own needs. We have opened this perspective with our adaptation of the aggregation framework to large fluorescence variations in light microscopy image sequences. Other biological problems could be addressed more specifically.

- We are also looking for biologically relevant applications of our variational diffusion coefficient estimation method.

- A unifying approach of regularized models and Temporal Image Correlation Spectroscopy techniques (TICS) for diffusion estimation could be investigated in the aggregation framework. The patch-based nature of TICS makes it well suited to produce candidates, and our regularized model of diffusion estimation can be used as a global aggregation model
- Specific quantitative protocols must be designed to evaluate the precision of motion estimation methods in biological imaging. This is challenging in live cell imaging since microscopy serves as an investigation tool. Reference frames cannot be controlled and ground truth is only available for simple dynamics not representative of the complexity observed in real data.

Part IV

Appendices

A Semi-local variational optical flow estimation

Published paper:
D. Fortun and C. Kervrann. Semi-local variational optical flow estimation. In *International Conference on Image Processing (ICIP)*, pages 77–80, 2012.

Global variational methods for optical flow estimation usually suffer from an over-smoothing effect. We propose a semi-local estimation framework designed to integrate and improve any variational method. The idea is to implicitly segment the minimization domain into coherently moving windows. In a first time, local variational estimations are performed in overlapping candidate square regions. Then, a global discrete optimization, non subject to the over-smoothing introduced by variational approaches, selects the optimal window for each pixel. Experimental results show an increasing of the sharpness of discontinuities and a significant improvement of global registration errors compared to the results of the baseline global variational method.

A.1 Introduction

The optical flow approximates the projection of the motion of a 3D scene on the image plane. Any optical flow estimation have thus to be based on a conservation assumption of some optical properties of the image able to capture the real motion (intensity, gradient, image descriptor ...). This data conservation constraint provides in general a single equation and is consequently insufficient to recover the two components of the motion field (*aperture problem*). The typical way to overcome this under-determination is to add to the data conservation constraint a spatial coherency constraint. Existing methods can be classified regarding their *local* or *global* strategy to impose such a constraint.

The spatial coherency of *local* approaches is ensured at a pixel $x \in \Omega \subset \mathbb{R}^2$ by the assumption of common parametric motion (translational in [Lucas and Kanade, 1981]) in a neighborhood $V(x) \subset \Omega$, where Ω is the image domain. The *global* approach allows to compute a dense motion field and expliciteply adds to a data potential $\rho_{data}(\cdot)$ which penalizes deviations from the data conservation constraint, a regularization potential $\rho_{reg}(\cdot)$ which penalizes high values of the norm of the gradient $\nabla \mathbf{w}$ of the velocity field

$\mathbf{w} : \Omega \to \mathbb{R}^2$. A global energy combining these two potentials is minimized [Horn and Schunck, 1981]:

$$E_{global}(\mathbf{w}) = \int_{\Omega} \rho_{data}(x, I, \mathbf{w}) + \lambda \, \rho_{reg}(x, \nabla \mathbf{w}) dx \tag{A.1}$$

where $I : \Omega \times [0, T] \to \mathbb{R}$ is an image sequence and λ is a balance parameter between data fitting and regularization.

The best results of the state-of-the-art are achieved by global variational motion estimation methods. However, this kind of methods still suffer from an over-smoothing effect. This phenomenon is particularly visible in the seminal work of [Horn and Schunck, 1981] which uses quadratic penalty function for the regularization potential. This shortcoming has been greatly reduced by the introduction of robust penalty functions [Black and Anandan, 1996; Mémin and Pérez, 1998], adaptation of the regularization along image discontinuities [Wedel et al., 2009a] or non-local regularization strategies [Werlberger et al., 2010], but it still remains. Indeed these methods are limited by the coarse-to-fine scheme [Mémin and Pérez, 1998; Brox et al., 2004], necessary to cope with large displacements. This approach avoids most local minima due to the non-convexity induced by the non-linearized data potential, at the price of an over-smoothing of the discontinuities.

We mention two non-variational approaches related to our method that have been investigated to reduce the over-smoothing effect of global variational methods: (i) parametric motion estimation based on motion field segmentation; (ii) discrete optimization of the energy (A.1). In the first case, a parametric model of the flow field is estimated inside coherently moving regions. The estimation of the discontinuities is thus transfered to the segmentation step [Sun et al., 2010b]. In the second case, discrete optimization of the energy (A.1) is able to find strong minima for non-convex functionals without coarse-to-fine schemes, but is limited by the quantization of the flow field range [Boykov et al., 2001].

In this paper, we present a method combining local estimations and discrete optimization to sharpen the discontinuities of a global variational method by implicitly segmenting the flow field. It is composed of two stages: first, local variational estimations are performed on a regular grid of overlapping windows; second, the resulting local motion vectors are used as candidates for a global discrete optimization. In this scheme, the discrete optimization module selects of the optimal spatial minimization domains, yielding an implicit segmentation of the flow field. It is worth noting that our framework is general and can be used to improve any baseline variational method. The results with the popular and representative method [Brox et al., 2004] show significant improvements over the global variational approach when applied on several sequences of the Middlebury database [Baker et al., 2011].

A.2 Variational optical flow estimation

A.2.1 Global variational approach

All global variational optical flow estimation methods are based on the minimization of the energy (A.1). The method described in [Brox et al., 2004] contains most of the basis concepts still used in the most recent methods. Therefore we used this method to evaluate our semi-local framework, which can integrate any other variational optical flow estimation methods. The main features of the method [Brox et al., 2004] are described in this section.

Global energy functional The data potential penalizes deviations from intensity and gradient conservation constraints with a L_1 penalty function:

$$\begin{aligned}\rho_{data}(x, I, \mathbf{w}) = \phi\left(|I(x+\mathbf{w}(x), t+1) - I(x,t)|^2\right) \\ + \gamma\phi\left(\|\nabla I(x+\mathbf{w}(x), t+1) - \nabla I(x,t)\|^2\right)\end{aligned} \tag{A.2}$$

where $\gamma > 0$ is a balance parameter and $\phi\left(z^2\right) = \sqrt{z^2+\epsilon^2}$ is a regularized form of the L_1 norm with ϵ a small constant.

The regularization potential penalizes high gradients with the same convex and discontinuity-preserving L_1 norm:

$$\rho_{reg}(x, \nabla\mathbf{w}) = \phi\left(\|\nabla u(x)\|^2 + \|\nabla v(x)\|^2\right) \tag{A.3}$$

Energy minimization The minimization of (A.1) is performed by solving the Euler-Lagrange equations. To make these equations tractable, the data potential (A.2) must be linearized, which limits the estimation to small displacements. Therefore all recent variational methods adopted a coarse-to-fine scheme to handle large displacements [Mémin and Pérez, 1998; Brox et al., 2004]. The coarse-to-fine levels are interpreted in [Brox et al., 2004] as fixed point iterations enabling the minimization of the initial non-linear energy. At each level, a second fixed point allows to cope with the remaining non-linearity of the Euler-Lagrange equations due to the L_1 norm. The resulting linear system is then solved with Successive Over Relaxation (SOR).

A.2.2 Restriction to local domains

The minimum reached by global variational methods is usually suboptimal. Indeed, variational optimization is proved to find the global minimum only for convex energies, which is not the case of (A.1) with the non-linearized data potential (A.2). Actually, the coarse-to-fine scheme transforms the problem into successive minimizations of convex approximations of (A.1) which tend to smooth the discontinuities of the flow field.

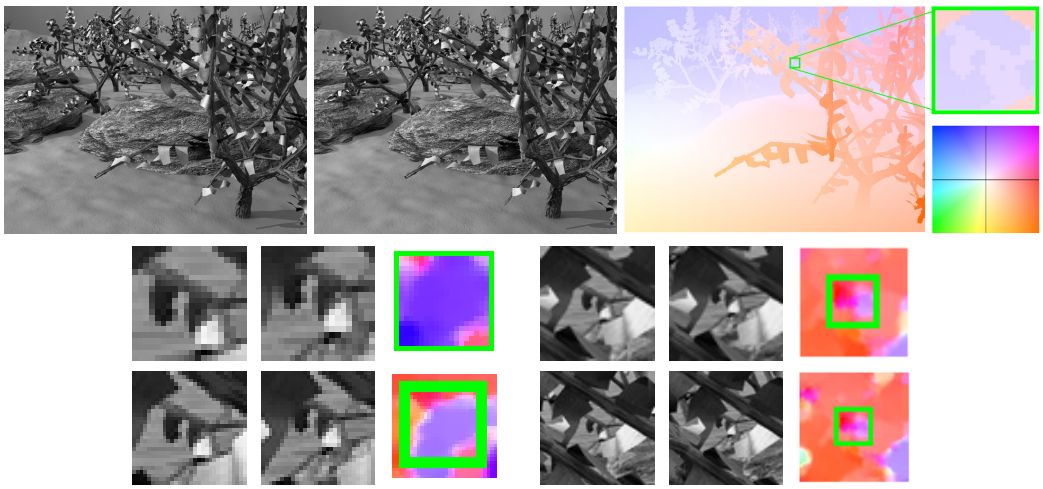


Figure A.1: Influence of the spatial minimization domain. 1^{st} row: two frames and ground truth of the motion field with a zoom on a region of interest (green square). 2^{nd} and 3^{d} rows: variational estimations over windows of different sizes centered on the same green square.

To reduce this over-smoothing effect we propose to minimize several energies of the form (A.1) over local regions, inspired by the localization of Total Variation for denoising [C. and L., 2011]. The local motion field
$w_V : V \subset \Omega \to \mathbb{R}^2$ is the minimizer of

$$E(\mathbf{w}_V) = \int_V \rho_{data}(x, I, w_V) + \lambda\, \rho_{reg}(x, \nabla w_V) dx \tag{A.4}$$

obtained with the method described in the previous section. If V is a coherently moving region without strong discontinuities, the over-smoothing does not occur. We emphasize that contrary to other approaches computing optical flow in local regions [Sun et al., 2010b; Zitnick et al., 2005], where the motion is restricted to a parametric model, we prefer to compute a regularized flow field. Thus we overcome the difficulty of local parametric approaches to handle complex motion fields, where the region V must be small enough to ensure the validity of the parametric assumption, and large enough to avoid the *aperture problem*. Figure 1 shows how the localization of variational methods can improve the accuracy of the global approach when the region is suitably chosen: as the window size increases, the details of the flow tends to disappear. Our strategy to select the optimal region of each pixel is presented in the next section.

A.3 Semi-local framework

Our aim is to estimate the motion at each pixel with a local variational model (A.4) in coherently moving regions. Existing similar approaches are restricted to parametric motion and are based either on an image gradient segmentation [Zitnick et al., 2005], subject to over-segmentation of the flow field, or on global variational minimization of non-convex functionals [Sun et al., 2010b] for joint estimation of regions and motion, suffering from the drawbacks described in the previous section. Our semi-local framework performs an *implicit* segmentation of the flow field, without global variational estimation or image segmentation.

Local estimations Local variational estimations are performed on overlapping square windows of different sizes. For a fixed size s, let $\mathcal{V}_{s,\alpha}$ be a set of regularly spaced windows covering the whole image, with an amount of overlap α between neighbors (see Fig. 2). For a set of varying sizes $\mathcal{S} = \{s_0, \dots, s_n\}$, we define $\mathcal{V}_{\mathcal{S},\alpha} = \bigcup_{s\in\mathcal{S}} \mathcal{V}_{s,\alpha}$. For each window $V \in \mathcal{V}_{\mathcal{S},\alpha}$, a local motion field $\mathbf{w}_V$ is estimated by minimization of (A.4).

One pixel is contained in several overlapping windows with different locations and sizes. We denote $\mathcal{N}_V(x)$ the set of windows containing the pixel x (see Fig. 2, $\mathcal{N}_V(x) = V_1, .., V_4\}$). The computed flow fields over these windows provide a set $\{\mathbf{w}_V(x)\}_{V\in\mathcal{N}_V(x)}$ of candidate motion vectors at each pixel x.

Global aggregation The aggregation step aims at combining the locally estimated flow fields to compute an optimal global flow field. The goal is to select at each pixel the most appropriate window. To this end, we consider the aggregation as a multi-label assignment problem, where local candidates $\{\mathbf{w}_V(x)\}_{V\in\mathcal{N}_V(x)}$ constitute the discrete label space at pixel x. The global flow field w_Ω resulting from the aggregation is then the minimizer of a global objective energy:

$$w_\Omega = \arg\min_{\mathbf{w}} E_\Omega(\mathbf{w}) \text{ s.t. } \mathbf{w}(x) \in \{\mathbf{w}_V(x)\}_{V\in\mathcal{N}_V(x)}. \tag{A.5}$$

Thus, the solution is found by selecting the best motion vector among the small set of candidates $\{\mathbf{w}(x)\}_{V\in\mathcal{N}_V(x)}$. We define E_Ω as:

$$E_\Omega(\mathbf{w}) = \sum_{x\in\Omega} \rho_{data}(x, I, \mathbf{w}) + \lambda\psi_{reg}(x, \mathbf{w}) \tag{A.6}$$

where $\psi_{reg}(\cdot)$ is a Markov Random Field prior:

$$\psi_{reg}(x, \mathbf{w}) = \sum_{y\in\Delta(x)} \phi\left(|u(x) - u(y)|^2 + |v(x) - v(y)|^2\right)$$

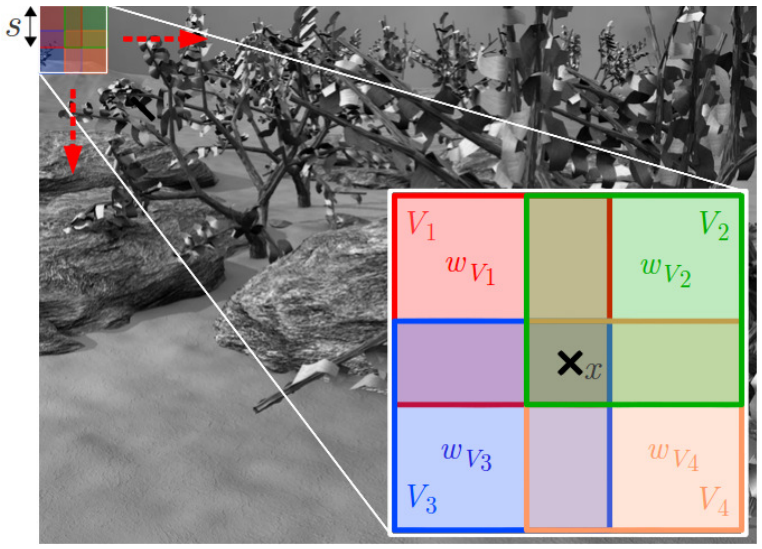


Figure A.2: Illustration of the patches distribution for a given size s and overlapping ratio $\alpha = 0.3$. The pixel x is contained in four patches $V_1, \ldots, V_4$ providing four candidates for x: $w_{V_1}(x), \ldots, w_{V_4}(x)$.

where $\Delta(x)$ is a 4 or 8 pixel neighborhood. The energy E_Ω is actually a discrete version of the energy (A.1). The difference with global variational minmization of (A.1) is that we impose the solution to belong to the set of local variational estimations $\{\mathbf{w}_V(x)\}_{V \in \mathcal{N}_V(x)}$. Our discrete optimization scheme does not suffer from the over-smoothing effect of the global variational approach. Consequently, it selects the candidates coming from coherently moving regions, which are not affected by this over-smoothing (see the smallest window in Fig. 1). This selection of the optimal window at each pixel can be seen as an implicit segmentation of the spatial minimization domain used for the baseline variational method.

Discrete optimization In this subsection we detail our approach for the discrete optimization problem (A.5). The formulation of (A.5) differs from the classical multi-label assignment scheme because the regularization is not applied to the discrete labels but to the continuous-valued motion vectors assigned to the labels. This problem has been addressed in the context of optical flow estimation in [Lempitsky et al., 2010] with candidates estimated by several global methods. The authors achieved the multi-label optimization with the *fusion-move* algorithm, which operates by successive fusions of proposal labellings (see [Lempitsky et al., 2010] for more details about *fusion-move* and its applications). We apply this technique for solving (A.5).

The *fusion-move* algorithm requires a set of global flow fields $\{w_{\Omega_1}, \ldots, w_{\Omega_N}\}$ to be fused. Let us consider that the set of candidates at each pixel $\{\mathbf{w}_V(x)\}_{V \in \mathcal{N}_V(x)}$ is

composed by a maximum number of N candidates. Then for $i \in \{1 \dots N\}$, we assign to $w_{\Omega_i}(x)$ one arbitrarily chosen candidate in $\{\mathbf{w}_V(x)\}_{V \in \mathcal{N}_V(x)}$, so that each candidate is assigned to at least one global flow field $w_{\Omega_i}(x)$. Our experiments showed that the assignment strategy of the local candidates in the global flow fields has negligible influence on the final result.

A.4 Results and discussion

We use the Average Angular Error (AAE) to evaluate the performance of our method on sequences of the Middlebury benchmark [Baker et al., 2011]. Two aggregation procedures are considered: *SL-fusion* performs the discrete optimization described in Section 3.2 and 3.3; *SL-Mean* performs a weighted mean of the local candidates, favoring central pixels with a gaussian filter centered on each window. In all our experiments we fix the parameters of [Brox et al., 2004] $\lambda = 40$ and $\gamma = 5$.

Table 1 compares the AAE of the baseline variational method [Brox et al., 2004] with *SL-fusion* and *SL-Mean*, for several sets of window sizes $\mathcal{S}$. The superior performance of *SL-fusion* over *SL-Mean* highlights that the selection of the best window is crucial to prevent the global flow field from being influenced by outliers coming from inappropriate regions. For *SL-fusion*, the errors obtained with single sizes are always higher than those obtained with their combination. This result shows that a single window size is not able to capture all types of coherently moving regions and that the aggregation procedure successfully combines the advantages of each size by selecting the best region. The results of *SL-fusion* with $\mathcal{S} = \{15, 49, 129\}$ are significantly better than those of the baseline variational method for the sequences *Grove3*, *Rubberwhale* and *Grove2*. This is due to the better preservation of the discontinuities illustrated in Fig. 3. We mention that even for very smooth sequences like *Dimetrodon*, large window sizes (here 129) ensure that the result cannot be worse than the global variational method, and is even slightly improved.

The influence of the amount of overlap α is shown in Fig. 4. As it can be intuitively expected, the error decreases when the overlap increase. Indeed, the overlap determines the number of candidate regions, and thus the probability that the windows fall in appropriate regions.

A.5 Conclusion and future work

We proposed a new approach for optical flow estimation, combining global methods with local minimization regions. Our experiments showed that our semi-local framework improves the estimation accuracy of a baseline variational method along discontinuities of the motion field, by performing an implicit segmentation of the spatial minimization

	Grove2	Grove3	RubberWhale	Dimetrodon
Variational [Brox et al., 2004]	**2.38**	**5.97**	**3.92**	**1.83**
SL-Mean				
$\mathcal{S} = \{15\}$	4.91	17.7	5.42	4.40
$\mathcal{S} = \{49\}$	2.43	6.11	4.04	1.91
$\mathcal{S} = \{129\}$	2.38	6.01	3.98	1.83
$\mathcal{S} = \{15, 49, 129\}$	4.21	16.2	5.05	3.54
SL-fusion				
$\mathcal{S} = \{15\}$	2.95	12.8	4.47	3.30
$\mathcal{S} = \{49\}$	2.27	5.85	3.69	1.87
$\mathcal{S} = \{129\}$	2.30	5.83	3.71	1.81
$\mathcal{S} = \{15, 49, 129\}$	**2.10**	**5.60**	**3.34**	**1.79**

Table A.1: Comparison of the results (AAE) obtained with our implementation of [Brox et al., 2004], *SL-Mean* and *SL-fusion* for $\alpha = 0.75$.

domain. In the future we plan to extend this general framework to non-variational baseline methods and adaptive region shapes.

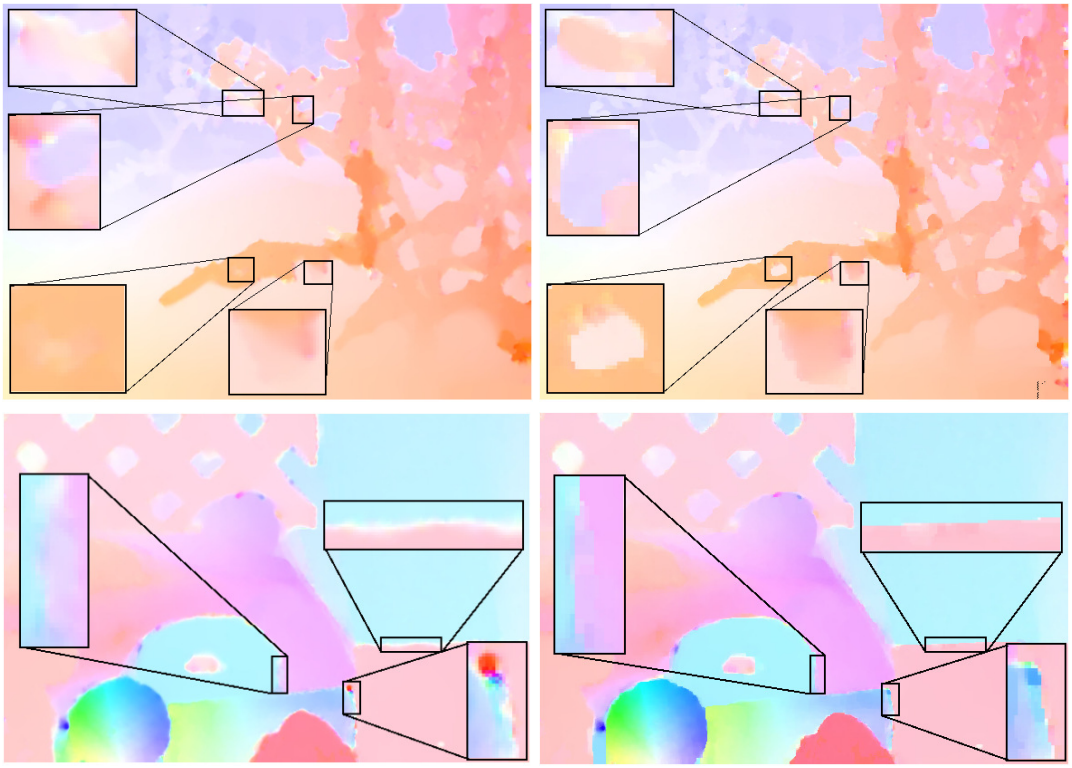


Figure A.3: Visual comparison between global variational estimation (1^{st} column) and its integration in *SL-fusion* with $\mathcal{S} = \{15, 49, 129\}$ and $\alpha = 0.75$ (2^{nd} column) on *Grove3* and *Rubberwhale*.

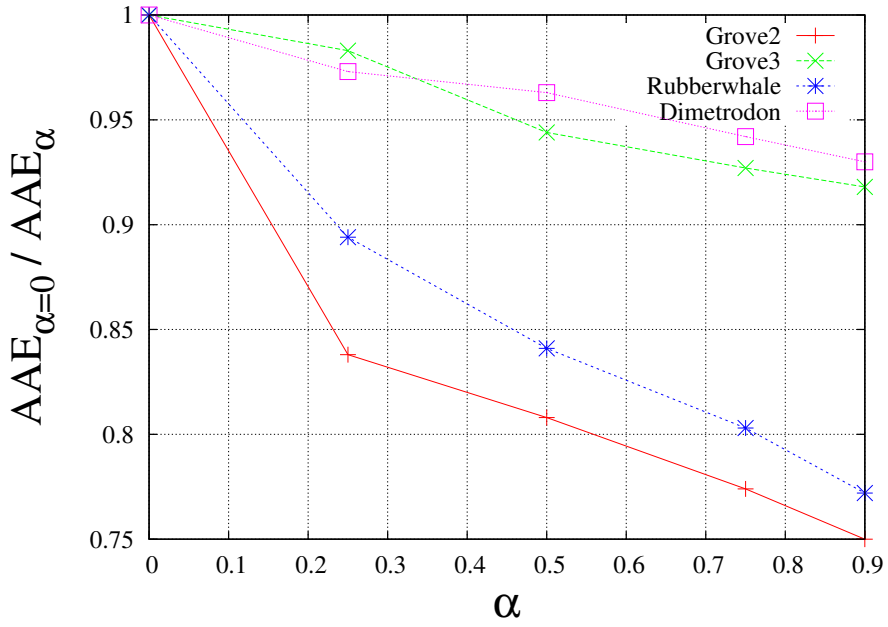


Figure A.4: Influence of α on the AAE. We plot $\mathrm{AAE}_{\alpha=0}/\mathrm{AAE}_{\alpha}$.

B Differentiation of (6.12)

We want to compute $\partial E/\partial\sigma(x)$, where

$$E(\mathbf{w},\sigma) = \sum_{x\in\Omega} \phi_d\left(\boldsymbol{\alpha}^T_{\mathbf{w},\sigma(x)}\mathbf{J}^0_{\sigma(x)}\boldsymbol{\alpha}_{\mathbf{w},\sigma(x)}\right) + \phi_d\left(\boldsymbol{\alpha}^T_{\mathbf{w},\sigma(x)}\mathbf{J}^{x_1x_2}_{\sigma(x)}\boldsymbol{\alpha}_{\mathbf{w},\sigma(x)}\right) + \beta \sum_{z\in\mathcal{N}(x)} \phi_r((\sigma(x)-\sigma(z))^2)$$

where we denote x_1, x_2 the vertical and horizontal axes, and

$$\boldsymbol{\alpha}_{\mathbf{w}}(x) = \begin{pmatrix} u(x) \\ v(x) \\ 1 \end{pmatrix},\ \mathbf{w} = \begin{pmatrix} u \\ v \end{pmatrix},\ \mathbf{J}^0_{\sigma(x)} = K_{\sigma(x)} * \begin{pmatrix} \eta_0 I^2_{x_1} & \eta_0 I_{x_1} I_{x_2} & \eta_0 I_{x_1} I_t \\ \text{“} & \eta_0 I^2_{x_2} & \eta_0 I_{x_2} I_t \\ \text{“} & \text{“} & \eta_0 I^2_t \end{pmatrix},$$

$$\mathbf{J}^{x_1x_2}_{\sigma(x)} = K_{\sigma(x)} * \begin{pmatrix} \eta_{x_1} I^2_{x_1x_1} + \eta_{x_2} I^2_{x_1x_2} & \eta_{x_1} I_{x_1x_1} I_{x_1x_2} + \eta_{x_2} I_{x_1x_2} I_{x_2x_2} & \eta_{x_1} I_{x_1x_1} I_{x_1t} + \eta_{x_1} I_{x_1x_2} I_{x_2t} \\ \text{“} & \eta_{x_1} I^2_{x_1x_2} + \eta_{x_2} I^2_{x_2x_2} & \eta_{x_1} I_{x_1x_2} I_{x_1t} + \eta_{x_2} I_{x_2x_2} I_{x_2t} \\ \text{“} & \text{“} & \eta_{x_1} I^2_{x_1t} + \eta_{x_2} I^2_{x_2t} \end{pmatrix}.$$

We rewrite $E(\mathbf{w},\sigma)$ by expliciting the convolution:

$$\begin{aligned} E(\mathbf{w},\sigma)) &= \sum_{x\in\Omega} \phi_d\left(k_\sigma * \left(\eta_0 \left(I_t + I_{x_1}u(x) + I_{x_2}v(x)\right)^2\right)\right) \\ &\quad + \gamma\phi_d\Big(k_\sigma * \Big(\eta_{x_1}\left(I_{x_1t} + I_{x_1x_1}u(x) + I_{x_1x_2}v(x)\right)^2 \\ &\qquad\qquad + \eta_{x_2}\left(I_{x_2t} + I_{x_2x_1}u(x) + I_{x_2x_2}v(x)\right)^2\Big)\Big) \\ &\quad + \beta \sum_{z\in\mathcal{N}(x)} \phi_r((\sigma(x)-\sigma(z))^2) \\ &= \sum_{x\in\Omega} \phi_d\left(\sum_{y\in\Omega} k_{\sigma(x)}(x,y)\left(\eta_0\left(I_t + I_{x_1}u(x) + I_{x_2}v(x)\right)^2\right)\right) \\ &\quad + \gamma\phi_d\Bigg(\sum_{y\in\Omega} k_{\sigma(x)}(x,y)\Big(\eta_{x_1}\left(I_{x_1t} + I_{x_1x_1}u(x) + I_{x_1x_2}v(x)\right)^2 \\ &\qquad\qquad + \eta_{x_2}\left(I_{x_2t} + I_{x_2x_1}u(x) + I_{x_2x_2}v(x)\right)^2\Big)\Bigg) \\ &\quad + \beta \sum_{z\in\mathcal{N}(x)} \phi_r((\sigma(x)-\sigma(z))^2) \end{aligned}$$

Case of $\phi_d(z^2) = z^2$:

$$\begin{aligned}
\frac{\partial E(\mathbf{w},\sigma)}{\partial \sigma(x)} &= \left(\sum_{y\in\Omega} \frac{\partial k_{\sigma(x)}(x,y)}{\partial \sigma(x)} \left(\eta_0 \left(I_t + I_{x_1} u(x) + I_{x_2} v(x) \right)^2 \right) \right. && \text{(B.1)} \\
&\quad + \frac{\partial k_{\sigma(x)}(x,y)}{\partial \sigma(x)} \Big(\eta_{x_1} \left(I_{x_1 t} + I_{x_1 x_1} u(x) + I_{x_1 x_2} v(x) \right)^2 && \text{(B.2)} \\
&\quad \left. + \eta_{x_2} \left(I_{x_2 t} + I_{x_2 x_1} u(x) + I_{x_2 x_2} v(x) \right)^2 \Big) \right) && \text{(B.3)} \\
&\quad -2\,\beta \sum_{z\in\mathcal{N}(x)} (\sigma(x) - \sigma(z)) \phi_r^{'}((\sigma(x) - \sigma(z))^2) && \text{(B.4)}
\end{aligned}$$

From (6.11), we have:

$$\frac{\partial k_{\sigma(x)}(x,y)}{\partial \sigma(x)} = \frac{k_{\sigma(x)}(x,y)}{\sigma(x)} \left(\frac{\|x-y\|^2}{\sigma^2(x)} - 1 \right) \tag{B.5}$$

Replacing (B.5) in (B.1), we obtain:

$$\begin{aligned}
\frac{\partial E(\mathbf{w},\sigma)}{\partial \sigma(x)} &= \left(\sum_{y\in\Omega} \frac{k_{\sigma(x)}(x,y)}{\sigma(x)} \left(\frac{\|x-y\|^2}{\sigma^2(x)} - 1 \right) \left(\eta_0 \left(I_t + I_{x_1} u(x) + I_{x_2} v(x) \right)^2 \right) \right. && \text{(B.6)} \\
&\quad + \frac{k_{\sigma(x)}(x,y)}{\sigma(x)} \left(\frac{\|x-y\|^2}{\sigma^2(x)} - 1 \right) \Big(\eta_{x_1} \left(I_{x_1 t} + I_{x_1 x_1} u(x) + I_{x_1 x_2} v(x) \right)^2 \\
&\quad \left. + \eta_{x_2} \left(I_{x_2 t} + I_{x_2 x_1} u(x) + I_{x_2 x_2} v(x) \right)^2 \Big) \right) \\
&\quad -2\,\beta \sum_{z\in\mathcal{N}(x)} (\sigma(x) - \sigma(z)) \phi_r^{'}((\sigma(x) - \sigma(z))^2).
\end{aligned}$$

We introduce the filter $h_{\sigma(x)}$ defined by:

$$h_{\sigma(x)} * I = \sum_{y\in\Omega} \frac{k_{\sigma(x)}(x,y)}{\sigma(x)} \left(\frac{\|x-y\|^2}{\sigma^2(x)} - 1 \right) I(y).$$

(B.7) can then be rewritten:

$$
\begin{aligned}
\frac{\partial E(\mathbf{w},\sigma)}{\partial \sigma(x)} &= h_{\sigma(x)} * \left(\eta_0 \left(I_t + I_{x_1} u(x) + I_{x_2} v(x)\right)^2\right) \\
&\quad + h_{\sigma(x)} * \Big(\left(I_{x_1 t} + I_{x_1 x_1} u(x) + I_{x_1 x_2} v(x)\right)^2 \\
&\qquad\qquad + \eta_{x_2} \left(I_{x_2 t} + I_{x_2 x_1} u(x) + I_{x_2 x_2} v(x)\right)^2 \Big) \\
&\quad -2\,\beta \sum_{z \in \mathcal{N}(x)} (\sigma(x) - \sigma(z)) \phi_r^{'}((\sigma(x) - \sigma(z))^2)
\end{aligned}
$$

Arbitrary ϕ_d:

$$
\begin{aligned}
\frac{\partial E(\mathbf{w},\sigma)}{\partial \sigma(x)} &= h_{\sigma(x)} * \left(\eta_0 \left(I_t + I_{x_1} u(x) + I_{x_2} v(x)\right)^2\right) \phi_d' \left(k_\sigma * \left(\eta_0 \left(I_t + I_{x_1} u(x) + I_{x_2} v(x)\right)^2\right)\right) \\
&\quad + h_{\sigma(x)} * \Big(\left(I_{x_1 t} + I_{x_1 x_1} u(x) + I_{x_1 x_2} v(x)\right)^2 \\
&\qquad\qquad + \eta_{x_2} \left(I_{x_2 t} + I_{x_2 x_1} u(x) + I_{x_2 x_2} v(x)\right)^2 \Big) \\
&\quad \phi_d' \Big(k_\sigma * \Big(\left(I_{x_1 t} + I_{x_1 x_1} u(x) + I_{x_1 x_2} v(x)\right)^2 \\
&\qquad\qquad + \eta_{x_2} \left(I_{x_2 t} + I_{x_2 x_1} u(x) + I_{x_2 x_2} v(x)\right)^2 \Big)\Big) \\
&\quad -2\,\beta \sum_{z \in \mathcal{N}(x)} (\sigma(x) - \sigma(z)) \phi_r^{'}((\sigma(x) - \sigma(z))^2)
\end{aligned}
$$

C Differentation of (6.19)

We want to compute $\partial E/\partial\sigma(x)$, where

$$E(\mathbf{w},\sigma) = \sum_{x\in\Omega} \phi\left(\boldsymbol{\alpha}_{\mathbf{w},\sigma(x)}^T \mathbf{J}_{\sigma(x)} \boldsymbol{\alpha}_{\mathbf{w},\sigma(x)}\right) dx + \beta \sum_{z\in\mathcal{N}(x)} \phi_r((\sigma(x)-\sigma(z))^2)$$

with

$$\boldsymbol{\alpha}_{\mathbf{w},\sigma(x)} = \begin{pmatrix} u(x) \\ v(x) \\ \sigma^2(x) \\ 1 \end{pmatrix},\ \mathbf{J}_{\sigma(x)} = k_{\sigma(x)} * \begin{pmatrix} I_{x_1}^2 & I_{x_1}I_{x_2} & I_{x_1}\frac{\Delta I}{2} & I_{x_1}I_t \\ \text{“} & I_{x_2}^2 & I_{x_2}\frac{\Delta I}{2} & I_{x_2}I_t \\ \text{“} & \text{“} & \left(\frac{\Delta I}{2}\right)^2 & \frac{\Delta I}{2}I_t \\ \text{“} & \text{“} & \text{“} & I_t^2 \end{pmatrix}$$

We rewrite $E(\mathbf{w},\sigma)$ to simplify calculations:

$$\hat{\sigma} = \arg\min_{\sigma} \sum_{x\in\Omega} \phi\left(\boldsymbol{\sigma}^T(x)\mathbf{J}_{2,\sigma(x)}\boldsymbol{\sigma}(x)\right) + \beta \sum_{z\in\mathcal{N}(x)} \phi_r((\sigma(x)-\sigma(z))^2)\, dx$$

with

$$\begin{aligned} \boldsymbol{\sigma}_{\mathbf{w},\sigma}(x) = \begin{pmatrix} \sigma^2(x) \\ 1 \end{pmatrix},\ \mathbf{J}_{2,\sigma(x)} &= k_{\sigma(x)} * \begin{pmatrix} \left(\frac{\Delta I}{2}\right)^2 & \left(\frac{\Delta I}{2}\right)(I_t + I_{x_1}\hat{u}(x) + I_{x_2}\hat{v}(x)) \\ \text{“} & (I_t + I_{x_1}\hat{u}(x) + I_{x_2}\hat{v}(x))^2 \end{pmatrix} \\ &= k_{\sigma(x)} * \begin{pmatrix} \mathbf{J}_2^{11} & \mathbf{J}_2^{12} \\ \text{“} & \mathbf{J}_2^{22} \end{pmatrix} \end{aligned}$$

We rewrite $E(\mathbf{w},\sigma)$ by expliciting the Gaussian convolution weights k:

$$\begin{aligned} E(\mathbf{w},\sigma) &= \sum_{x\in\Omega} \phi\left(k_{\sigma(x)} * \left(\mathbf{J}_2^{11}\sigma^4(x) + 2\mathbf{J}_2^{12}\sigma^2(x) + \mathbf{J}_2^{22}\right)\right) dx + \beta \sum_{z\in\mathcal{N}(x)} \phi_r((\sigma(x)-\sigma(z))^2) \\ &= \sum_{x\in\Omega} \phi\left(\sum_{y\in\Omega} k_{\sigma(x)}(x,y)\left(\mathbf{J}_2^{11}(y)\sigma^4(x) + 2\mathbf{J}_2^{12}(y)\sigma^2(x) + \mathbf{J}_2^{22}(y)\right)\right) \\ &\qquad + \beta \sum_{z\in\mathcal{N}(x)} \phi_r((\sigma(x)-\sigma(z))^2) \end{aligned}$$

Case of $\phi(z^2) = z^2$:

$$\frac{\partial E(\mathbf{w},\sigma)}{\partial \sigma(x)} = \Bigg(\sum_{y\in\Omega} \frac{\partial k_{\sigma(x)}(x,y)}{\partial \sigma(x)} \left(\mathbf{J}_2^{11}(y)\sigma^4(x) + 2\mathbf{J}_2^{12}(y))\sigma^2(x) + \mathbf{J}_2^{22}(y) \right) \tag{C.1}$$
$$+ k_{\sigma(x)}(x,y) \left(4\mathbf{J}_2^{11}(y)\sigma^3(x) + 4\mathbf{J}_2^{12}(y)\sigma(x)) \right) \Bigg)$$
$$-2\,\beta \sum_{z\in\mathcal{N}(x)} (\sigma(x) - \sigma(z))\phi_r^{'}((\sigma(x) - \sigma(z))^2)$$

From (6.11), we have:

$$\frac{\partial k_{\sigma(x)}(x,y)}{\partial \sigma(x)} = \frac{k_{\sigma(x)}(x,y)}{\sigma(x)} \left(\frac{\|x-y\|^2}{\sigma^2(x)} - 1 \right) \tag{C.2}$$

Replacing (C.2) in (C.1), we obtain:

$$\begin{aligned}
\frac{\partial E(\mathbf{w},\sigma)}{\partial \sigma(x)} &= \Bigg(\sum_{y\in\Omega} k_{\sigma(x)}(x,y) \left(\frac{\|x-y\|^2}{\sigma^2(x)} - 1 \right) \left(\mathbf{J}_2^{11}(y)\sigma^3(x) + 2\mathbf{J}_2^{12}(y)\sigma(x)) + \frac{\mathbf{J}_2^{22}(y)}{\sigma(x)} \right) \\
&\quad + k_{\sigma(x)}(x,y) \left(4\mathbf{J}_2^{11}(y)\sigma^3(x) + 4\mathbf{J}_2^{12}(y)\sigma(x)) \right) \Bigg) \\
&\quad - 2\,\beta\,\mathrm{div}(\nabla\sigma\phi'(|\nabla\sigma(x)|^2)) \\
&= \Bigg(\sum_{y\in\Omega} k_{\sigma(x)}(x,y)\|x-y\|^2 \left(\mathbf{J}_2^{11}(y)\sigma(x) + 2\mathbf{J}_2^{12}(y)\sigma^{-1}(x)) + \mathbf{J}_2^{22}(y)\sigma^{-3}(x) \right) \\
&\quad + k_{\sigma(x)}(x,y) \left(3\mathbf{J}_2^{11}(y)\sigma^3(x) + 2\mathbf{J}_2^{12}(y)\sigma(x) - \mathbf{J}_2^{22}(y)\sigma^{-1}(x) \right) \Bigg) \\
&\quad -2\,\beta \sum_{z\in\mathcal{N}(x)} (\sigma(x) - \sigma(z))\phi_r^{'}((\sigma(x) - \sigma(z))^2)
\end{aligned}$$

We introduce the filter $h_{2_{\sigma(x)}}$ defined by:

$$h_{2_{\sigma(x)}} * I = \sum_{y\in\Omega} k_{\sigma(x)}(x,y) \left(\|x-y\|^2 \right) I(y).$$

(B.7) can then be rewritten:

$$\begin{aligned}\frac{\partial E(\mathbf{w},\sigma)}{\partial \sigma(x)} &= h_{2_{\sigma(x)}} * \left(\mathbf{J}_2^{11}\sigma(x) + 2\mathbf{J}_2^{12}\sigma^{-1}(x)) + \mathbf{J}_2^{22}\sigma^{-3}(x)\right) \\ &+ k_{\sigma(x)} * \left(3\mathbf{J}_2^{11}\sigma^3(x) + 2\mathbf{J}_2^{12}\sigma(x)) - \mathbf{J}_2^{22}\sigma^{-1}(x)\right) \\ &-2\,\beta \sum_{z\in\mathcal{N}(x)} (\sigma(x)-\sigma(z))\phi_r^{'}((\sigma(x)-\sigma(z))^2)\end{aligned}$$

Arbitraryϕ:

$$\begin{aligned}\frac{\partial E(\mathbf{w},\sigma)}{\partial \sigma(x)} &= \Bigg(h_{2_{\sigma(x)}} * \left(\mathbf{J}_2^{11}\sigma(x) + 2\mathbf{J}_2^{12}\sigma^{-1}(x)) + \mathbf{J}_2^{22}\sigma^{-3}(x)\right) \\ &+ k_{\sigma(x)} * \left(3\mathbf{J}_2^{11}\sigma^3(x) + 2\mathbf{J}_2^{12}\sigma(x)) - \mathbf{J}_2^{22}\sigma^{-1}(x)\right)\Bigg) \cdot \\ &\phi'\left(k_{\sigma(x)} * \left(\mathbf{J}_2^{11}\sigma^4(x) + 2\mathbf{J}_2^{12}\sigma^2(x) + \mathbf{J}_2^{22}\right)\right) \\ &-2\,\beta \sum_{z\in\mathcal{N}(x)} (\sigma(x)-\sigma(z))\phi_r^{'}((\sigma(x)-\sigma(z))^2)\end{aligned}$$

List of publications

International journals

- D. Fortun, P.Bouthemy, C. Kervrann. *Aggregation of local parametric candidates and exemplar-based occlusion handling for optical flow.* Submitted to Transaction on Image Processing, 2014.
- D. Fortun, P.Bouthemy, P. Paul-Gilloteaux, C. Kervrann. *Aggregation of patch-based estimations for optical flow in live cell imaging with intensity changes.* Submitted to Medical Image Analysis, Special Issue on Discrete Graphical Models in Biomedical Image Analysis, 2014.

International conferences

- D. Fortun, C. Chen, P. Paul-Gilloteaux, F. Waharte, J. Salamero, C. Kervrann. *Correlation and variational approaches for motion and diffusion estimation in fluorescence imaging.* European Signal Processing Conference (EUSIPCO'13), Marrakech, September 2013.
- D. Fortun, P. Bouthemy, P. Paul-Gilloteaux, C. Kervrann. *Aggregation of patch-based estimations for illumination-invariant optical flow in live cell imaging.* International Symposium on Biomedical Imaging (ISBI'13), San Francisco, California, April 2013. *Finalist of the best student paper award*
- D. Fortun, C. Kervrann. *Semi-local variational optical flow estimation.* International Conference on Image Processing (ICIP'12), Orlando, Florida, September 2012.

National conferences

- D. Fortun, P. Bouthemy, C. Kervrann. *Agrégation d,Äôestimations semi-locales pour le flot optique.* Groupe de Recherche et d'Etude du Traitement du Signal et de l'Image (GRETSI'13) , Brest, September 2013.

Bibliography

A. Alba, E. Arce-Santana, and M. Riviera. Optical flow estimation with prior models obtained from phase correlation. *Advances in Visual Computing*, pages 417–426, 2010.

L. Alvarez, J. Esclarín, M. Lefebure, and S. Javier. A pde model for computing the optical flow. In *Proc. XVI Congreso de ecuaciones diferenciales y aplicaciones*, pages 1349–1356, Las Palmas de Gran Canaria, Spain, September 1999.

T. Amiaz and N. Kiryati. Piecewise-smooth dense optical flow via level sets. *International Journal of Computer Vision*, 68(2):111–124, 2006.

P. Arias, G. Facciolo, V. Caselles, and G. Sapiro. A variational framework for exemplar-based image inpainting. *International Journal of Computer Vision*, 93(3): 319–347, 2011.

G. Aubert, R. Deriche, and P. Kornprobst. Computing optical flow via variational techniques. *SIAM Journal on Applied Mathematics*, 60(1):156–182, 1999.

J.-F. Aujol, G. Gilboa, T. Chan, and S. Osher. Structure-texture image decomposition - modeling, algorithms, and parameter selection. *International Journal of Computer Vision*, 67(1):111–136, 2006.

A. Ayvaci, M. Raptis, and S. Soatto. Sparse occlusion detection with optical flow. *International Journal of Computer Vision*, 97(3):322–338, 2012.

N. Azzabou, N. Paragios, F. Guichard, and F. Cao. Variable bandwidth image denoising using image-based noise models. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1–7, 2007.

S. Baheerathan, F. Albregtsen, and H. Danielsen. Registration of serial sections of mouse liver cell nuclei. *Journal of microscopy*, 192(1):37–53, 1998.

S. Baker and I. Matthews. Lucas-kanade 20 years on: A unifying framework. *International Journal of Computer Vision*, 56(3):221–255, 2004.

S. Baker, M. J. Black, J. Lewis, S. Roth, D. Scharstein, and R. Szeliski. A database and evaluation methodology for optical flow. In *International Conference on Computer Vision (ICCV)*, pages 1–8, Rio de Janeiro, Brazil, October 2007.

S. Baker, D. Scharstein, J. Lewis, S. Roth, M. Black, and R. Szeliski. A database and evaluation methodology for optical flow. *International Journal of Computer Vision*, 92 (1):1–31, 2011.

S. Balay, M. F. Adams, J. Brown, P. Brune, K. Buschelman, V. Eijkhout, W. D. Gropp, D. Kaushik, M. G. Knepley, L. C. McInnes, K. Rupp, B. F. Smith, and H. Zhang. PETSc Web page. http://www.mcs.anl.gov/petsc, 2014. URL http://www.mcs.anl.gov/petsc.

C. Ballester, L. Garrido, V. Lazcano, and V. Caselles. A tv-l1 optical flow method with occlusion detection. In *Pattern Recognition*, pages 31–40, 2012.

L. Bao, Q. Yang, and H. Jin. Fast edge-preserving patchmatch for large displacement optical flow. In *Computer Vision and Pattern Recognition (CVPR)*, Columbus, June 2014.

C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman. Patchmatch: a randomized correspondence algorithm for structural image editing. In *ACM Transactions On Graphics*, volume 28, page 24. ACM, 2009.

C. Barnes, E. Shechtman, D. B. Goldman, and A. Finkelstein. The generalized patchmatch correspondence algorithm. In *European Conference on Computer Vision (ECCV)*, pages 29–43, 2010.

J. Barron, D. Fleet, and S. Beauchemin. Evaluation of optical flow. *International Journal of Computer Vision*, 12(1):43–77, 1994.

D. Batra and P. Kohli. Making the right moves: Guiding alpha-expansion using local primal-dual gaps. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1865–1872, 2011.

H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *European Conference Computer Vision (ECCV)*, pages 404–417, 2006.

F. Becker, B. Wieneke, S. Petra, A. Schroder, and C. Schnorr. Variational adaptive correlation method for flow estimation. *Image Processing*, 21(6):3053–3065, 2012.

J. R. Bergen, P. Anandan, K. J. Hanna, and R. Hingorani. Hierarchical model-based motion estimation. In *European Conference on Computer Vision (ECCV)*, pages 237–252. Springer, 1992.

M. Bertalmio, G. Sapiro, V. Caselles, and C. Ballester. Image inpainting. In *Proceedings of the 27th annual conference on Computer graphics and interactive techniques*, pages 417–424. ACM Press/Addison-Wesley Publishing Co., 2000.

J. Besag. On the statistical analysis of dirty pictures. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 259–302, 1986.

J. Bigun, G. H. Granlund, and J. Wiklund. Multidimensional orientation estimation with applications to texture analysis and optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 13(8):775–790, 1991.

M. Black and P. Anandan. The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer Vision and Image Understanding*, 63(1):75–104, 1996.

M. Black and A. Jepson. Estimating optical flow in segmented images using variable-order parametric models with local deformations. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(10):972–986, 1996.

M. J. Black. Recursive non-linear estimation of discontinuous flow fields. In *European Conference on Computer Vision (ECCV)*, pages 138–145, 1994.

M. J. Black and P. Anandan. A framework for the robust estimation of optical flow. In *International Conference on Computer Vision (ICCV)*, pages 231–236, 1993.

M. J. Black, Y. Yacoob, A. D. Jepson, and D. J. Fleet. Learning parameterized models of image motion. In *Computer Vision and Pattern Recognition (CVPR)*, pages 561–567, 1997.

A. Blake and A. Zisserman. *Visual reconstruction*. MIT press Cambridge, 1987.

M. Bleyer, C. Rhemann, and M. Gelautz. Segmentation-based motion with occlusions using graph-cut optimization. In *DAGM symposium on Pattern Recognotion*, pages 465–474, Berlin, Germany, September 2006.

G. Boncompain, S. Divoux, N. Gareil, H. de Forges, A. Lescure, L. Latreche, V. Mercanti, F. Jollivet, G. Raposo, and F. Perez. Synchronization of secretory protein traffic in populations of cells. *Nature Methods*, 9(5):493–498, 2012.

H. Bornfleth, P. Edelmann, D. Zink, T. Cremer, and C. Cremer. Quantitative motion analysis of subchromosomal foci in living cells using four-dimensional microscopy. *Biophysical journal*, 77(5):2871–2886, 1999.

E. Boros, P. Hammer, and X. Sun. Network flows and minimization of quadratic pseudo-boolean functions. Technical report, Technical Report RRR 17-1991, RUTCOR, 1991.

P. Bouthemy and E. François. Motion segmentation and qualitative dynamic scene analysis from an image sequence. *International Journal of Computer Vision*, 10(2): 157–182, 1993.

Y. Boykov, O. Veksler, and R. Zabih. Markov random fields with efficient approximations. In *Computer Vision and Pattern Recognition (CVPR)*, pages 648–655, 1998.

Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(11):1222–1239, 2001.

G. Bradski. Opencv. *Dr. Dobb's Journal of Software Tools*, 2000.

J. Braux-Zin, R. Dupont, and A. Bartoli. A general dense image matching framework combining direct and feature-based costs. In *Intenational Conference on Computer Vision (ICCV)*, 2013.

K. Bredies, K. Kunisch, and T. Pock. Total generalized variation. *SIAM Journal on Imaging Sciences*, 3(3):492–526, 2010.

T. Brox. *From pixels to regions: partial differential equations in image analysis*. PhD dissertation, Saarland University, 2005.

T. Brox and J. Malik. Large displacement optical flow: descriptor matching in variational motion estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 33(3):500–513, 2011.

T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *European Conference on Computer Vision (ECCV)*, pages 25–36, Prague, Czech Republic, May 2004.

T. Brox, A. Bruhn, and J. Weickert. Variational motion segmentation with level sets. *European Conference on Computer Vision (ECCV)*, pages 471–483, 2006.

A. Bruhn and W. Weickert. A confidence measure for variational optic flow methods. *Geometric Properties for Incomplete Data*, pages 283 – 298, 2006.

A. Bruhn, J. Weickert, and C. Schnörr. Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *International Journal of Computer Vision*, 61(3): 211–231, 2005.

A. Bugeau, M. Bertalmío, V. Caselles, and G. Sapiro. A comprehensive framework for image inpainting. *IEEE Transactions on Image Processing*, 19(10):2634–2645, 2010.

D. J. Butler, J. Wulff, G. B. Stanley, and M. J. Black. A naturalistic open source movie for optical flow evaluation. In *European Conference on Computer Vision (ECCV)*, pages 611–625. Springer-Verlag, 2012.

L. C. and M. L. Total variation as a local filter. *SIAM J. Imaging Sci.*, pages 651–694, 2011.

A. Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1-2):89–97, 2004.

A. Chambolle and T. Pock. A first-order primal-dual algorithm for convex problems with applications to imaging. *Journal of Mathematical Imaging and Vision*, 40(1):120–145, 2011.

T. F. Chan, S. H. Kang, and J. Shen. Euler's elastica and curvature-based inpainting. *SIAM Journal on Applied Mathematics*, pages 564–592, 2002.

P. Charbonnier, L. Blanc-Féraud, G. Aubert, and M. Barlaud. Deterministic edge-preserving regularization in computed imaging. *IEEE Transactions on Image Processing*, 6(2):298–311, 1997.

Z. Chen, H. Jin, Z. Lin, S. Cohen, and Y. Wu. Large displacement optical flow from nearest neighbor fields. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2443–2450, 2013.

T. M. Chin, W. C. Karl, and A. S. Willsky. Probabilistic and sequential computation of optical flow using temporal coherence. *IEEE Transactions on Image Processing*, 3(6): 773–788, 1994.

P. G. Ciarlet. *The finite element method for elliptic problems.* Elsevier, 1978.

D. Comaniciu and P. Meer. Mean shift: A robust approach toward feature space analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(5):603–619, 2002.

T. Cooke. Two applications of graph-cuts to image processing. In *Digital Image Computing: Techniques and Applications (DICTA)*, pages 498–504, 2008.

T. Corpetti and E. Mémin. Stochastic uncertainty models for the luminance consistency assumption. *IEEE Transactions on Image Processing*, 21(2):481–493, 2012.

T. Corpetti, É. Mémin, and P. Pérez. Dense estimation of fluid flows. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(3):365–380, 2002.

D. Cremers and S. Soatto. Motion competition: A variational approach to piecewise parametric motion segmentation. *International Journal of Computer Vision*, 62(3): 249–265, 2005.

A. Criminisi, P. Pérez, and K. Toyama. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on Image Processing*, 13(9): 1200–1212, 2004.

M. Daisy, D. Tschumperlé, and O. Lézoray. A fast spatial patch blending algorithm for artefact reduction in pattern-based image inpainting. In *SIGGRAPH Asia 2013 Technical Briefs*, pages 8:1–8:4, New York, NY, USA, 2013.

N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 886–893, 2005.

M. Dawood, F. Buther, X. Jiang, and K. P. Schafers. Respiratory motion correction in 3-d pet data with advanced optical flow algorithms. *IEEE Transactions on Medical Imaging*, 27(8):1164–1175, 2008.

J. Delon. Midway image equalization. *Journal of Mathematical Imaging and Vision*, 21 (2):119–134, 2004.

J. Delon and A. Desolneux. Stabilization of flicker-like effects in image sequences through local contrast correction. *SIAM Journal on Imaging Sciences*, 3(4):703–734, 2010.

J. Delon and B. Rougé. Small baseline stereovision. *Journal of Mathematical Imaging and Vision (JMIV*, 28(3):209–223, 2007.

J. Delpiano, J. Jara, J. Scheer, O. Ramírez, J. Ruiz-del Solar, and S. Härtel. Performance of optical flow techniques for motion analysis of fluorescent point signals in confocal microscopy. *Machine Vision Applications*, 23(4):675–689, 2011.

J. Delpiano, J. Jara, J. Scheer, O. A. Ramírez, J. Ruiz-del Solar, and S. Härtel. Performance of optical flow techniques for motion analysis of fluorescent point signals in confocal microscopy. *Machine Vision and Applications*, 23(4):675–689, 2012.

P. Dérian, P. Héas, C. Herzet, and É. Mémin. Wavelet-based fluid motion estimation. In *Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 737–748, 2011.

P. Dérian, P. Héas, C. Herzet, E. Memin, et al. Wavelets and optical flow motion estimation. *Numerical Mathematics: Theory, Methods and Applications*, 2012.

R. Deriche, P. Kornprobst, and G. Aubert. Optical-flow estimation while preserving its discontinuities: a variational approach. *Recent Developments in Computer Vision*, pages 69–80, 1996.

M. Drulea and S. Nedevschi. Total variation regularization of local-global optical flow. In *Intelligent Transportation Systems Conference (ITSC)*, pages 318–323, 2011.

M. Drulea and S. Nedevschi. Motion estimation using the correlation transform. *IEEE Transactions on Image Processing*, 2013.

R. Dupont, N. Paragios, R. Keriven, and P. Fuchs. Extraction of layers of similar motion through combinatorial techniques. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 220–234. Springer, 2005.

W. Enkelmann. Investigations of multigrid algorithms for the estimation of optical flow fields in image sequences. *Computer Vision, Graphics, and Image Processing*, 43(2): 150–177, 1988.

G. Facciolo, N. Limare, and E. Meinhardt. Integral images for block matching. *Image Processing On Line*, 2013.

P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *International journal of computer vision*, 70(1):41–54, 2006.

A. Fix, A. Gruber, E. Boros, and R. Zabih. A graph cut algorithm for higher-order markov random fields. In *International Conference on Computer Vision (ICCV)*, pages 1020–1027, 2011.

D. J. Fleet, M. J. Black, Y. Yacoob, and A. D. Jepson. Design and use of linear models for image motion analysis. *International Journal of Computer Vision*, 36(3):171–193, 2000.

G. Forbin, B. Besserer, J. BoldyÀás, D. Tschumperlé, et al. Temporal extension to exemplar-based inpainting applied to scratch correction in damaged images sequences. *Visualization, Imaging, and Image Processing (VIIP 2005)*, pages 12–20, 2005.

D. R. Fulkerson. *Flows in networks*. Princeton Princeton University Press, 1962.

R. Garg, A. Roussos, and L. Agapito. Robust trajectory-space tv-l1 optical flow for non-rigid sequences. In *Energy Minimization Methods in Computer Vision and Pattern Recognition (EMMCVPR)*, pages 300–314. Springer, 2011.

A. Geiger, P. Lenz, and R. Urtasun. Are we ready for autonomous driving? the kitti vision benchmark suite. In *Computer Vision and Pattern Recognition (CVPR)*, pages 3354–3361, 2012.

M. Gelgon and P. Bouthemy. A region-level motion-based graph representation and labeling for tracking a spatial image partition. *Pattern Recognition*, 33(4):725–740, 2000.

M. Gelgon, P. Bouthemy, and T. Dubois. A region tracking method with failure detection for an interactive video indexing environment. In *Visual Information and Information Systems*, pages 261–269. Springer, 1999.

D. Geman and G. Reynolds. Constrained restoration and the recovery of discontinuities. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(3):367–383, 1992.

S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (6):721–741, 1984.

B. Glocker, N. Komodakis, N. Paragios, G. Tziritas, and N. Navab. Inter and intra-modal deformable registration: Continuous deformations meet efficient optimal linear programming. In *Information Processing in Medical Imaging (IPMI)*, pages 408–420. Springer, 2007.

B. Glocker, N. Paragios, N. Komodakis, G. Tziritas, and N. Navab. Optical flow estimation with uncertainties through dynamic mrfs. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, Anchorage, Alaska, June 2008.

B. Glocker, T. Heibel, N. Navab, P. Kohli, and C. Rother. Triangleflow: Optical flow with triangulation-based higher-order likelihoods. *European Conference on Computer Vision (ECCV)*, pages 272–285, 2010.

A. V. Goldberg and R. E. Tarjan. A new approach to the maximum-flow problem. *Journal of the ACM (JACM)*, 35(4):921–940, 1988.

P. Golland and A. M. Bruckstein. Motion from color. *Computer Vision and Image Understanding*, 68(3):346–362, 1997.

A. P. Goobic, J. Tang, and S. T. Acton. Image stabilization and registration for tracking cells in the microvasculature. *Transactions on Biomedical Engineering*, 52(2):287–299, 2005.

S. Grauer-Gray and C. Kambhamettu. Hierarchical belief propagation to reduce search space using cuda for stereo and motion estimation. In *Workshop on Applications of Computer Vision (WACV)*, pages 1–8, 2009.

S. Grewenig, J. Weickert, and A. Bruhn. From box filtering to fast explicit diffusion. In *DAGM symposium on Pattern Recognition*, pages 533–542, 2010.

P. Gwosdek, H. Zimmer, S. Grewenig, A. Bruhn, and J. Weickert. A highly efficient gpu implementation for variational optic flow based on the euler-lagrange framework. In *Trends and Topics in Computer Vision*, pages 372–383, 2012.

D. Hafner, O. Demetz, and J. Weickert. Why is the census transform good for robust optic flow computation? In *Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 210–221, 2013.

H. W. Haussecker and D. J. Fleet. Computing optical flow with physical models of brightness variation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(6):661–673, 2001.

K. He, J. Sun, and X. Tang. Guided image filtering. In *European Conference on Computer Vision (ECCV)*, pages 1–14, 2010.

P. Héas and E. Mémin. Three-dimensional motion estimation of atmospheric layers from image sequences. *Transactions on Geoscience and Remote Sensing*, 46(8):2385–2396, 2008.

P. Héas, E. Memin, N. Papadakis, and A. Szantai. Layered estimation of atmospheric mesoscale dynamics from satellite imagery. *Transactions on Geoscience and Remote Sensing*, 45(12):4087–4104, 2007.

P. Héas, C. Herzet, and E. Memin. Bayesian inference of models and hyperparameters for robust optical-flow estimation. *IEEE Transactions on Image Processing*, 21(4): 1437–1451, 2012.

B. Hebert, S. Costantino, and P. Wiseman. Spatiotemporal image correlation spectroscopy (stics) theory, verification, and application to protein velocity mapping in living cho cells. *Biophysical J.*, 88(5):3601–3614, 2005.

P. Heise, S. Klose, B. Jensen, and A. Knoll. Pm-huber: Patchmatch with huber regularization for stereo matching. In *Intenational Conference on Computer Vision (ICCV)*, 2013.

F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1993.

P. Hellier and C. Barillot. Coupling dense and landmark-based approaches for nonrigid registration. *IEEE Transactions on Medical Imaging*, 22(2):217–227, 2003.

S. Hermann and R. Klette. Hierarchical scan-line dynamic programming for optical flow using semi-global matching. In *Workshops of Asian Conference Computer Vision*, pages 556–567. Springer, 2013.

P. W. Holland and R. E. Welsch. Robust regression using iteratively reweighted least-squares. *Communications in Statistics-Theory and Methods*, 6(9):813–827, 1977.

B. Horn and B. Schunck. Determining optical flow. *Artificial intelligence*, 17(1-3):185–203, 1981.

A. Hosni, C. Rhemann, M. Bleyer, C. Rother, and M. Gelautz. Fast cost-volume filtering for visual correspondence and beyond. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2013.

P. J. Huber. *Robust statistics*. Springer, 1981.

A. Humayun, O. Mac Aodha, and G. J. Brostow. Learning to find occlusion regions. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2161–2168, 2011.

S. Ince and J. Konrad. Occlusion-aware optical flow estimation. *IEEE Transactions on Image Processing*, 17(8):1443–1451, 2008.

H. Ishikawa. Higher-order clique reduction in binary graph cut. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2993–3000, 2009.

L. Ji and G. Danuser. Tracking quasi-stationary flow of weak fluorescent signals by adaptive multi-frame correlation. *Journal of Microscopy*, 220(3):150–167, 2005.

P.-M. Jodoin and M. Mignotte. Optical-flow based on an edge-avoidance procedure. *Computer Vision and Image Understanding*, 113(4):511–531, 2009.

S. Ju, M. Black, and A. Jepson. Skin and bones: Multi-layer, locally affine, optical flow and regularization with transparency. In *Computer Vision and Pattern Recognition (CVPR)*, pages 307–314, 1996.

E. M. Kalmoun and L. Garrido. Trust region versus line search for computing the optical flow. *Multiscale Modeling & Simulation*, 11(3):890–906, 2013.

E. M. Kalmoun, L. Garrido, and V. Caselles. Line search multilevel optimization as computational methods for dense optical flow. *SIAM Journal on Imaging Sciences*, 4 (2):695–722, 2011.

J. H. Kappes, B. Andres, F. A. Hamprecht, C. Schnörr, S. Nowozin, D. Batra, S. Kim, B. X. Kausler, J. Lellmann, N. Komodakis, et al. A comparative study of modern

inference techniques for discrete energy minimization problems. In *Computer Vision and Pattern Recognition (CVPR)*, 2013.

C. Kervrann, J. Boulanger, T. Pecot, P. Perez, J. Salamero, et al. Multiscale neighborhood-wise decision fusion for redundancy detection in image pairs. *SIAM Journal Multiscale Modeling & Simulation*, 2011.

I.-H. Kim, C. Y.-C.M., S. D.L., E. R., and R. K. Nonrigid registration of 2-d and 3-d dynamic cell nuclei images for improved classification of subcellular particle motion. *IEEE Transactions on Image Processing*, 2011.

T. H. Kim, H. S. Lee, and K. M. Lee. Optical flow via locally adaptive fusion of complementary data costs. *International Conference on Computer Vision (ICCV)*, 2013.

Y. Kim, A. Martínez, and A. Kak. A local approach for robust optical flow estimation under varying illumination. In *British Machine Vision Conference (BMVC)*, pages 91–1, 2004.

Y.-H. Kim, A. M. Martínez, and A. C. Kak. Robust motion estimation under varying illumination. *Image and Vision Computing*, 23(4):365–375, 2005.

P. Kohli and P. H. Torr. Dynamic graph cuts for efficient inference in markov random fields. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(12):2079–2088, 2007.

P. Kohli, M. P. Kumar, and P. H. Torr. P^3 & beyond: Move making algorithms for solving higher order functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(9):1645–1656, 2009.

D. L. Kolin and P. W. Wiseman. Advances in image correlation spectroscopy: measuring number densities, aggregation states, and dynamics of fluorescently labeled macromolecules in cells. *Cell biochemistry and biophysics*, 49(3):141–164, 2007.

V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(10):1568–1583, 2006.

V. Kolmogorov and C. Rother. Minimizing nonsubmodular functions with graph cuts-a review. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(7): 1274–1279, 2007.

N. Komodakis and N. Paragios. Beyond pairwise energies: Efficient optimization for higher-order mrfs. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2985–2992, 2009.

N. Komodakis and G. Tziritas. Image completion using efficient belief propagation via priority scheduling and dynamic pruning. *IEEE Transactions on Image Processing*, 16 (11):2649–2661, 2007.

N. Komodakis, G. Tziritas, and N. Paragios. Performance vs computational efficiency for optimizing single and dynamic mrfs: Setting the state of the art with primal-dual strategies. *Computer Vision and Image Understanding*, 112(1):14–29, 2008.

C. Kondermann, M. Rudolf, and C. Garbe. A statistical confidence measure for optical flows. In *European Conference on Computer Vision (ECCV)*, pages 290–301, Marseille, France, October 2008.

P. Krähenbühl and V. Koltun. Efficient nonlocal regularization for optical flow. In *European Conference on Computer Vision (ECCV)*, pages 356–369, 2012.

K. Krajsek and R. Mester. On the equivalence of variational and statistical differential motion estimation. In *Southwest Symposium on Image Analysis and Interpretation*, pages 11–15, 1996.

K. Krajsek and R. Mester. Bayesian model selection for optical flow estimation. *Pattern Recognition*, pages 142–151, 2007.

F. R. Kschischang, B. J. Frey, and H.-A. Loeliger. Factor graphs and the sum-product algorithm. *Transactions on Information Theory*, 47(2):498–519, 2001.

N. Kumar, L. Zhang, and S. Nayar. What is a good nearest neighbors algorithm for finding similar patches in images? In *European Conference on Computer Vision (ECCV)*, pages 364–378, 2008.

J. Kybic and C. Nieuwenhuis. Bootstrap optical flow confidence and uncertainty measure. *Computer Vision and Image Understanding*, 115(10):1449–1462, 2011.

S.-H. Lai. Robust image matching under partial occlusion and spatially varying illumination change. *Computer Vision and Image Understanding*, 78(1):84–98, 2000.

T. Lecomte, R. Thibeaux, N. Guillen, A. Dufour, and J.-C. Olivo-Marin. Fluid optical flow for forces and pressure field estimation in cellular biology. In *International Conference on Image Processing (ICIP)*, pages 69–72, Orlando, FL, USA, 2012.

K. Lee, D. Kwon, I. Yun, and S. Lee. Optical flow estimation with adaptive convolution kernel prior on discrete framework. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2504–2511, San Fransisco, June 2010.

V. Lempitsky, S. Roth, and C. Rother. Fusionflow: discrete-continuous optimization for optical flow estimation. In *Computer Vision and Pattern Recognition (CVPR)*, 2008.

V. Lempitsky, C. Rother, S. Roth, and A. Blake. Fusion moves for markov random field optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8): 1392–1405, 2010.

M. Leordeanu, A. Zanfir, and C. Sminchisescu. Locally affine sparse-to-dense matching for motion and occlusion estimation. In *International Conference on Computer Vision (ICCV)*, 2013.

J. Lewis. Fast normalized cross-correlation. *Vision Interface*, pages 120–123, 1995.

W. Li, D. Cosker, M. Brown, and R. Tang. Optical flow estimation using laplacian mesh energy. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2435–2442. IEEE, June 2013.

C. Liu et al. *Beyond pixels: exploring new representations and applications for motion analysis*. PhD thesis, Massachusetts Institute of Technology, 2009.

K. Liu, S. Lienkamp, A. Shindo, J. Wallingford, G. Walz, and O. Ronneberger. Optical flow guided cell segmentation and tracking in developing tissue. In *IEEE International Symposium on Biomedical Imaging (ISBI)*, 2014.

D. G. Lowe. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *International Joint Conference on Artificial Intelligence*, pages 674–679, 1981.

J. Luo and E. E. Konofagou. A fast normalized cross-correlation calculation method for motion estimation. *Transactions on Ultrasonics, Ferroelectrics and Frequency Control*, 57(6):1347–1357, 2010.

Z. Ma, K. He, Y. Wei, J. Sun, and E. Wu. Constant time weighted median filtering for stereo matching and beyond. *Intenational Conference on Computer Vision (ICCV)*, 2013.

M. Maurizot, P. Bouthemy, B. Delyon, A. Juditski, and J.-M. Odobez. Determination of singular points in 2d deformable flow fields. In *In International Conference on Image Processing (ICIP)*, volume 3, pages 488–491, 1995.

E. Meijering, I. Smal, and G. Danuser. Tracking in molecular bioimaging. *Signal Processing Magazine*, 23(3):46–53, 2006.

E. Mémin and P. Pérez. Dense estimation and object-based segmentation of the optical flow with robust techniques. *IEEE Transactions on Image Processing*, 7(5):703–719, 1998.

E. Memin and P. Perez. Hierarchical estimation and segmentation of dense motion fields. *International Journal of Computer Vision*, 46(2):129–155, 2002.

Y. Mileva, A. Bruhn, and J. Weickert. Illumination-robust variational optical flow with photometric invariants. *Pattern Recognition*, pages 152–162, 2007.

M. Mohamed, H. Rashwan, B. Mertsching, M. Garcia, and D. Puig. Illumination-robust optical flow approach using local directional pattern. *Transactions on Circuits and Systems for Video Technology*, 2014.

M. A. Mohamed and B. Mertsching. Tv-l1 optical flow estimation with image details recovering based on modified census transform. In *Advances in Visual Computing*, pages 482–491, 2012.

J. Molnár, D. Chetverikov, and S. Fazekas. Illumination-robust variational optical flow using cross-correlation. *Computer Vision and Image Understanding*, 114(10):1104–1114, 2010.

C. Mota, L. Stuke, and E. Barth. Analytic solutions for multiple motions. In *International Conference on Image Processing (ICIP)*, pages 917–920, Thessaloniki, Greece, October 2001.

M. Mozerov. Constrained optical flow estimation as a matching problem. *IEEE Transactions on Image Processing*, 2013.

T. Müller, C. Rabe, J. Rannacher, U. Franke, and R. Mester. Illumination-robust dense optical flow using census signatures. *Pattern Recognition*, pages 236–245, 2011.

T. Muller, J. Rannacher, C. Rabe, and U. Franke. Feature-and depth-supported modified total variation optical flow for 3d motion field estimation in real scenes. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1193–1200, 2011.

D. Mumford and J. Shah. Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on pure and applied mathematics*, 42 (5):577–685, 1989.

D. W. Murray and B. F. Buxton. Scene segmentation from visual motion using global optimization. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (2): 220–228, 1987.

H. Nagel and W. Enkelmann. An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, (5):565–593, 1986.

H.-H. Nagel. Extending the ‚Äòoriented smoothness constraint‚Äôinto the temporal domain and the estimation of derivatives of optical flow. In *European Conference on Computer Vision (ECCV)*, pages 139–148, 1990.

S. Negahdaripour. Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(9):961–979, 1998.

L. Ng and V. Solo. Errors-in-variables modelling in optical flow problems. In *International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 5, pages 2773–2776, 1998.

C. Nieuwenhuis, D. Kondermann, and C. S. Garbe. Complex motion models for simple optical flow estimation. *Pattern Recognition*, pages 141–150, 2010.

C. Nieuwenhuis, E. Töppe, and D. Cremers. A survey and comparison of discrete and continuous multi-label optimization approaches for the potts model. *International Journal of Computer Vision*, pages 1–18, 2013.

T. Nir, A. M. Bruckstein, and R. Kimmel. Over-parameterized variational optical flow. *International Journal of Computer Vision*, 76(2):205–216, 2008.

J. Nocedal. Updating quasi-newton matrices with limited storage. *Mathematics of computation*, 35(151):773–782, 1980.

P. Ochs, Y. Chen, T. Brox, and T. Pock. ipiano: Inertial proximal algorithm for non-convex optimization. *SIAM journal on imaging science*, 2013.

J. Odobez and P. Bouthemy. Robust multiresolution estimation of parametric motion models. *Journal of visual communication and image representation*, 6(4):348–365, 1995.

J. Odobez and P. Bouthemy. Direct incremental model-based image motion segmentation for video analysis. *Signal Processing*, 66(2):143–155, 1998.

G. Panin. Mutual information for multi-modal, discontinuity-preserving image registration. In *Advances in Visual Computing*, pages 70–81, 2012.

N. Papadakis, R. Yildizoglu, J.-F. Aujol, and V. Caselles. High-dimension multilabel problems: Convex or nonconvex relaxation? *SIAM Journal on Imaging Sciences*, 6(4): 2603–2639, 2013.

N. Papenberg, A. Bruhn, T. Brox, S. Didas, and J. Weickert. Highly accurate optic flow computation with theoretically justified warping. *International Journal of Computer Vision*, 67(2):141–158, 2006.

N. Paragios and R. Deriche. Geodesic active regions and level set methods for motion estimation and tracking. *Computer Vision and Image Understanding*, 97(3):259–282, 2005.

J. Pearl. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 1988.

M. Pinot, V. Steiner, B. Dehapiot, B. Yoo, F. Chesnel, L. Blanchoin, C. Kervrann, and Z. Gueroui. Confinement induces actin flow in a meiotic cytoplasm. *Proc. Nat. Acd. Sci. (PNAS)*, 109(29):11705–11710, 2012.

L. Pizarro, J. Delpiano, P. Aljabar, J. Ruiz-del Solar, and D. Rueckert. Towards dense motion estimation in light and electron microscopy. In *International Symposium on Biological Imaging (ISBI)*, pages 1939–1942, 2011.

T. Pock, M. Pock, and H. Bischof. Algorithmic differentiation: Application to variational problems in computer vision. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(7):1180–1193, 2007.

R. Ranftl, S. Gehrig, T. Pock, and H. Bischof. Pushing the limits of stereo using variational stereo estimation. In *Intelligent Vehicles Symposium (IV)*, pages 401–407, 2012.

H. A. Rashwan, M. A. García, and D. Puig. Variational optical flow estimation based on stick tensor voting. *IEEE Transactions on Image Processing*, 22(7):2589–2599, 2013.

K. Rohr, W. J. Godinez, N. Harder, S. Wörz, J. Mattes, W. Tvaruskó, and R. Eils. Tracking and quantitative analysis of dynamic movements of cells and particles. *Live Cell Imaging*, 2010(6):239–256, 2010.

S. Roth and M. Black. On the spatial statistics of optical flow. *International Journal of Computer Vision*, 74(1):33–50, 2007.

C. Rother, V. Kolmogorov, V. Lempitsky, and M. Szummer. Optimizing binary mrfs via extended roof duality. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007.

L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D: Nonlinear Phenomena*, 60(1):259–268, 1992.

D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *Pattern Analysis and Machine Intelligence*, 18(8):712–721, 1999.

J. A. Schnabel, D. Rueckert, M. Quist, J. M. Blackall, A. D. Castellano-Smith, T. Hartkens, G. P. Penney, W. A. Hall, H. Liu, C. L. Truwit, et al. A generic framework for non-rigid registration based on non-uniform multi-level free-form deformations. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 573–581. Springer, 2001.

C. Schnörr and W. Peckar. Motion-based identification of deformable templates. In *International Conference Computer Analysis of Images and Patterns (CAIP)*, pages 122–129, Prague, Czech Republic, 1995.

T. Schoenemann and D. Cremers. Near real-time motion segmentation using graph cuts. In *DAGM symposium on Pattern Recognition*, pages 455–464, 2006.

P. Schwille. Fluorescence correlation spectroscopy. *Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine*, pages 576–578, 2006.

T. Senst, V. Eiselen, and T. Sikora. Robust local optical flow for feature tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(9):1377–1387, 2012.

M. Sergeev. *High Order Autocorrelation Analysis in Image Correlation Spectroscopy*. PhD thesis, McGill University, 2004.

X. Shen and Y. Wu. Sparsity model for robust optical flow estimation at motion discontinuities. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2456–2463, 2010.

J. Shi and C. Tomasi. Good features to track. In *Computer Vision and Pattern Recognition (CVPR)*, pages 593–600, 1994.

W. Shi, X. Zhuang, L. Pizarro, W. Bai, H. Wang, K.-P. Tung, P. Edwards, and D. Rueckert. Registration using sparse free-form deformations. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 659–666, 2012.

D. Shulman and J.-Y. Herve. Regularization of discontinuous flow fields. In *Workshop on Visual Motion*, pages 81–86, 1989.

E. P. Simoncelli, E. H. Adelson, and D. J. Heeger. Probability distributions of optical flow. In *Computer Vision and Pattern Recognition (CVPR)*, pages 310–315, 1991.

S. N. Sinha, J.-M. Frahm, M. Pollefeys, and Y. Genc. Feature tracking and matching in video using programmable graphics hardware. *Machine Vision and Applications*, 22(1): 207–217, 2011.

A. Sotiras, C. Davatzikos, and N. Paragios. Deformable medical image registration: A survey. *IEEE Transactions on Medical Imaging*, 2013.

A. N. Stein and M. Hebert. Occlusion boundaries from motion: Low-level detection and mid-level reasoning. *International Journal of Computer Vision*, 82(3):325–357, 2009.

F. Stein. Efficient computation of optical flow using the census transform. *Pattern Recognition*, pages 79–86, 2004.

F. Steinbrucker, T. Pock, and D. Cremers. Advanced data terms for variational optic flow estimation. In *Vision, Modeling, and Visualization Workshop*, 2009.

D. Sun, S. Roth, J. Lewis, and M. Black. Learning optical flow. In *European Conference on Computer Vision (ECCV)*, pages 83–97, 2008.

D. Sun, S. Roth, and M. Black. Secrets of optical flow estimation and their principles. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2432–2439, San Fransisco, June 2010a.

D. Sun, E. Sudderth, and M. Black. Layered image motion with explicit occlusions, temporal consistency, and depth ordering. In *Advances in Neural Information Processing Systems (NIPS)*, pages 2226–2234, Vancouver, Canada, December 2010b.

D. Sun, E. B. Sudderth, and M. J. Black. Layered segmentation and optical flow estimation over time. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1768–1775, 2012.

D. Sun, S. Roth, and M. J. Black. A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision*, 106(2):115–137, 2014.

R. Szeliski and H.-Y. Shum. Motion estimation with quadtree splines. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1199–1210, 1996.

R. Szeliski, R. Zabih, D. Scharstein, O. Veksler, V. Kolmogorov, A. Agarwala, M. Tappen, and C. Rother. A comparative study of energy minimization methods for markov random fields with smoothness-based priors. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(6):1068–1080, 2008.

M. Tao, J. Bai, P. Kohli, and S. Paris. Simpleflow: A non-iterative, sublinear optical flow algorithm. In *Computer Graphics Forum*, volume 31, pages 345–353, 2012.

C.-H. Teng, S.-H. Lai, Y.-S. Chen, and W.-H. Hsu. Accurate optical flow computation under non-uniform brightness variations. *Computer vision and image understanding*, 97(3):315–346, 2005.

M. Tistarelli. Multiple constraints to compute optical flow. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(12):1243–1250, 1996.

Trobin, T. Pock, D. Cremers, and H. Bischof. An unbiased second-order prior for high-accuracy motion estimation. *DAGM symposium on Pattern Recognition*, pages 396–405, 2008a.

W. Trobin, T. Pock, D. Cremers, and H. Bischof. Continuous energy minimization via repeated binary fusion. In *European Conference on Computer Vision (ECCV)*, pages 677–690, Marseille, France, October 2008b.

D. Tschumperlé and L. Brun. Non-local regularization and registration of multi-valued images by pde,Äôs and variational methods on higher dimensional spaces. In *Mathematical Image Processing*, pages 181–197, 2011.

D. Tzovaras, M. G. Strintzis, and H. Sahinoglou. Evaluation of multiresolution block matching techniques for motion and disparity estimation. *Signal Processing: Image Communication*, 6(1):59–67, 1994.

J. Ulén and C. Olsson. Simultaneous fusion moves for 3d-label stereo. In *Energy Minimization Methods in Computer Vision and Pattern Recognition*, pages 80–93. Springer, 2013.

M. Unger, M. Werlberger, T. Pock, and H. Bischof. Joint motion estimation and segmentation of complex scenes with label costs and occlusion modeling. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1878–1885, 2012.

S. Uras, F. Girosi, A. Verri, and V. Torre. A computational approach to motion perception. *Biological Cybernetics*, 60(2):79–87, 1988.

J. van de Weijer and T. Gevers. Robust optical flow from photometric invariants. In *International Conference on Image Processing (ICIP)*, volume 3, pages 1835–1838, 2004.

O. Veksler. *Efficient graph-based energy minimization methods in computer vision.* PhD thesis, Cornell University, 1999.

O. Veksler. Graph cut based optimization for mrfs with truncated convex priors. In *Computer Vision and Pattern Recognition (CVPR)*, pages 1–8, 2007.

C. Vogel, S. Roth, and K. Schindler. An evaluation of data costs for optical flow. In *DAGM symposium on Pattern Recognition*, pages 343–353, 2013.

S. Volz, A. Bruhn, L. Valgaerts, and H. Zimmer. Modeling temporal coherence for optical flow. In *International Conference on Computer Vision (ICCV)*, pages 1116–1123, 2011.

M. J. Wainwright, T. S. Jaakkola, and A. S. Willsky. Map estimation via agreement on trees: message-passing and linear programming. *Transactions on Information Theory*, 51(11):3697–3717, 2005.

C. Wang, N. Komodakis, and N. Paragios. Markov random field modeling, inference & learning in computer vision & image understanding: A survey. *Computer Vision and Image Understanding*, 117(11):1610–1627, 2013.

Z. Wang, F. Wu, and Z. Hu. Msld: A robust descriptor for line matching. *Pattern Recognition*, 42(5):941–953, 2009.

A. Wedel, T. Pock, J. Braun, U. Franke, and D. Cremers. Duality tv-l1 flow with fundamental matrix prior. In *Image and Vision Computing (IVC)*, pages 1–6, 2008.

A. Wedel, D. Cremers, T. Pock, and H. Bischof. Structure-and motion-adaptive regularization for high accuracy optic flow. In *International Conference on Computer Vision (ICCV)*, pages 1663–1668, Kyoto,Japan, October 2009a.

A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers. An improved algorithm for tv-l 1 optical flow. In *Statistical and Geometrical Approaches to Visual Motion Analysis*, pages 23–45, 2009b.

J. Weickert and C. Schnorr. A theoretical framework for convex regularizers in pde-based computation of image motion. *International Journal of Computer Vision*, 45(3):245–264, 2001.

J. Weickert and C. Schnörr. Variational optic flow computation with a spatio-temporal smoothness constraint. *Journal of Mathematical Imaging and Vision*, 14(3):245–255, 2001.

J. Weickert, J. Heers, C. Schnörr, K. J. Zuiderveld, O. Scherzer, and H. Siegfried Stiehl. Fast parallel algorithms for a broad class of nonlinear variational diffusion approaches. *Real-Time Imaging*, 7(1):31–45, 2001.

P. Weinzaepfel, J. Revaud, Z. Harchaoui, C. Schmid, et al. Deepflow: Large displacement optical flow with deep matching. In *Intenational Conference on Computer Vision (ICCV)*, 2013.

M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof. Anisotropic huber-l1 optical flow. In *British Machine Vision Conference (BMVC)*, 2009.

M. Werlberger, T. Pock, and H. Bischof. Motion estimation with non-local total variation regularization. In *Computer Vision and Pattern Recognition (CVPR)*, pages 2464–2471, San Fransisco, June 2010.

M. Werlberger, M. Unger, T. Pock, and H. Bischof. Efficient minimization of the non-local potts model. In *Scale Space and Variational Methods in Computer Vision (SSVM)*, pages 314–325, 2012.

Y.-T. Wu, T. Kanade, C.-C. Li, and J. Cohn. Image registration using wavelet-based motion model. *International Journal of Computer Vision*, 38(2):129–152, 2000.

T. Würflinger, J. Stockhausen, D. Meyer-Ebrecht, and A. Böcking. Robust automatic coregistration, segmentation, and classification of cell nuclei in multimodal cytopathological microscopic images. *Computerized Medical Imaging and Graphics*, 28 (1):87–98, 2004.

J. Xiao, H. Cheng, H. Sawhney, C. Rao, and M. Isnardi. Bilateral filtering-based optical flow estimation with occlusion detection. In *European Conference on Computer Vision (ECCV)*, pages 211–224, 2006.

L. Xu, J. Chen, and J. Jia. A segmentation based variational model for accurate optical flow estimation. In *European Conference on Computer Vision (ECCV)*, pages 671–684, Marseille, France, October 2008.

L. Xu, C. Lu, Y. Xu, and J. Jia. Image smoothing via l 0 gradient minimization. *ACM Transactions on Graphics*, 30(6):174, 2011.

L. Xu, Z. Dai, and J. Jia. Scale invariant optical flow. In *European Conference on Computer Vision (ECCV)*, pages 385–399, 2012a.

L. Xu, J. Jia, and Y. Matsushita. Motion detail preserving optical flow estimation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 34(9):1744–1757, 2012b.

K.-J. Yoon and I. S. Kweon. Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):650–656, 2006.

G. Yu and J.-M. Morel. Asift: an algorithm for fully affine invariant comparison. *SIAM journal on imaging science*, 2:438–469, 2009.

R. Zabih and J. Woodfill. Non-parametric local transforms for computing visual correspondence. In *Europena Conference on Computer Vision (ECCV)*, pages 151–158, 1994.

C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime tv-l 1 optical flow. In *DAGM symposium on Pattern Recognition*, pages 214–223, 2007.

C. Zach, D. Gallup, and J.-M. Frahm. Fast gain-adaptive klt tracking on the gpu. In *Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1–7, 2008.

H. Zimmer, A. Bruhn, and J. Weickert. Optic flow in harmony. *International Journal of Computer Vision*, 93(3):1–21, 2011.

C. Zitnick, N. Jojic, and S. Kang. Consistent segmentation for optical flow estimation. In *International Conference on Computer Vision (ICCV)*, pages 1308–1315, Beijing, China, October 2005.