Multimodal image registration in 2D and 3D correlative microscopy

Bertha Mayela Toledo Acosta

ANNÉE 2018

UNIVERSITÉ DE RENNES 1

UNIVERSITE BRETAGNE LOIRE

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

pour le grade de

DOCTEUR DE L'UNIVERSITÉ DE RENNES 1

Mention : Signal, image et vision

Ecole doctorale MathSTIC

présentée par

Bertha Mayela TOLEDO-ACOSTA

préparée au centre de recherche INRIA - Rennes
Bretagne-Atlantique

Multimodal image registration in 2D and 3D correlative microscopy

Soutenance de thèse à Rennes prévue le 23 mai 2018

devant le jury composé de :

Su RUAN
Professeure Université de Rouen/ rapporteur

Michael LIEBLING
Senior researcher IDIAP, Suisse/ rapporteur

Christian BARILLOT
Directeur de recherche CNRS/ examinateur

Yannick SCHWAB
Team leader EMBL, Allemagne/ examinateur

Xavier HEILIGENSTEIN
Chief Scientific Officer de CryoCapCell et Institut Curie, Paris/ examinateur

Grégoire MALANDAIN
Directeur de recherche INRIA/ examinateur

Patrick BOUTHEMY
Directeur de recherche INRIA / directeur de thèse

RÉSUMÉ EN FRANÇAIS

Introduction

La microscopie corrélative, ou “correlative microscopy” en anglais (CM), combine différentes techniques de microscopie pour étudier un échantillon biologique. Imager un objet selon deux ou plusieurs approches différentes permet de fournir des informations importantes et complémentaires sur cet échantillon.

Les techniques utilisées en CM relèvent généralement de la microscopie optique (par exemple, microscopie de contraste de phase, microscopie de fluorescence, etc.), dénommée "light microscopy" en anglais (LM), et de la microscopie électronique, "electron microscopy" en anglais (EM) (par exemple, microscopie électronique à balayage, microscopie électronique à transmission, etc.), constituant un schéma particulier, dénommé “correlative light and electron microscopy” (CLEM) en anglais. LM permet l’imagerie d’échantillons vivants, fournissant des informations dynamiques à une résolution spatiale relativement faible, tandis que EM acquiert des images extrêmement détaillées au prix de la fixation de l’échantillon, chimiquement ou par congélation. EM et LM sont deux techniques fondamentales de microscopie dont les caractéristiques se complètent, ce qui est la raison pour laquelle elles sont couramment adoptées en CM. Cependant, d’autres techniques de microscopie peuvent être employées dans CM, à deux et trois dimensions.

L’acquisition d’images CLEM est une tâche difficile où la préparation de l’échantillon est cruciale. Elle dépend du type d’études biologiques et des techniques de microscopie utilisées [26, 145, 149]. Habituellement, l’échantillon est d’abord préparé et coloré pour LM. Les images optiques sont alors acquises. A la suite de l’identification d’un événement ou d’un contenue d’intérêt, l’échantillon est fixé pour le préparer à l’acquisition EM. La figure 1 illustre un pipeline d’acquisition CLEM usuel. Certes, cette description est une vue simplifiée. Le processus complet est généralement étalé sur plusieurs jours.

Les spécificités de chaque étape du pipeline ont des effets différents sur l'échantillon, certaines d'entre elles introduisant des artefacts ou des déformations, rendant la correspondance entre les images plus difficile à réaliser. De plus, les acquisitions LM et EM sont généralement effectuées sur deux équipements différents et l'échantillon doit être transporté de l'un à l'autre, ce qui provoque des différences significatives dans la disposition et l'orientation de l'échantillon pendant les acquisitions.

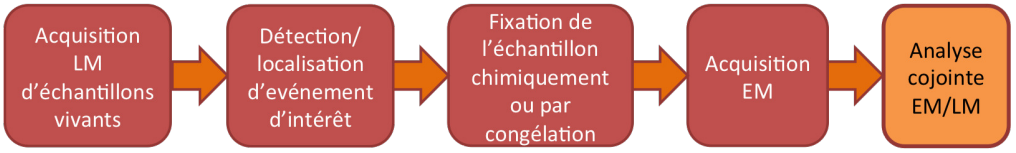


Fig. 1: *Pipeline d'acquisition CLEM.*

L'association de la microscopie optique et de la microscopie électronique délivre des informations fonctionnelles et dynamiques ainsi que structurelles à haute résolution, ce qui peut aider à la compréhension des mécanismes cellulaires et intracellulaires. Cependant, les images EM et LM sont très différentes, en termes de contenu, d'apparence, de résolution, de taille d'image et de champ de vue; par conséquent, le recalage des images EM et LM nécessaire à la fusion des informations acquises n'est pas simple. Une illustration des différences mentionnées entre les images EM et LM est fournie à la figure 6, où le cadre orange dans l'image 2D LM englobe toute la zone imagée en 2D EM. Nous pouvons également en déduire la grande différence dans les orientations des images EM et LM, dans ce cas, proche d'une rotation de 90°. La taille de cette image LM est de 1392 ×1040 alors que la taille de l'image EM est de 4008×2664. La résolution du pixel LM est environ dix fois inférieure à celle de l'image EM. L'image EM peut être acquise à un encore plus fort grossissement, ce qui augmente d'autant l'écart de résolution. Enfin, la dissemblance des intensités et des structures présentes dans les images est remarquable.

La combinaison de différentes modalités d'imagerie offre de nombreuses possibilités mais laisse à l'utilisateur le défi d'apparier et de superposer les images résultantes. En pratique aujourd'hui, de interventions manuelles sont nécessaires à différentes étapes du pipeline, par exemple lors de la localisation d'une région d'intérêt dans l'échantillon pour l'acquisition EM après avoir acquis des images en LM, ou dans la sélection et l'appariement d'amers pour la calcul du recalage. Une fois le calcul du recalage effectué les images sont généralement superposées pour une bonne compréhension de leur contenus respectifs et de leur juxtaposition. Ces différentes interventions manuelles sont exigeantes et difficiles et le résultat final peut être insatisfaisant. De plus, les dispositifs de CLEM vont se multiplier et l'acquisition d'un toujours plus grand nombre d'images

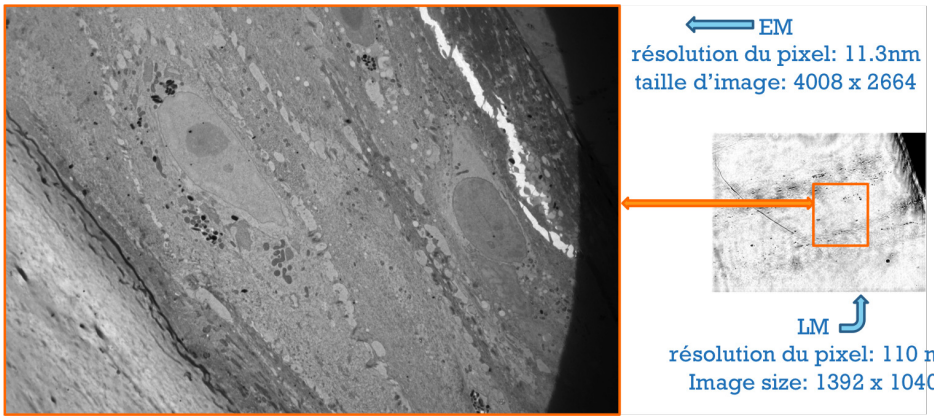


Fig. 2: *Différences de champ de vue, d'apparence, de contenu et de résolution entre les images EM et LM dans une expérience CLEM.*

posera le problème du passage à l'échelle. Pour cette raison, la nécessité d'automatiser la phase de recalage s'impose de plus en plus.

Recalage automatique d'images CLEM 2D

Au cours de ce travail, nous avons défini plusieurs méthodes visant à créer une procédure automatique pour aligner correctement les images CLEM en 2D et en 3D. Cet alignement ou recalage résulte du calcul de transformations géométriques entre les images LM et EM. Il existe deux grandes approches de recalage. La première est de nature géométrique. Elle extrait d'abord des points caractéristiques des deux images, les apparie, puis à partir de ces appariements, calcule la transformation géométrique établissant le recalage. La seconde approche s'appuie directement sur les intensités des images et cherche à estimer la transformation géométrique à travers la minimisation d'une mesure de similarité.

Compte tenu des différences entre les images EM et LM, illustrées sur la figure 6, un important écart est attendu entre les positions et orientations des échantillons dans les deux images. Bien que les méthodes de recalage basées sur des points caractéristiques soient capables de traiter de grands écarts initiaux, l'extraction et l'appariement automatique de points communs entre images EM et LM est une tâche très ardue. Par ailleurs, la plupart des méthodes de recalage basées directement sur l'intensité ne sont pas capables de gérer de gros écarts initiaux. En conséquence, une étape de pré-recalage est impérative afin d'aligner grossièrement les images EM et LM et de faciliter ensuite l'estimation d'une transformation pour le recalage final. Actuellement, une telle étape de pré-recalage est effectuée manuellement.

Quand on traite des images multimodales, deux stratégies peuvent être mises en œuvre. La première est l'utilisation de métriques de similarité capables de gérer les intensités multimodales des images, telle que l'information mutuelle, "mutual information" (MI) en anglais. L'autre approche consiste à adapter une image à la modalité de l'autre, ou à transformer les deux images en une troisième modalité commune. Nous adoptons ce dernier schéma pour l'étape de pré-recalage, en exploitant le Laplacien de Gaussien (LoG). La transformation LoG a l'avantage supplémentaire de lisser le bruit tout en mettant en évidence les structures importantes, ce qui en fait un choix efficace [7].

Le recalage entre images EM et LM est souvent défini autour d'une région d'intérêt ("region of interest" -ROI- en anglais) contenant des structures partagées. La définition d'une ROI est nécessaire puisque les images CLEM ont des contenus souvent très différents. Généralement, le ROI est déterminé dans l'une des images CLEM, alors que son emplacement correspondant dans l'autre image est inconnu. Ces images sont respectivement la source et la cible.

Comme il n'y a en général aucune information *a priori* concernant l'emplacement du ROI dans l'image cible, ni sur son orientation, nous proposons de mettre en œuvre une recherche exhaustive par patch dans l'image cible, du ROI prédéfini délimité dans l'image source. En comparant le contenu du ROI source avec le contenu des patchs testés dans l'image cible, le patch correspondant au ROI peut être localisé. Pour faciliter cette étape de pré-recalage, nous utilisons pour cette recherche exhaustive la représentation LoG des images CLEM. Un descripteur invariant à l'échelle et à la rotation est nécessaire pour comparer chaque patch avec la ROI initiale. Ainsi, deux méthodes à base d'histogrammes ont été définies. L'écart entre l'emplacement du patch sélectionné dans la cible et celle du ROI fournit une première estimation de la translation entre les images LM et EM.

Pour la suite des étapes de recalage, il nous faudra tenir en compte de la nature multimodale du problème, ainsi que des différences précédemment mentionnées entre les images EM et LM. Par conséquent, nous adoptons l'information mutuelle comme mesure de similarité. Une fois que le patch correspondant au ROI a été localisé, l'image source est mise à l'échelle et déplacée vers l'emplacement de ce patch. L'échelle utilisée est connue à partir du rapport de résolution entre les images EM et LM fourni dans les métadonnées associées aux images acquises. Ensuite, une première rotation entre les images est calculée par une approche exhaustive dans un ensemble de valeurs quantifiées d'angle de rotation 2D dans l'image. Elle est suivie de l'estimation, toujours par minimisation de l'information mutuelle, d'une transformation rigide puis affine pour obtenir le recalage final des images EM et LM.

Notre approche a été testée avec succès sur sept paires d'images CLEM 2D réelles fournies par l'institut Curie, comprenant une variété de contenus et de configurations en termes de taille d'image, de champ de vue, de résolution et de structures imagées [157].

Détection multi-échelle de spots

Parallèlement à notre recherche principale sur le recalage CLEM, nous avons étudié deux problèmes complémentaires. Le premier concerne une méthode de segmentation multi-échelle dans les images de microscopie. La détection de structures ponctuelles de différentes tailles est nécessaire pour de nombreuses applications de traitement d'image. Les images peuvent contenir des objets tels que des voitures dans la surveillance du trafic routier, des bateaux sur l'océan en télédétection, des étoiles et des objets célestes en astronomie, des animaux dans des vidéos de scènes naturelles, des éléments cellulaires et subcellulaires en imagerie microscopique, entre autres. Ces objets peuvent apparaître comme des structures locales, qu'on nommera "spots", de taille et de forme partagées. Pour segmenter de façon fiable ces spots il devient crucial de connaître l'échelle appropriée pour les détecter correspondant à leur taille. En conséquence, nous avons introduit une méthode efficace capable de détecter des spots de différentes tailles contenues dans des images corrompues par le bruit. Nous avons défini un critère original de type *a contrario* pour sélectionner automatiquement les échelles significatives, c'est-à-dire les échelles correspondant aux tailles des structures pertinentes dans l'image. Nous avons de plus élaboré une méthode de segmentation multi-échelles afin d'extraire les spots associés à chaque échelle significative. Des images simulées et réelles ont été utilisées pour tester l'efficacité et la précision de notre méthode. Nous avons montré qu'elle surpasse les méthodes de segmentation multi-échelles existantes [156].

Nous avons exploité notre méthode de segmentation de spots pour compléter notre méthode de recalage de CLEM 2D par une dernière étape de recalage affine à partir de points. Comme les images sont très proches à ce stade, nous pouvons apparier les points extraits respectivement de l'image LM et de l'image EM selon le critère du plus proche voisin. Ensuite, il est aisé de calculer une transformation affine complémentaire. Cela a permis d'améliorer encore la précision obtenue dans le recalage 2D des images CLEM. Nous adoptons la méthode "random sampling consensus" (RANSAC) en anglais, pour calculer cette dernière transformation. Notre approche a été testée avec succès sur cinq paires d'images de CLEM 2D.

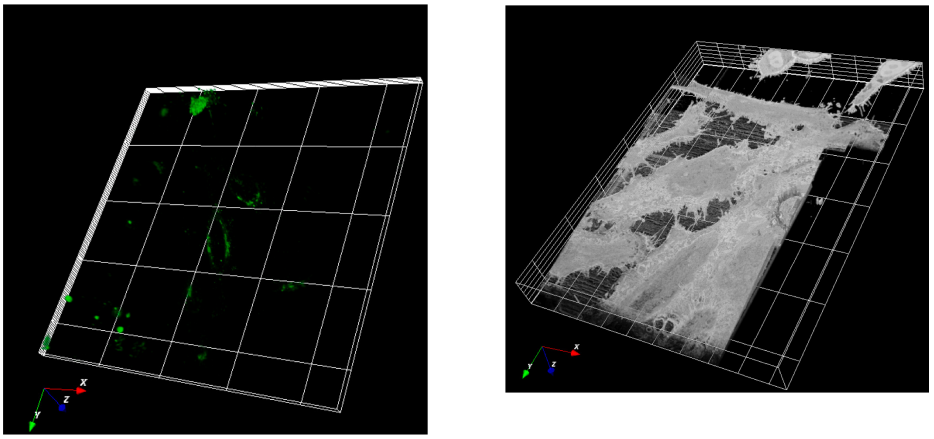
Sélection robuste de modèles de mouvement

Le deuxième module que nous avons développé était motivé par le besoin de savoir quel modèle paramétrique peut le mieux décrire la transformation géométrique entre deux images. Nous avons plus généralement attaqué le problème du modèle de mouvement global entre deux images successives dans une séquence d'images. Des modèles paramétriques de mouvement sont couramment utilisés dans le recalage d'images, l'analyse du mouvement, le suivi, la stabilisation vidéo, entre autres applications. Pour estimer les paramètres d'un modèle global de mouvement, un cadre robuste d'estimation du mouvement est nécessaire pour appréhender la présence fréquente d'outliers vis-à-vis du mouvement dominant, ce dernier étant généralement dû au mouvement de la caméra et les premiers à des objets mobiles indépendants de la scène. Décider du modèle paramétrique de mouvement le plus adapté pour estimer le mouvement dans une séquence est une tâche importante. Certains critères traitent déjà de ce problème, mais la plupart d'entre eux ne sont pas conçus dans un cadre d'estimation robuste. Nous proposons deux critères statistiques de sélection robuste de modèles de mouvement. Le premier est basé sur la statistique de Fisher, tandis que le second est une extension robuste du critère d'information de Takeuchi (TIC). Les deux critères ont été largement testés sur des séquences d'images générées par calcul avec vérité-terrain, ainsi que sur de nombreuses séquences vidéo réelles, démontrant leur efficacité. Enfin, dans le but de tester davantage les performances de nos critères de sélection robuste de modèles de mouvement et d'analyser si le recalage final est amélioré, nous l'avons appliqué sur quelques unes de nos paires d'images CLEM 2D recalées. Comme l'estimation des modèles de mouvement paramétriques est basée sur l'hypothèse de conservation de l'intensité, nous avons utilisé la représentation LoG des patchs d'images EM et LM.

Recalage d'images CLEM 3D

Pour étudier plus d'aspects du problème de recalage pour la CLEM, nous avons étendu notre travail à la CLEM 3D. L'alignement manuel de volumes de CLEM 3D est une tâche beaucoup plus difficile et fastidieuse que l'alignement des images CLEM 2D. Elle peut prendre des dizaines de minutes pour obtenir un résultat satisfaisant. Dans la plupart des problèmes de recalage multimodal 3D, il existe à nouveau deux approches principales pour réaliser l'alignement des volumes ou piles d'images : soit à partir de points caractéristiques, soit par l'utilisation directe de l'intensité. Cependant, des difficultés similaires au cas de la CLEM 2D apparaissent. Une illustration de piles d'images CLEM est montrée à la figure 3, où des différences flagrantes dans le champ de vue, le contenu

de l'image et l'apparence rendent l'application des schémas de recalage classiques peu évidente. L'extraction et l'appariement de caractéristiques est une tâche compliquée car de nombreux éléments présents dans la pile d'images EM sont sans correspondants dans la pile d'images LM. D'autre part, la mise en œuvre directe d'une méthode de recalage itérative basée sur l'intensité sera difficile compte tenu des écarts importants d'orientation et de localisation des contenus respectifs des volumes d'images LM et EM. De plus, contrairement aux situations multimodales en imagerie médicale, il existe une forte anisotropie des volumes ou piles d'images de microscopie. Les volumes comportent peu de couches en Z comparé à la résolution latérale en X et Y des images.



(a) Pile d'images LM (b) Pile d'images EM

Fig. 3: *Piles d'images 3D CLEM (images provenant de The Francis Crick Institute, UK)*

Motivés par notre méthode de recalage en CLEM 2D, nous avons conçu un schéma de recalage CLEM 3D étendant certaines des étapes définies pour le problème 2D. Comme en CLEM 2D, la grande différence de localisation des structures d'intérêt rend nécessaire la mise en œuvre d'une étape de pré-recalage entre les piles d'images EM et LM. Dans de nombreuses acquisitions CLEM 3D, la différence majeure entre les emplacements des ROI se situe dans le plan XY, tandis que les emplacements respectifs le long de l'axe Z sont plus proches. Par conséquent, nous proposons de traiter d'abord l'écart de recalage en XY. Pour pouvoir appliquer la méthode de recherche exhaustive, nous devons associer les piles LM et EM à une modalité commune. Nous exploitons à nouveau la transformation LoG compte tenu des avantages précédemment mentionnés. Une fois que les intensités des piles LM et EM sont ainsi transformées, nous réduisons les piles d'images à deux dimensions en projetant la pile LM en 2D et en sélectionnant un ensemble de couches EM 2D intersectant une ROI 3D prédéfinie. La projection de la pile LM en une carte 2D est obtenue par projection sur le plan XY du maximum d'intensité

en chaque point (x, y) le long de l'axe des Z, dit en anglais "Maximum Intensity Projection" (MIP). Ensuite, nous implémentons une recherche exhaustive par patch similaire à celle du schéma de CLEM 2D, en essayant de localiser les ROI 2D extraits des couches EM 2D dans la projection LM-MIP. Nous obtenons ainsi une ensemble de candidats de localisation, et une moyenne pondérée robuste de l'emplacement des patchs candidats fournit le patch sélectionné. Les coordonnées de son centre sont utilisées pour déplacer la pile LM au centre du ROI EM. Une fois les piles pré-recalées, une méthode de recalage 3D à plusieurs étapes (transformation rigide puis affine) peut être appliquée en utilisant l'information mutuelle comme mesure de similarité. Notre méthode a été testée sur trois jeux de données de CLEM 3D, avec des résultats prometteurs.

Conclusion

Il n'existe à ce jour que peu d'études d'automatisation du recalage CLEM et encore moins en 3D. Notre travail représente une contribution novatrice à la fois pour la CLEM 2D et la CLEM 3D. Nous avons obtenu des résultats satisfaisants, tant visuellement que quantitativement. De plus, nos résultats ont été validés par des biologistes impliqués dans l'acquisition des jeux de données CLEM utilisés dans nos expériences. Néanmoins, la conception de méthodes supplémentaires pour améliorer encore l'alignement final des images CLEM 2D et surtout 3D, tout en diminuant encore plus le temps de calcul est toujours nécessaire. Actuellement, le temps de calcul du recalage en CLEM 2D est en moyenne de 1,2 minutes et en CLEM 3D de 3,5 minutes sur un processeur Intel Core i7 2,8 GHz avec RAM de 16 GB. Idéalement, les futures méthodes de recalage CLEM devront être plus robustes au manque de structures correspondantes et aux différences d'orientation, d'échelle et de champ de vue. Des expériences supplémentaires sont également nécessaires pour valider davantage notre méthode de recalage pour la CLEM 3D, néanmoins l'accès aux données reste délicat, et la disponibilité de dispositifs de CLEM 3D encore limitée.

CONTENTS

INTRODUCTION

Many microscopy techniques currently exist to study small objects and structures in live sequences. Each technique has specificities, benefits and limitations and their application is usually dependent on the biological problem. The main types of microscopes can be divided in optical (light) and electron microscopes, although other kinds of microscopes also exist (micro-computer tomography or microCT, x-ray microscope, ultrasonic microscope, etc.). Given the particular characteristics of each technique it is natural to envision that fusing different techniques could generate new information to help researchers solve problems in many different fields. Still, merging microscopy techniques is not straightforward and new challenges emerge. The combination of different microscopy techniques to study a single object is denominated correlative microscopy (CM) and it has become a widely used method in disciplines such as cell biology [105, 111], cancer research [69, 109], neuroscience [4, 21] and marine microbiology [113]. Our work will be primarily connected to cell biology applications. The most commonly techniques incorporated in CM are light microscopy(LM) and electron microscopy(EM), creating a new technique denominated correlative light and electron microscopy (CLEM).

Light microscopy

The light microscope is the most widely used tool in microbiology laboratories, with applications going back to the 19th century. The phase contrast microscope was invented in 1932 allowing to clearly observe microorganisms and it is still extensively used nowadays, since it requires no staining of the sample. From there, many different light microscopy techniques have been developed. A detailed listing of existing light microscopy techniques can be found in [78, 124]. Some of the current light microscopy techniques include:

1. **Transmitted light microscopy:** in this kind of microscopes, the light is sent from a source on the opposite side of the sample to the objective lens. A magnified image of the illumined sample is created after the light passes through the sample and the lens. Transmitted-light microscopes are used to study thin sections of transparent or semitransparent samples. Other microscopy techniques such as phase contrast, dark and bright field, among others, depend upon the transmitted light path.

2. **Bright field microscopy:** one of the most widely used microscopy imaging techniques, bright field microscopy captures the magnified image of the sample after applying a strong illumination. Some of the light is absorbed by dense areas of the sample, stained structures and other elements, which creates contrast. Then, the image appears as a collection of dark or highly colored structures against a bright background, from which the name of the technique is derived.

3. **Dark field microscopy:** to imagine the sample, oblique rays of light illuminate the sample and no direct light enters the lens. Then, the only light captured is the one diffracted by the sample, resulting in a bright sample against dark background.

4. **Phase contrast microscopy**: in this technique, phase shifts due to light diffracted by an object are converted to amplitude changes, creating contrast for imaging a sample. Staining of the sample is not required.

5. **Fluorescence microscopy (FM):** one of the most important light microscopy tools is the fluorescence microscopy. As stated by its name, fluorescence microscopy is used for imaging objects that emit fluorescent light. However, most biological molecules do not fluorescence on their own. Then, biological molecules are linked with fluorescent molecules named fluorochromes or fluorophores. These molecules can emit light upon light excitation. Several varieties of fluorophores exist and can selectively stain tissues or cells, which is called fluorescent labeling. Fluorophores can be found in the literature as markers, tags, dyes, indicators, etc. Depending on the kind of fluorophores used, a specific wavelength of light is used to illuminate the sample, causing the fluorophores to emit the energy as a longer wavelength. The imaging is performed by either transmitted or reflected illumination (epi-illumination). Many microscopy techniques derived from fluorescence microscopy, such as confocal laser scanning microscopy, two-photon microscopy, scanning disk confocal microscopy, total internal reflection and super-resolution microscopy. A more comprehensive description of these methods can be found in [136].

6. **Light sheet fluorescence microscopy:** is an alternative to traditional light microscopy techniques, where the optical illumination and detection share the same

path. By decoupling the optical illumination and detection pathways, unnecessary excitation of out-of-focus fluorescent elements is avoided. In addition, since lower illumination intensities are required, the signal to noise ratio is improved and faster acquisition is achieved. Further information can be found in [168].

Two main benefits are provided by light microscopy in the study of biological problems. Using light microscopy, living cellular and subcellular specimens can be imagined, producing abundant and valuable details on fundamental life processes such as cellular proliferation, cell mitosis and apoptosis, and cell migration [71]. Additionally, fluorescent labeling provides dynamic contrast and helps to localize events of interest and to distinguish similar structures. Nevertheless, light microscopy techniques are limited in spatial resolution to approximately $0.2\mu m$ due to the wavelength of light. Recently developed superresolution optical microscopy techniques are able to increase the resolution of optical-based microscopy imaging methods, allowing to capture images with a higher resolution than the diffraction limit. Stimulated emission depletion (STED) microscopy [52] creates super-resolution images by employing a second laser to selectively deactive fluorophores localized off the center of the excitation, reaching a resolution of 50nm-90nm. High-precision localization of single fluorophores was developed in [31, 102], which led to microscopy techniques such as photoactivated localization microscopy [13] employs fluorescent probes that can switch between light and dark states. Then, two close probes can be distinguished by being activated at different times with resolution levels in the tens of nm. The joint advances in this new technologies were recently awarded a Nobel Prize in 2014 in the field of chemistry for the works presented in [13, 52, 102]. Still, details of the ultrastructure of cells are still not effectively resolved with optical microscopy approaches, even with super-resolution microscopy techniques [58].

Electron microscopy

Electron microscopes differ from light microscopes by producing an image of a sample using a beam of electrons instead of a light beam. Given that electrons have shorter wavelength than light, they are capable of reaching very high resolutions, allowing to examine from samples containing multiple cells to subcellular ultrastructures, molecules and even atomic structures. To imagine a sample using a electron microscope the sample must be placed in a vacuum along to being subjected to fixation or freezing, making it impossible to imagine living samples. Electron microscopy can be divided in two types: scanning electron microscopy (SEM) and transmission electron microscopy (TEM).

- **Transmission electron microscopy:** in this microscopy technique, a beam of electrons is transmitted through a sample to produce an image. The contrast necessary for imaging is given by the electron scattering. However, since most biological elements do not scatter electrons, additional preparation steps must be undertaken, such as adding heavy metals to the sample. The sample is cut into ultrathin slices to be imagined by TEM. The sectioning of the sample along with the additional steps needed such as fixation, embedding or freezing, makes preserving the original structure of the sample a challenge. Several different protocols and methods have been developed for TEM acquisition. TEM is often used to imagine detailed ultrastructure of cells. Extensive information can be found in [172].
- **Scanning electron microscopy:** to produce an image using this technique, a beam of electrons is moved back and forth across the surface of a sample. Energy is lost when the electron beam interacts with the sample. This loss is converted to other kinds of signals, which are detected and mapped to create an image. Since SEM scans the surface of the sample, sample preparation for sectioning is reduced. Additionally, SEM is capable of producing good 3D images due to its great depth of field. Further information can be found in [43].

Important advances in the sample preparation process improve the preservation of the sample for EM acquisition. Cryo-electron microscopy (CryoEM) is a method where the sample is rapidly cooled, allowing the imaging of molecules near their native state. A Nobel prize in chemistry in 2017 was awarded for the works presented in [32, 37, 53].

Although high spatial resolution is achieved using electron microscopy techniques, it is limited to taking only still images of key stages in biological processes, lacking the capacity of observing real-time dynamic events. Additionally, structures similar to other structures or structures with unknown characteristics can be hard to distinguish [77].

CLEM

It is evident that neither light microscopy or electron microscopy techniques are able to provide all the information a biologist might need. Given the complex nature of biological processes, integrating two different microscopy imaging approaches can help in increasing the knowledge of cellular dynamic behavior and the associated morphology to decipher life mechanisms. As previously mentioned, LM is fast and allows the study of dynamic events as well as fixed samples, labeling structures of interest for localization and analysis at low resolution, while EM visualizes different elements such as membrane shapes, organelle architecture and cellular ultrastructure, both labeled and unlabeled at

high resolution. Then, analyzing events of interest with LM followed by ultrastructural EM could help to address important cell biological issues. A list of general fundamental characteristics of transmitted light microscopy, fluorescence microscopy and electron microscopy is presented in Table 1 to better visualize the complementary qualities of these microscopy techniques.

	Fluorescence microscopy	Transmitted light microscopy	Electron microscopy
Spatial resolution	Low	Low	High
Observation of living cells	Yes	Yes	No
Identification of labeled structures	Strong	Regular	Regular
Identification of unlabeled structures	Not possible	Regular	Regular

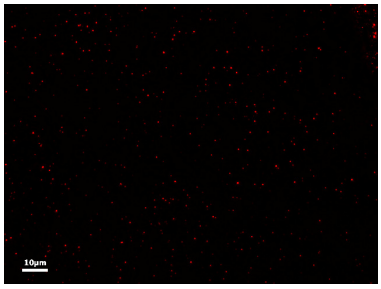
Table 1: Imaging features of fluorescence and electron microscopy

An example of a CLEM acquisition [1] is illustrated in Fig. 4. Reconstructed epidermis samples are imaged following the protocol described in [51]. After live LM acquisition, a quick freeze substitution was conducted, allowing the preservation of fluorescence in the sample. Afterwards, the sample is sectioned. The LM images are acquired with epifluorescence Nikon Te2000, in multiple different channels, including transmitted light. The section was then post-stained with Lead Citrate and imaged by a Tecnai Spirit FEI transmission electron microscopy (TEM) [50]. Four fluorescent channels including mCherry, green fluorescent protein (GFP), Hoechst and Cy5 are used to label different parts of the sample. Along with the FM acquisition, a transmitted light microscopy image is also captured to help in the correlation of the LM image with the EM image.

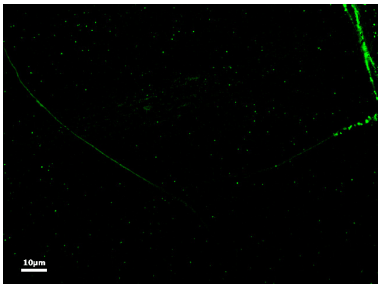
As observed in Fig. 4, the information provided by each LM channel and the EM image is vastly different, highlighting the appeal of combining such different modalities to draw new knowledge. The fluorescent channels depict only the labeled structures while the rest of the sample is not visible. Transmitted light microscopy images all the sample at low resolution, but without the additional information provided by the fluorescent labels.

CLEM has been developed in the last decades by combining a broad range of light

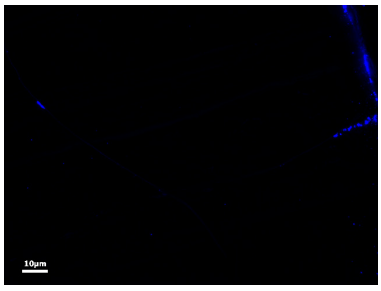
[1] We thank X. Heiligenstein and P. Paul-Gilloteaux from UMR 144, Institut Curie, Paris, for providing us the images and the associated explanations.



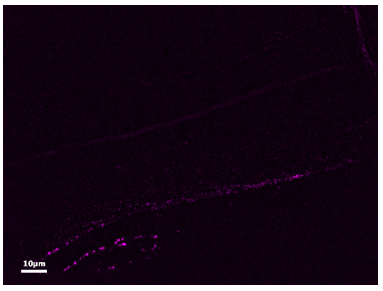
(a) mCherry fluorescence channel



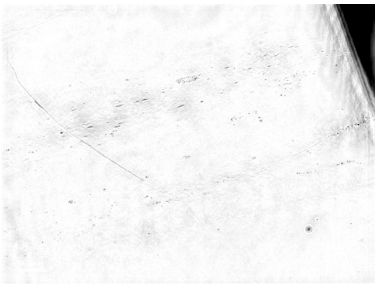
(b) GFP fluorescence channel



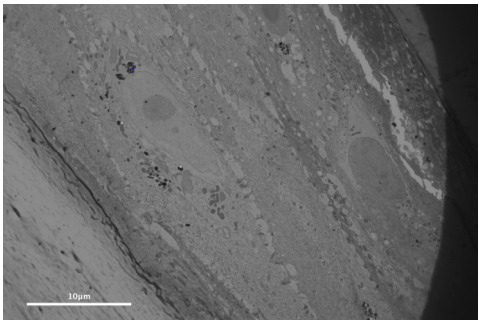
(c) Hoeschst fluorescence channel



(d) Cy5 fluorescence channel



(e) Transmitted light channel



(f) EM image

Fig. 4: *CLEM acquisition of a single sample. (a) - (e) Four fluorescent channels plus a transmitted light image make the LM acquisiton. (f) EM image of the same sample.*

and electron microscopy techniques, both in 2D and 3D. Very specific workflows are designed depending on the microscopy techniques adopted for CLEM and the biological problem to be studied. In most CLEM experiments, the usual process is to first acquire a LM image of a sample labeled using fluorescent tags, followed by preparation of the sample for EM imaging, such as fixation/freezing, embedding, inmunolabeling and sectioning, always depending of the specific subject of study. Then, the resulting images are aligned to be jointly analyzed, generally creating an overlay of the images. A detailed workflow of the most popular CLEM techniques is illustrated in Fig. 5. General information on CLEM workflow is presented in [19]. Further information on CLEM acquisition and sample preparation is reviewed in [29].

Although the concept of CLEM is associated with the joint analysis of the EM and LM images at the final stage of the CLEM framework, the combination of LM and EM images is exploited in several different ways. The most simple and conventional way LM is used along with EM, is when transmission light microscopy is used as a guide for EM acquisition by identifying specific regions or structures of interest in a sample [145]. LM is used to locate a region of interest (ROI) in [87], then a Focus Ion Beam Scanning Electron Microscopy (FIBSEM) is performed. Afterwards, axons and dendrites are imaged using manual selection and seed growing.

CLEM is also implemented when acquiring large-scale LM and EM images [145]. Imaging large areas at high magnification is usually done by capturing several EM and LM images at different locations and stitching them together. The samples are correlated by locating parts of the holder visible in the images or by identifying structures in the sample. Then, the LM and EM images are aligned by partly manually scaling, rotating and shifting the images. Large-scale studies generate great quantities of valuable information, and some more automated approaches have been developed [4, 21].

A CLEM protocol is clearly not as straightforward as illustrated in Fig. 5. Each stage is usually subdivided into several highly specific consecutive steps, involving different physical setups and equipment. Therefore, it is expected that the handling of the sample through all these steps will introduce changes in the orientation of the sample, making it hard to locate and follow a ROI between each step. Multiple efforts have been made in retrieving the ROI, most of them focused on introducing visible marks on the sample [98] or the use of marks or grids in the holders [162]. However, adding marks to the sample can introduce artifacts, and the use of grids may partially occlude the sample [137]. A different localization technique implements a laser dissection, creating a mapping system to locate the ROI [72].

Other mechanisms are exploited to retrieve the ROI during the CLEM process, such

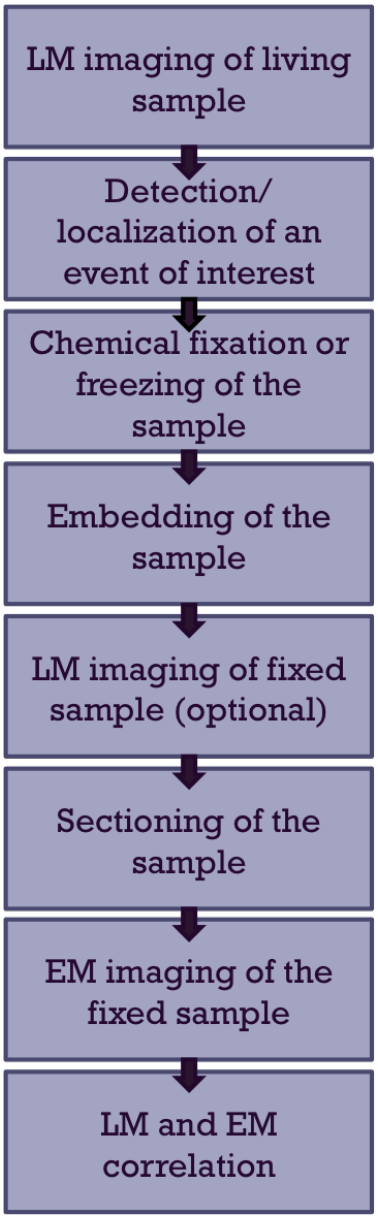


Fig. 5: *General CLEM workflow. LM may occur at two different stages, imaging the sample when it is alive and after fixation for EM.*

as photoconversion [97], inmunolabeling [164] and the use of fluorescent proteins. The latter is one of the most widely used methods of correlating EM and LM images. Many current CLEM workflows use fluorescent proteins for live-cell imaging followed by EM acquisition thanks to the preservation of the fluorescence for EM imaging by cryo-freezing of the sample [50, 77], chemical fixation [167] and high-pressure freezing [64]. Specialized software for CLEM alignment by exploiting fluorescent proteins and other markers have been developed [51, 137].

Recently, advances have been made in performing LM and EM imaging in a single physical setup. Integrating two microscopy modalities in one setup allows the LM imaging of a sample following immediately by EM imaging, consequently producing images that are already aligned. A module to set the sample is developed in [2], in a way that FM is first performed and then the sample is rotated 90° for EM acquisition. This way, the ROI can be easily localized in both modalities. In another integrated setup for 3D CLEM acquisition, a laser is used to create patterns that facilitate the orientation and localization of a structure of interest in both LM and EM images [151]. However, integrated setups are highly specific, limiting the options a biologist has to combine different imaging modalities and to select diverse sample preparation techniques [55].

Integrating two microscopy imaging techniques by correlating the images produced by each one generates additional and substantial information. However, CLEM is a complex process that can take several days. A detailed protocol of CLEM acquisition with specific times delineated for each step of the process is presented in [87]. The analysis and alignment (also called registration) of CLEM images involves the relocation of the ROI identified in LM, in the EM (or vice versa). As previously stated, this relocation is more frequently done manually, with the help of grids, fiducial markers, laser marks and cell features. In some CLEM workflows, the sample is first analyzed at low magnification in order to identify the particular cell of interest, before increasing the magnification to visualize the ultrastructure within that cell and the precise location of fluorescent labels. What all the CLEM works previously cited have in common, is that the LM and EM images are registered at least partly manually, most of the time creating image overlays. Manually relocating and aligning or registering the LM and EM images is usually a difficult and time consuming task in an already long and complicated process. Furthermore, a certain level of expertise is required for the manual registration of CLEM images, as well as previous knowledge of the current dataset. Even then, the resulting alignment may not be entirely satisfactory. In consequence, the development of accurate automatic CLEM registration tools is crucial.

The manipulation of the sample during the different stages of the CLEM framework most certainly will introduce distortions and alter the orientation of the sample. Along

with the inherited differences between the LM and EM images in terms of intensity, pixel resolution, imaged structures, image size and field of view, automatic CLEM registration is a highly challenging problem. An example of this is illustrated in Fig. 6. In this experiment, significant differences can be easily observed between the EM and LM images. Intensities are remarkably dissimilar, as well the level of detail present in the images. A rotation of more than 90° is present. Some of the few shared structures of interest are the dark spots enclosed in orange in the LM image. Additionally, the orange frame indicates the region that contains the whole area imagined by EM. All the listed discrepancies furthermore establish automatic CLEM registration as a complex problem.

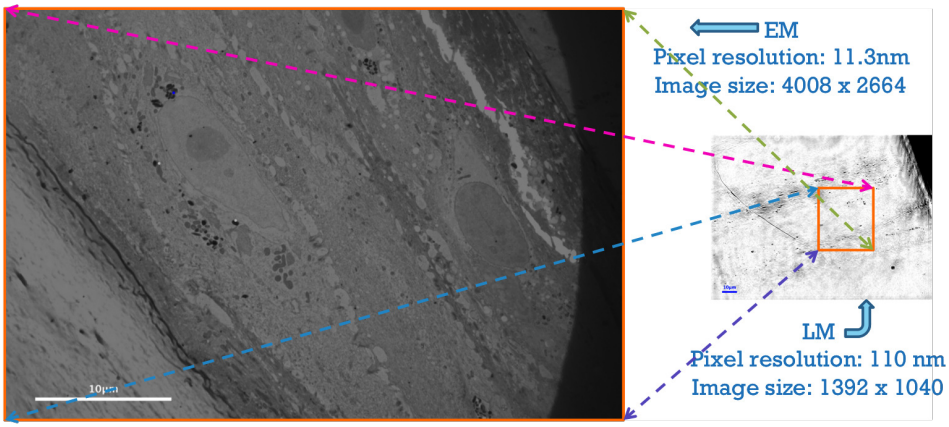


Fig. 6: *Differences between EM and transmitted-LM images in a CLEM experiment in terms of orientation, field of view, pixel resolution, imaged structures, image size and intensity. The width of the EM image is 45 µm and the width of the LM image is 153 µm.*

The different sections of this doctoral thesis work are focused in developing a method to automatize the registration of LM and EM images, focusing on the different areas of the problem: pre-alignment of the images, 2D multimodal image registration, feature detection, and the estimation of the optimal transformation between LM and EM images. We deal with a frequently implemented CLEM workflow described in Fig. 5, although other workflows also exist where our method could be exploitable.

Contributions

Various contributions to CLEM image registration and image analysis in general are presented in this thesis manuscript.

- A novel 2D CLEM registration method capable of overcoming multiple large differences in terms of appearance, pixel resolution, image size, image content, field of view and orientation between EM and LM images. Large initial discrepancies in intensity distribution and sample position (frequently expected in CLEM acquisitions) can be significantly diminished by mapping the multimodal CLEM images to a common modality and implementing a search stage of a pre-defined ROI. Using this approach, an overlay of EM and LM images can be generated, allowing the simultaneous analysis of functional, dynamic and ultrastructural information of biological samples. This contribution was published in the proceedings of one of the most important conferences in the biomedical image field, the 2016 IEEE International Symposium on Biomedical Imaging (ISBI) [157].

- An approach for 3D CLEM registration was also developed. This approach, resembling to the one developed for 2D CLEM, allows the alignment of 3D LM and EM image stacks while dealing with large discrepancies between the stacks. Our approach comprises two major steps, involving a two-dimensional pre-registration step followed by a three-dimensional registration refinement. This important tool can facilitate the analysis of 3D CLEM datasets, saving biologists significant time and effort. This contribution will be presented in the 2018 IEEE International Symposium on Biomedical Imaging (ISBI) [158].

- An important contribution in the microscopy imaging field as well as in the general image processing field is the multiscale spot segmentation method developed in this thesis work. By exploiting the LoG transform, the meaningful scales representing the sizes of the spots present in an image can be automatically selected. Multiple applications can be benefited from this approach, particularly the analysis of fluorescence microscopy images of live cells. This contribution was published in the proceedings of the 2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP) [156].

- A final contribution is the development of two robust motion model selection criteria for parametric image motion estimation. The first criterion is based on the Takeuchi information criterion while the second is based on the Fisher statistic. This new criteria allow to establish which motion model better describes the dominant motion in an image sequence. This provides valuable information exploitable in fields such as motion detection, image stitching, video stabilization, tracking, crowd analysis, among many others. We applied it to the 2D CLEM registration issue. The Fisher statistic-based part of the contribution was published in the proceedings of the 2016 IEEE International Conference on Image Processing (ICIP) [16].

Organization of the thesis manuscript

The structure of the thesis manuscript is organized as follows, knowing that a general introduction and a conclusion along with perspectives of the work envelop the five chapters.

Chapter 1. This chapter introduces existing methods for semi-automatic and automatic CLEM registration. The most commonly used approaches for medical image registration are addressed, since they are may be exploited also for biological images. Given the fact that EM and LM are different modalities, multimodality tools are reviewed.

Chapter 2. A detailed description of the methods we developed for 2D CLEM automated registration is presented in Chapter 2. Our 2D CLEM registration approach has two main parts: a pre-registration step and a fine registration step. A patch-based exhaustive search of a ROI pre-defined in one image is implemented in the other image to compute a pre-alignment of the EM and LM images. The pre-registration part involved four sub-steps: a common representation of the two microscopy modalities, the definition of the patch geometry, the search of the patch corresponding to the ROI and a coarse computation of the rotation between the CLEM images. Afterwards, we introduce the refine registration part, where a set of consecutive transformations are estimated to refine the alignment of the EM and LM images. Results on seven real CLEM experiments are presented, disclosing the outcome of each step in the registration workflow, along with a visual and quantitative evaluation of the resulting alignments.

Chapter 3. A novel method for the extraction of multiscale spots from images is presented in Chapter 3. This chapter introduces a spot detection method based on the Laplacian of Gaussian transform, which allows the estimation of a set of scales representing the spots sizes present in the processed image. A coarse-to-fine scheme for the segmentation of the spots is also defined. Extensive synthetic experiments to assess the performance of our method are reported, as well as results of the method on real images. Then, a robust point-based registration refinement for 2D CLEM is presented, exploiting this multiscale spot detection methods. Quantitative evaluation is included for all experiments.

Chapter 4. This chapter presents two robust motion model selection criteria for 2D image sequences. The goal is to select which model from a set of parametric motion models better describes the motion between two frames in an image sequence. Two criteria for the selection of motion models are presented, one based on the Takeuchi information criterion and the other one based on the Fisher statistic. The chapter includes

a thorough evaluation of the criteria on a large set of synthetic experiments and on real images. Finally, an implementation of the robust motion model selection approach is applied on 2D CLEM pre-registered images aiming to generate information regarding the most appropriate motion model that could be used for improving the final registration.

Chapter 5. An extension of CLEM registration to 3D is introduced in this final chapter. Here, an approach for registering EM and LM image stacks using an intensity-based method is presented. This approach, analogous to the work presented in Chapter 2, is based on a common representation of the EM and LM image stacks to reduce the complexity of the problem. A pre-registration step is defined to improve the initial misalignment of the stacks, followed by a 3D registration by estimating successively rigid and affine 3D transformations. Experimental results are included in this chapter, where visual and quantitative evaluation assess the performance of the method.

CHAPTER 1

OVERVIEW OF CORRELATIVE MICROSCOPY REGISTRATION METHODS

Correlative light and electron microscopy (CLEM) is a widely used technique that combines two different imaging modalities, electron microscopy (EM) and light microscopy (LM), in a to overcome their respective limitations. LM microscopy, and especially with fluorescence microscopy (FM), allows the analysis of live biological samples and the identification of labeled cellular structures. However, these techniques are limited by the resolution of the light microscope, preventing the imaging of fine structural details. Conversely, EM offers much-improved resolution, revealing crucial ultrastructural information at the expense of fixing the sample and with a limited field of view. Then, by fusing images from these two imaging techniques, it is possible to relate dynamic biological events to structural information at high resolution.

The analysis of CLEM images is frequently performed by aligning the resulting EM and LM images containing the same structures of interest. Such alignment, also called registration or overlay, can be defined as determining the transformation that maps all the elements in an image (known as source) to the space of another image (known as target). CLEM registration is not a straightforward task, as stressed in the introduction chapter of this thesis manuscript. Large differences in terms of appearance, image content, pixel resolution, image size, orientation and field of view complicate the location of the features shared between the two CLEM images and their registration. Currently, this challenging task involves strong manual interactions, requiring significant time, expertise and previous knowledge of the dataset and its acquisition. Yet, sometimes the resulting alignment may not be completely accurate or satisfactory. Then, the need for

an automatic registration method to assist in the analysis stage of CLEM is of great relevance.

1.1 Correlation methods

Alignment between LM and EM images is mostly performed at two different stages of the CLEM workflow: first, after acquisition of LM and after preparation of the sample for EM acquisition, where the sample is relocated manually by visually identifying grids or marks on the sample holder, fiducial markers added to the sample or features within the sample itself. The other stage where CLEM alignment is performed is at the final stage of the workflow, where the EM and LM images are registered and overlayed to be jointly analyzed. Typically, the alignment at any of these two stages is currently carried out with manual intervention of the biologist. For instance, to observe magnetotactic bacteria, the correlation and overlay between FM and EM images is done manually using ImageJ [173], with the EM image rotated and the FM image scaled and rendered partially transparent.

Some semi-automatic registration tools have been developed, most of them through the user selection of points of interest (points of reference, fluorescent markers, cell features, etc.).

1.1.1 Coarse CLEM alignment

Alignment of the EM and LM images is sometimes done during the CLEM acquisition. Marks on the sample holder are exploited for location of an LM-ROI to manually orient and align the sample for EM acquisition at low-magnification [55]. 2D coordinate systems exploiting marks and grids on the holder are designed to facilitate the location of ROIs after coarse manual alignment of the LM image with a low-magnification EM using grids on the holder [39].

1.1.2 Manual feature selection for semi-automatic registration

Most semi-automatic registration approaches are based on the manual selection of features visible on both modalities. A fluorescent fiducial-based correlation is developed to correlate fluorescence with ultrastructural information in [76]. Fluorescent microspheres are used as fiducial markers, which are manually picked and paired by the user

to correlate the EM and LM images. A geometric transformation is then computed through a point-based method using the paired fiducials. Several transformations are computed using a minimum of three fiducial pairs. To select the optimal transformation, the location of the predicted positions of the fiducials is compared to the actual location of the fiducials. Then, the transformation is applied to the LM image to align it with the EM image. The software for CLEM registration developed in this work has been used in other investigations [141].

In a similar approach, three reference points are chosen, for instance, the corners of an identified square of the holder grid, and a transformation is computed to relocate the LM features in the EM image [137]. A combination of different techniques is presented in [47], where microscopic X-ray computed tomography (microCT) is used along with LM and EM for the analysis of small biological samples. Using the software Amira®, the three modalities are observed within the same view, after manual registration of LM and EM images, and semi-automatic registration of LM and microCT images.

A comparison between the fusion of confocal laser scanning microscopy (CLSM) and focused ion beam - scanning electron microscopy (FIB-SEM), and the fusion of computed array tomography (CAT) and scanning electron microscopy shows that any of the two can replace ultrastuctural investigations performed by transmitted electron microscopy (TEM) in 3D studies [85]. Once again, manual pre-alignment is performed on the datasets, and the fine alignment is improved using Amira®.

A CLEM study over HeLa cells using FM and TEM in [164] achieves registration by manually scaling and applying rigid rotation over the FM in order to align it with the TEM image.

Besides the software presented in [76], other software have been developed for feature-base registration. *CorrelativeJ* [46] is an ImageJ plugin with two procedures. The first procedure consist of transferring the positions of features selected by the user from an FM image to an EM image to act as system coordinates. The second procedure consists of directly registering the transmitted light microscopy images and the transmitted electron microscopy images using the grid squares observed in both modalities as reference objects. *ecCLEM* [51] is a novel software that allows the user to manually select features from the LM and EM images to compute a transformation to register them. It contains an additional feature where an automatic registration of CLEM images is possible, based on the automatic extraction of features by an external software.

1.1.3 Automatic CLEM registration

Very few investigations have been undertaken so far from an image processing perspective to make the registration more automatic. A method for feature-based registration is proposed in [104], where structural information such as cell centers or nucleus is exploited as descriptive features to compute a transformation. The cell centers are detected using a gradient-based method, where a shifted Gaussian kernel and mean shift are applied onto a single-pass voting algorithm. Yet, the registration accuracy heavily depends on the presence of sufficiently salient structures and their correct segmentation, and the method is designed for pre-aligned CLEM images. In another study, an automated refined registration for 3D CLEM is proposed based on mutual information to compute a rigid transformation after a manual coarse registration using Amira® [83].

In [114], the image superposition is achieved by exploiting the auto-fluorescence which exhibits strong enough contours. Following a template-based approach, affine transformations are then applied until the one supplying the highest cross-correlation score is found, first with an EM image at low magnification, then for progressively higher magnifications. A very different approach is proposed in [20]. It aims at transforming the two-modality issue into a single modality image registration problem. They introduce the so-called paradigm of image analogies and exploit a sparse representation model to obtain these image analogies. Convincing results are reported but this scheme requires a prior supervised learning stage which is not always affordable.

1.2 Multimodal biomedical image registration

The study of strategies suitable for generating automatic methods for CLEM registration is closely related to the study of medical image registration. Medical imaging comprises multiple imaging modalities, such as X-ray, ultrasound (US), computed tomography (CT), magnetic resonance imaging (MRI), positron emission tomography (PET), single-photon emission computed tomography (SPECT), functional MRI (fMRI), among others. The information provided by each medical imaging technique is different, therefore, it is logical that combining two different modalities will offer very unique and valuable insight, useful for medical diagnosis, treatment or surgery. Hence, the two images must be aligned in order to analyze them and gather knowledge from their combination [54].

Roughly speaking, medical, biological or traditional image registration methods can be classified as intensity-based methods and feature-based methods [178].

1.2.1 Feature-based registration

Feature-based registration methods involve the extraction of corresponding 2D or 3D features in the images to be aligned, inferring the transformation needed. Usually, the features are represented by the coordinates of representative points, which can be either the centers of mass, the geometrical centers, corners, etc. The usual process for feature-based registration involves three main stages: feature detection, feature matching, and estimation of the transformation. Each step of the process has their own challenges depending on the kind of images to be registered, and selecting the right tool for each one is crucial. Feature-based methods are typically fast while being able of handling large initial misregistrations, however, they are highly dependent on the correct extraction and matching of significant features.

The extraction of reliable features from biomedical images is a complex task with many active investigations. In the computer vision field, countless studies in the extraction of features currently exist. A review of the detection of key points in images can be found in [140, 160]. In biomedical image processing applications, feature detection methods are usually developed for specific problems, but the basis for most of the developed methods comes from computer vision techniques, specially from object detection and segmentation methods. Detection of blob-like features has been proposed by [82], where a scale-space is built by applying a Laplacian of Gaussian (LoG) transform to an image. Another approach for detecting blob-like features is by thresholding with different values and tracking the stability of regions in the image through the process [93].

A widely used image descriptor, denominated scale invariant feature transform (SIFT), was proposed by [84]. In SIFT, feature points are detected by exploiting a scale-space representation built using the Difference of Gaussians. Many SIFT variants have been proposed [42], which have been applied to many different image processing problems, both in computer vision and biomedical imaging [24, 175]. Several detection and segmentation methods for different medical imaging problems can be found in [90] and [107]. In [6], a survey on automatic segmentation evaluates their performance on optical images of mammalian cells. However, most detection and extraction of features of interest on medical images is tailored to specific organs and structures that are not applicable for CLEM registration. A survey and experimental comparison of spot detectors dedicated to bioimaging can be found in [146]

Once features are extracted from an image, they must be matched. The aim is to assign every feature from one image its corresponding feature from the other image. The correspondence between points usually relies on the closeness of their locations. In addition, geometrical constraints can be established to avoid mismatching [150]. An evaluation of spot detection methods and matching strategies can be found in [99].

Finally, after detection and matching of features, the transformation can be computed. Many algorithms incorporate the matching and computation of the transformation in a single scheme. One of the most well-known approaches for inferring the matching and transformation is the iterative closest point method (ICP) [12], where a correspondence is established based on a closest neighbor principle and the transformation can be computed based on the pairings. The HAMMER algorithm [144] computes the matching of features based on a local search for the best matching. A robust method for feature matching and computation of the transformation is the Random Sample Consensus (RANSAC) [34]. This method computes an optimal transformation by implementing a voting scheme, given a set of features containing inliers and outliers. RANSAC-based methods for segmentation and registration have been vastly investigated. For instance, serial confocal microscopy images of axons are registered using RANSAC in [56]. A multimodality registration of human retina captured using Fundus imaging and spectral domain-optical coherence tomography (SD-OCT), also exploits RANSAC and histograms of oriented gradients (HOG) [100].

Although feature-based registration methods are expected to be robust to initial conditions and large misregistrations [150], its application on CLEM image pairs may not be sufficient to successfully register them, due to the large discrepancies between the orientation, field of view and content of both images, which complicate the correct detection and matching of significant features. At the beginning of our CLEM investigation, we tested several widely used methods for feature detection on two sets of 2D CLEM images with diverse results. SIFT detected a large number of features on both EM and LM images, complicating the matching necessary for computing the transformation. Harris corner detector [48], a well known feature detector based on spatial derivatives, is invariant to rotation, scale and illumination variation. When applied to CLEM images, it was able to detect spots of interest along with many other spurious detections. Similar results were obtained with the FAST (features from accelerated segment test) method [127], which analyzes the intensities of the pixels around a pixel $\mathbf{p}$ to determine if $\mathbf{p}$ is a corner. Speed-Up Robust Features (SURF) [10], a feature detector inspired by SIFT, detects a large number of features on EM but failed to detect all the features of interest in LM images. All these feature detection methods would require an additional criterion to discriminate detections when applied to 2D CLEM images.

1.2.2 Intensity-based registration

Registration of two images based on their intensity values is the process of searching iteratively the values of the parameters of the transformation by optimizing some similarity measure. Intensity-based methods take into account all the information in an image, consequently increasing the computation time.

Selecting the similarity measure is the most important part of an intensity-based registration method, since it must be tailored to the intensity relation between the images to be registered. Similarity measures can be divided on two main fields: those suitable for images from a single modality, and those designed for multimodality problems.

In single modality image registration, the corresponding structures in the images are expected to have similar intensity values. Then, methods such as the sum of squared differences (SSD) or the sum of absolute differences (SAD) can be used as similarity criteria. Other widely used similarity measures are cross correlation (CCor) and correlation coefficient (CCoef) [150]. However, simply using similarity metrics may not be sufficient to obtain satisfactory registration. Then, the incorporation of attribute-based methods and regularization constraints allow to improve registration. A regularization scheme for nonrigid image registration to account for the preservation of topology is presented in [79], where the proposed scheme is applied to MRI brain images. In microscopy image registration, other problems arise due to the specific characteristics of the acquisition. For instance, when implementing intensity-based registration methods, it is important to account for the anisotropic image formation process. An intensity-based registration method for 3D fluorescence microscopy is introduced in [23], where the source and target images are re-blurred with transformed forms of the point-spread-function (PSF), mapping the images to a more comparable representation in order to optimize a mean squared intensity difference to compute the transformation to register them.

Multimodal registration is more challenging, and the selection of the appropriate methods and similarity criteria is not simple. Two main ways to deal with multimodality are commonly adopted. The first one is the use of information theory for similarity computation. Mutual information (MI) [88, 166] is one of the most important similarity measures designed for multimodality registration, since it does not assume any relationship between the images intensities. Ample information about MI and its many variants and applications can be found in [117]. Multiple techniques incorporate MI with spatial information to create more robust metrics [116, 129, 130, 152]. Other information theory-based metrics, such as Renyi Entropy [120] have been adapted for image registration [49].

The second approach is reducing the multimodality to a single modality. It is possible to apply an intensity mapping function to a source image to create a new image with similar intensity characteristics to the target image. Then, single modality similarity metrics can be used for registration. Rigid registration of US images mapped to the intensity space of MRI images is presented in [123], and in [170]. US images are mapped to CT intensities for rigid and affine registration.

A similar approach is to create new representations of the source and target images such that, although the target and source image are very different from each other, their new representation are quite similar. One way to implement this is by applying scale-space derivatives to both images in order to generate new representations that highlight their edges [33, 89]. This way, registration can be achieved by using single modality similarity metrics such as cross-correlation.

Intensity-based methods are sensitive to large initial misalignments. Therefore, applying a pre-registration transformation to decrease the initial misalignment between the source and target images, will increase the likelihood of finding an optimal final transformation [107].

The existing semi-automatic CLEM registration methods exploit some of the strategies previously listed: feature-based registration [104], which is specific to certain types of CLEM images and dependent on the presence and correct extraction of cell features; mapping of multimodality to single modality [20], which requires a training step which is not always affordable; cross-correlation of the FM and auto-fluorescence in the EM [114], relying on the existence of highly similar template-like features; and MI for computing a rigid transformation [83], which, as most of these methods, is applied after manual pre-alignment of the image, which can be complicated and time-consuming. In some cases, the CLEM protocol has to be specifically modified to include this manual step.

Then, we observe the need for a fully automatic CLEM registration framework capable of dealing with the significant differences between EM and LM images, aiming to eliminate manual intervention in both the pre-alignment of the images and estimation of a fine registration. In the next chapters we will explore ways to adapt different multimodal image registration strategies to build an automatic CLEM registration method. We will validate each developed approach on real 2D and 3D CLEM images, demonstrating the efficacy and potential use of our method.

CHAPTER 2

2D CORRELATIVE LIGHT ELECTRON MICROSCOPY REGISTRATION

Correlative light-electron microscopy (CLEM) allows to relate dynamic and functional information provided by different light microscopy (LM) techniques, with structural information given by electron microscopy (EM) methods for a better understanding of cellular and subcellular mechanisms. Registration has many occurrences in 2D/2D, 2D/3D and 3D/3D CLEM frameworks, depending on the subject to be studied. However, as previously mentioned, LM and EM images are of very different size, spatial resolution, field of view, content and appearance. Therefore, registration of LM and EM modalities is a timely, important but difficult open problem, which still requires some manual assistance in current applications. In this chapter, we will focus on 2D CLEM registration.

As the main objective of our thesis work, we have designed an original automated CLEM search-and-registration method. The search stage involves a common representation of EM and LM images with an adaptive associated scale (or blurring), the specification of the search patch geometry and the appropriate descriptors and similarity criterion for the EM/LM patch search. The registration stage concerns the use of multimodality registration techniques to align and overlay the EM and LM images, allowing the simultaneous analysis of the information provided by both techniques. The efficiency of our method is demonstrated on a set of real and diverse CLEM images.

2.1 Introduction

Light microscopy and electron microscopy are two fundamental investigation tools in cell biology. LM imaging allows the visualization of live cellular and subcellular behaviors but at relatively low spatial resolution, although recently developed super-resolution LM methods have been improving the attainable resolution [59], such as stimulated emission depletion (STED) microscopy [52] and photoactivated localization microscopy [13].

Different probes, dyes, markers and fluorescent labels are usually employed to make the elements of interest visible. In certain LM techniques, such as fluorescent light microscopy (FLM), the unlabeled structures are usually not perceivable and even more so not recognizable. Phase-contrast LM images are frequently acquired along with FLM images to help in the visualization and localization of unlabeled structures. LM is usually acquired on live samples, allowing the generation of temporal sequences in both two and three dimensions. However, LM is also exploitable on fixed samples, as long as the labels are correctly preserved for techniques such as FLM.

In contrast, EM imaging enables the examination of the whole cellular ultrastructure at very high spatial resolution, but at the cost of fixating or freezing the sample. The main downside of EM is its inherent inability to perform live-cell imaging, since the fixation or freezing stage will cease normal life functions in the sample.
Correlative light-electron microscopy (CLEM) is then of key interest to combine the respective complementary strengths of LM and EM. However, managing to do it is not that straightforward. In recent years, research has been active in biology and microscopy to design efficient setups and protocols to achieve effective consecutive LM and EM acquisitions of the same sample [50, 114, 137, 151].

As described in Chapter 1, broadly speaking, the usual workflow is to first acquire a LM image sequence of the living sample, and after detecting and locating the event of interest, the sample is then chemically fixated or frozen to acquire the EM images. Usually, another LM acquisition is performed after fixation or freezing, acting as an intermediate between living LM and EM acquisitions. This additional step is helpful considering that the freezing or fixation process can alter the sample, making the correlation between living LM and EM images more difficult.

Registering LM and EM images stands as a very difficult problem due to the big differences in spatial resolution, field of view and appearance. CLEM registration may respond to different needs: guiding the EM acquisition with the LM location of interest, retracing the object of interest from one image to the other, overlaying the two images for simultaneous analysis. CLEM registration usually requires manual interventions and the use of artificial landmarks (fiducial markers as beads, marker grids) at different stages of the overall CLEM workflow. The manual relocation and registration of LM and EM images is time consuming and demanding, therefore, it is important to develop accurate automated registration tools [55].

We present an original automated approach to search for the EM (or LM) patch corresponding to the LM (or EM) region of interest (ROI) and to register them. It comprises two main stages: a pre-registration stage and a registration refinement stage. The pre-registration can be divided in two main components: the shift given by the retracing of the patch containing the ROI, and the rotation computed between the matched EM and LM patches. The search stage involves a common representation for the LM and EM images with an adaptive associated scale (or blurring), the determination of the geometry of the search patch and the similarity criterion for the search. The second stage concerns a multimodality registration process of the selected patch and the ROI. Our method was tested on a series of real CLEM images with highly dissimilar contents, demonstrating its efficiency and accuracy on different scenarios.

2.2 Automated 2D CLEM Registration

We propose a novel method to achieve automatic pairing and registration between 2D LM and EM images. This method is able to manage significant differences in appearance, field of view, spatial resolution and content scale between both modalities. As in any registration process, there is a target image I_T and a source I_S image. Initially, we explored the 2D CLEM registration problem by setting the EM image as target, and the LM image as source for the retracing step [157]. However, since then we have also explored the application of our method using LM image as target and EM image as source. The overall workflow is divided as follows:

1. Pre-registration

    (a) Common LoG representation of 2D EM and LM images with scale adaptation.

(b) Determination of the search-patch geometry given a ROI and a resolution ratio between EM and LM.

(c) Search of ROI extracted from the source image in the target image.

(d) Computation of the rotation angle between the ROI and the paired patch in the other image.

2. Registration refinement

(a) Estimation of rigid transformation between paired patch in the target image and ROI in the source image.

(b) Estimation of affine transformation between patch and ROI.

The diagram of Figure 2.1 summarizes the overall framework and the interactions between the listed steps of the 2D CLEM registration process. Each step will be described in the following subsections.

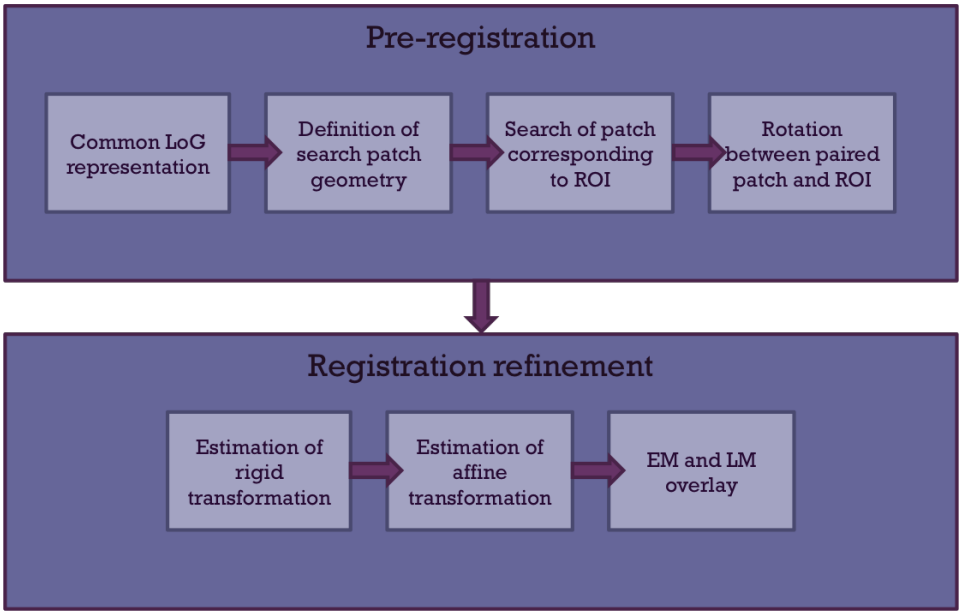


Fig. 2.1: *2D CLEM registration workflow.*

2.2.1 Common LOG representation for EM and LM images

The multimodality nature of the CLEM images causes a great difference in terms of appearance, content and resolution between LM and EM images, thus the selection of an appropriate matching criterion for the search is a hard task. When facing multimodality registration problems, there are usually two main strategies: the use of information

theory-based methods [117, 123] and the reduction of the multimodality problem to a "single-modality" one [150]. Both strategies have been explored during our thesis work. We adopted the latter for the first pre-registration stage. The former will be used in the second registration stage as explained in subsection 2.2.3. Indeed, the initial EM and LM images are likely to be badly misregistered, making it extremely difficult to directly estimate any kind of transformation between them.

By boiling down the multimodality issue to something equivalent to a single-modality issue, a simplification of the problem is expected. There are two ways to achieve this reduction. One is to simulate one modality from another based on a learning stage, in such a way that both images are expressed in the modality of one of them, as illustrated in [20]. The other way is to map both images to a third common modality, hoping that structures present in both images will provide the necessary correspondences to achieve their registration. We adapt this idea and propose to map CLEM images to a common representation, which will allow us to retrace the ROI and roughly register EM and LM images.

The idea is to find an intensity transformation of both 2D LM and EM images which produces images of closer appearance. When mapping two different modalities to a single one, it is advantageous to apply a filter that can help to extract geometrical information. The Laplacian of Gaussian (LoG) filter is a good candidate, since it is well known for suppressing local linear intensity variations and enhancing high spatial frequencies, while being linear, simple to apply, producing scalar values and enabling scale adaptation. LoG successively applies two filters: a Gaussian filter of variance σ^2 to smooth the image, and a Laplacian filter to enhance contrast. This filter was employed as a second-order edge detector [9]. It is also the most commonly used spot detector [7]. The LoG filter of an image I with domain Ω can be written as follows:

$$\forall p \in \Omega, \qquad \mathcal{L}_\sigma(I(p)) = \Delta(G_\sigma * I(p)), \tag{2.1}$$

Where Δ denotes the Laplacian operator and G_σ the Gaussian filter of standard deviation σ.

To compute the LoG representation of LM and EM images and select the right scales s_{LM} and s_{EM} corresponding to the Gaussian variance σ^2, we use the first stage of the spot detection method ATLAS [7]. ATLAS automatically detects the characteristic scale of the objects of interest from a set of scales S, by exploiting the Lindeberg's scale-space approach [81]. The scale-space representation L_I of an image I is defined as

$$\forall (p,s) \in \Omega \times S, L_I(p,s) = K_s * K_s^\top * \nabla^2 I(p) \tag{2.2}$$

where K_s is a Gaussian kernel of scale s, $\top$ denotes the transposition operation, ∇^2 is the Laplacian operator. Then, a normalized discrete LoG filter H_I is adopted as follows:

$$\begin{aligned} \forall (p,s) \in \Omega \times S, \quad H_I(p,s) &= \alpha(s)\nabla^2 L_I(p,s) \\ &= \alpha(s)(K_s * K_s^{\top} * \nabla^2 I)(p) \end{aligned} \tag{2.3}$$

where α is a normalization function. ATLAS follows the concept of blob given in [82]. A blob $b = (p_b, s_b) \in \Omega \times S$ is defined as as a local minimum (conversely maximum, depending on the problem) of H_I:

$$\forall (p,s) \in \nu(b) \quad H_I(p_b, s_b) \leq H_I(p,s) \tag{2.4}$$

where $\nu(b)$ is a neighborhood in $\Omega \times S$. Let $\mathcal{B}_s(I)$ be the set of blobs at scale s of image I, the blob density is then given by:

$$\forall s \in S, \quad \rho_s(I) = \frac{\|\mathcal{B}_s(I)\|}{\mathcal{A}(I)} \tag{2.5}$$

where $\|\|$ denotes the cardinality and $\mathcal{A}(I) = |\Omega|$ is the area of I in square pixels.

The scale at which the number of blobs is the highest is the sought one, but we need also to counterbalance noise influence. Assuming that the input image I is affected by additive Gaussian noise, then I can be separated into a noise-free image I_0 containing only structures and a noise image ε:

$$I = I_0 + \varepsilon. \tag{2.6}$$

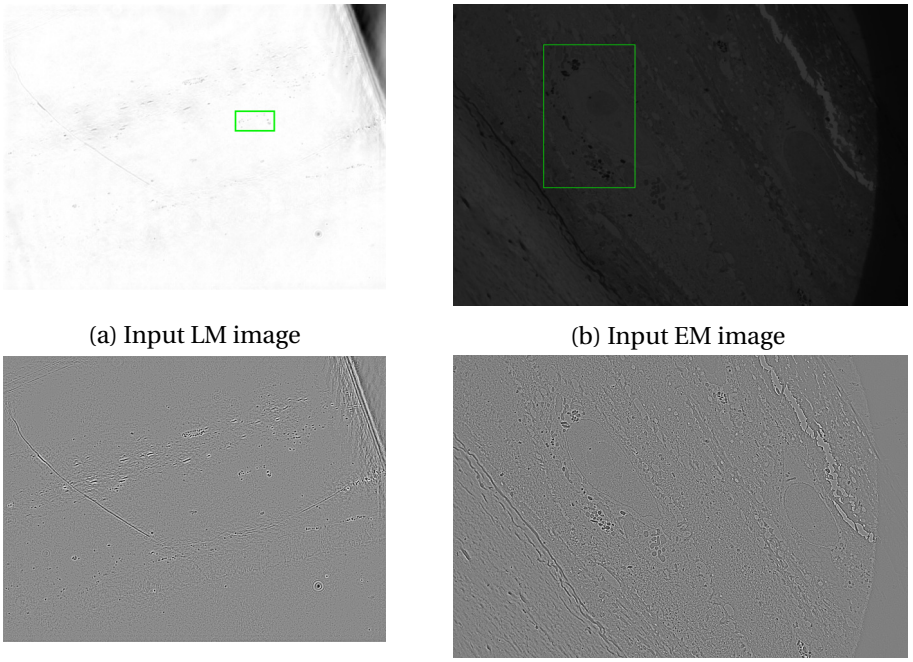
If I_0 does not contain any structures, the blob density at each scale s would be $\rho_s(\varepsilon)$. Consequently, the scale s at which the difference between $\rho_s(I)$ and $\rho_s(\varepsilon)$ is the largest will be the one where more structures are present in I. Then, we adopt the following criterion [7]:

$$s^{\star} = \arg\max_{s \in S} \frac{\rho_s(I)}{\rho_s(\varepsilon)} \tag{2.7}$$

The set of scales S is defined as $S = \{s_0 r^n, n \in [1, \mathbb{N}]\}$, where s_0 is a strictly positive real number and r is an odd integer.

The ability of fixing σ^2 values adapted to the LM and EM content respectively, will allow us to mitigate the difference in content scale of LM and EM images. The selected scale is expected to be higher for the EM image, yielding a stronger blurring of the EM image. These scales should not be confused with the pixel resolution ratio between the EM and LM images.

The impact the application of the LoG filter has over a set of CLEM images is illustrated in Figure 2.2. The difference between the intensities of LM and EM images (Figures 2.2a and 2.2b, respectively) is visibly large, while their LoG representations exhibit a much closer appearance (Figures 2.2c and 2.2d).



(a) Input LM image

(b) Input EM image

(c) LoG of input LM image

(d) LoG of input EM image

Fig. 2.2: *Impact of the LoG filter applied to a CLEM image set.*

In the experiment presented in Fig.2.2, the ROIs are enclosed by a green frame. By observing the shape of the rectangles containing the ROI and the distribution of the structures inside of it, we can infer that the initially acquired images are greatly misregistered. The location of the respective ROIs are far apart from each other and their orientations are very different, indicating a big rotation between both images. In any registration problem where such big misalignments are present a pre-registration step is required.

To pre-align the LM and EM images, we developed a method based on the matching of the ROI from the source image with a patch in the target image containing the same structures. First, we need to delineate the ROI in the LoG-source image. This ROI is supposed to encompass a dynamic event of interest, only observable through LM, or it can also contain meaningful structures highly detailed by EM. In a given application, such an event could be automatically detected using an appropriate method. In our study we assume that the ROI is already available and it can be delineated it either on the LM or the EM image, depending on the subject of the study for which they were acquired.

2.2.2 Searching LoG-I_S-ROI in LoG-I_T

To automatically retrace the LoG-I_S-ROI in the LoG-I_T image, we carry out a patch-based exhaustive search. There is a considerable geometrical transformation (combination of rotation, shift, magnification,...) between LM and EM images. Hence, we resort to histograms since they are invariant to rotation and scale. We designed two histogram-based methods to find the patch in the LoG-I_T image corresponding to the LoG-I_S-ROI. The first one involves the LoG-value histograms and the second one exploits Local Directional Pattern (LDP) [61].

We thoroughly tested several histogram distances to measure the similarity between LoG-I_S-ROI and LoG-I_T patch: KL-divergence, Jeffrey-divergence, cosine similarity, histogram intersection, χ^2 distance, Bhattacharyya distance and Kolmogorov-Smirnov distance [22]. The best results were obtained with the cosine similarity and the histogram intersection, respectively defined as:

$$d_{cos}(H_{LM}, H_{EM}) = \frac{\sum_{v=1}^{V} H_{LM}(v) H_{EM}(v)}{\sqrt{\sum_{v=1}^{V} H_{LM}^2(v)} \sqrt{\sum_{v=1}^{V} H_{EM}^2(v)}},$$
$$d_{hi}(H_{LM}, H_{EM}) = \sum_{v=1}^{V} \min(H_{LM}(v), H_{EM}(v)), \tag{2.8}$$

where H_{LM} is the histogram computed over the LoG-I_S-ROI, and H_{EM} is the histogram computed on the LoG-I_T patch, v denotes the histogram bin and V the total number of bins. Histogram values are normalized over $[0,1]$.

Histograms of LoG values can be computed as explained above. A more elaborate descriptor may be necessary to capture more information. LDP is a local descriptor, used primarily for recognizing human faces [61]. It creates a pattern from an eight-bit binary code at each pixel of the input image I. Using Kirsch masks defined in Fig.2.3, the code is calculated by comparing the edge response values of a pixel in eight different orientations. The pattern records only the k most significant directions (with the corresponding bits) and the rest of the bits are set to 0. With k=3, the LDP may generate up to 56 distinct patterns, creating an encoded image $\mathcal{D}_I$. We apply this operation to the LoG-I_S-ROI and the LoG-I_T patch, producing encoded subimages, respectively denoted $\mathcal{D}_{I_S}$ and $\mathcal{D}_{I_T}$, and histograms of these values are computed. We use the same histogram distances as in (2.8).

Since no a priori information is available on the possible location of the patch corresponding to the source ROI in the target image and the geometric transformation of the

$$\underset{\text{M1}}{\begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix}} \quad \underset{\text{M2}}{\begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix}} \quad \underset{\text{M3}}{\begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}} \quad \underset{\text{M4}}{\begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix}}$$

$$\underset{\text{M5}}{\begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix}} \quad \underset{\text{M6}}{\begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix}} \quad \underset{\text{M7}}{\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix}} \quad \underset{\text{M8}}{\begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix}}$$

Fig. 2.3: *Kirsch masks used to create a pattern for LDP.*

ROI locations between EM and LM images, can be considerable, there is no straightforward way to reduce the search space. Therefore, we implement a patch-based exhaustive search to explore all potential locations in the LoG-I_T image. Initially, to specify the patch to be searched in the LoG-I_T image, we started from the rectangular shape of the I_S-ROI of length l and width w. Then, we took into account the (known) resolution ratio η between the EM and LM images. We conducted two searches with two orientations of LoG patches (with a very coarse quantization of the rotation angle at this stage, but easy to implement): horizontal rectangle of dimensions $r\,l \times r\,w$ and vertical rectangle of dimensions $r\,w \times r\,l$ (i.e., after a $\pi/2$ rotation). The selected patch was the optimal one over the two searches. However, in a later version, we implemented the slightly different process to establish the geometry of the search patch. Now, the patch is defined as a disk, in order to avoid the double exhaustive search for the shift.

2.2.3 Mutual information-based for image registration

Information theory-based methods are commonly adopted when dealing with multi-modality registration problems in medical image registration problems [117]. Since there is not physical relationship between LM and EM intensities, intensity-based are not really appropriate to compute the geometric transformation between the ROI and the selected patch (SP) of the target image. We have explore Mutual Information (MI) [25, 171] to determine if it is a method suitable to our problem, given that MI does not assume any explicit one-to-one relationship between the intensities of the images. MI is a measure of the mutual dependence between two random variables.

To define MI it is necessary to first define entropy. Entropy is a measure of information. It is low when a distribution has a few sharply defined dominant peaks, and it is high when all outcomes have the same chance of occurring. An image with low entropy

will have a small variation of its intensities, indicating that the information it contains is poor. On the other hand, an image rich in information will have a dispersed distribution of its intensity values and in consequence, a high entropy value. The entropy of two images or joint entropy is defined as a 2D joint histogram combining the gray values in each of the two images for all corresponding points. When two images are correctly aligned, the corresponding structures overlap and the join histogram will display clusters corresponding to their gray values. This implies that the dispersion in the distribution will be low, meaning the joint entropy of the two images is low when two images are correctly registered [117].

To measure the dispersion of the joint histogram of two random variables A and B with marginal probability distributions $p(a)$ and $p(b)$ and joint probability distribution $p(a,b)$, the approach most commonly used is the Shannon entropy:

$$-\sum_{a,b} p(a,b)\log p(a,b) \tag{2.9}$$

MI can be explained as the measure of the amount of information two images share. Let $\mathrm{MI}(I_A, I_B)$ be the mutual information between images I_A and I_B expressed in terms of the entropy of the images as:

$$\mathrm{MI}(I_A, I_B) = H(I_A) + H(I_B) - H(I_A, I_B) \tag{2.10}$$

where $H(I_A)$ and $H(I_B)$ are the entropies of images I_A and I_B, and $H(I_A, I_B)$ their joint entropy. The goal is then to maximize $\mathrm{MI}(I_A, I_B)$, that is, minimizing the joint entropy $H(I_A, I_B)$. An equivalent representation of MI uses the Kullback-Leibler distance. Then, the mutual information of images I_A and I_B in terms of their distribution is given by:

$$\mathrm{MI}(I_A, I_B) = \sum_{a,b} p(a,b)\log\frac{p(a,b)}{p(a)p(b)} \tag{2.11}$$

where a and and b are intensity values of images I_A and I_B, respectively.

This definition measures the distance between the joint distribution of the gray values $p(a,b)$ and the product of the distributions of the images $p(a)$ and $p(b)$. When the images are correctly aligned the maximum value of MI is reached.

The number of bins used to compute the histograms for MI can be set arbitrarily or using some criterion. We decided to implement Sturge's rule [153], since it is a well established method to estimate the number of bins, known to be reliable when the number of samples is large. This rule is defined by:

$$k = [1 + \log_2 n] \tag{2.12}$$

where k is the number of bins and n is the number of samples.

2.2.4 MI-based multi-step image registration

Following the search step, several geometrical transformations are computed in turn, in order to complete the registration between the EM and LM images. These transformations are computed between the ROI and the patch selected by the search step in their original intensity values. First, the orientation between the two CLEM images has to be estimated, since the CLEM images are expected to be largely misaligned. A rigid transformation is first computed and then a non-rigid transformation to achieve a faster convergence and robustness [107].

1. **Rotation and pre-registration**. A large difference in between the orientations of the EM and LM images is commonly expected in CLEM registration problems, as illustrated in Fig.2.2. Since there is no a priori information regarding this issue, it is necessary to implement an additional exhaustive step to estimate the rotation angle between the ROI and the patch selected by the search step. First, the selected patch is magnified according to the known resolution ratio η between the LM and EM images. The rotation angle α is inferred by applying a rotation centered on the selected patch with different angle values and computing the MI value between the patch and the rotated ROI. For all experiments, the tested angles in degrees are $\alpha \in \{0°, 5°, 10°, \ldots, 355°\}$. The angle at which the MI is maximum is selected and it is used to apply the rotation to the source image, along with the known resolution ratio, after shifting the center of the ROI $\mathbf{c}_{ROI} = (xc_{ROI}, yc_{ROI})$ to the selected patch $\mathbf{c}_{SP} = (xc_{SP}, yc_{SP})$. The center coordinates will be expressed in homogeneous coordinates. The first part of the transformation is related to the shift given by the search step. It relates the center of the ROI $\mathbf{c}_{ROI}$ with the center $\mathbf{c}_{SP}$ of the SP and is given by:

$$\mathbf{T} = \mathbf{c}_{ROI} - \mathbf{c}_{SP} \tag{2.13}$$

The second part of the transformation is given by the rotation angle α and the size ratio η, as defined:

$$\mathbf{SR} = \begin{bmatrix} \eta\cos\alpha & \eta\sin\alpha & 0 \\ -\eta\sin\alpha & \eta\cos\alpha & 0 \\ 0 & 0 & 1 \end{bmatrix} \tag{2.14}$$

Then, any points $\mathbf{p} = [x \quad y \quad 1]^\top$ of the source image I_S will be mapped its corresponding point in the target image I_T by:

$$\mathbf{p}' = \mathbf{SRp} + \mathbf{T} \tag{2.15}$$

where $\mathbf{p}' = [x' \quad y' \quad 1]^\top$. This transformation acts as a pre-registration step. The following transformations will be computed between the selected patch and the transformed ROI.

2. **Rigid transformation**. A rigid transformation comprising a 2D rotation and 2D translation is computed between the pre-registered patch and the ROI. The rigid transformation is defined as:

$$G_R(\mathbf{p}, \mathbf{a}) = \begin{bmatrix} \cos a_0 & \sin a_0 & a_1 \\ -\sin a_0 & \cos a_0 & a_2 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} x\cos a_0 + y\sin a_0 + a_1 \\ -x\sin a_0 + y\cos a_0 + a_2 \\ 1 \end{bmatrix} \tag{2.16}$$

where $\mathbf{a} = [a_0 \quad a_1 \quad a_2]$ are the parameters of the rigid transformation and $\mathbf{p} = [x \quad y \quad 1]^\top$. Now, we have estimated a more precise rotation while refining the translation, since we previously used discrete values and now the estimation is performed using real values.

3. **Affine transformation**. A 6-parameter 2D affine transformation is then computed after applying the rigid transformation. This transformation is defined as $\mathbf{p}' = \mathbf{Ap}$, where $\mathbf{A}$ is a 3×3 matrix containing parameters representing rotation, scaling and shearing transformations, and $\mathbf{b}$ is a 2D translation. The affine transformation preserves the parallelism of lines, but not their lengths or their angles, introducing a given level of non rigid transformation. The affine transformation is defined by:

$$A = \begin{bmatrix} a_0 & a_1 & a_2 \\ a_3 & a_4 & a_5 \\ 0 & 0 & 1 \end{bmatrix}, \tag{2.17}$$

The parameters of the affine transformation are given by the vector $\mathbf{a} = [a_0 \quad a_1 \quad \dots \quad a_5]$.

A multi-resolution workflow was implemented in ITK [66] for the estimation of the rigid and affine transformations using MI as defined in [95] and a gradient descent optimization. Multi-resolution or pyramidal strategies are often exploited with the aim of achieving a faster convergence while avoiding local optima. Usually, pyramidal strategies down-sample the source and target images at different scales. The registration is computed from the lowest to the highest resolution images, using the transformation found in the previous step as the initial transformation for the next step. Although MI is by definition a maximization problem, the MI-based registration presented in [95] employs the negative of MI to transform the registration into a minimization problem.

Once the rigid and affine registrations are estimated, we apply them to the source image I_S to overlay it over the target image I_T around the ROI area.

Our full 2D registration framework was tested on a set of real CLEM image pairs and the results will be presented in the next section.

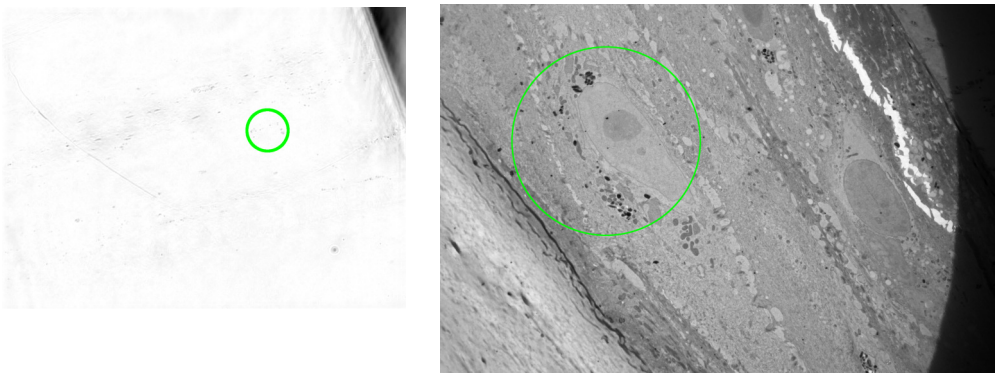
2.3 Experimental results

We tested our method on seven sets of LM and EM images acquired at Institut Curie[1]. Reconstructed epidermis samples are imaged following the procol described in [51]. After live LM acquisition, a quick freeze substitution was conducted, allowing the preservation of fluorescence in the sample. Afterwards, the sample is sectioned. The LM images are acquired with epifluorescence Nikon Te2000, in multiple different channels, including transmitted light. The section was then post-stained with Lead Citrate and imaged by a Tecnai Spirit FEI transmission electron microscopy (TEM) [50].

For the following series of experiments, we use the transmitted light channel, since it contains the most exploitable information to correlate the image content between EM and LM images. A throughout demonstration of the performance of every stage of our method is reported for all the experiments. To evaluate the registration accuracy, EM and LM spots are manually selected and paired from the overlay images, and the distance between the corresponding pairs of spots is calculated. The mean, median, standard deviation (SD), maximum and minimum registration errors are listed for each experiment.

In the first experiment, the studied cell elements are endosomes, which are dark, dense biological material stored in round bags, visible both in transmitted light and EM. The LM image depicts several cells while the EM image also contains several cells at different magnification and orientation. This first experiment involves a 1392×1040 LM image and a 4008×2664 EM image. The input image pair is displayed in Fig.2.4. In this experiment, the ROI is delineated on the LM image to be searched in the EM image. We have framed the LM-ROI and the ground-truth EM corresponding patch in green, both being defined by a biologist. The difference between LM and EM image contents is remarkable, both in appearance, size, content and orientation. If we observe the aligned dark semi-circular structures inside the LM-ROI (Fig. 5.1a), and then analyze the corresponding EM ground truth patch, we can visualize the same alignment of dark structures (Fig. 5.1b) with a far different orientation, indicating the presence of a large rotation between the EM and LM images.

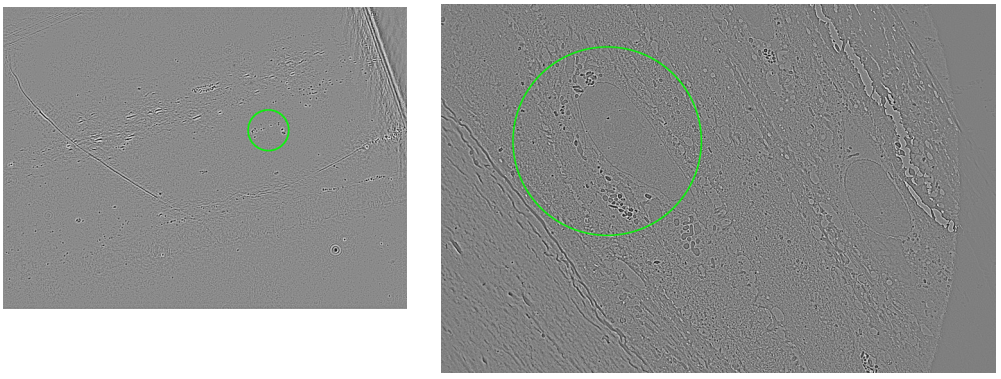
[1] We thank X. Heiligenstein from UMR 144, Institut Curie, Paris, for providing us the images and the associated explanations.



(a) Input LM image (b) Input EM image

Fig. 2.4: *Experiment 1. Input LM and EM images with LM-ROI and ground-truth EM patch marked with green circles.*

As explained in the description of the method, we aim to map both EM and LM images to a common LoG representation that minimizes the difference of appearance between them. We use ATLAS to get the appropriate scale for the LoG transform of the EM and LM images. For this first experiment, they are respectively estimated as $s_{EM} = 8.91$ and $s_{LM} = 2.48$. The LoG representation can be observed in Fig. 2.40 where, as expected, the difference of appearance is significantly decreased.



(a) LoG of input LM image (b) LoG of input EM image

Fig. 2.5: *Experiment 1. LoG representation of input images.*

For the first shift and rotation steps with exhaustive search, the source is the LM image and the target the EM image, since the ROI is defined in the LM image. To locate the EM patch corresponding to the LM-ROI, we set the EM searched patch with a disk shape, with a radius defined by the radius of the LM-ROI magnified by the resolution ratio η, which is available as a metadata in the image files. As the LM pixel resolution is 110 nm, while the EM one is 11.3 nm, we get $\eta = 9.7$ to define the EM patch dimensions.

In the search stage, we applied a patch-based exhaustive search to locate the patch in LoG-EM image corresponding to the LoG-LM-ROI, by comparing a patch from the LoG-EM image and the LoG-ROI with both LoG-value histograms and LDP. We first tested EM patch locations at intervals of 2.3% of the x-dimension and 3.6% of the y-dimension (with 95% overlap). Then, we refined the search around the primarily selected location. By comparing Fig. 2.5b and 2.6 we can visually state that the EM patch location was correctly retrieved. The distance between the center of the ground-truth patch and the finally selected one is 17 EM pixels (0.35% of the image diagonal). The best retracing result was given by the LDP and the histogram intersection distance, although similar results were obtained with the cosine distance as well with a rectangle-shaped patch.

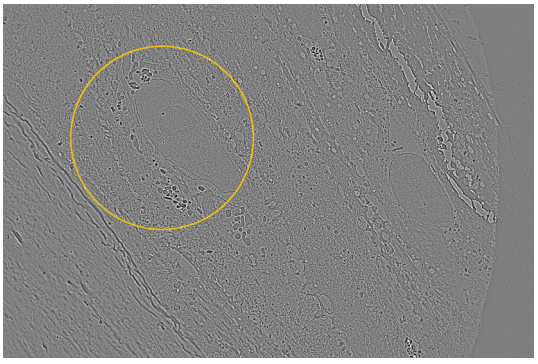
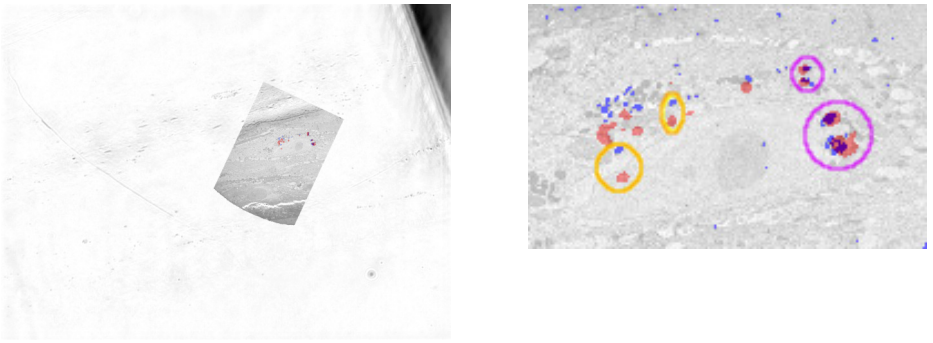


Fig. 2.6: *Experiment 1. Patch selected on EM by the search step.*

From this point on, all transformations are computed using the original intensity values of LM and EM images, the LM-ROI now as target image and the EM selected patch as source, since the final goal is to overlay the EM image onto the LM image around the ROI. First, the EM-SP is downscaled using η, then the rotation angle is computed using the coarse exhaustive approach previously described. For all experiments, the tested angles in degrees are $\alpha \in \{0°, 5°, 10°, \ldots, 355°\}$. The angle obtained between the LM-ROI and the EM-SP is 115°. Figure 2.7 shows the EM image overlaid on the LM-ROI after applying the pre-registration step, along with a closeup of the overlay between the LM-ROI and the selected EM patch using the same coordinates of the LM-ROI. For a better visualization of the alignment, the spots of interest in the LM-ROI are marked in red while the spots in the transformed EM patch are marked in blue. The purple circles indicate a good alignment of the spots, while the yellow circles show a misalignment of the spots.

Afterwards, a rigid transformation followed by an affine transformation are computed between the EM patch and the LM-ROI using mutual information (MI). The resulting overlay of each step is displayed in Fig.2.8. Inside the purple circles in both the

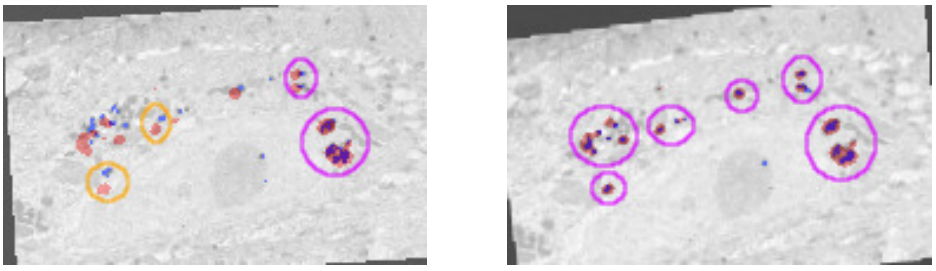


(a) Overlay of full EM and LM images

(b) Overlay of EM patch and LM-ROI

Fig. 2.7: *Experiment 1. Pre-registration overlay of EM and LM images.*

rigid and affine registrations, the correct alignment of the blue (EM) and red (LM) spots can be observed, while the yellow circles indicate misregistrations. We can notice that after affine registration, no distinguishable misalignments are present in the overlay.



(a) Overlay of EM patch and LM-ROI after rigid registration

(b) Overlay of EM patch and LM-ROI after affine registration

Fig. 2.8: *Experiment 1. Registration overlay of EM and LM patches after rigid and affine transformation steps.*

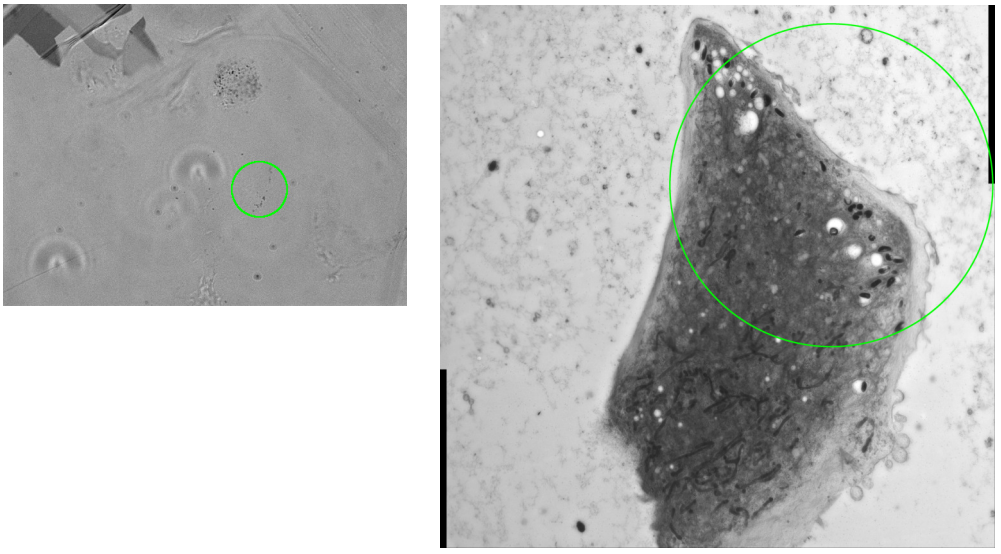
Registration errors of experiment 1 are collected in Table 2.1. Errors are expressed in terms of LM-pixels. In Table 2.1, we can observe that the error decreases significantly after each successive registration steps, supporting the visual evaluation of the overlays in Figures 2.7b, 2.8a and 2.8b.

In the second experiment, the subcellular elements of interest are endosomes. The LM image contains several structures and some acquisition artifacts while the EM is focused on the cell of interest. In this experiment, the dimensions of the LM image and the EM image are 512×383 and 4056×3970, respectively. Fig. 2.9 shows the input images, where the LM-ROI and the corresponding EM ground truth patch are marked with a green circle. The diameter of the patch is defined by magnifying the diameter of the LM-ROI by $\eta = 38.7$. Similar to experiment 1, if we observe the structures aligned inside

	Pre-registration	Rigid registration	Affine registration
Mean	4.80	2.54	0.22
SD	4.64	2.79	0.27
Median	3.40	1.40	0.17
Max.	12.55	7.13	0.67
Min.	0.47	0.33	0.0

Table 2.1: Experiment 1. Registration errors in LM pixels.

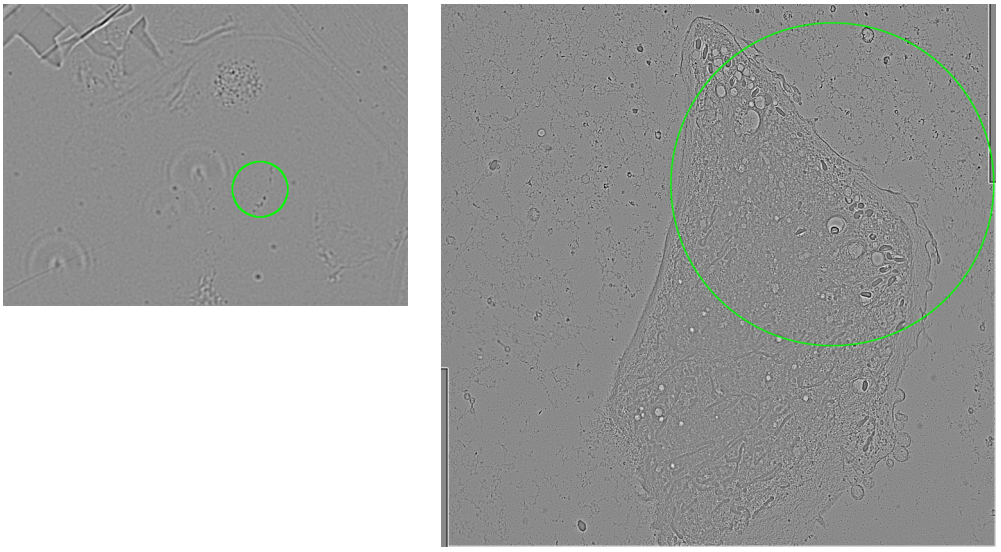
the LM-ROI and the EM-ROI, we can infer a significant rotation between the images. Additionally, the resolution ratio between both images is considerable, as well as the difference in image content. This is more apparent by noticing the absence of the delineation of the cell of interest inside the LM-ROI, where only the structures of interest (melanosomes) are discernible (Fig. 5.1a).



(a) Input LM image (b) Input EM image

Fig. 2.9: *Experiment 2. Input LM and EM images with LM-ROI and ground-truth EM patch.*

For this second experiment the scales selected by ATLAS are $s_{EM} = 6.19$ and $s_{LM} = 1.44$. The LoG representation of the EM and LM images are shown in Fig.2.10. The appearance difference between the EM and LM images is significantly reduced, while the shared structures of interest are highlighted. A two step exhaustive search was completed. First, we tested EM patch locations at intervals of 5.8% of the x-dimension and 6.9% of the y-dimension, with 90% overlap. Then, we refined the search around the primarily selected location. The LoG-EM patch containing the elements of interest is



(a) LoG of input LM image (b) LoG of input EM image

Fig. 2.10: *Experiment 2. LoG representation of input images.*

correctly retrieved using LoG-value histograms and the cosine distance, with both a rectangular and a disk-shaped patch. This selection can be visually assessed by comparing Fig. 2.10b and 2.11. The distance between the center of the ground-truth patch and the finally selected one is 66 EM pixels (1.18% of the image diagonal).

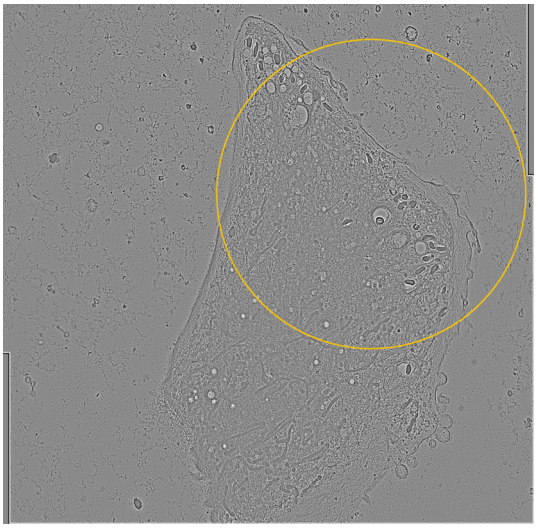
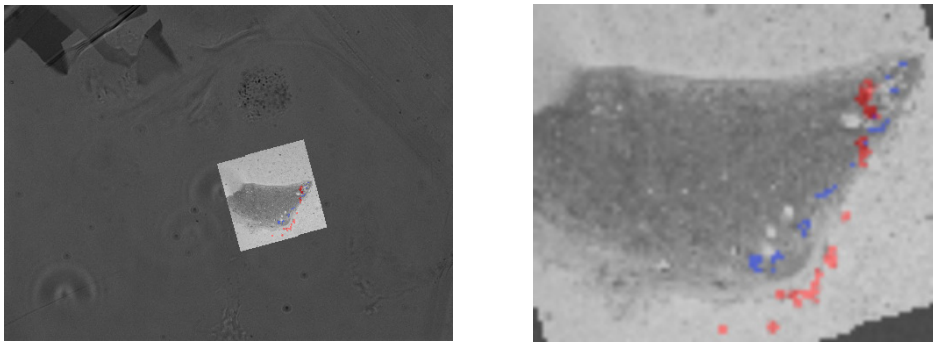


Fig. 2.11: *Experiment 2. Patch selected on EM by the search step.*

From now on, the source image will be the EM-SP and the target image the LM-ROI. To compute all the needed transformations, the EM-SP is downscaled by $\eta = 38.7$. The

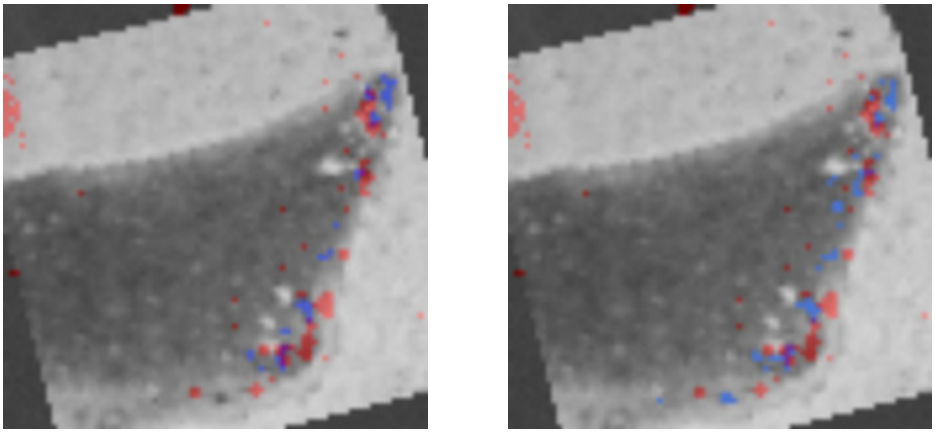
rotation angle computed between the EM-SP and the LM-ROI by the exhaustive search is 75°. The overlay of EM image on LM-ROI after this first pre-registration step is shown in Fig. 2.12, in addition to a close-up of the ROI. Following the distribution of the elements of interest, marked in red in the LM image and in blue in the EM image, it is easy to distinguish a small misalignment in the pre-registration.



(a) Overlay of full EM and LM images (b) Overlay of EM patch and LM-ROI

Fig. 2.12: *Experiment 2. Pre-registration overlay of EM and LM images.*

For the final registration, rigid an affine transformations are computed in turn between the EM patch and the LM-ROI. The overlay of the two registrations is shown in Fig. 2.13. We can observe that the resulting alignment is good, although not completely accurate. Registration errors are reported in Table 2.2.



(a) Overlay of EM patch and LM-ROI after rigid registration (b) Overlay of EM patch and LM-ROI after affine registration

Fig. 2.13: *Experiment 2. Registration overlay of EM and LM patches after rigid and affine registration.*

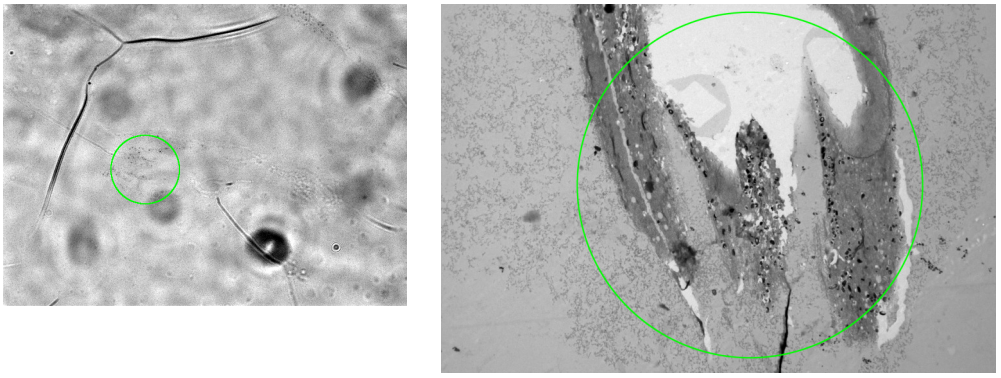
Registration errors in Table 2.2 are expressed in terms of LM-pixels. From Table 2.2

and the visual evaluation of Figures 2.12b, 2.13a and 2.13b, we can confirm that the error decreases after each registration step, although the difference between the result of the rigid and the affine registration is small.

	Pre-registration	Rigid registration	Affine registration
Mean	8.36	2.28	2.17
SD	3.97	1.46	1.36
Median	8.03	2.00	1.70
Max.	14.47	4.53	4.24
Min.	3.07	0.74	0.67

Table 2.2: Experiment 2. Registration errors in LM pixels.

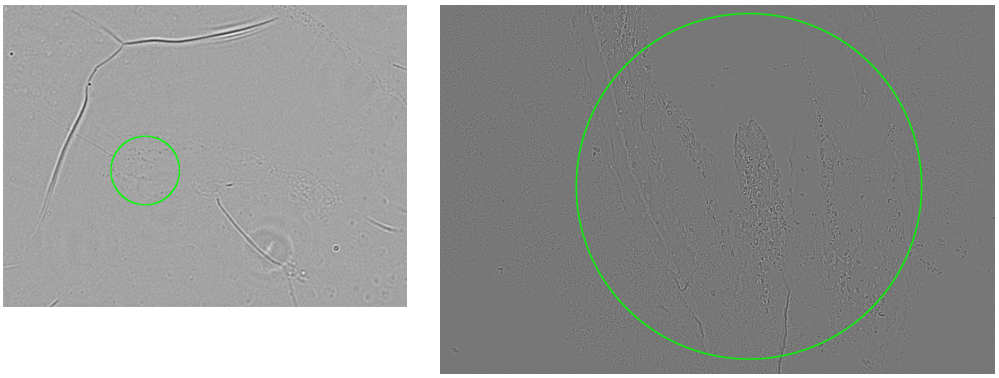
MNT1 cells are studied in the CLEM images of the third experiment, with endosomes are elements of interest. Like most of the following experiments, the LM image contains several cells while the EM image is a high magnification of a single cell. Hence, the ROI will be delineated in the EM image and the corresponding patch will be searched in the LM image. This experiment involves a 696×520 LM image and a 4008×2664 EM image, both displayed in Fig. 2.14, where the EM-ROI and the corresponding LM ground truth patch are circled in green. The diameter of the LM-patch is defined by dividing the diameter of the EM-ROI by $\eta = 20.96$. Several acquisition artifacts can be noticed in the input LM image (Fig. 2.14a), making the elements of interest hard to perceive and increasing the vast difference in terms of content between the EM and LM images.



(a) Input LM image (b) Input EM image

Fig. 2.14: *Experiment 3. Input LM and EM images with EM-ROI and ground-truth LM patch.*

The LoG representation of the EM and LM images is shown in Fig.5.7. The scales selected by ATLAS are respectively $s_{EM} = 6.19$ and $s_{LM} = 1.44$. By applying the LoG transform, many of the artifacts present in the input LM image (Fig. 2.14a) are smoothed and the structures of interest are more salient.



(a) LoG of input LM image (b) LoG of input EM image

Fig. 2.15: *Experiment 3. LoG representation of input images.*

The LoG-LM patch corresponding to the LoG-EM-ROI is correctly retrieved using LoG-value histograms with cosine distance, with both a rectangular and a disk-shaped patch. Similar results are obtained with the histogram intersection distance. The first exhaustive search step was implemented with LM patch locations at intervals of 1.7% of the x-dimension and 2.3% of the y-dimension (with 95% overlap), followed by a refined search around the selected location. The selected patch is displayed in Fig. 2.16, whose location is close to the ground truth displayed in Fig. 2.15a. The distance between the center of the ground-truth patch and the finally selected one is 22.36 LM pixels (2.67% of the image diagonal).

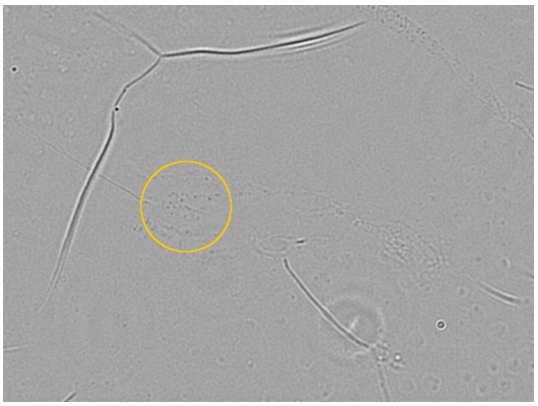
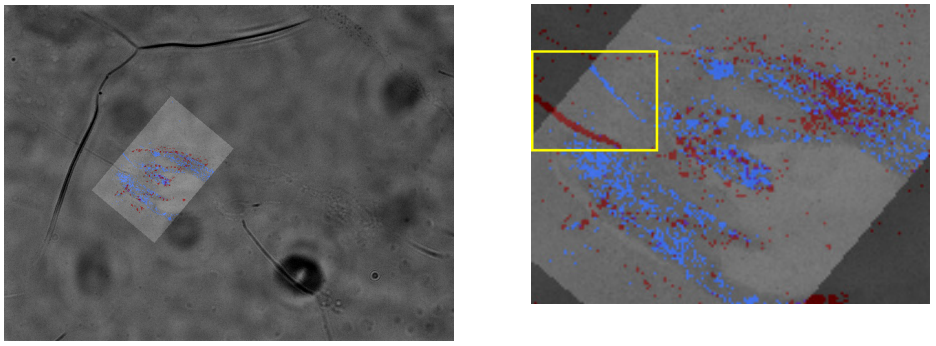


Fig. 2.16: *Experiment 3. Patch selected on LM by the search step.*

The EM-ROI is downscaled by $\eta = 20.96$ to compute the rest of the transformations. The rotation angle computed between the EM-ROI and the LM-SP is 130°. The overlay of the EM image on LM-SP after the pre-registration step is shown in Fig. 2.17, along with a closeup of the ROI. The elements of interest are marked in red in the LM image and in blue in the EM image. The alignment is harder to observe in this experiment, given the little saliency of the spots of interest and the lack of clear distributions of the structures. However, the registration can be evaluated by following the lines enclosed in yellow in Fig. 2.17b.

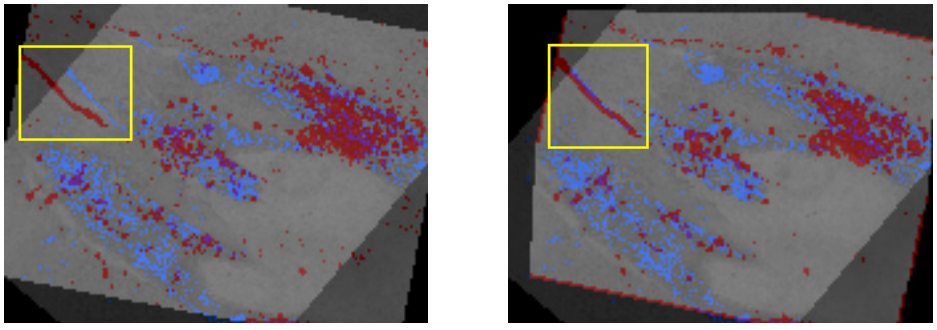


(a) Overlay of full EM and LM images

(b) Overlay of EM-ROI and LM patch

Fig. 2.17: *Experiment 3. Pre-registration overlay of EM and LM images.*

For the final registration, consecutive rigid an affine transformations are computed between the EM-ROI and the LM-SP. The overlay of the two registrations is shown in Fig. 2.18. The misalignment of the lines inside the yellow square decreases from the rigid (Fig. 2.18a) to the affine registrations (Fig. 2.18b). Registration errors are reported in Table 2.3.



(a) Overlay of EM-ROI and LM-SP after rigid registration

(b) Overlay of EM-ROI and LM-SP after affine registration

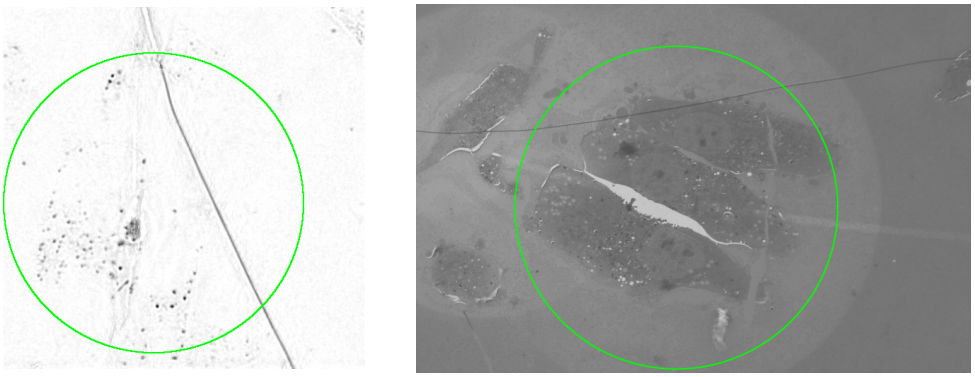
Fig. 2.18: *Experiment 3. Registration overlay of EM and LM patches after rigid and affine registration.*

Registration errors are expressed in terms of LM-pixels. From the visual evaluation of Figures 2.17b, 2.18a and 2.18b and the errors in Table 2.3, we can confirm that the error decreases after each registration step, generating an accurate alignment after the affine registration step.

	Pre-registration	Rigid registration	Affine registration
Mean	8.19	6.76	1.09
SD	7.65	7.01	0.76
Median	4.61	3.30	0.70
Max.	21.21	16.92	2.12
Min.	2.00	1.12	0.50

Table 2.3: Experiment 3. Registration errors in LM pixels.

Experiment 4 is different from the previous three experiments in terms of the resolution ratio between the EM and LM images, with $\eta = 8.5$ indicating the lowest difference in resolution of all the experiments, due to the EM image being of low magnification. However, the difference in image content is significant, since just some of the elements of interest present in the LM image are visible in the EM image, as observable in Fig. 2.19, where the ROI and the corresponding ground truth location are circled in green. Again in this experiment, the ROI is delineated in the EM image and it will be sought in the LM image. Both images contain several cells. LM image dimensions are 512×512 while EM image dimensions are 4008×2664.



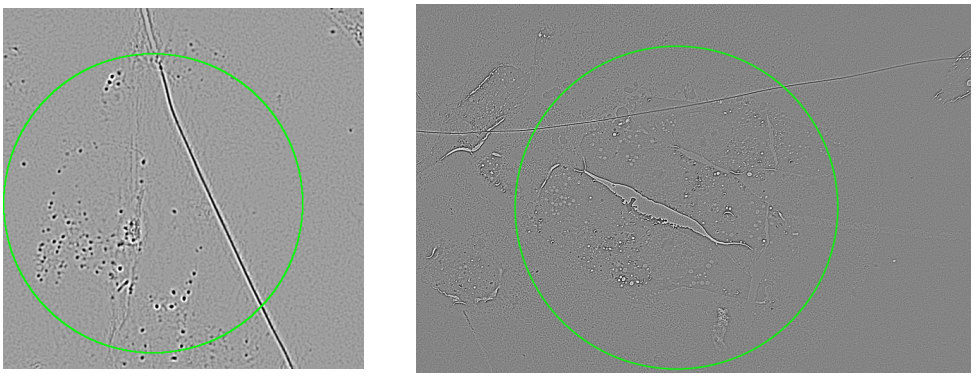
(a) Input LM image

(b) Input EM image

Fig. 2.19: *Experiment 4. Input LM and EM images with EM-ROI and ground-truth LM patch.*

The common LoG representation of the EM and LM images are shown in Fig.2.20, with LoG scales $s_{EM} = 6.19$ and $s_{LM} = 1.44$. LoG transform brings the appearance of

EM and LM images closer, while highlighting the structures of interest of each image, as observed in Figures 2.20a and 2.20b.



(a) LoG of input LM image

(b) LoG of input EM image

Fig. 2.20: *Experiment 4. LoG representation of input images.*

For the searching step of the EM-ROI in the LM image, the diameter of the searched patch is defined by dividing the diameter of the EM-ROI by $\eta = 8.5$. The first exhaustive search step was implemented with LM patch locations at intervals of 8.2% of the x-dimension and 8.2% of the y-dimension (with 95% overlap), followed by a refined search around the selected location. The LM patch containing the elements of interest is correctly retrieved by both LoG-value histograms and LDP using cosine and histogram intersection distances. As observed in Figures 2.20a and 2.21, the search area is significantly smaller compared to the rest of the experiments. However, the use of the search step provides a good initialization of the final registration. The distance between the center of the ground-truth patch and the selected patch is 20 LM pixels (2.77% of the image diagonal).

The EM-ROI is downscaled by $\eta = 8.5$ to compute the rotation between the EM-ROI and the LM-SP, giving an angle of 85°. The pre-registration overlay is shown in Fig. 2.22. To visually assess the pre-registration, we can take the line crossing the top of images, marked in blue in the EM image and in red in the LM image. We can notice that there is still a significant misalignment; nonetheless, the pre-registration step is still useful to fully register the EM and LM images, as it will be demonstrated in the results of the next steps of the registration process.

Rigid an affine transformations are consecutively estimated between the EM-ROI patch and the LM-SP. The overlay of the two registered images is given in Fig. 2.23. The rigid registration is capable of compensating the misalignment of the pre-registered images.

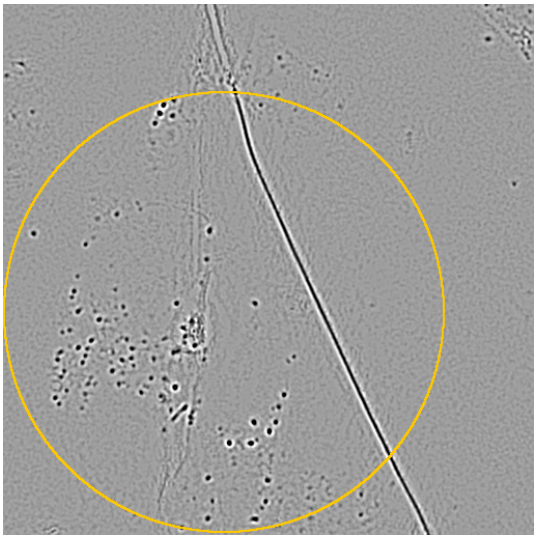


Fig. 2.21: *Experiment 4. Patch selected on EM by the search step.*

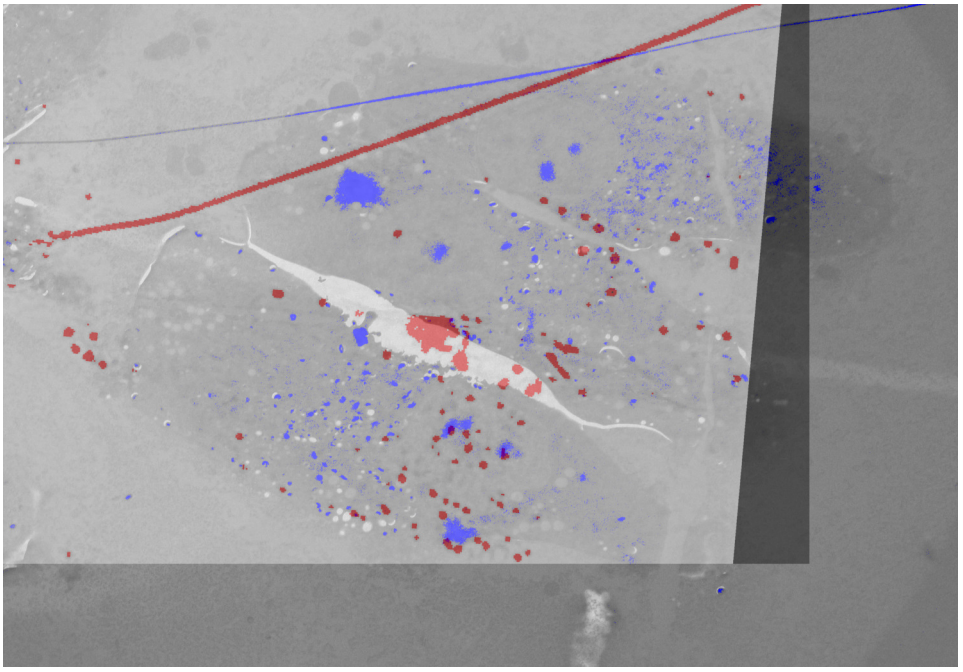
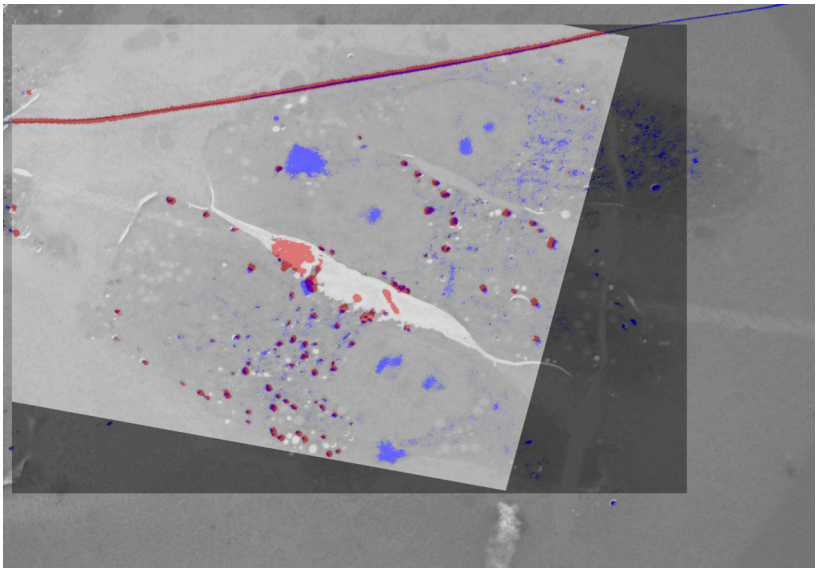
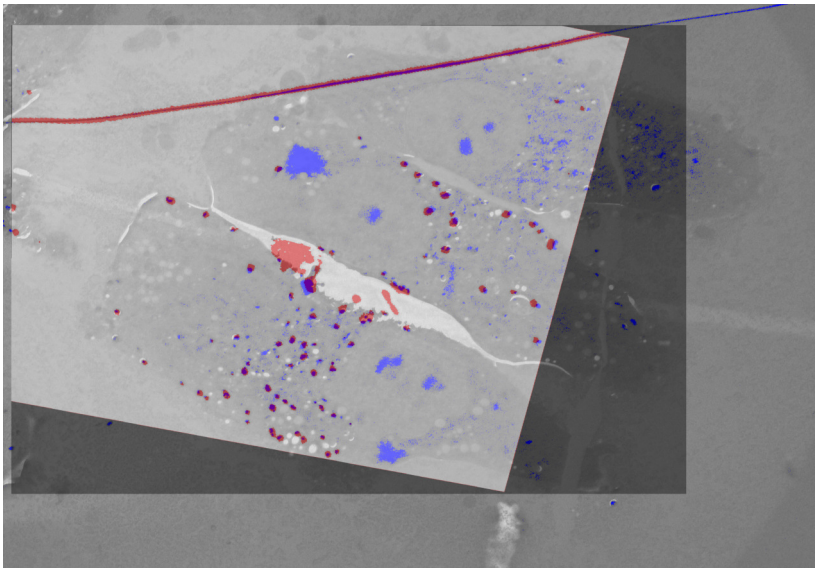


Fig. 2.22: *Experiment 4. Pre-registration overlay of EM and LM images.*



(a) Overlay of EM-ROI and LM-SP after rigid registration



(b) Overlay of EM-ROI and LM-SP after affine registration

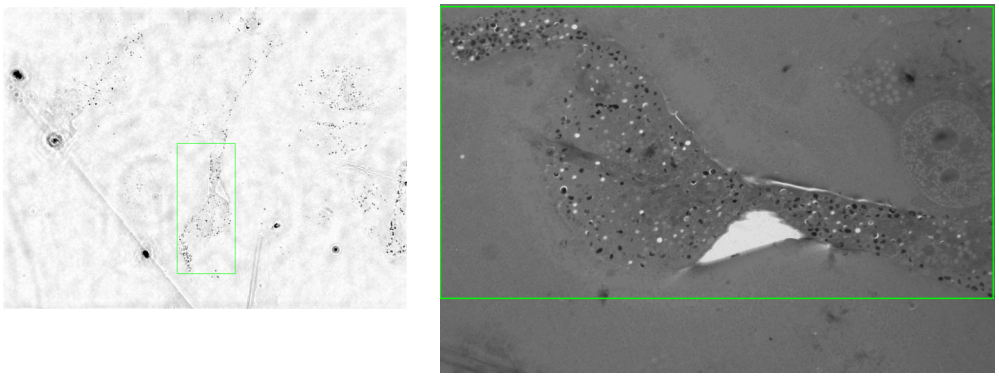
Fig. 2.23: *Experiment 4. Registration overlay of EM and LM patches after rigid and affine transformation steps.*

Registration errors in Table 2.4 are expressed in terms of EM-pixels. The misalignment of the pre-registered LM and the EM is large as outlined by visually evaluating the overlay of Fig. 2.22 and the errors of the pre-registration step in Table 2.4. However, the MI-based rigid registration is capable of overcoming this, as shown in Fig. 2.23a. Then, the affine registration step has minimum effect over the final registration, as seen in Fig. 2.23b.

	Pre-registration	Rigid registration	Affine registration
Mean	171.2	8.43	6.60
SD	67.9	5.30	3.61
Median	196.4	6.70	6.70
Max.	209.90	16.96	10.81
Min.	50.26	3.00	2.99

Table 2.4: Experiment 4. Registration errors in EM pixels.

Experiment 5 contains a structure that is considerably different from the rest of the experiments, as depicted in Fig. 2.24. The shape of the cell is elongated, preventing the use of disk-shaped patches. Therefore, the EM-ROI will be defined as a rectangle as plotted in Fig. 2.24b, along with the corresponding LM patch. The LM-patch ground truth is presented in Fig. 2.24a. The resolution ratio is $\eta = 8.9$ and the LM and EM image dimensions are 1392×1040 and 4008×2664 respectively. Following the shape of the structure in the EM-ROI and the LM-patch ground truth, a large difference in terms of orientation and size obviously appears, while their content is rather similar.



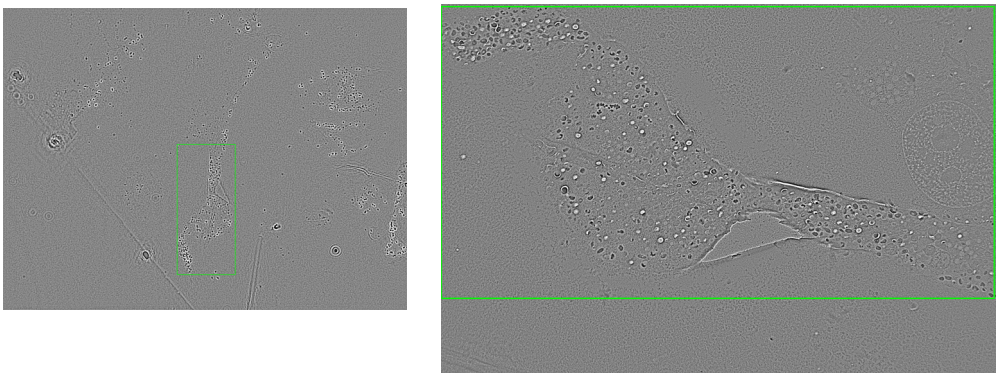
(a) Input LM image

(b) Input EM image

Fig. 2.24: *Experiment 5. Input LM and EM images with EM-ROI and ground-truth LM patch.*

The common LoG representation of the EM and LM images is shown in Fig.2.25, with

scales $s_{EM} = 6.19$ and $s_{LM} = 1.44$. As expected, the appearance difference is significantly reduced.



(a) LoG of input LM image

(b) LoG of input EM image

Fig. 2.25: *Experiment 5. LoG representation of input images.*

For the search of the EM-ROI in the LM image, the dimensions of the search patch are defined by downscaling the dimensions of the EM-ROI by $\eta = 8.9$. The LM patch containing the elements of interest is correctly retrieved by LDP and histogram intersection distance. The first exhaustive search step was implemented with LM patch locations at intervals of 1.4% of the x-dimension and 1.9% of the y-dimension (with 95% overlap), followed by a refined search around the selected location. This selection can be visually assessed by comparing Fig. 2.25a and 2.26. The distance between the center of the ground-truth patch and the finally selected one is 37.85 LM pixels (2.17% of the image diagonal).

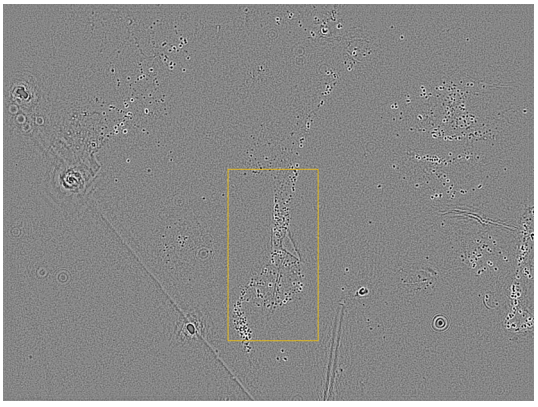
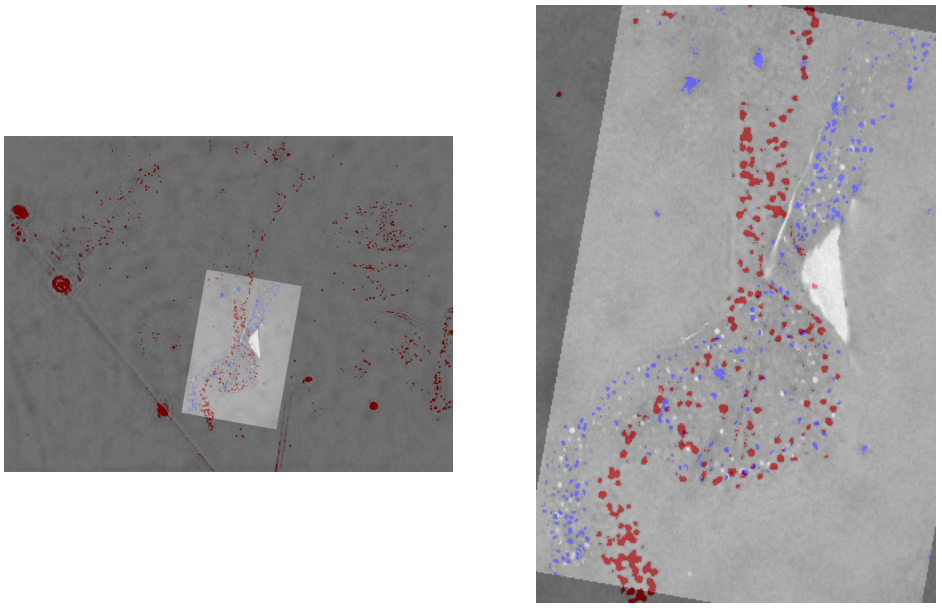


Fig. 2.26: *Experiment 5. Patch selected on LM by the retracing step.*

To further compute all the rigid and affine transformations, the EM-ROI is downscaled by the resolution ratio $\eta = 8.9$. The rotation angle found with the exhaustive search step between EM-ROI and LM-SP is 80°. The overlay of the EM image on the LM-ROI after this first pre-registration step is shown in Fig. 2.27. By comparing the overlaid structures of interest, marked in blue in the EM image and in red in the LM image, we can visually evaluate the quality of the pre-registration, noting that a misalignment is present between the two images.



(a) Overlay of full EM and LM images

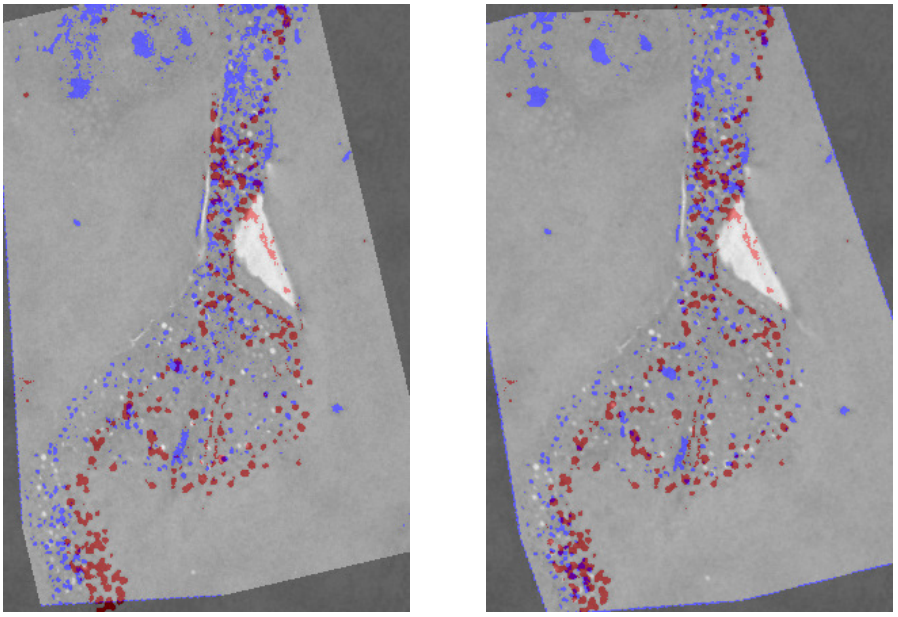
(b) Overlay of EM-ROI and LM-SP

Fig. 2.27: *Experiment 5. Pre-registration overlay of EM and LM images.*

For the final registration stage, rigid an affine transformations are consecutively computed between the EM-ROI and the LM-SP. The overlay of the two registered images is displayed in Fig. 2.28. The rigid registration is mostly capable of compensating the misalignment of the pre-registration, although there is still a small misalignment noticeable in the white structure in the EM image.

Registration errors of experiment 5 are collected in Table 2.5. The errors are expressed in terms of LM-pixels. The registration error decreases after each successive registration step, although the final registration could still be improved, as noticeable in the last affine registration displayed in Fig. 2.28b and their corresponding errors.

MNT1 cells corresponding to human skin cells are the subject of the CLEM images in experiment 6. The resolution ratio between the EM and LM images is $\eta = 17.5$. LM



(a) Overlay of EM-ROI and LM-SP after rigid registration

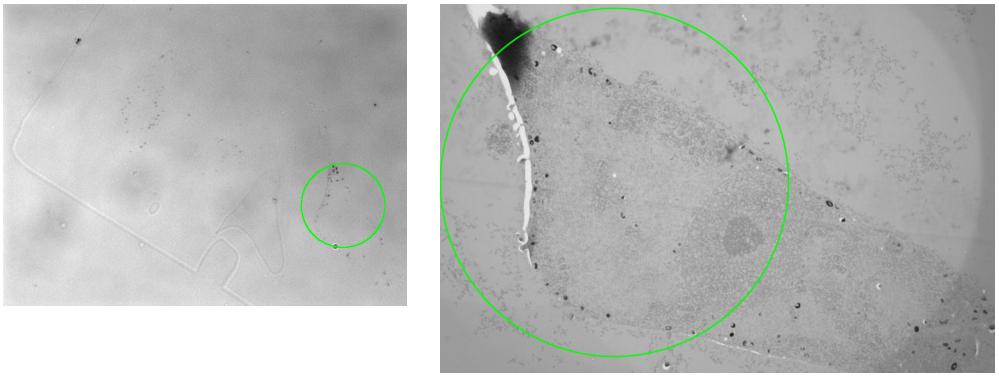
(b) Overlay of EM-ROI and LM-SP after affine registration

Fig. 2.28: *Experiment 5. Registration overlay of EM and LM patches after rigid and affine transformation steps.*

	Pre-registration	Rigid registration	Affine registration
Mean	35.21	18.39	4.98
SD	24.10	6.90	3.61
Median	33.41	16.52	4.00
Max.	69.26	29.55	9.43
Min.	5.83	9.85	1.00

Table 2.5: Experiment 5. Registration errors in LM pixels.

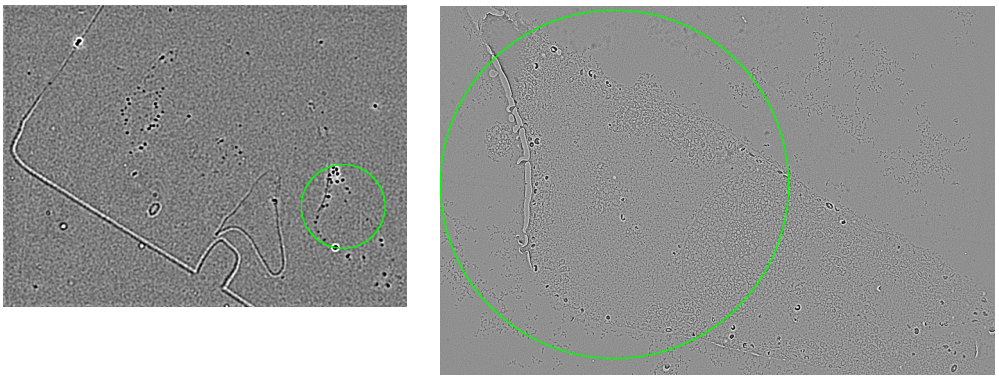
image dimensions are 696 × 520 while EM image dimensions are 4008 × 2664. As in previous experiments, the high magnification EM image is focused on a single cell while the LM image contains multiple cells. Contrary to the rest of the experiments, the intensity appearance of these CLEM images is similar. Still, large differences in terms of size, orientation, and visible structures of interest are prevalent, as confirmed in Fig. 2.29.



(a) Input LM image (b) Input EM image

Fig. 2.29: *Experiment 6. Input LM and EM images with EM-ROI and ground-truth LM patch circled in green.*

For this sixth experiment the LoG scales selected by ATLAS are $s_{EM} = 6.19$ and $s_{LM} = 3.58$. The LoG representation of the EM and LM images is displayed in Fig.2.30, where the use of the LoG transform visibly highlights the structures present in both images, specially in the LM image.



(a) LoG of input LM image (b) LoG of input EM image

Fig. 2.30: *Experiment 6. LoG representation of input images.*

For the search of the EM-ROI in the LM image, the diameter of the searched patch

is defined by downscaling the diameter of the EM-ROI by $\eta = 17.5$. The first exhaustive search step was implemented with LM patch locations at intervals of 2.0% of the x-dimension and 2.7% of the y-dimension (with 95% overlap), followed by a refined search around the selected location. The LM patch containing the elements of interest is correctly retrieved using LoG-value histograms and LDP with both cosine similarity and histogram intersection distance, either using a rectangular or a disk-shaped patch. This selection can be visually assessed by comparing Fig. 2.30a and 2.31. The distance between the center of the ground-truth patch and the finally selected one is 10.82 LM pixels (1.24% of the image diagonal).

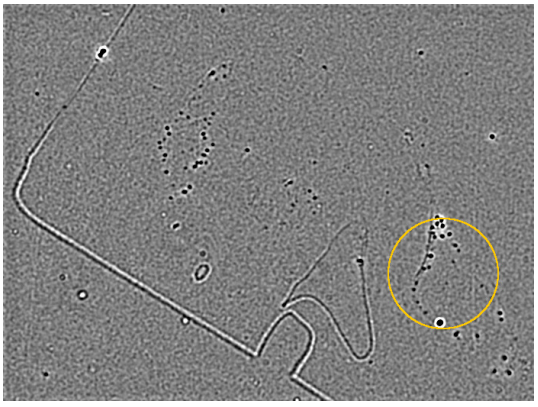


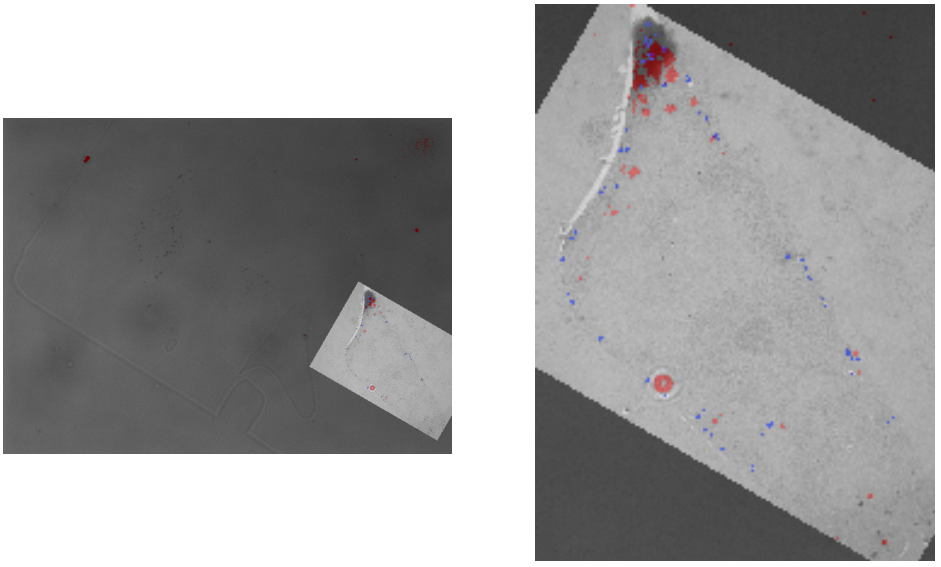
Fig. 2.31: *Experiment 6. Patch selected in LM image by the search step.*

Then, the EM-ROI is downscaled by the resolution ratio $\eta = 17.5$. The rotation angle coarsely found between EM-ROI and LM-SP is 30°. The overlay of EM on LM after the pre-registration step is shown in Fig. 2.32. By looking at the overlay of the red structures of the LM image with the dark region of the EM image at the top of the ROI, as well as the location of the adjacent blue and red structures, we can visually evaluate the alignment of the images.

For the final registration, rigid and affine transformations are computed in turn between the EM-ROI and the LM-SP. The overlay of the registered images is depicted in Fig. 2.33. A good alignment is achieved at every registration step, with a small improvement after each step.

Registration errors of experiment 6 are expressed in terms of LM-pixels in Table 2.6. As mentioned before, a good alignment was already obtained in the pre-registration step, and every subsequent step slightly improves the alignment of the EM and LM images.

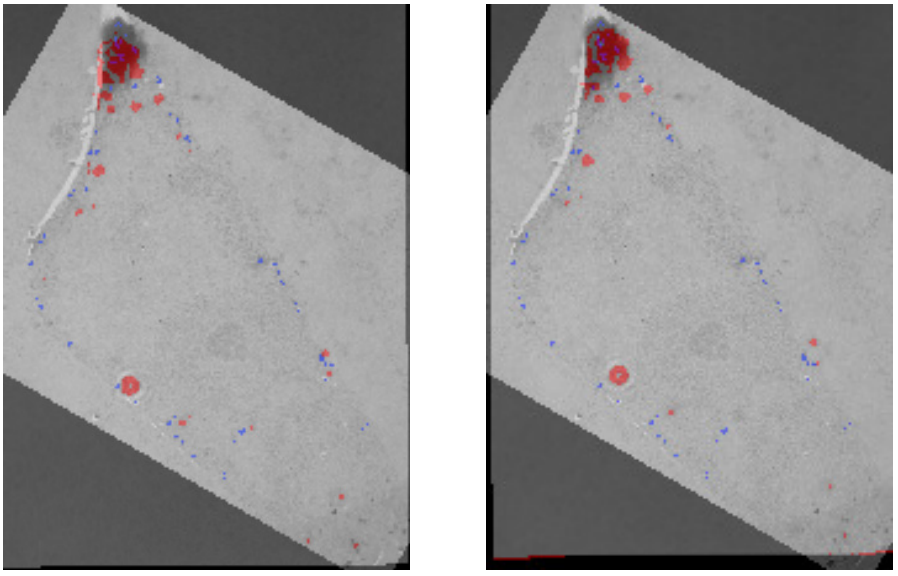
Experiment 7 involves the biggest resolution ratio between the EM and LM images,



(a) Overlay of full EM and LM images

(b) Overlay of EM-ROI and LM-SP

Fig. 2.32: *Experiment 6. Pre-registration overlay of EM and LM.*



(a) Overlay of EM-ROI and LM-SP after rigid registration

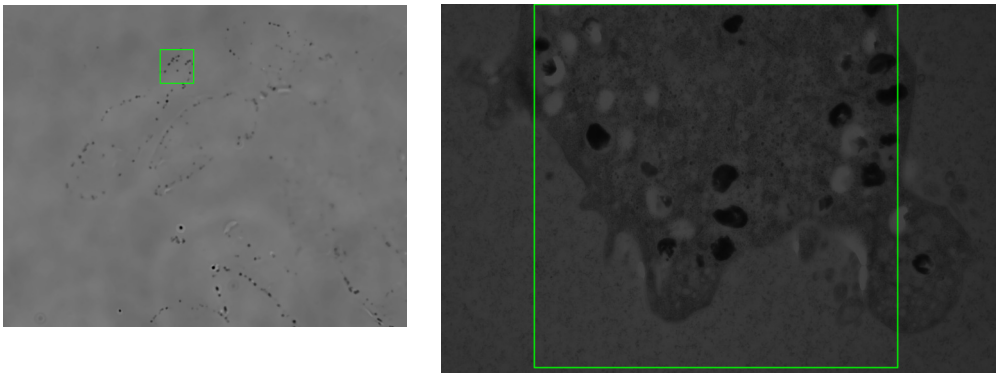
(b) Overlay of EM-ROI and LM-SP after affine registration

Fig. 2.33: *Experiment 6. Registration overlay of EM and LM patches.*

	Pre-registration	Rigid registration	Affine registration
Mean	5.45	5.44	3.76
SD	3.92	2.25	1.72
Median	5.58	5.75	4.06
Max.	9.82	8.02	5.59
Min.	1.41	1.80	1.00

Table 2.6: Experiment 6. Registration errors in LM pixels.

with $\eta = 41$. As for previous experiments, the biological structures imaged are MNT1 cells containing melanosomes. In this experiment, the LM image contains many configurations of melanosomes, which makes the retracing of the EM-ROI more challenging, as seen in Fig. 2.34. LM image dimensions are 771×613, while EM image dimensions are 4008×2664.

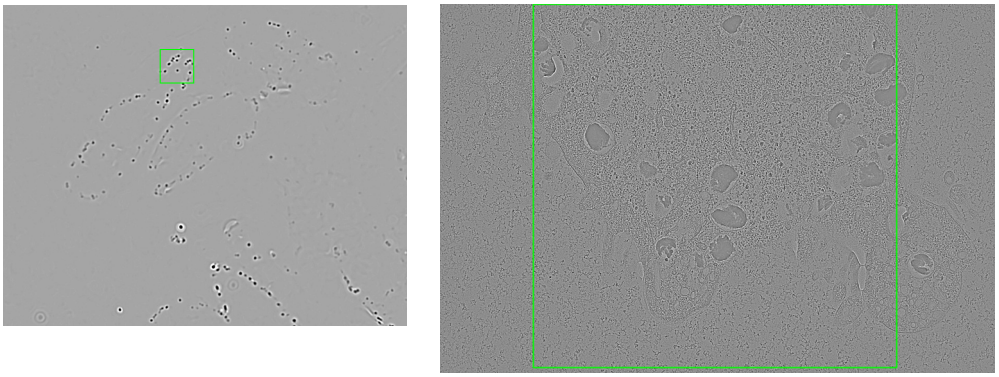


(a) Input LM image (b) Input EM image

Fig. 2.34: *Experiment 7. Input LM and EM images with EM-ROI and ground-truth LM patch framed in green.*

The scales selected by ATLAS are $s_{EM} = 6.19$ and $s_{LM} = 1.73$ and the LoG representation of the EM and LM images is shown in Fig.2.35.

For the search of the EM-ROI in the LM image, the diameter of the searched patch is defined by downscaling the diameter of the EM-ROI by $\eta = 41$. A two-step exhaustive search was carried out. The first exhaustive search step was implemented with LM patch locations at intervals of 0.9% of the x-dimension and 1.4% of the y-dimension (with 95% overlap), followed by a refined search around the selected location. The LM patch corresponding to the EM-ROI is correctly retrieved using LoG-value histograms with both cosine similarity and histogram intersection distance, using both a rectangular and a disk-shaped patch. This selection can be visually assessed by comparing Fig. 2.30a and



(a) LoG of input LM image

(b) LoG of input EM image

Fig. 2.35: *Experiment 7. LoG representation of input images.*

2.36. The distance between the center of the ground-truth patch and the finally selected one is 7.28 LM pixels (0.74% of the image diagonal).

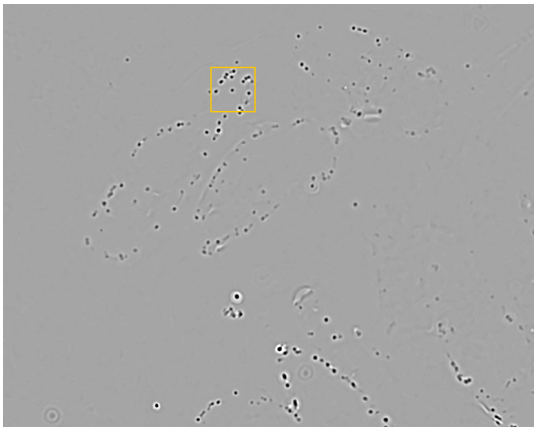
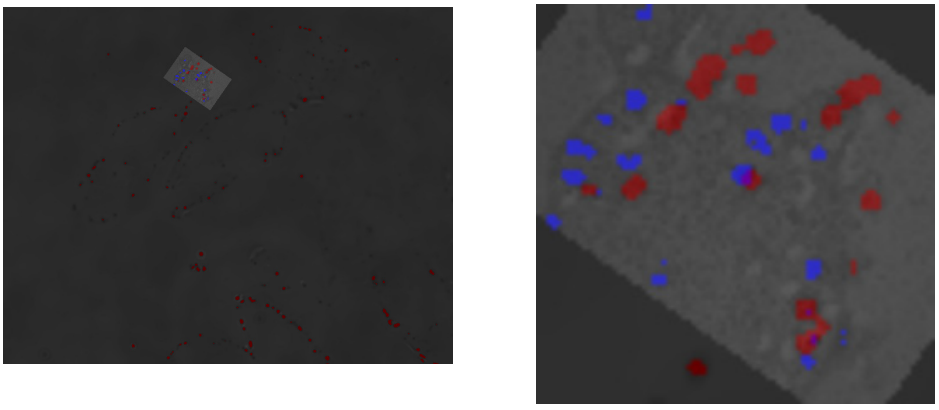


Fig. 2.36: *Experiment 7. Patch selected in the EM image by the search step.*

Then, EM-ROI is downscaled by the resolution ratio $\eta = 17.5$. The rotation angle found with the exhaustive search between EM-ROI and LM-SP is 145°. The pre-registration overlay of the full EM and LM images is shown in Fig. 2.37, along with a close-up of the ROI. The configuration of the melanosomes resembles an uppercase "M" letter, marked in blue in the EM image and in red in the LM image. A small noticed of the structures can be observed in the pre-registration overlay.

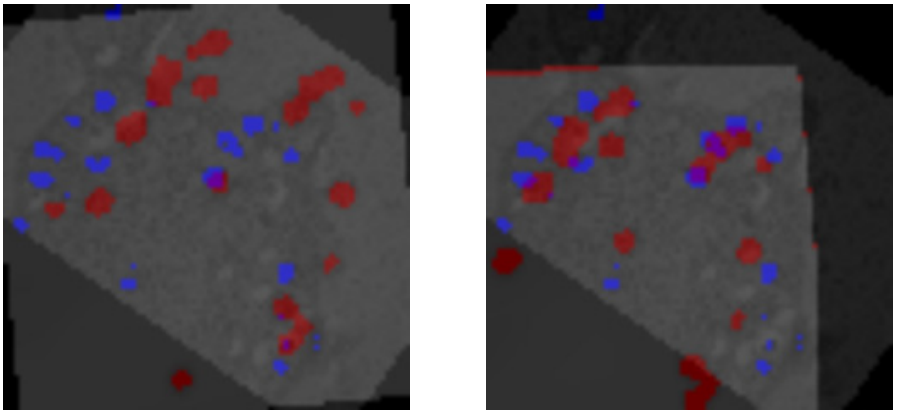
Finally, rigid an affine transformations are successively computed between the EM-ROI and the LM-SP. The resulting overlays are displayed in Fig. 2.38. Once again, the registration improves after each step, as confirmed by the registration errors collected in Table 2.7.



(a) Overlay of full EM and LM images

(b) Overlay of EM-ROI and LM-SP

Fig. 2.37: *Experiment 7. Pre-registration overlay of EM and LM images.*



(a) Overlay of EM-ROI and LM-SP after rigid registration

(b) Overlay of EM-ROI and LM-SP after affine registration

Fig. 2.38: *Experiment 7. Registration overlay of EM and LM patches after rigid and affine transformation steps.*

Registration errors of experiment 7 are expressed in Table 2.7 in terms of LM-pixels. As illustrated in Figures 2.37, 2.38a and 2.38b, the alignment improves after each step, however, leading to a good accuracy after the affine registration.

	Pre-registration	Rigid registration	Affine registration
Mean	15.86	14.91	1.66
SD	5.02	4.86	1.60
Median	16.10	16.60	1.08
Max.	21.02	19.89	4.74
Min.	9.55	6.57	0.47

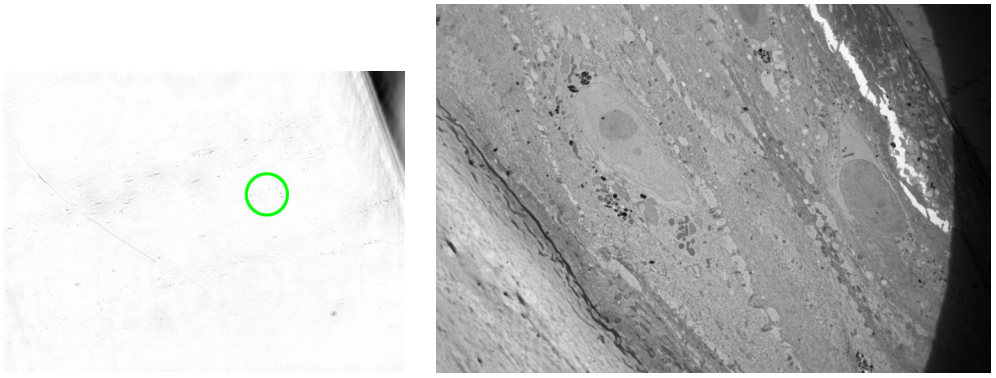
Table 2.7: Experiment 7. Registration errors in LM pixels.

2.4 LM-guided low to high magnification EM acquisition

Imaging a sample using EM is a complex process that comprises many stages. EM acquisition is usually performed from lower to higher magnification. During this process, location and relocation of the region of interest is frequently needed. A different microscopy modality is used for this purpose in many CLEM acquisition frameworks. Normally, LM is used as guide for EM acquisition at different magnifications. First, the LM image is acquired, providing a wide field of view of the sample. Afterwards, a low-magnification EM image is captured covering roughly the same area imaged by LM. A region of interest is visually localized in the LM image and then is visually retrieved in the low-magnification EM image, and a high-magnification EM acquisition is performed next. Multiple EM images with higher magnifications can be acquired focused in the same ROI located in the EM, with magnifications increasing as high as necessary, depending on the particular biological subject. Our pre-registration stage can be exploited within this part of the CLEM framework. As previously mentioned, the pre-alignment of EM and LM images is achieved by the implementation of a patch-based exhaustive search of a pre-defined ROI contained in one image in the other image. The images intensities are transformed to a common representation using the LoG transform, facilitating the implementation of intensity-based similarity metrics for comparing a given EM patch with the LM-ROI. In this case, the ROI would be delineated in the LM image and the search would be carried out in the EM image. Once the EM patch corresponding to the LM-ROI is selected, a high-magnification acquisition would be performed focused in the selected EM patch, followed by automatic registration and overlaying of the high-magnification EM and LM images.

Experiment 1 is an example emulating these type of application. The LM image

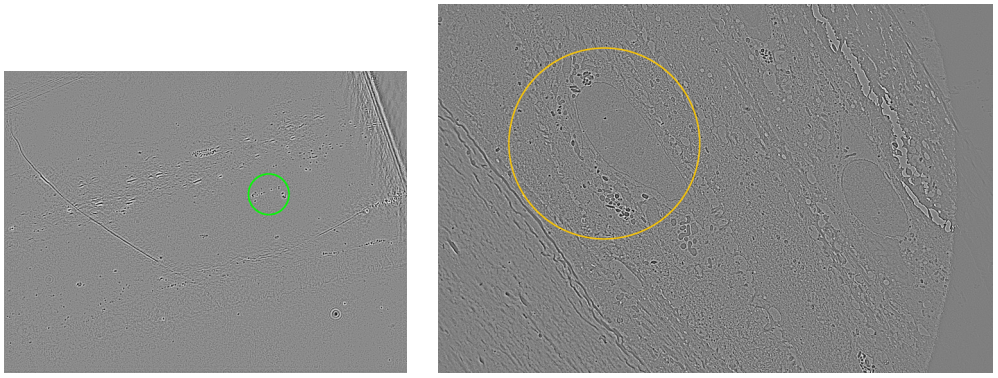
presents a wide field of view of the sample, as observed in Fig. 2.39a. A low-magnification EM image is then acquired, as seen in Fig. 2.39b. A ROI is delineated in the LoG-LM image, enclosed in green in Fig. 2.40a. The corresponding patch is located in the LoG representation of the low-magnification EM, enclosed in orange in Fig. 2.40b.



(a) Input LM image with ROI marked with green circle.

(b) Input low-magnification EM image

Fig. 2.39: *Experiment 1. Input LM image and low-magnification EM image.*



(a) LoG-LM image with ROI marked with green circle.

(b) LoG-EM image. The selected patch is enclosed in the orange circle.

Fig. 2.40: *Experiment 1. LoG representation of input images.*

A shift is computed from the difference between the center of the LM-ROI and the center of the EM-SP. Then, the rotation is computed as explained in Subsection 2.2.4 and the LM is transformed by applying the shift, a magnification of $\eta = 9.7$, and the computed rotation. Finally, a registration refinement is computed by estimating rigid and affine transformations between the pre-registered LM image and the high-magnification EM image, using mutual information as similarity criterion. The final overlay is displayed

in Fig. 2.41, with the structures of interest in the LM image in red color and the same structures in the EM image in blue color.

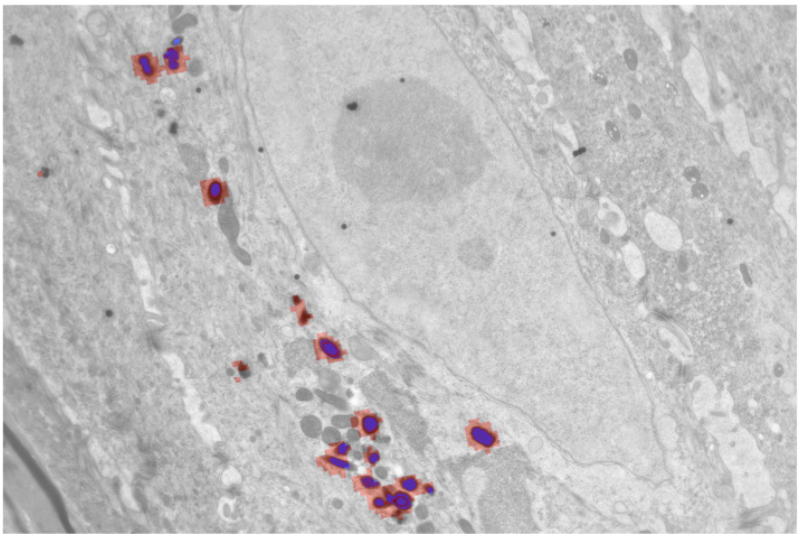


Fig. 2.41: *Experiment 1. Overlay of high-magnification EM and registered LM images.*

This approach could be implemented within different CLEM setups to assist in the acquisition of consecutive low-to-high magnification EM images guided by LM images. Further testing is necessary to assess the extended usefulness or our method within different CLEM acquisition frameworks.

2.5 Conclusion

We have defined a fully automated method for the search and registration stages in the CLEM framework, which is a very challenging task due the great differences in appearance, resolution, image size, content and field of view between LM and EM images. Given these vast differences, the estimation of transformations to directly register CLEM images is not possible, making a pre-registration stage necessary. For that reason, we have proposed to implement a patch-based search step where a predefined ROI will be searched in the target image. The pre-registration is composed by the shift given by the search step, a coarse rotation computed with a MI-based exhaustive method, bringing the EM and LM images closer in order to estimate the final registration. To make matching manageable, we map the multimodal LM and EM images to a common representation supplied by the LoG transform, which provides additional advantages as decreasing noise in the image while highlighting structures in the images. To be invariant to size, scaling and rotation, we adopted normalized histograms of LoG values or LDP values,

in the search step. This pre-registration stage was successfully tested on all the experiments.

We reported convincing results on seven real CLEM image sets for the full registration and the final overlay of LM and EM images. Each of the CLEM image pairs present a different challenge regarding differences of content, image size, appearance, resolution and orientation. Accurate registration results evaluated both visually and quantitatively are obtained for almost all experiments. Apart from experiment 5, the overall final registration accuracy is between 0.2 and 2.1 pixels, assessing the good performance of our registration method for 2D CLEM images. Given the variety of the experiments carried out, the achieved results demonstrate the good performance of our novel approach. The pre-registration step is essential to the registration process, due to the challenges stated throughout the introduction to the CLEM registration problem. Starting from the pre-registration, different techniques can be applied to complete the final registration stage. We adopted a classical mutual information-based registration framework, given its highly proved efficiency in multimodal registration problems.

Further improvements and additions of the registration framework will be presented in Chapters 3 and 4, respectively, a point-based geometrical refinement step exploiting our multiscale spot detection method, and the choice of the most appropriate transformation based on our robust motion model selection method.

CHAPTER 3

MULTISCALE SPOT SEGMENTATION WITH SELECTION OF IMAGE SCALES

Detecting spot structures of different sizes in images is required for many applications. Multiple image scales must then be considered for reliable spot segmentation. Microscopy image analysis is our primary goal but the developed method can be applied to other types of images and videos as well. We present an efficient method able to detect spots of various sizes corrupted by noise. We define an original criterion based on the *a contrario* approach to automatically select the meaningful scales. Our coarse-to-fine spot segmentation scheme performs well even for spots in close proximity while preserving their true size and involving minimal parameter setting. We demonstrate on simulated and real images of different types that our method outperforms other existing methods.

In many cases, image content may consist of a collection of elements, such as cars in traffic monitoring, boats on the ocean in remote sensing, stars in astronomy, animals in videos of natural scenes, cells and subcellular elements in microscopy imaging, to name a few. If they are small enough or seen from a distance, they usually appear as similar spots of a more or less regular shape. Thus, detecting spots in images is a common prerequesite in many applications. In order to countervail noise resulting from the image acquisition or the presence of spurious elements, selecting the right image scale is required to correctly detect the spots of interest. For a given scale, a spot detection framework is usually divided in three sub-steps : first, preprocessing of the image to reduce noise; second, signal enhancement to highlight the spots to be detected; third, spot detection by thresholding; the two first ones being often merged in a single operator.

However, elements of interest may be not all of the same image scale, if the collection includes subgroups of different sizes or if perspective effects occur. Then, the need is not merely the selection of the optimal scale, but of all the meaningful scales. We will deal with the problem of multiscale spot detection with an automated selection of the meaningful scales. As already pointed out, our primary interest is to detect particles in microscopy images, but our method can be applied to other types of images as well. Our method adopts the *a contrario* approach for multiscale selection, and performs locally adaptive thresholding across scales for spot segmentation.

The remainder of the chapter will be organized as follows. In Section 5.2, we will provide a brief review of spot detection methods. Section 3.2 will be devoted to the presentation of our multiscale spot detection method. In Section 3.3, we will report experimental results on simulated data with an objective comparative evaluation, and on real images of different kinds. After, an important application of our segmentation method is presented, we will exploit our multiscale spot detection method to construct a point-based registration framework to be applied on 2D CLEM images. Finally, Section 3.5 will contain concluding remarks.

3.1 Related work

Many efforts have been made towards automatic spot detection in images, specially in biological microscopy images. Detailed information on existing methods along with experimental comparative evaluation of several spot detection methods can be found in [7, 132, 146]. Spot detection methods can be divided in single-scale and multiscale approaches.

Single-scale methods [119, 133, 155] extract spots from an image, corresponding to one given size. The scale parameter is usually predefined according to the application. In [155] and [110], a mixture of Gaussian models is used to detect overlapping spots, while approaches based on the top hat scheme are used in [17] and [148]. Methods based on h-dome [121, 147, 165] can deal with close particles by detecting domes or local intensity maxima. In [133], a single-scale spot-enhancing filter (SEF) is presented, where the Laplacian of Gaussian (LoG) filter is used to detect spots in fluorescence microscopy images. The LoG filter enhances structures of a determined size, corresponding to the variance of the Gaussian filter involved, while smoothing the image, and removing (to a certain extent) background structures.

However, the standard deviation of the Gaussian filter within the LoG transform

needs to be adapted to the size of the particles to be detected. A statistical criterion was introduced in [7] to automatically select one single optimal scale, based on the scale-space paradigm of [80] and a discrete version of the Gaussian filter. The selected scale is the one corresponding to the maximum number of spots normalized by the number of blobs in a pure noise image at the same scale, where blobs designate local extrema in the scale-space domain.The resulting spot detection method, named ATLAS, was proven to outperform existing methods on several benchmarks. ATLAS was applied to vesicle segmentation in total internal reflection fluorescence microscopy (TIRFM) images [7].

Conversely, multiscale detection methods are able to extract spots of different sizes from an image. In [82] a general Gaussian scale-space framework was investigated to select multiple scales for blob and junction detection. [115] locally selects the most salient scale for region contour points to drive the PDE-based image segmentation. In [45] a non-linear scale-space representation employing a differential morphological decomposition, is used for multiscale corner detection. In [177], an isotropic undecimated wavelet method is designed to detect biological particles of different sizes, exploiting the wavelet multiscale product introduced in [108]. In [73], a generalized Laplacian of Gaussian allows to detect circular and elongated structures, while estimating their dimensions and orientations. In [63], a multiscale spot detection scheme, exploiting the LoG transform, is developed and used together with a multi-frame association algorithm to track virus particles. Multiscale LoG scheme was also adopted for pulmonary nodule detection in [65], with 150 predefined LoG kernels of incrementally increasing sizes. In most of these approaches, several parameters must be predetermined to optimize their performance.

We propose a new multiscale spot detection method based on the LoG transform, able to automatically select the meaningful scales in the processed image. Moreover, we will design a locally adaptive thresholding process across scales to come up with the final map of segmented spots.

3.2 Multiscale spot detection

In this section, we present our multiscale spot detection method. It is divided in two main stages: the multiscale selection step, where we recover the meaningful scales corresponding to the significant objects in the image, and the detection step, where we exploit the selected scales to compose the binary map of segmented spots of different sizes. For both stages, we will rely on the LoG transform. First, it is the main component of the

scale-space framework of [80] that we will adopt. Second, LoG filtering of the input image has been proven to be efficient enough for spot detection. LoG is a bandpass filter able to enhance spots of a specific size. For multiscale spot detection, we will combine the LoG output obtained at different scales to create a single segmented binary image.

3.2.1 A contrario selection of multiple scales

We rely on the *a contrario* approach [30], to select the meaningful scales. Briefly speaking, the *a contrario* approach can be viewed as a hypotheses test, where only the null hypothesis H_0 needs to be specified. H_0 is called the background model accounting for randomness. A structured element is likely to appear under H_0 only with a very low probability. This approach has been successful for several pattern detection problems, as motion detection [163]. To our knowledge, it is applied to scale selection for the first time. In [44], it was used to predict the detectability of spots in textured images.

Let us consider an image f over the domain $\Omega \subset \mathbf{Z}^2$, containing spots of various sizes and corrupted by Gaussian noise. The issue is how to automatically select the meaningful scales. We start from a set of scales $\mathcal{S} \subset \mathbf{R}_+^*$, as in [7], defined by $\mathcal{S} = \{s_0 r^n, n \in [0, \nu]\}$, where s_0 is taken equal to 1, r is close to 1 (e.g., 1.2), and ν depends on the range of possible scales in the given application. Given $\mathcal{S}$, we build a scale-space representation of the image following [80], with a LoG transform based on a discrete analogous of the Gaussian filter to be able to deal with arbitrary scale values. We come up with a 3-dimensional map H_f, where each slice corresponds to the LoG filtered image for a given scale $s \in \mathcal{S}$:

$$\forall (p, s) \in \Omega \times \mathcal{S}, H_f(p, s) = (K_s * f)(p, s), \tag{3.1}$$

where K_s denotes the LoG kernel of variance s. The response of a bright spot of size ς and located at point p, to the multiscale LoG transform is be minimum at $p \in \Omega$ and scale $s \in \mathcal{S}$, where s is the closest value to ς. Such a scale-space minimum is named *blob*, following [80].

To detect spots as reliably as possible, we need to find the scales at which LoG best enhances them, while reducing noise. To do so, we elaborate a probability measure to account for the ability of the LoG to distinguish noise and spots. We have no prior information on the spots, but we suppose that the noise is Gaussian. Thus, we construct a model representing the situation where no spots are present in the image (H_0 hypothesis), that is, an image containing only uncorrelated Gaussian noise. Then, we denote by N_s the random variable representing the number of blobs at scale s in such a random image.

We can assume that the probability for any point $p \in \Omega$ to be a blob at scale s follows a binomial distribution of mean μ_s. Then, the variable N_s of the number of blobs at scale s is Poisson-distributed of mean $\lambda_s = \mu_s|\Omega|$, where $|.|$ denotes the cardinality of the set. Let $\mathcal{G} = \{g_i, 1 \leq i \leq K\}$ be a set of K such random images, and let $n_s(g_i)$ be the computed number of blobs in g_i at scale s. It was shown in [7] that $n_s(g_i)$ is unchanged when adding any constant to g_i or multiplying g_i by any positive number. Therefore, we merely consider a normalized Gaussian noise, $\forall p \in \Omega, g_i(p) \sim \mathcal{N}(0,1)$. We empirically estimate λ_s as the average number of blobs at scale s in the set $\mathcal{G}$ of noise image samples:

$$\hat{\lambda}_s = \frac{1}{K} \sum_{i=1}^{K} n_s(g_i). \tag{3.2}$$

Meaningful scales in image f will be those for which the probability of the measured number of blobs in H_f is the least likely to be high under the "no-spot" H_0 hypothesis, hence, the name of *a contrario* approach. To do so, we count the number $n_s(f)$ of blobs in H_f at every scale $s \in \mathcal{S}$, and we evaluate the probability that so many blobs may exist under the "no-spot" H_0 hypothesis. We refer to it as the probability of false alarm PFA(s, f), which can be estimated as:

$$\begin{aligned} \text{PFA}(s,f) &= \mathbb{P}(N_s \geq n_s(f)) \\ &= 1 - \Phi_{\lambda_s}(n_s(f)) \approx 1 - \Phi_{\hat{\lambda}_s}(n_s(f)) \end{aligned} \tag{3.3}$$

where $\Phi_{\hat{\lambda}_s}$ is the cumulative density function (CDF) of the Poisson distribution of mean $\hat{\lambda}_s$:

$$\Phi_{\hat{\lambda}_s} = e^{-\hat{\lambda}_s} \sum_{i=0}^{n_s(f)} \frac{\hat{\lambda}_s^i}{i!}. \tag{3.4}$$

We come up with a set of probabilities $\{\text{PFA}(s,f), s \in \mathcal{S}\}$, and we can simply select the subset of ϵ-meaningful scales $\mathcal{S}^\star \subset \mathcal{S}$ as given by:

$$\mathcal{S}^\star = \{s \in \mathcal{S} \quad | \quad \text{PFA}(s,f) < \epsilon\}. \tag{3.5}$$

Let denote $|\mathcal{S}^\star| = \eta$. In practice, since we look for very low PFA(s,f), we arbitrarily fix ϵ to 0.1. Alternatively, in case we know *a priori* the number η of relevant spot sizes, we can select the scales corresponding to the η lowest PFAs.

3.2.2 Spot detection at a given scale

Once the set of scales $\mathcal{S}^\star$ is determined, we can build a spot detection binary map $\Delta_s : \Omega \to \{0,1\}$ for each scale $s \in \mathcal{S}^\star$. We will again exploit the LoG transform, since it smooths noise while enhancing spots. This will be achieved by thresholding the lowest (resp. highest) values of the corresponding LoG map $H_f(\cdot, s)$, $s \in \mathcal{S}^\star$, if spots are bright

(resp. dark) in the image. We will automatically inferred the threshold value τ_s which will be adapted for every point $p \in \Omega$ from local statistics of the LoG map $H_f(\cdot, s)$ in the vicinity of p. It is assumed that the local background in the LoG map is smooth and corrupted by a white Gaussian noise. For every point $p \in \Omega$, we estimate the local mean $\mu_s(p)$ and variance $\sigma_s^2(p)$ over a Gaussian window $W_s(p)$. The likelihood $\mathcal{L}_s$ of belonging to the background of the LoG map in the vicinity of p at scale $s \in \mathcal{S}^\star$ is then given by:

$$\mathcal{L}(p) = \Phi\left(\frac{H_f(p,s) - \mu_s(p)}{\sigma_s(p)}\right) \tag{3.6}$$

where Φ denotes the Gaussian density function. Equation (3.6) is inverted to get a threshold value below which a point is detected as spot, according to a user-selected p-value α:

$$\tau_s \quad : \quad \Omega \to \mathbb{R};\ p \mapsto \sigma(p)\phi^{-1}(\alpha) + \mu(p) \tag{3.7}$$

α is not related to the local image properties, but only depends on the reliability level required in the application. It can be fixed by the user once for all for the whole experiment. In our case, we fixed α to 1e-3 for all experiments.

3.2.3 Multiscale spot detection

When detecting spots of different sizes, it is important to correctly combine results of spot segmentation obtained at different scales. Similarly to [63], we adopt a coarse-to-fine approach. The scheme is defined as follows. Let us consider the input image f and the set of the η meaningful scales selected at the first stage of our method, $\mathcal{S}^\star = \{s_l, l = 1, ..., \eta\}$, ranked in decreasing order. We define a filtered image $\psi(p, s_l)$ with $s_l \in \mathcal{S}^\star$ as:

$$\psi(p, s_l) = H_f(p, s_l)\Delta_{s_{l-1}}(p) \tag{3.8}$$

where $\Delta_{s_{l-1}}(p)$ is the spot detection binary map computed at scale s_{l-1}. For $l = 1$, corresponding to the coarsest scale or level, by definition we take $\Delta_{s_0}(p) = 1, \forall p \in \Omega$. The spot detection binary map at a given scale operates as a mask for spot detection at the subsequent finer scale. Indeed, this way, spurious spot detections are avoided at coarser scales, while at finer scales spots in close proximity can be further resolved. Thus, we can perform the multiscale spot detection map for several scales automatically selected on the processed image f, and there is just one single user-friendly parameter to set for the segmentation step, that is, the p-value α.

In contrast, the max, min and number of scales (for a regular scale sampling) have to be predefined by the user in [63], along with parameters in the thresholding and masking operations. This is also the case for [177], where the user has to set the threshold, max and min scales and the false discovery rate.

In addition, for comparison purpose we have implemented a variant, denoted AS-MSSEF, combining the coarse-to-fine spot detection framework of [63] with our automated scale selection. It will allows us to evaluate the improvement brought by our automated multiscale selection paradigm on its own.

The overall multiscale spot detection framework is illustrated in Fig. 3.1, where the different stages for generating the final binary map are delineated.

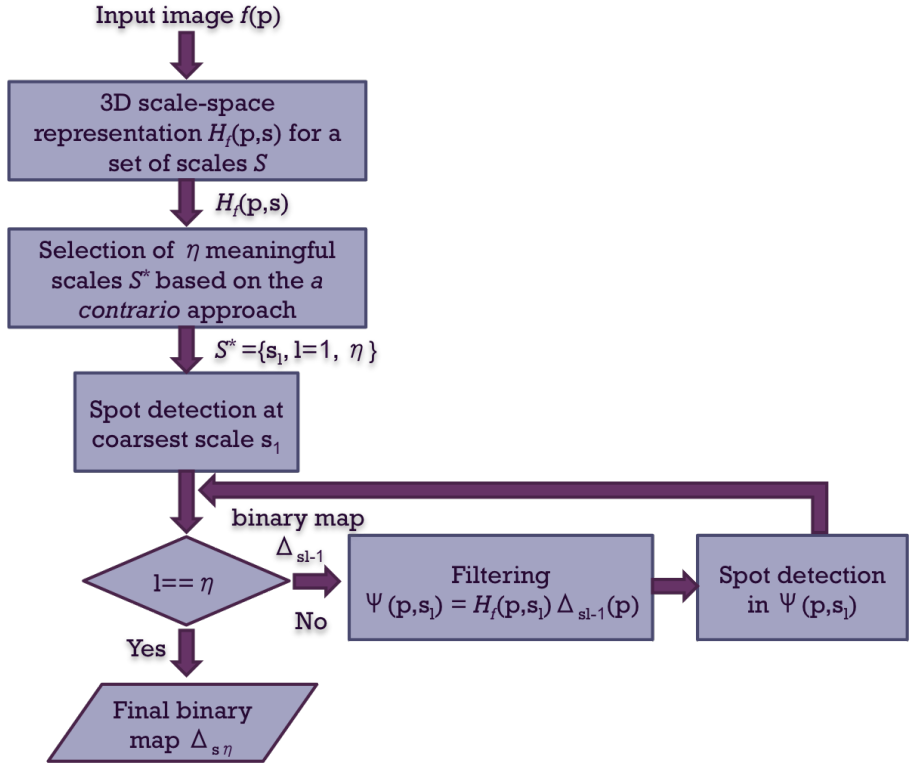
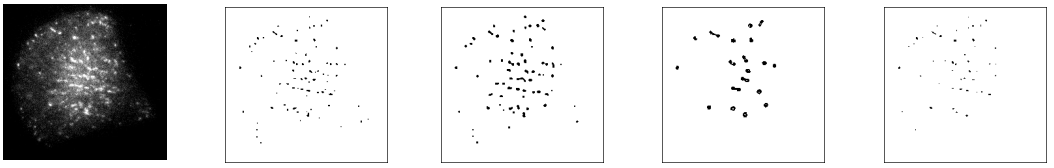


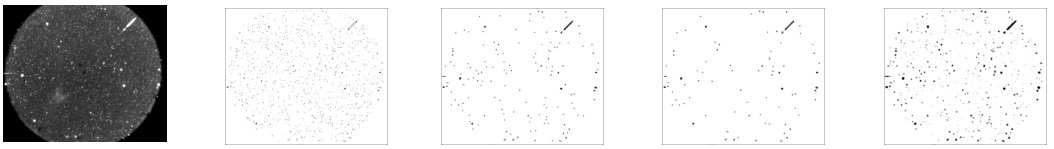
Fig. 3.1: *Multiscale spot segmentation pipeline.*

3.3 Experimental results

We have evaluated the performance of our method on both synthetic and real images. We have compared our multiscale method with other multiscale methods: MSSEF [63], MS-VST [177], and our variant AS-MSSEF.



(a) Input image (b) Our method (c) Variant AS-MSSEF (d) MSSEF [63] (e) MS-VST [177]



(a) Input image (b) Our method (c) Variant AS-MSSEF (d) MSSEF [63] (e) MS-VST [177]

Fig. 3.2: *Spots segmented (in black) by the four compared methods on a real TIRFM cell image (top row), and on a real astronomy image (bottom row).*

3.3.1 Simulated data

We report two series of experiments, each one comprising 20 images containing three different spot sizes, with respectively $S = \{2.6, 4, 6\}$, and $S = \{3, 5, 7\}$. 100 Gaussian spots are randomly sampled in each simulated image over a uniform zero-valued background. We added Gaussian noise to the image, with $\mu = 2$ and $\sigma = 0.6$, the spot peak intensity value is 10. The objective evaluation is divided in three steps: multiple scale selection, spot detection (spot center location), and spot segmentation. For the two last steps, we compare our method to the three other methods.

To assess multiscale selection, we compute precision and recall scores. 100% recall means that all the true scales are correctly selected. 100% precision means that all the selected scales correspond to true scales. Since we start with a set of 18 predefined scales $\mathcal{S} = \{1, 1.2, 1.44, ..., 18.49, 22.19\}$, a true scale is stated as recovered if the scale the closest to it is selected among the tested ones. In the first experiment, precision of our multiscale selection method amounts to 100% and recall to 90%. In the second experiment, we get precision of 95% and recall of 95%.

To evaluate spot detection, we compute F-measure scores on the binary maps supplied by the tested methods. Detection is stated as correct if the distance between the detected spot center and its corresponding ground truth center is less than four pixels following [147]. We report statistics on results of the two experiments in Table 3.1. For the forty images, detected spot locations were the most accurate with our method. In several images, the variant AS-MSSEF has the same performance as our method, indicating that our multiscale selection has a strong impact on the whole process.

	Our method		AS-MSSEF		MSSEF		MS-VST	
	F-m.	Jacc.	F-m.	Jacc.	F-m.	Jacc.	F-m.	Jacc.
Mean	**0.982**	**0.724**	0.978	0.664	0.937	0.645	0.961	0.357
SD	**0.008**	0.052	0.009	0.068	0.037	0.048	0.015	**0.019**
Min.	**0.966**	**0.641**	0.955	0.565	0.866	0.589	0.926	0.331
Max.	**0.995**	**0.790**	**0.995**	0.745	0.989	0.708	0.989	0.386

Table 3.1: *Statistics over the 40 simulated images of the two experiments on the F-measures and Jaccard index, obtained with the four compared methods.*

Regarding spot segmentation assessment, we compute the overlap between the segmented spots in the binary map delivered by each method and the ground truth using the Jaccard index, defined as $J(A,B) = |A \cap B|/|A \cup B|$. As reported in Table 3.1, our method yields the best scores, meaning that our method better recovers size and shape of spots.

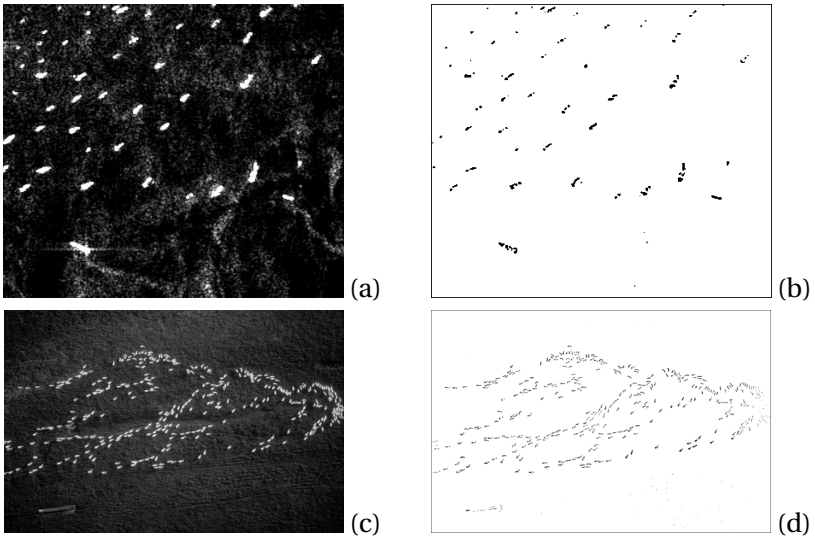


Fig. 3.3: *Spots segmented (in black) by our method on a SAR satellite image including ships (top row) and an aerial color image depicting a sheep herd in a meadow (bottom row).*

3.3.2 Real images

We have also carried out experiments on a set of diverse real images. Since there is no ground truth available, we rely only on visual assessment of spot segmentation, when comparing our method to others. We report two of them: on a total internal reflection fluorescence microscopy (TIRFM) cell image and on an astronomical image. Fig.3.2 contains the TIRFM input image (resp. astronomy image) and the maps of segmented spots

obtained with our method, the variant AS-MSSEF, MSSEF and MS-VST. Our method selects 5 scales, $S^\star = \{1.728, 2.073, 2.488, 3.583, 6.191\}$, in the first real image (resp. 5 scales, $S^\star = \{1.440, 1.728, 5.159, 8.916, 15.407\}$, in the second one) as representative scales of the objects present in the image. Clearly, our method outperforms the three others, since we are closer to the right amount of objects to detect (after visual examination) and are able to more accurately segment them.

We report additional results in Fig.3.3, showing that our method can handle very different types of images with the same accuracy. Regarding computation time, it takes $1.25s$ to select the meaningful scales and segment the associated spots in an image of size 500×500 on a laptop with 2,8 GHz Intel Core i7 processor and 16 GB memory.

3.4 Point-based 2D CLEM registration refinement

A relevant application of our multiscale spot segmentation method is its use to segment 2D CLEM images with the purpose of implementing a point-based registration. Blob-like objects of different sizes can be expected to be present in 2D CLEM images, due to the nature of the biological samples from which they are acquired. These blobs can be exploited to register the electron microscopy (EM) and light microscopy (LM) images leading to the need for a method to locate and extract them.

First, we applied our multiscale segmentation method to the unprocessed EM and LM images. In Figures 3.5 and 3.4 we can observe the results of our multiscale spot detection method on two CLEM images. After visual analysis, it becomes apparent that the existence of highly similar configurations of sets of blobs would probably cause existing point-based registration methods to fail, since the correct matching of corresponding spots would be difficult to implement in a unsupervised way. This is illustrated in Figure 3.5, where the distribution of the spots inside the regions outlined in red is similar to the distribution of the spots inside the region of interest (ROI) outlined in blue. In addition, the transformation mapping the spot distribution in the LM-ROI onto the spot distribution in the corresponding EM-ROI is obviously very large.

Although the segmentation of blobs in CLEM images may not be directly exploitable for automated registration, it can be employed further in our CLEM registration process as a final refinement step. Indeed, applying our multiscale spot segmentation method on the aligned CLEM images can help to improve their final registration and overlay. In this part of our work, we apply our segmentation method on the registered EM and LM

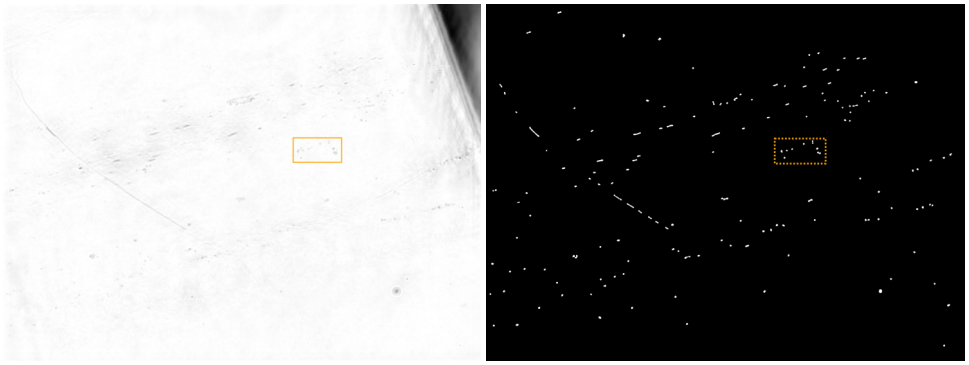


Fig. 3.4: *Multiscale spot segmentation (right) of the LM image(left) of experiment 1. ROI is enclosed in the orange rectangle.*

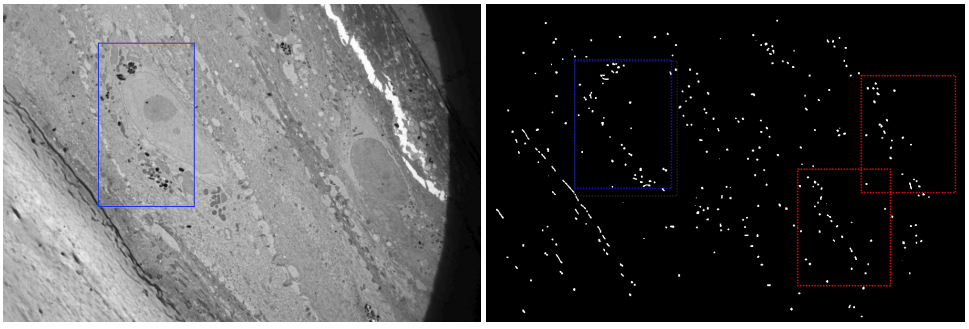


Fig. 3.5: *Multiscale spot segmentation (right) of the EM image(left) of experiment 1. ROI is enclosed in the blue rectangle. Red rectangles denote regions where the blob distribution is visually similar to the one inside the ROI.*

patches obtained with the registration framework described in Chapter 2. Then, employing a point-based algorithm, we compute an affine transformation. We can expect that a point-based method, when automatically exploitable, can be more accurate than an intensity-based method.

The location of the segmented blobs is given by their centroid coordinates. The centroids of the spots are extracted using connected components [135]. Connected components is a widely known algorithm for the labeling of an image. It is used in image processing to detect connected regions in binary images, although it can also be used in color images. All the pixels in an image are grouped into regions of connected pixels and assigned a label. Once the different connected components of the registered EM and LM binary images are identified, the centroid of each component is computed using the moments of the binary images.

The 2D image moments are defined as:

$$M_{i,j} = \iint_{\Omega} x^i y^j I(x,y)\, dx\, dy \tag{3.9}$$

From 3.9, the centroid is computed as $\{\overline{x}, \overline{y}\} = \{M_{10}/M_{00}, M_{01}/M_{00}\}$.

Registration refinement is achieved by implementing a Random sample consensus (RANSAC) image registration procedure. RANSAC [34] is an iterative algorithm to estimate the parameters of a mathematical model to register two sets of data containing outliers ($\mathcal{S}_0$ and $\mathcal{S}_1$). RANSAC algorithm can be described by steps of the pseudo-code of Algorithm 1.

Two sets of matching pairs of points are used to estimate a geometric transformation. Each set of points corresponds to the coordinates of the centroids of the detected spots in the EM and LM images. The detected spots are matched using a nearest neighbor approach. In an iterative way, random points are selected from each set and are used to estimate the affine transformation. The estimated transformation is applied to the rest of the points on the set corresponding to the source image and its fitness is evaluated by measuring the distance between the transformed points and their corresponding match in the set of points associated to the target image.

Algorithm 1 RANSAC algorithm

procedure RANSAC($\mathcal{S}_0, \mathcal{S}_1, N_{\text{iter}}, \eta_{\text{in}}, Trsh$)
  $i \leftarrow 0$ ▷ Iteration number
  $N_{\text{best}} \leftarrow 0$ ▷ Best inlier number
  $\mathbf{M}_{\text{best}} \leftarrow 0$ ▷ Best estimated transformation
  while ($i < N_{\text{iter}}$) **do** ▷ Iterative process
    $N_{\text{in}} \leftarrow 0$ ▷ Inlier points number
    $\chi \leftarrow \emptyset$
    $\mathbf{s}_0 \in \mathcal{S}_0$, $\mathbf{s}_1 \in \mathcal{S}_1$ ▷ Select randomly N probable inlier points for both data sets
    $\mathbf{M}$ = transformation($\mathbf{s}_0$, $\mathbf{s}_1$) ▷ Estimate the possible transformation using $\mathbf{s}_0$ and $\mathbf{s}_1$
    for all ($\mathbf{p}_0 \in \mathcal{S}_0 | \mathbf{p}_0 \neq \mathbf{s}_0$) and ($\mathbf{p}_1 \in \mathcal{S}_1 | \mathbf{p}_1 \neq \mathbf{s}_1$) **do**
      $\mathbf{p}_2$ =transform($\mathbf{p}_0$,$\mathbf{M}$)
      $\mathbf{d} = \mathbf{p}_1 - \mathbf{p}_2$
      if ($|\mathbf{d}| < Trsh$) **then**
        $\chi \leftarrow \mathbf{p}_0, \mathbf{p}_0$ ▷ $\mathbf{p}_0$ and $\mathbf{p}_1$ are inlier points
        $N_{in} \leftarrow N_{in} + 1$ ▷ Increase the number of inlier points
    if ($N_{\text{in}} \geq \eta_{\text{in}} N_{\text{pts}}$) and ($N_{\text{in}} > N_{\text{best}}$) **then**
      $N_{\text{best}} \leftarrow N_{\text{in}}$
      $\mathbf{M}_{\text{best}} \leftarrow \mathbf{M}$
    $i \leftarrow i + 1$
  return $\mathbf{M}_{\text{best}}$

where N_{iter} is the maximum number of iterations, $\eta_{\text{in}} \in (0,1]$ is the ratio of minimum inlier points and $Trsh$ is the threshold of minimum distance to consider an inlier point.

In our case, we propose to estimate an affine model such that the data set $\mathcal{S}_0$, corresponding to the centroids of the segmented blobs of the LM pre-registered ROI, matches the data set $\mathcal{S}_1$, corresponding to the centroids of the segmented blobs of the EM ROI. We randomly select three points as $\mathbf{s}_0$ and $\mathbf{s}_1$:

$$\mathbf{s}_0 = \begin{bmatrix} x_0 & y_0 \\ x_1 & y_1 \\ x_2 & y_2 \end{bmatrix} \tag{3.10}$$

$$\mathbf{s}_1 = \begin{bmatrix} x'_0 & y'_0 \\ x'_1 & y'_1 \\ x'_2 & y'_2 \end{bmatrix} \tag{3.11}$$

The affine transformation is defined as a matrix $\mathbf{M} \in \mathbb{R}^{2\times 3}$:

$$\mathbf{M} = \begin{bmatrix} a & b & c \\ d & e & f \end{bmatrix} \tag{3.12}$$

We applied our point-based registration method directly on the spot maps shown in Figures 3.4 and 3.5. Knowing that the initial misalignment is very large, it is likely that the estimation of the geometric transform between these images will not yield good results, since the matching of the spots is likely to be incorrect. The resulting overlay is shown in Figure 3.6, where the ROI in the EM image is framed in blue and the ROI in the LM image is framed in red. As expected, the registration is incorrect, suggesting that the direct point-based registration between the segmented input EM and LM images might not be able to generate satisfactory results. Then, point-based registration is only applicable as a refinement step in an automated framework, and an intensity-based registration framework is more suitable for the correct alignment of EM and LM images.

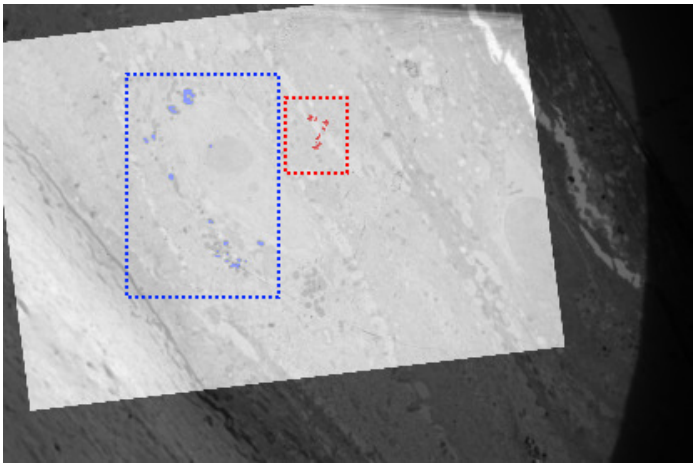


Fig. 3.6: *Overlay of the EM and LM images after direct point-based registration computed from the spots maps of Figures 3.4 and 3.4.*

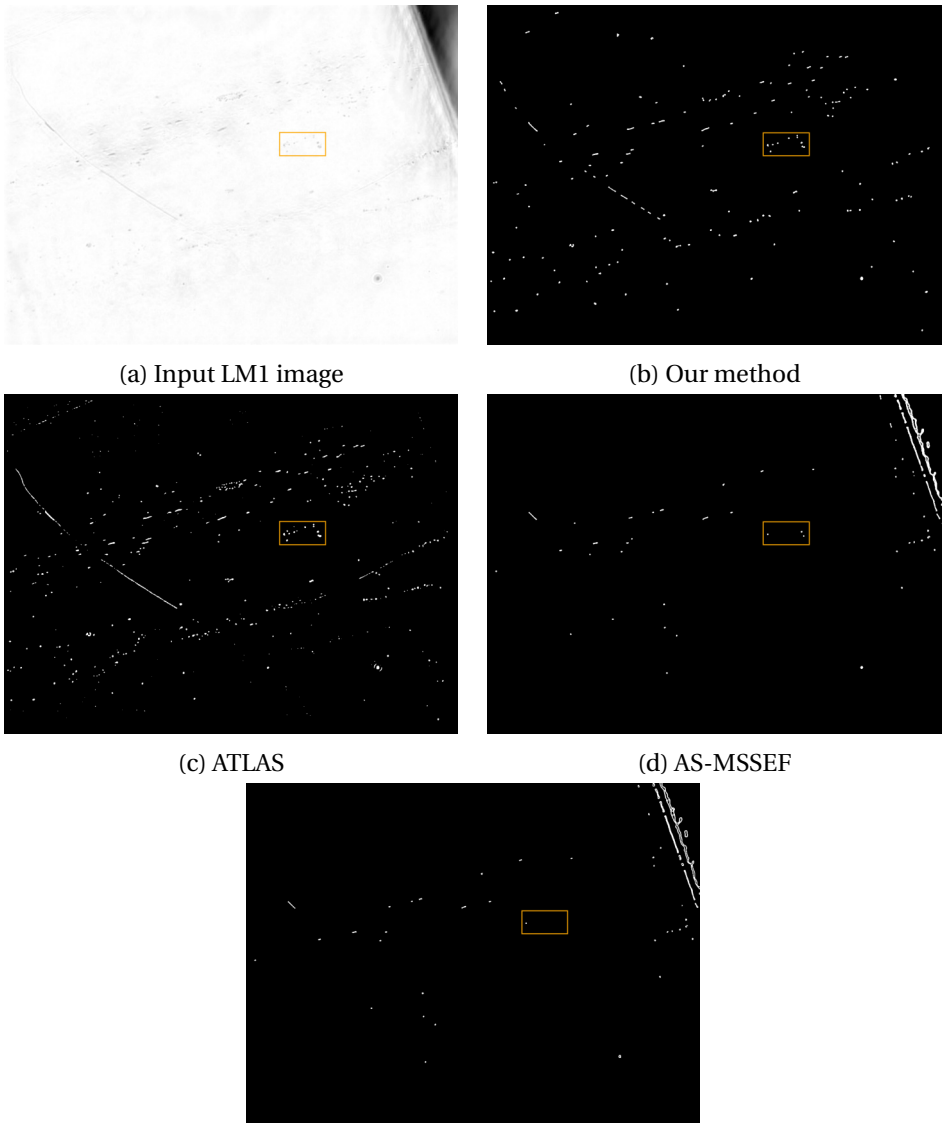
The method was tested on four 2D CLEM pre-registered image pairs, and the results are presented in the next subsection.

3.4.1 Spot segmentation on real 2D CLEM datasets

To demonstrate the performance of our multiscale spot segmentation method, we apply it on a subset of the 2D CLEM datasets. Figures 3.7 to 3.16 present for several examples the unprocessed whole EM and LM images and the segmented spots given by our method, along with a visual comparison with ATLAS (single scale spot detection), MSSEF and the variant AS-MSSEF.

Figure 3.7 displays the input LM image of experiment 1, where the ROI is enclosed in the orange frame. From left to right and top to bottom, we can observe the binary maps

obtained with our multiscale spot segmentation method, the single scale segmentation method ATLAS [7], the variant of our method AS-MSSEF, and the segmentation given by MSSEF. The ROI is enclosed in orange in all of these binary maps. By visual examination, we can notice that the most distinct blobs within the ROI of the LM image are correctly segmented by both our multiscale method and the single scale method ATLAS, however, the single scale method also includes many detections that can be considered as noise. Both AS-MSSEF and MSSEF fail at segmenting all the blobs within the ROI, as well as other noticeable blobs outside the ROI.



(a) Input LM1 image

(b) Our method

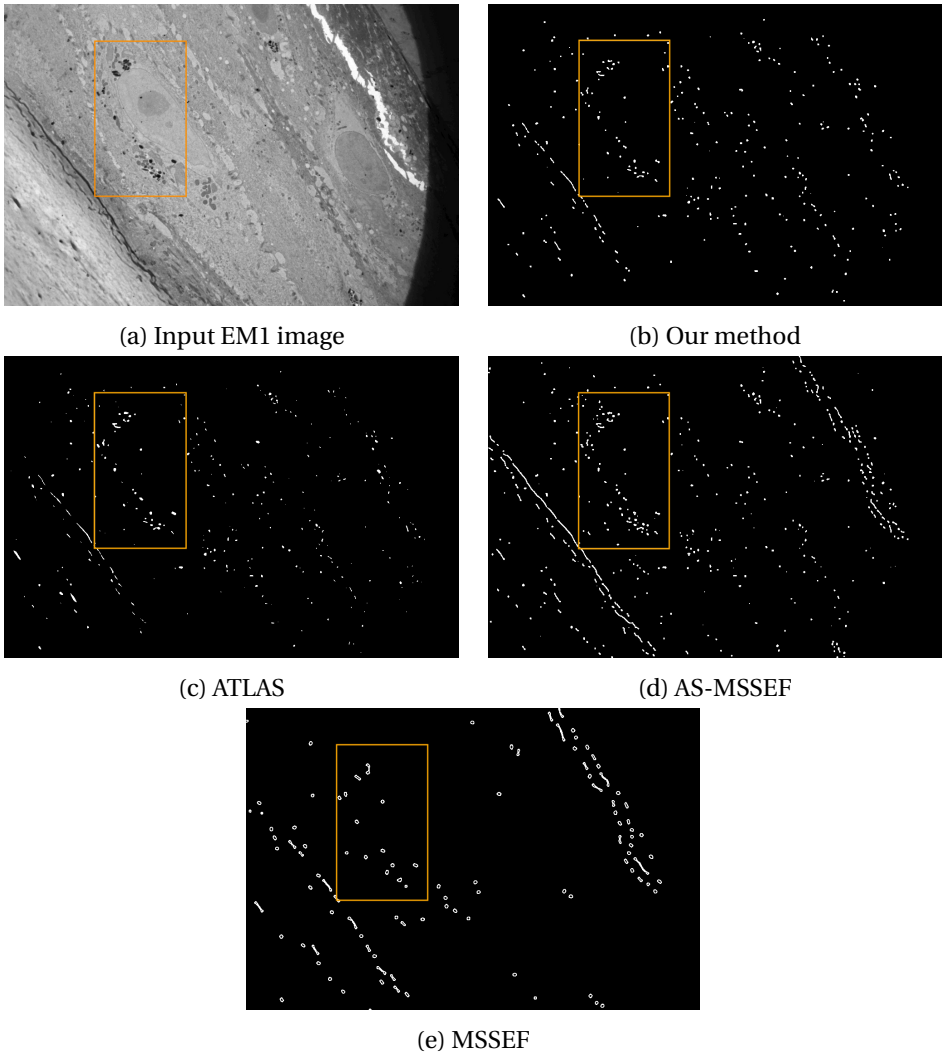
(c) ATLAS

(d) AS-MSSEF

(e) MSSEF

Fig. 3.7: *From top to bottom and left to right: input LM image of experiment 1, spot segmentation using respectively our method, ATLAS, AS-MSSEF and MSSEF.*

The input EM image of experiment 1 is displayed in Figure 3.8 along the binary maps given by our multiscale spot segmentation method, ATLAS, AS-MSSEF and MSSEF (from top to bottom, left to right). The ROI in all images is enclosed in the orange frame. Similar to the results of the previous LM image of the same experiment 1, both our method and ATLAS correctly segment the noticeable blobs inside the ROI, however, ATLAS displays more spurious detections, which can affect the spot-based registration refinement. The variant AS-MSSEF also detects the noticeable blobs of the EM-ROI but produces more noise than the previous two methods. MSSEF detects some of the distinct blobs within the EM-ROI but does not correctly preserve their size or shape.



(a) Input EM1 image

(b) Our method

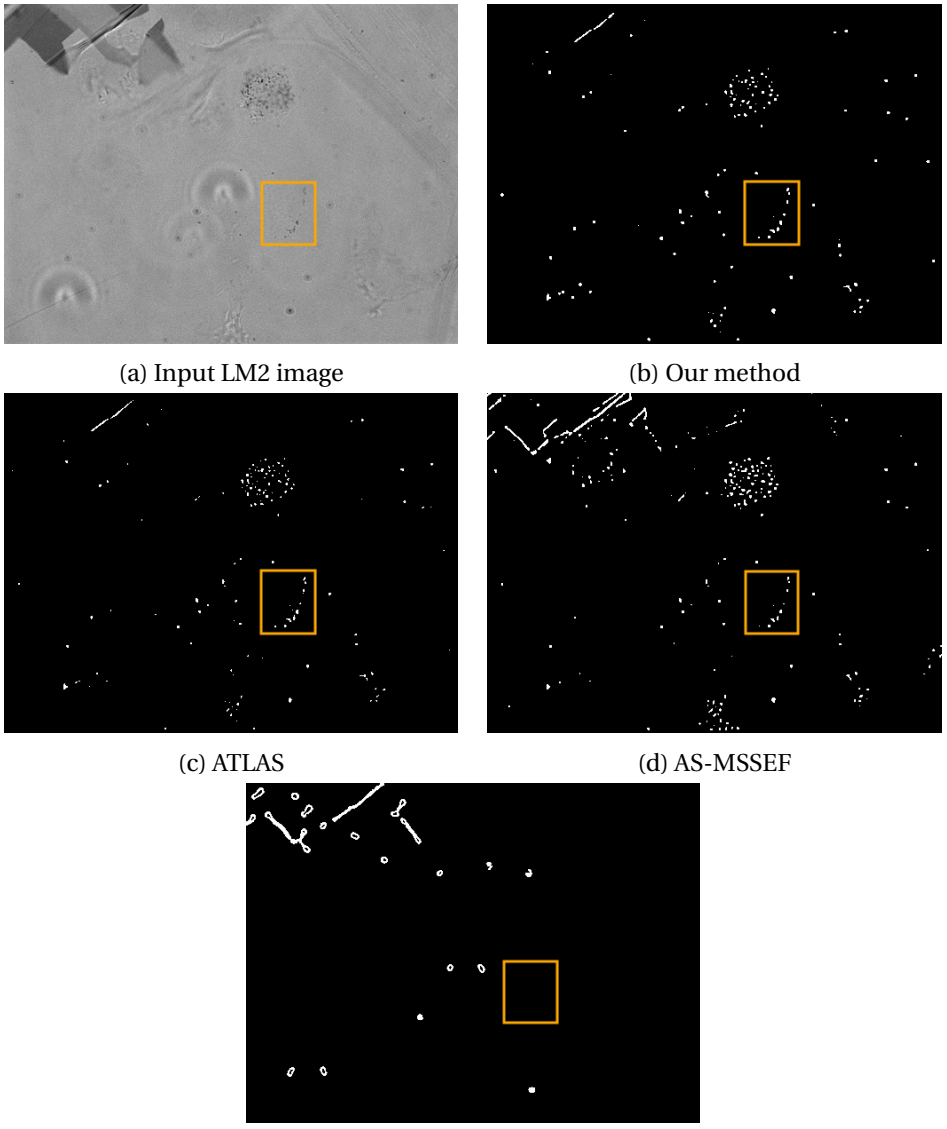
(c) ATLAS

(d) AS-MSSEF

(e) MSSEF

Fig. 3.8: *From top to bottom and left to right: input EM image of experiment 1, spot segmentation using respectively our method, ATLAS, AS-MSSEF and MSSEF.*

Figure 3.9 contains the input LM image corresponding to the second experiment, and the spot maps detected respectively with our method, ATLAS, AS-MSSEF and MSSEF. In this case, our method, ATLAS and AS-MSSEF yield similar spots maps, with the three of them extracting correctly the blobs of interest within the ROI, enclosed in orange in all images. MSSEF does not detect any of the blobs within the ROI, most likely due to its inability to automatically adapt the range of scales it employs to segment an image.



(a) Input LM2 image

(b) Our method

(c) ATLAS

(d) AS-MSSEF

(e) MSSEF

Fig. 3.9: *From top to bottom and left to right: input LM image of experiment 2, spot segmentation using respectively our method, ATLAS, AS-MSSEF and MSSEF.*

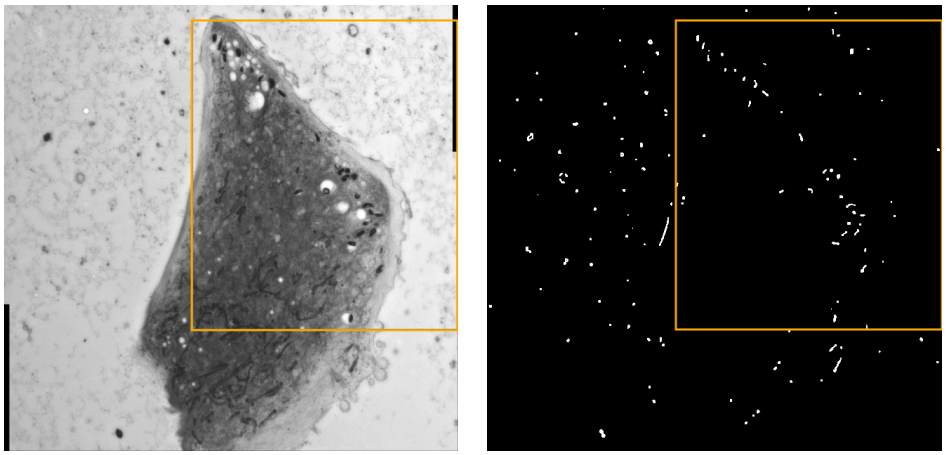
For the EM image of experiment 2, Figure 3.10 displays the input EM image and the outputs given by the four compared spot segmentation methods. In this example, our multiscale spot segmentation method clearly outperforms all the other segmentation methods, segmenting and preserving the right size and shape of the blobs of interest inside the ROI while discarding small noisy spots.

It is more difficult to visualize the spots inside the ROI framed in orange in the input LM image of experiment 3 in Figure 3.11a. However, we can observe in the spots maps shown in Figures 3.11b - 3.11d that the extracted spots pinpoint to the location of the spots in the input LM. Our method, ATLAS, and AS-MSSEF generate similar results, and our method is clearly able to preserve the size and shape of the spots of interest while eliminating noisy points. AS-MSSEF misses several spots. As in the input LM image of experiment 2, MSSEF fails to segment the spots of interest.

For the EM image of experiment 3 (Fig. 3.12a), our method, ATLAS and AS-MSSEF present similar results, however, our method once again is able to extract the spots with the right shape and size. ATLAS and AS-MSSEF have slightly worse results and MSSEF fails in both detecting many of the spots of interest and preserving their size and shape.

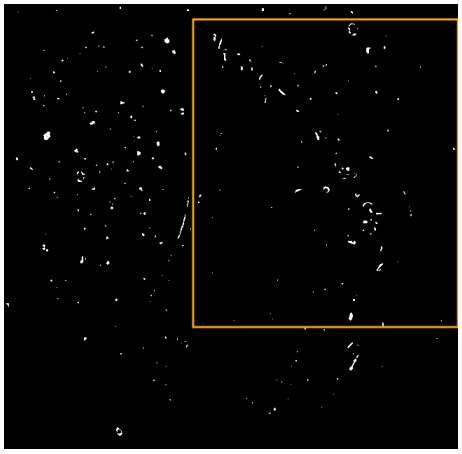
In experiment 4, for both EM and LM images shown in Figures 3.13a and 3.14a, our method is able to extract the blobs of interest inside the ROI, followed closely by ATLAS and AS-MSSEF, both also detecting the majority of the spots of interest. When segmenting the EM image, our method misses some spots in the upper left part of the image, which are correctly detected by AS-MSSEF. As in the previous experiments, MSSEF omits some detections for both EM and LM images, and in the case of the EM image, it loses important morphological information.

In the last experiment, both EM and LM images (Fig. 3.15a and Fig. 3.16a) are correctly segmented by our method, as well by ATLAS and AS-MSSEF. In the two cases, our method delivers good extraction of the spots present in the image, eliminating spots that could be considered noisy. ATLAS has a very close performance to our method, and AS-MSSEF has some blobs of big size that could be separated in smaller blobs. In both EM and LM images, MSSEF detects some of the spots of interest but misses many important ones, and fails to preserve both the size and shape of the spots. Once again, this is probably due to the fixed range of scales that are used to detect and segment the images.

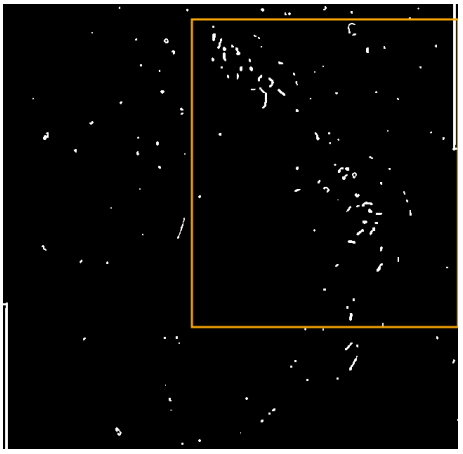


(a) Full EM2

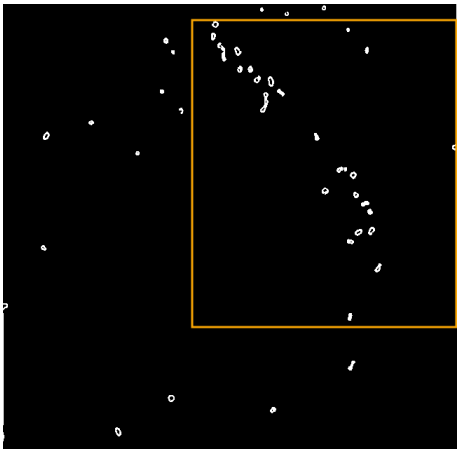
(b) Our method



(c) ATLAS

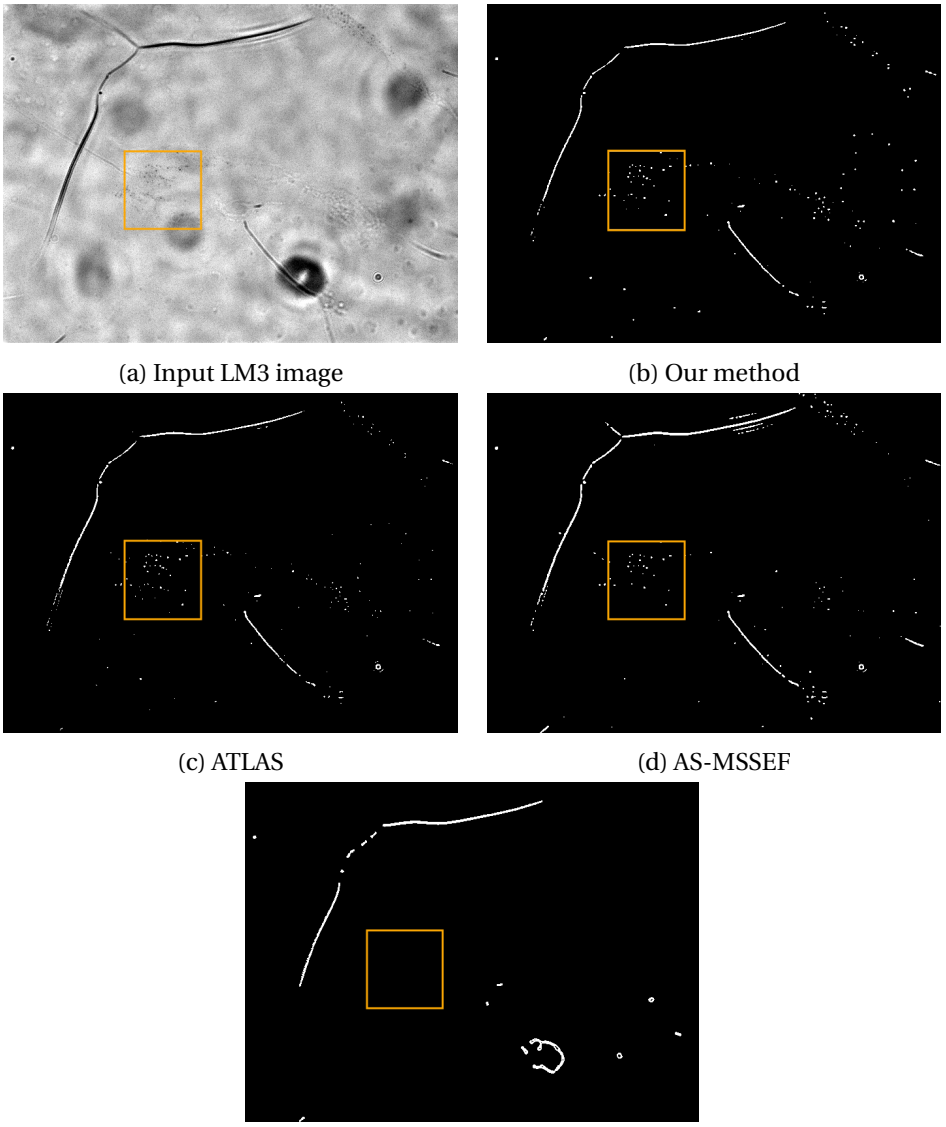


(d) AS-MSSEF



(e) MSSEF

Fig. 3.10: *From top to botton and left to right: input EM image of experiment 2, spot segmentation using respectively our method, ATLAS, AS-MSSEF and MSSEF.*



(a) Input LM3 image

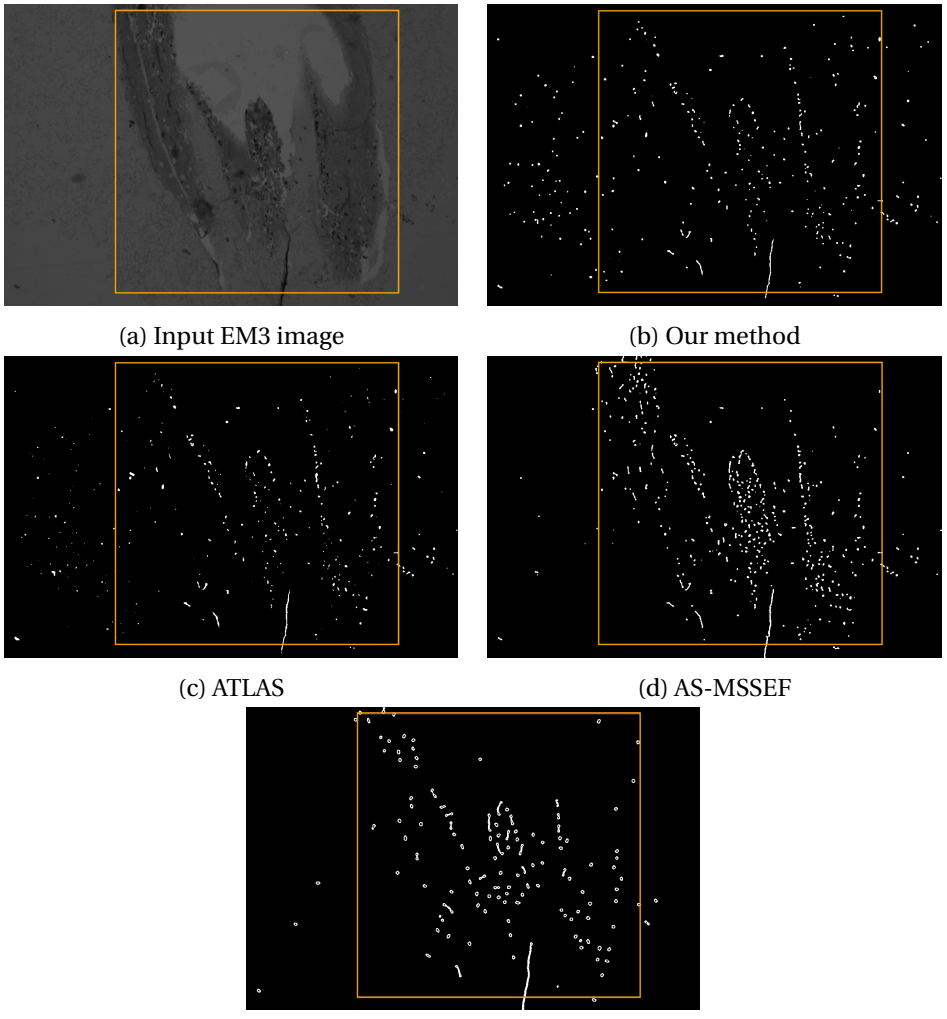
(b) Our method

(c) ATLAS

(d) AS-MSSEF

(e) MSSEF

Fig. 3.11: *From top to botton and left to right: input LM image of experiment 3, spot segmentation using respectively our method, ATLAS, AS-MSSEF and MSSEF.*



(a) Input EM3 image

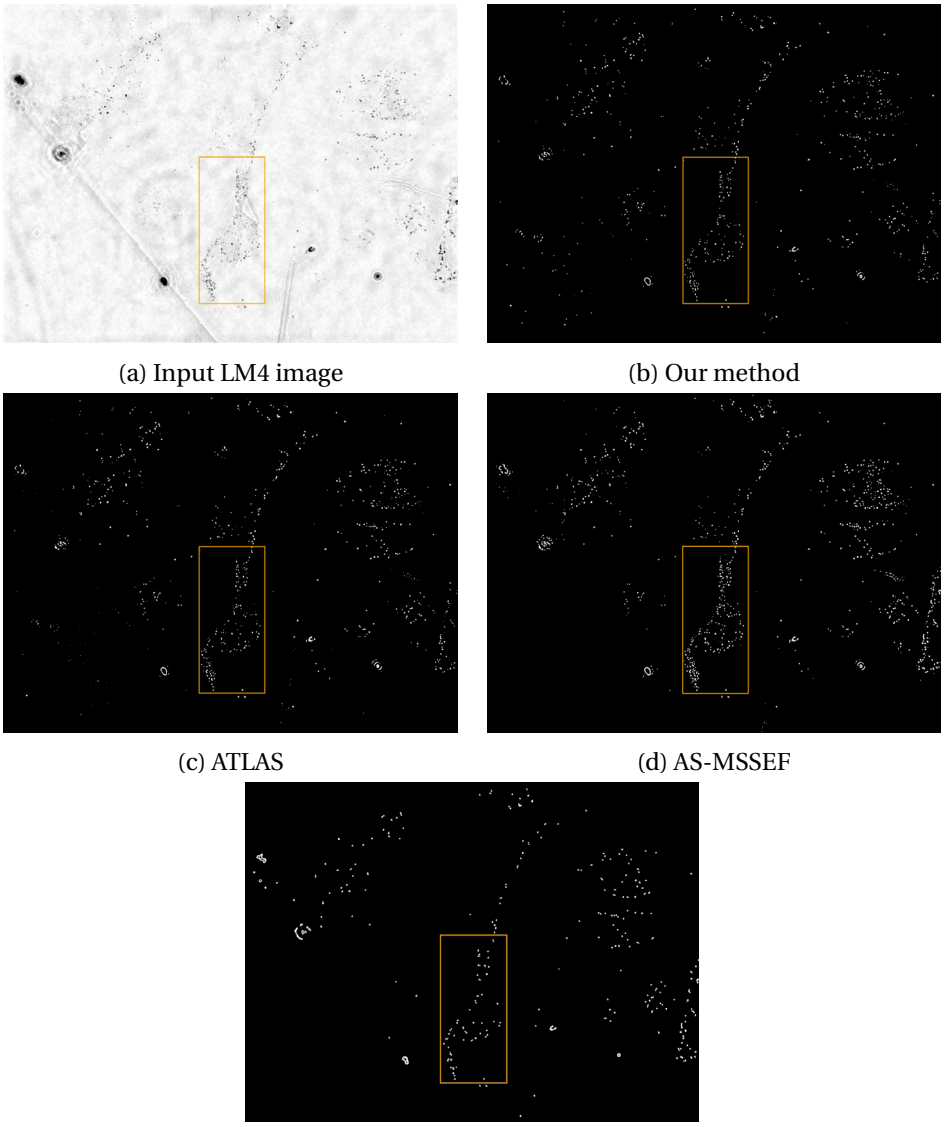
(b) Our method

(c) ATLAS

(d) AS-MSSEF

(e) MSSEF

Fig. 3.12: *From top to bottom and left to right: input EM image of experiment 3, spot segmentation using respectively our method, ATLAS, AS-MSSEF and MSSEF.*



(a) Input LM4 image

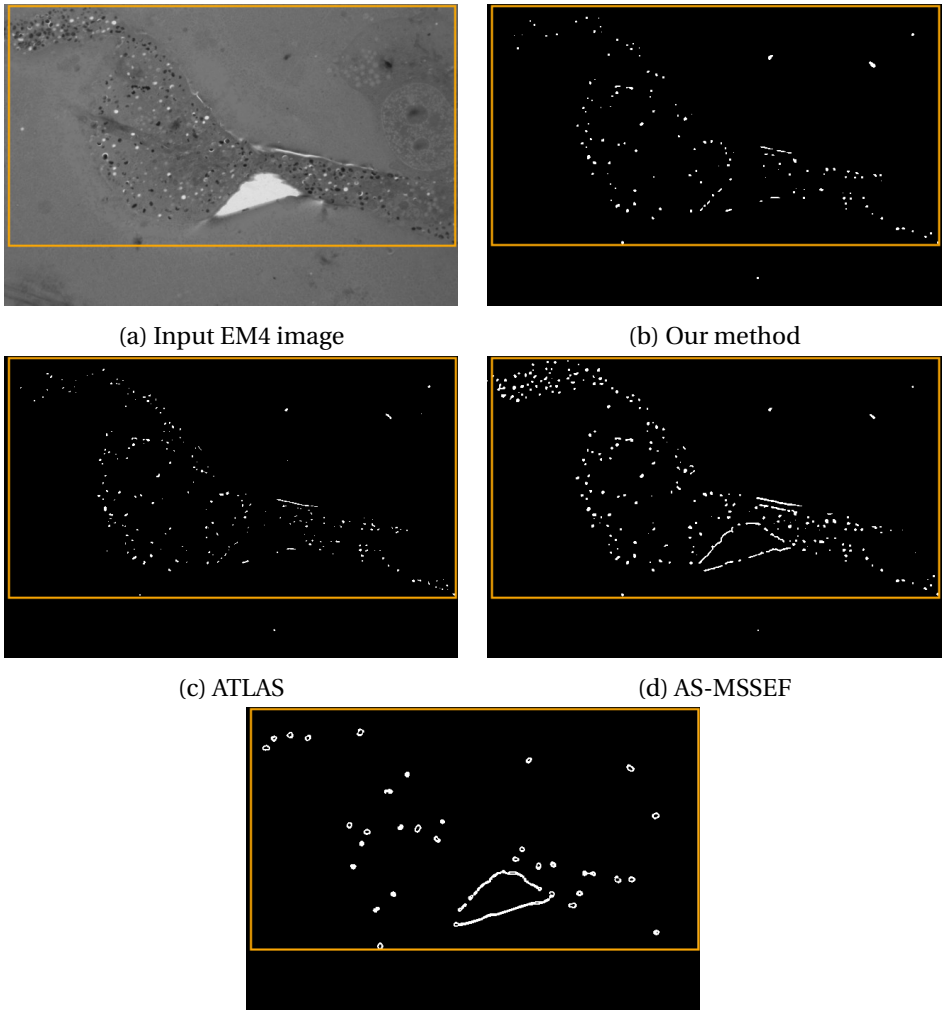
(b) Our method

(c) ATLAS

(d) AS-MSSEF

(e) MSSEF

Fig. 3.13: *From top to bottom and left to right: input LM image of experiment 4, spot segmentation using respectively our method, ATLAS, AS-MSSEF and MSSEF.*



(a) Input EM4 image

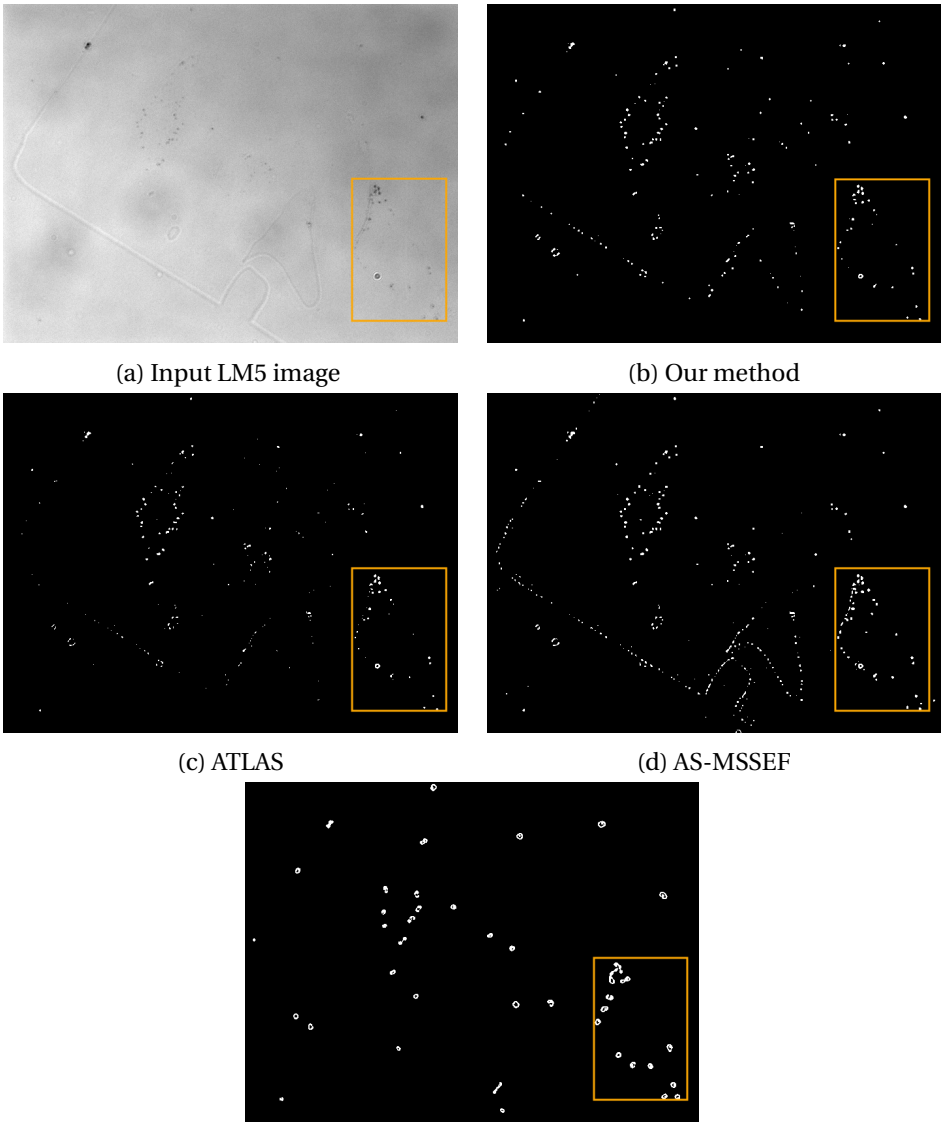
(b) Our method

(c) ATLAS

(d) AS-MSSEF

(e) MSSEF

Fig. 3.14: *From top to bottom and left to right: input EM image of experiment 4, spot segmentation using respectively our method, ATLAS, AS-MSSEF and MSSEF.*



(a) Input LM5 image

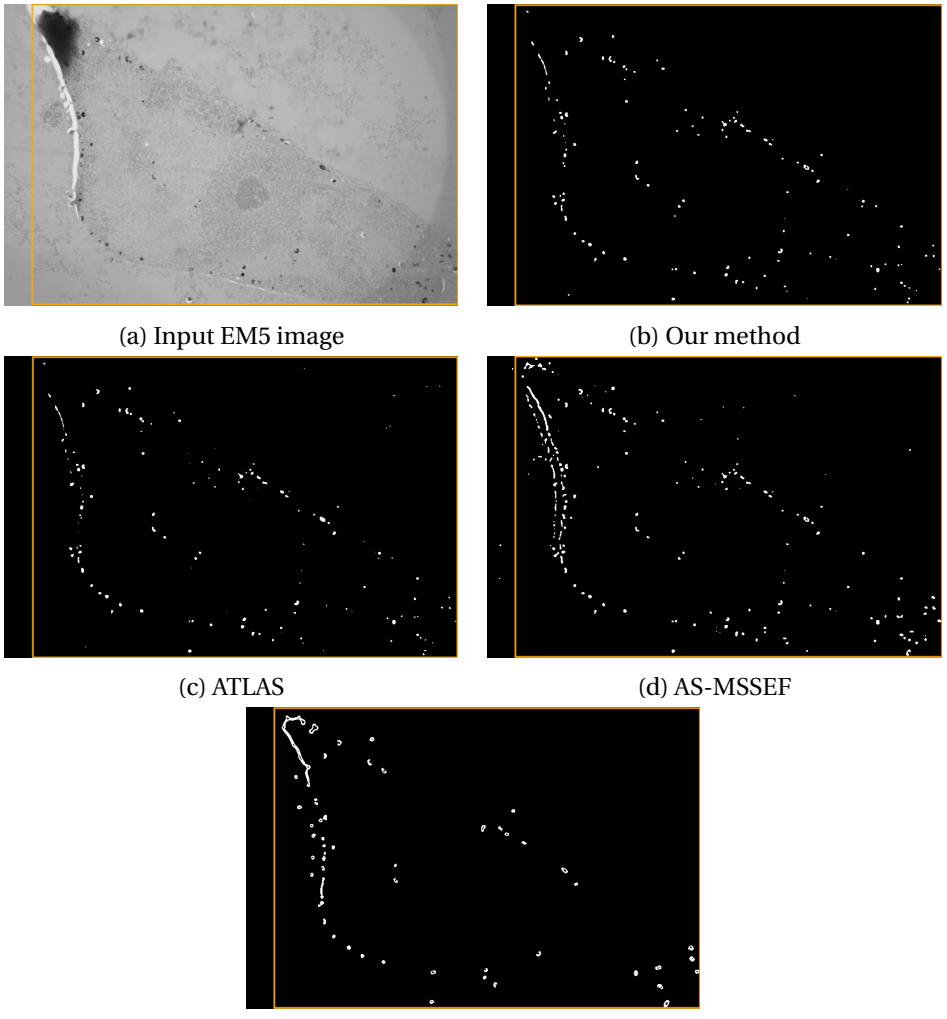
(b) Our method

(c) ATLAS

(d) AS-MSSEF

(e) MSSEF

Fig. 3.15: *From top to bottom and left to right: input LM image of experiment 5, spot segmentation using respectively our method, ATLAS, AS-MSSEF and MSSEF.*



(a) Input EM5 image

(b) Our method

(c) ATLAS

(d) AS-MSSEF

(e) MSSEF

Fig. 3.16: *From top to bottom and left to right: full EM image of experiment 5, spot segmentation using respectively our method, ATLAS, AS-MSSEF and MSSEF.*

3.4.2 Point-based registration refinement

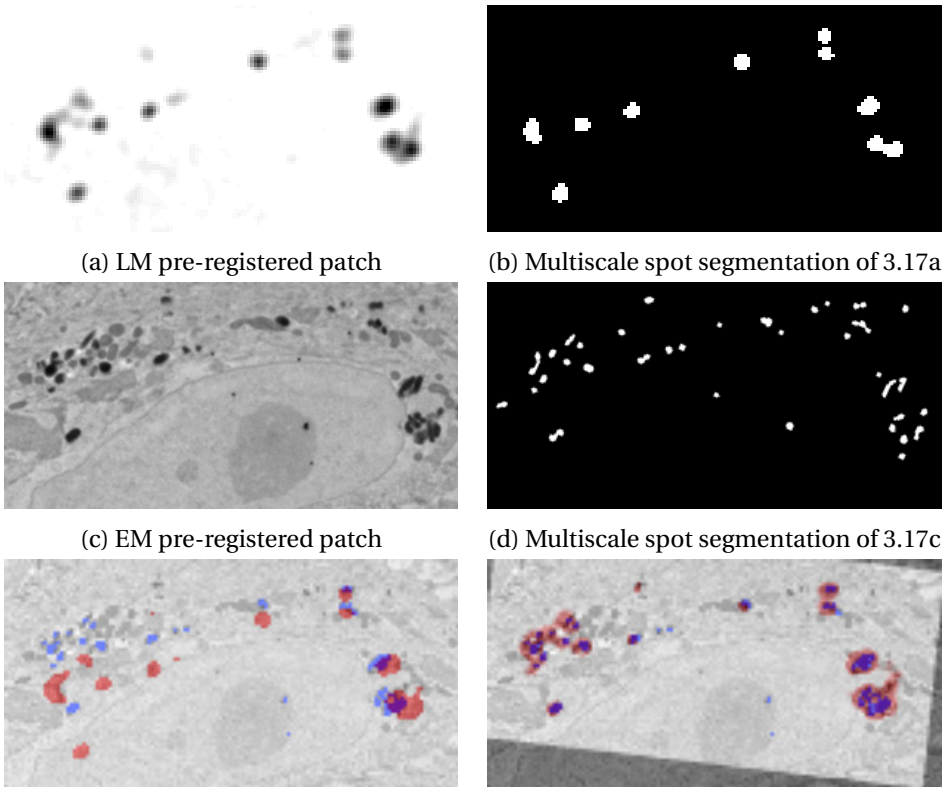
Results of the point-based registration refinement in 2D CLEM images are presented in this subsection. The point-based registration algorithm was implemented to improve the alignment between pre-registered EM and LM patches, exploiting our multiscale spot detection method and the fact that spots of interest are constantly present in these 2D CLEM images. This is a prerequisite to run the point-base registration. We take now as input the pre-registered EM and LM patches issued from the intensity-based registration described in Chapter 2, and we apply our spot segmentation method on both patches.

In Figures 3.17 to 3.21 the spot segmentation of the pre-registered patches and their overlay are displayed, as well as the resulting refinement after applying the point-based registration.

Through visual analysis of Figures 3.17 to 3.21, we can assess than the five experiments the registration between the pre-registered LM and EM patches is noticeably improved, demonstrating that our point-based registration refinement framework using multiscale spot segmentation can overcome small and large misregistrations.

In experiment 1 (Fig. 3.17), experiment 2 (Fig. 3.18), experiment 3 (Fig. 3.19), and experiment 5 (Fig. 3.19), we can observe that the initial misalignment between the EM and LM patches is small and the refinement step aligns almost perfectly the images. In contrast, in experiment 4 (Fig. 3.20) the gap is greater and our method is capable of improving the registration, bringing the LM and EM patches much closer than the pre-registration step.

To further evaluate the registration refinement, we computed the registration errors collected in Table 3.2. The errors are computed by manually selecting corresponding points in the LM and EM images, both pre-registered and after the registration refinement. We can infer from the computed errors that our point-based registration refinement is capable of significantly improving the final registration between the pre-registered patches of the 5 experiments.



(a) LM pre-registered patch

(b) Multiscale spot segmentation of 3.17a

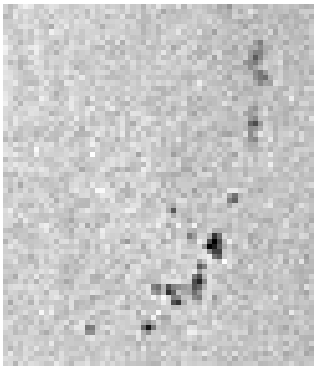
(c) EM pre-registered patch

(d) Multiscale spot segmentation of 3.17c

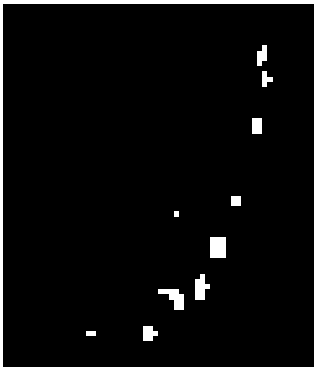
(e) Overlay of pre-registered EM and LM patches

(f) Overlay of EM and LM patches after point-based refinement of registration

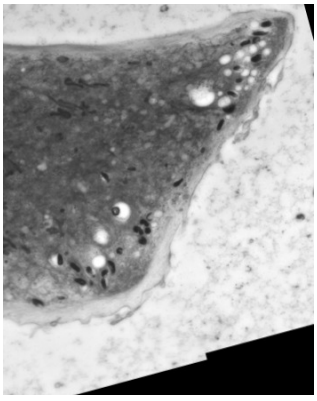
Fig. 3.17: *Experiment 1. From top to bottom and left to right: pre-registered LM patch and its spot segmentation using our multiscale approach. Pre-registered EM and its spot segmentation. Overlay of pre-registered LM and EM patches. Overlay after point-based refinement of the registration.*



(a) LM pre-registered patch

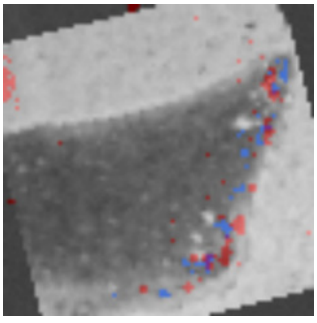


(b) Multiscale spot segmentation of 3.18a

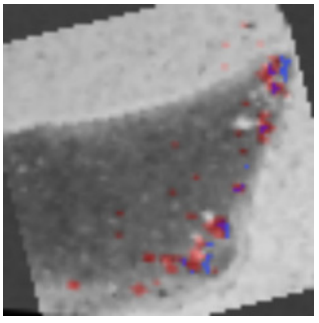


(c) EM pre-registered patch

(d) Multiscale spot segmentation of 3.18c

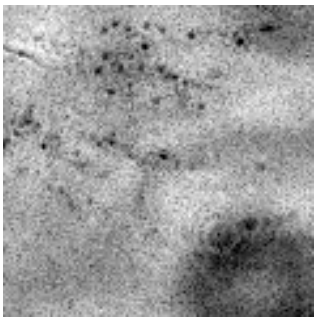


(e) Overlay of pre-registered EM and LM patches



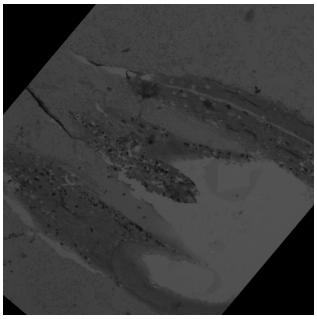
(f) Overlay of EM and LM patches after point-based refinement of registration

Fig. 3.18: *Experiment 2. From top to bottom and left to right: pre-registered LM patch and its spot segmentation using our multiscale approach. Pre-registered EM and its spot segmentation. Overlay of pre-registered LM and EM patches. Overlay after point-based refinement of the registration.*



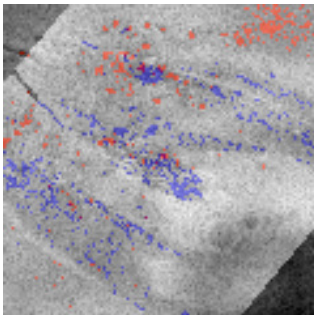
(a) LM pre-registered patch

(b) Multiscale spot segmentation of 3.19a

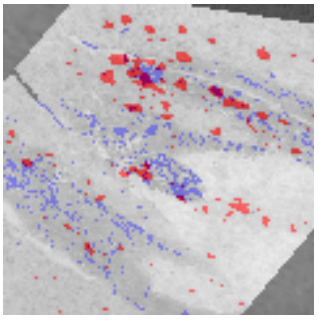


(c) EM pre-registered patch

(d) Multiscale spot segmentation of 3.19c

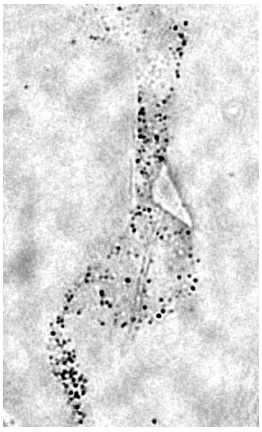


(e) Overlay of pre-registered EM and LM patches

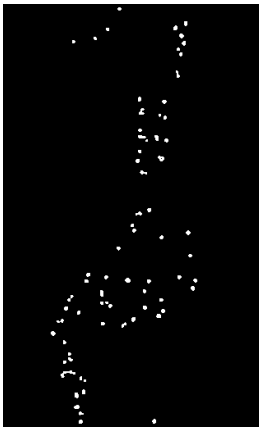


(f) Overlay of EM and LM patches after point-based refinement of registration

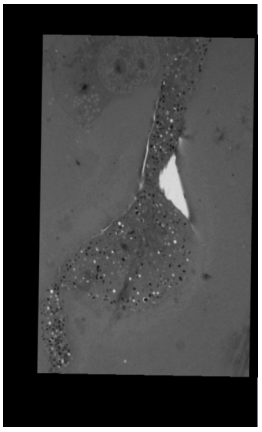
Fig. 3.19: *Experiment 3. From top to bottom and left to right: pre-registered LM patch and its spot segmentation using our multiscale approach. Pre-registered EM and its spot segmentation. Overlay of pre-registered LM and EM patches. Overlay after point-based refinement of the registration.*



(a) LM pre-registered patch

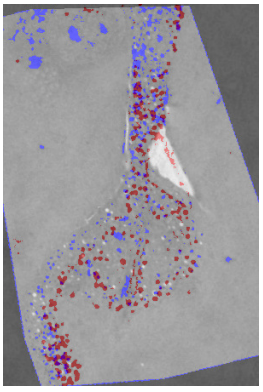


(b) Multiscale spot segmentation of 3.20a

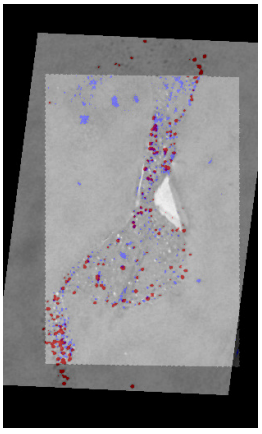


(c) EM pre-registered patch

(d) Multiscale spot segmentation of 3.20c

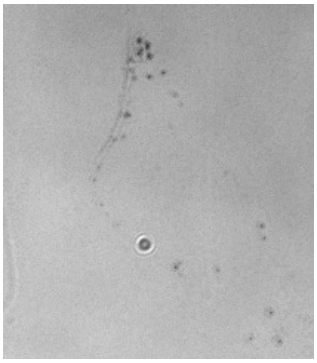


(e) Overlay of pre-registered EM and LM patches

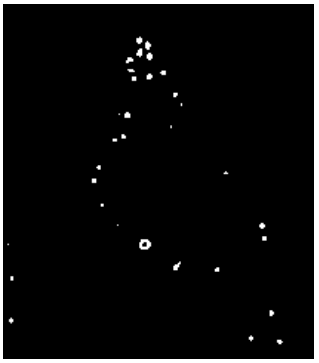


(f) Overlay of EM and LM patches after point-based refinement of registration

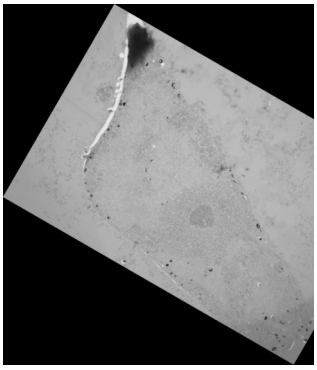
Fig. 3.20: *Experiment 4. From top to bottom and left to right: pre-registered LM patch and its spot segmentation using our multiscale approach. Pre-registered EM and its spot segmentation. Overlay of pre-registered LM and EM patches. Overlay after point-based refinement of the registration.*



(a) LM pre-registered patch

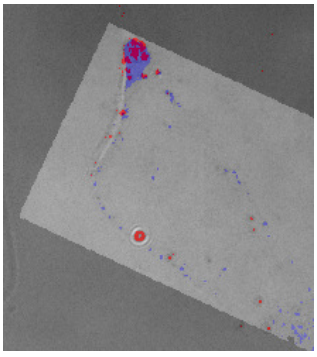


(b) Multiscale spot segmentation of 3.21a

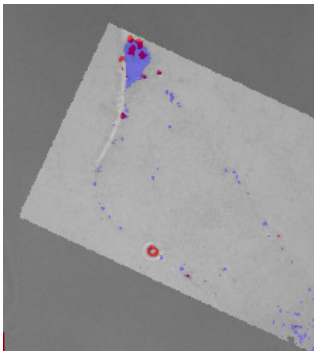


(c) EM pre-registered patch

(d) Multiscale spot segmentation of 3.21c



(e) Overlay of pre-registered EM5 and LM5 patches



(f) Overlay of EM and LM patches after point-based refinement of registration

Fig. 3.21: *Experiment 5. From top to bottom and left to right: pre-registered LM patch and its spot segmentation using our multiscale approach. Pre-registered EM and its spot segmentation. Overlay of pre-registered LM and EM patches. Overlay after point-based refinement of the registration.*

	Exp. 1		Exp. 2		Exp. 3		Exp. 4		Exp. 5	
Error	Pre-reg.	Refined	Pre-reg.	Refined	Pre-reg.	Refined	Pre-reg.	Refined	Pre-reg.	Refined
Mean	4.80	0.48	2.17	1.76	1.09	1.05	4.98	4.79	3.76	1.02
SD	4.64	0.68	1.36	0. 95	0.76	0.74	3.61	2.08	1.72	1.10
Median	3.40	0.20	1.70	1.37	0.70	0.68	4.00	5.39	4.06	1.00
Max.	12.55	1.41	4.24	3.35	2.12	2.10	9.43	7.00	5.59	2.05
Min.	0.47	0.00	0.67	1.00	0.50	0.48	1.00	2.24	1.00	0.00

Table 3.2: Final registration errors for experiments 1-5 in LM pixels.

3.5 Conclusion

We have defined an original coarse-to-fine multiscale spot detection method where we automatically select multiple scales according to a criterion based on the *a contrario* approach. Experiments on simulated images with objective evaluation, and on real images with visual assessment, demonstrated that our method outperforms existing spot segmentation methods. Without introducing any critical parameter setting, our method is able to automatically select the relevant scales corresponding to spot-like objects of different sizes in the image, and to correctly segment spots, even in close proximity. Tests performed on real images containing different types of objects illustrate the possible diverse application of our method. Additionally, we exploit our multiscale spot segmentation method to further register CLEM images pre-aligned with the intensity-based registration framework of Chapter 2. We extract spots of interest present in the EM and LM pre-registered patches and compute a point-based registration refinement, which allow us to still improve the CLEM registration. We tested this framework on five different datasets with good results, indicating that our spot segmentation and point-based registration refinement framework can be an additional tool in the automated CLEM registration problem.

CHAPTER 4

TWO ROBUST MODEL SELECTION CRITERIA FOR PARAMETRIC IMAGE MOTION ESTIMATION

Parametric motion models are commonly used in image sequence analysis for different tasks. A robust estimation framework is usually required to reliably compute the motion model over the estimation support in presence of outliers, while the choice of the right motion model is also important to properly perform the task. However, dealing simultaneously with both issues remains an open question. We have defined two new robust motion model selection methods. The first one is an extension of the Takeuchi information criterion (TIC). The second one is built from the Fisher statistic. We also derive an interpretation of the latter as a robust Mallows' C_P criterion. Both robust motion model selection criteria are straightforward to compute. We have conducted a comparative objective evaluation on computer-generated image sequences with ground truth, along with experiments on real videos, for the parametric estimation of the dominant motion corresponding to the 2D apparent motion of the static scene background due to the camera motion. They demonstrate the interest and the efficiency of the proposed robust model selection methods.

Resorting to parametric models is a common practice in motion analysis, image registration, and more generally in dynamic scene analysis. Video stabilization [94], image stitching [154], motion detection [163], motion layer segmentation [28], optical flow computation [14, 35, 174], tracking [143, 176], action recognition and localization [62],

crowd motion analysis [112], to name a few, all may rely on parametric motion estimation. Two key issues then arise: how to choose the right motion model and how to reliably estimate it?

The former issue can be circumvented by settling for empirical choice. The affine motion model is often claimed as a good trade-off between efficiency and representativeness without any available information on the dynamic scene. However, a principled method is more satisfying to properly solve the motion model selection problem [36, 41, 159] The most commonly-used model selection criteria are undoubtedly Akaike information criterion (AIC) [3], Bayesian information criterion (BIC) [142], or Takeuchi information criterion (TIC) [18]. Broadly speaking, it starts from the maximum likelihood, and amounts to add to the quadratic model fit, a weighted penalty term on the model complexity, or model dimension, given by the number of the model parameters. The definition of the weight depends on the statistical information criterion. The likelihood term accounts for a Gaussian distribution of the residuals involved in the regression issue. A comparative study of several of them is reported in [161] for classification in pattern recognition. Let us add the Mallows' C_P criterion [91] and the Minimum Description Length criterion (MDL) [122] respectively equivalent to AIC and BIC under certain hypotheses. Finally, let us mention the Akaike criterion with a correction for finite sample sizes (AICc) [18].

Regarding the latter issue, robust motion estimation [14, 96, 106, 143] is the required framework to cope with the presence of outliers in the model estimation support, due to local independent motions, occlusions, or any local violation of the assumptions associated with motion computation. Indeed, least-square estimation is generally biased. As a consequence, the aforementioned information criteria involving a quadratic (or Gaussian) likelihood term are not in line with robust regression.

However, combining model selection and robust estimation for parametric motion computation has rarely been investigated [159]. In this chapter, we propose two statistical criteria for robust motion model selection. The first one is an extension of the Takeuchi information criterion. The second one tackles this problem from a different perspective based on the Fisher statistic. An interpretation as a robust version of the Mallows' C_P criterion [91] is also provided. A preliminary work was presented in the conference paper [16]. The present chapter is a significant extension of the latter. We have added a new criterion, the robust TIC, improvements of the Fisher-based criterion, an augmented related-work section, a revisited objective comparative evaluation, more experiments on real videos, and more detailed comments throughout the sections.

The remainder of the chapter is organized as follows. Section 2 is devoted to related work and positioning of our approach. In Section 3, we recall a robust estimation method of motion models, and formulate the model fit. Section 4 describes our first robust motion model selection method called Robust Takeuchi Information Criterion (RTIC). In Section 5, we present our second original method for robust motion model selection, called Fisher-based robust information criterion (FRIC). Objective comparisons on computer-generated examples with ground-truth are reported in Section 6, along with experiments on real videos, to assess the performance of our two criteria. Concluding remarks are given in Section 7.

4.1 Related work and positioning

4.1.1 State of the art

Statistical information criteria have been exploited in computer vision for years [41], sometimes with specific formulations and characteristics. Geometric counterparts of AIC and MDL, respectively termed GAIC and GMDL, were proposed in [68] to take into account a different formulation of model fitting along with the dimension of the manifold involved in a 3D geometric transformation. AIC and BIC were tested in [36] for 2D affine motion model classification, but they were experimentally proven less efficient than a succession of hypothesis tests deciding in turn on the nullity of each affine motion parameter. Indeed, AIC tends to overestimate the complexity of the underlying model. In [169], the most appropriate model among 2D polynomial motion models for motion estimation from normal flows, was selected with a penalization factor given by the Vapnik's measure; the resulting algorithm was favourably compared to AIC, BIC and generalized cross-validation. In [139], a MDL-based criterion was designed for model selection in 3D multibody structure-and-motion from images. A MDL principle is also adopted in [92] for non rigid image registration. On the other hand, AICc was used in [8] for a pixel-wise motion model selection with a view to crowd motion analysis in video sequences.

Robust model selection on its own was explored in the robust statistics literature along several directions [1, 74, 101, 118, 125, 134]. In [125], a robust extension of AIC (RAIC), was defined, coming up with substituting a general robust estimator ρ of the model parameters θ for the maximum likelihood estimator. M-estimators are incorporated in BIC and the asymptotic performance is studied in [86]. A special case is the use of the Huber robust function [60], leading to the RBIC criterion. The Mallows' C_P

criterion is revisited in [126] to yield a robust version. The generalized information criterion (GIC), described in [74], can be applied to evaluate statistical models constructed by other procedures than maximum likelihood, such as robust estimation or maximum penalized likelihood.

In contrast, only a few investigations of that type have been undertaken regarding motion estimation and segmentation. In [5], the authors designed a global energy function for both the robust estimation of mixture models and the validation of a MDL criterion. The overall goal is to get a layering representation of the moving content of an image sequence. The MDL encoding acts on the overall cost of the representation comprising the number of layers, residuals and motion parameters. However, the primary purpose was parsimonious motion segmentation, and not motion model selection per se. In [159], a robust extension of the Geometric Information Criterion (GIC) [67], termed GRIC, is proposed in the vein of RAIC. It was applied to the selection of the 3D geometric transformation attached to a rigid motion and estimated through the matching of image interest points. Geometrical and physical constraints are also explored in [40] for image motion segmentation, with the so-called surface selection criterion (SSC) primarily designed by the authors for range data segmentation. Better performance is reported than with several information criteria, but the use of SSC in this work is comparable to a regularization approach.

4.1.2 Our approach

However, all the previous approaches were not explicitly concerned with jointly maximizing the inlier set size attached to the motion model. The problem of robust motion model selection is in fact three-fold: *i)* to maximize the motion model fit to the data, *ii)* to penalize the motion model complexity, *iii)* to account for the motion of the largest possible set of points in the estimation support (the whole image or a given image region depending on the motion analysis problem). Indeed, the two latter ones must be simultaneously satisfied, which might be contradictory. By definition, this is an issue specific to robust model selection. It apparently did not draw interest in the robust statistics literature, while it is of key importance in motion analysis. In this chapter, we introduce two robust motion model selection methods in that perspective. In addition, the second method follows a novel approach based on Fisher statistic.

To make the robust motion model selection problem concrete, we will deal with the dominant image motion estimation issue. The dominant (or global) image motion is usually due to the camera motion, and then corresponds to the background motion, i.e., the apparent motion of the static scene in the image sequence. Computing the dominant

Motion model	*Dimension*	*Mathematical expression*
Translation (T)	2	$\mathbf{w}_\theta = (a_1, a_4)^T$
Pan-tilt (PT)	2	$\mathbf{w}_\theta(p) = (a_1 + a_1x^2 + a_4xy, a_4 + a_1xy + a_4y^2)^T$
Translation + Rotation (TR)	3	$\mathbf{w}_\theta(p) = (a_1 + a_3y, a_4 - a_3x)^T$
Translation + Scale (TS)	3	$\mathbf{w}_\theta(p) = (a_1 + a_2x, a_4 + a_2y)^T$
Translation + Rotation + Scale (TRS)	4	$\mathbf{w}_\theta(p) = (a_1 + a_2x + a_3y, a_4 - a_3x + a_2y)^T$
Full affine (FA)	6	$\mathbf{w}_\theta(p) = (a_1 + a_2x + a_3y, a_4 + a_5x + a_6y)^T$
		$a_4 + a_2y + a_1xy + a_4y^2)^T$
Planar surface	8	$\mathbf{w}_\theta(p) = (a_1 + a_2x + a_3y + a_7x^2 + a_8xy,$
rigid motion (PSRM)		$a_4 + a_5x + a_6y + a_7xy + a_8y^2)^T$
Full quadratic (FQ)	12	$\mathbf{w}_\theta(p) = (a_1 + a_2x + a_3y + a_7x^2 + a_8xy + a_9y^2,$
		$a_4 + a_5x + a_6y + a_{10}x^2 + a_{11}xy + a_{12}y^2)^T$

Table 4.1: Set of 2D polynomial motion models

motion has numerous applications such as image stabilization, background subtraction for a free moving camera, image stitching or simply image registration. Nevertheless, the proposed scheme can be applied to other issues as well, for instance to select the right motion model on each image region for motion layer segmentation.

4.2 Robust motion model estimation

We will first recall the main principles of the robust estimation of parametric motion models. The estimation process relies on the brightness constancy assumption, and is embedded in a coarse-to-fine scheme to handle large displacements. We will present it in the frame of the motion model computation over the whole image domain Ω, but it can be straightforwardly adapted to the computation of the motion model over a given area in the image. Then, we will define the motion model fit for the estimated motion model parameters. Finally, we will describe the set of 2D parametric motion models that will be considered, appertaining to the category of polynomial models.

4.2.1 Computation of motion model parameters

We consider a set of 2D polynomial motion models. They will be precisely defined in Section 4.2.3 and Table 4.1. Let θ_m denote the parameters of model m, that is, the polynomial coefficients for the two components of the velocity vector. The full model will be denoted by M, if we have M models to test. $\mathbf{w}_{\theta_m}(p)$ is the velocity vector supplied by the motion model m at point $p = (x, y)$ of the image domain Ω.

To estimate the parameters of the motion model, we exploit the usual brightness

constancy assumption [35], leading to the linear regression equation relating the motion model parameters, through the velocity vector, and the spatio-temporal derivatives of the image intensity I:

$$\nabla I(p).\mathbf{w}_{\theta_m}(p) + I_t(p) = 0. \tag{4.1}$$

Let us denote $r_{\theta_m}(p)$ the left member of (4.1). The estimation of the motion model parameters is formulated as:

$$\hat{\theta}_m = \arg\min_{\theta_m} \sum_{p\in\Omega} \rho(r_{\theta_m}(p)), \tag{4.2}$$

where ρ denotes a robust penalty function. As examples, the Lorentzian function is used in [14], in [143] the Hampel estimator is preferred, and in [106] the Tukey's function is adopted.

Equation (4.1) is in fact the linearization of the more general constraint $I(p + \mathbf{w}_{\theta_m}(p)) - I(p,t) = 0$. As a consequence, it only holds for small displacements. A usual way to overcome this problem, is to resort to a coarse-to-fine scheme based on image multiresolution and incremental motion estimation [35]. The minimization of (4.2) is achieved by an iterative algorithm, the Iterated Reweighting Least Squares (IRLS) method [57]. The IRLS method iteratively updates weights at every point $p \in \Omega$. The weights express the influence of each point p in the estimation of the motion model parameters. These weights can be further exploited to determine the inlier set associated with the estimated motion model m.

4.2.2 Motion model fit

Once we compute an estimate $\hat{\theta}_m$ of the motion model parameters, we get the residuals $r_{\hat{\theta}_m}(p)$, for all $p \in \Omega$, measuring the discrepancy between the input data and the estimated motion model. To evaluate how the estimated motion model fits the input data over the associated inlier set, we consider the residual sum of squares (RSS) obtained for the robustly estimated parameters $\hat{\theta}_m$ of the motion model m given by:

$$\mathrm{RSS}_m = \sum_{p\in\mathcal{I}_m} r^2_{\hat{\theta}_m}(p), \tag{4.3}$$

where $\mathcal{I}_m$ represents the set of inliers associated with the estimated motion model m. The residual is formally defined by:

$$r_{\hat{\theta}_m}(p) = I(p + \mathbf{w}_{\hat{\theta}_m}(p), t+1) - I(p,t), \tag{4.4}$$

knowing that the left member of (4.1) is a linearized version of (4.4).

We compute RSS_m on the inlier set $\mathcal{I}_m$ and not on the overall domain Ω, to obtain the model fit evaluation precisely on the subset of points whose motion conforms with the estimated motion model.

Furthermore, we introduce the expression RSS_m^+, which represents the residual sum of squares computed over the inlier set $\mathcal{I}_m$ attached to model m, but for the full model M:

$$\text{RSS}_m^+ = \sum_{p \in \mathcal{I}_m} r^2_{\hat{\theta}_M}(p). \tag{4.5}$$

As aforementioned, solving for the minimization (4.2) amounts to apply the Iteratively Reweighted Least Squares algorithm within a coarse-to-fine framework. At convergence, the final weights $\alpha_m(p), p \in \Omega$, are given by:

$$\alpha_m(p) = \frac{\psi(r_{\hat{\theta}_m}(p))}{r_{\hat{\theta}_m}(p)}, \tag{4.6}$$

where the influence function $\psi(.)$ is the derivative of the robust function $\rho(.)$, $\psi(r) = \frac{d\rho(r)}{dr}$. In practice, we adopt the robust estimation method defined in [106], and we use the publicly available Motion2D[1] software implementing this method.

4.2.3 Set of parametric motion models

We are dealing with 2D polynomial motion models ranging from translation (polynomial of degree 0) to quadratic models (polynomials of degree 2), including different affine models (polynomials of degree 1). They are forming a set of models which is *partly nested*. The model complexity ranges from dimension 2 to dimension 12. The full set of motion models is given in Table 4.1 with their main features. The explicit equivalence, when available, between the 2D polynomial models and physical motions assumes a perspective projection for the image formation.

Let us mention that the constant part of the affine motion model does not necessarily mean that the underlying physical motion has actually a translation component. Indeed, the constant part is merely the velocity vector given by the motion model at the origin of the image coordinate system. The in-plane rotation is not necessarily centered at the origin. The same holds for the focus of expansion in case of scale motion, knowing that the scale motion in the image is due to a translation of the camera along its axis of view. The 8-parameter quadratic motion model exactly corresponds to a rigid motion between

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

a planar surface and the camera. The 2-parameter quadratic model also exactly accounts for a pan-tilt camera motion.

4.3 Robust motion model selection with RTIC

4.3.1 TIC criterion

We start from the Takeuchi information criterion (TIC) which is a more general derivation of Akaike's information criterion [18], and we will derive a robust version of it in the vein of RAIC and RBIC ones. TIC can be written as follows:

$$\mathrm{TIC}(m) = 2\mathcal{K}(\hat{\theta}_m) + 2\,\mathrm{tr}(P(\theta)Q(\theta)^{-1}), \tag{4.7}$$

where $\mathcal{K}$ denotes the contrast function (or equivalently the negated logarithm of the likelihood or pseudo-likelihood function), and "tr" the trace of the matrix. The two mxm matrices $P(\theta)$ and $Q(\theta)$ respectively involve first and second partial derivatives of the likelihood function w.r.t. model parameters. In the regression case, the two matrices P and Q are defined by:

$$\begin{aligned} P(\theta) &= E[\frac{\partial}{\partial_\theta} g(r_\theta(p)) \frac{\partial}{\partial_\theta} g(r_\theta(p))^T] \\ Q(\theta) &= E[\frac{\partial^2}{\partial_\theta \partial_\theta} g(r_\theta(p))], \end{aligned} \tag{4.8}$$

where $g(.)$ is defined by $\mathcal{K}(\theta_m) = \sum_p g(r_{\theta_m}(p))$, $r_{\theta_m}(p)$ acts as the regression residual, and E denotes expectation.

4.3.2 Robust TIC

Let us move to the robust estimation context. The g function is specified by a robust penalty function $\rho(.)$, as introduced in eq.(4.2). We come out with the following expression of the Takeuchi information criterion, which we will call Robust Takeuchi Information Criterion (RTIC) to make it short:

$$\mathrm{RTIC}(m) = 2\sum_{p\in\Omega} \rho(r_{\hat{\theta}_m}(p)) + 2q_m \frac{E[\psi(r_\theta(p))^2]}{E[\psi'(r_\theta(p))]}_{|\theta=\hat{\theta}_m}, \tag{4.9}$$

where q_m is the dimension (i.e., number of parameters) of model m, $\psi(.)$ is the influence function as defined in subsection 4.2.2, and ψ' its derivative.

We will develop two versions of RTIC for two robust penalty functions, the Talwar and Huber penalty functions. The Talwar function is defined by:

$$\rho_{tal}(r) = \begin{cases} r^2/2 & \text{if } |r| \leq \alpha \\ \alpha^2/2 & \text{if } |r| > \alpha \end{cases} \tag{4.10}$$

Knowing that the inlier set $\mathcal{I}_m$ corresponds to the points p such that $|r_{\theta_m}(p)| \leq \alpha$, we estimate the expectation as:

$$E[\psi(r_\theta(p))^2] \simeq \frac{1}{|\Omega|} \sum_{p \in \Omega} \psi(r_\theta(p))^2 = \frac{1}{|\Omega|} \sum_{p \in \mathcal{I}_m} r_\theta^2(p),$$
$$\text{and } E[\psi'(r_\theta(p))] \simeq \frac{|\mathcal{I}_m|}{|\Omega|}, \tag{4.11}$$

where $|.|$ denotes the cardinality of the set. We come up with the following expression of RTIC for the Talwar penalty function:

$$\text{RTIC}_{tal}(m) = 2 \sum_{p \in \Omega} \rho(r_{\hat{\theta}_m}(p)) + \frac{2q_m}{|\mathcal{I}_m|} \sum_{p \in \mathcal{I}_m} r^2_{\hat{\theta}_m}(p). \tag{4.12}$$

We make the same development for the Huber function defined as follows:

$$\rho_{hub}(r) = \begin{cases} \frac{1}{2} r^2 & \text{if } |r| \leq \alpha \\ \alpha(|r| - \frac{1}{2}\alpha) & \text{if } |r| > \alpha \end{cases} \tag{4.13}$$

and we obtain as RTIC expression:

$$\text{RTIC}_{hub}(m) = 2 \sum_{p \in \Omega} \rho(r_{\hat{\theta}_m}(p)) + \frac{2q_m}{|\mathcal{I}_m|} \Bigg(\sum_{p \in \mathcal{I}_m} r^2_{\hat{\theta}_m}(p) + \sum_{p \in \Omega \setminus \mathcal{I}_m} \alpha^2 \Bigg). \tag{4.14}$$

The selected model $\tilde{m}$ will be the one minimizing $\text{RTIC}_{tal}(m)$ (respectively $\text{RTIC}_{hub}(m)$) among the tested models. The parameters $\hat{\theta}_m$ are obtained from 4.2 with the Talwar (resp. Huber) ρ-penalty function. In contrast to RAIC and RBIC, the size of the inlier set, $|\mathcal{I}_m|$, explicitly intervenes in the second term of the expression of the two RTIC variants (4.12) and (4.14). Minimizing RTIC implies to maximize the size of the inlier set.

4.4 Robust motion model selection with FRIC

We propose a second robust motion model selection criterion which follows a very different approach than those proposed so far for robust model selection. As in [36], we adopt a two-class hypothesis test approach. This is motivated in particular by the fact

that we are dealing with an unnested set of parametric motion models. For instance, both the rotation and the scaling models involve three parameters as described in Section 4.2.3, but respectively account for quite different motions. In addition, we aim to select the model m which explains the motion of the maximum number of points in the estimation support (which may be the whole image domain Ω or a given image region), that is, with the largest possible inlier set. Indeed, we are not interested in selecting a simple motion model if it can only account for a too limited part of the estimation support. Let us remind that retaining the as less complex as possible motion model, while maximizing the size of the inlier set, may be contradictory.

4.4.1 Fisher statistic

First, we want to compare any model m of the set of tested models to the full model M. To this end, we resort to the Fisher statistic [128] formulated in our case as:

$$\mathcal{F}(m) = \frac{(\mathrm{RSS}_m - \mathrm{RSS}_m^+)/(q_M - q_m)}{\mathrm{RSS}_m^+/(|\mathcal{I}_m| - q_M)}, \tag{4.15}$$

where again $|.|$ designates the set cardinality, q_m represents the number of parameters of model m. Both RSS_m and RSS_m^+ are evaluated on the inlier set $\mathcal{I}_m$ attached to the tested model m. To really deal with Fisher statistic, both model parameters, θ_m and θ_M, must be estimated on the same set too. Therefore, we re-estimate θ_m and θ_M over $\mathcal{I}_m$ in a least-square setting, before evaluating $\mathcal{F}(m)$. By the way, it also improves the estimated parameters of model m, and consequently, the model fit.

To make (4.15) an actual Fisher statistic, we must also check that $|\mathcal{I}_m|$ can be considered as a deterministic variable. We can at least empirically assess it, by providing histograms of weights computed for the estimation of the motion parameters in the IRLS procedure, as explained in subsection 4.2.1. Examples are supplied in Fig.4.1 for the use of the Tukey function in the motion model parameter estimation. The plots show that the weight histograms are clearly bimodal, with one mode close to 0 and the second one close to 1. Then, the inlier set $\mathcal{I}_m$ is easy to get, and this step does not introduce any randomness. It is even simpler in the case of the Talwar function since the weights are null for outliers and strictly positive for inliers.

The denominator of expression (4.15) can be interpreted as a non-biased empirical estimate of the full model variance computed on $\mathcal{I}_m$, which will be denoted by:

$$\hat{\sigma}_M^2(\mathcal{I}_m) = \frac{\mathrm{RSS}_m^+}{|\mathcal{I}_m| - q_M}. \tag{4.16}$$

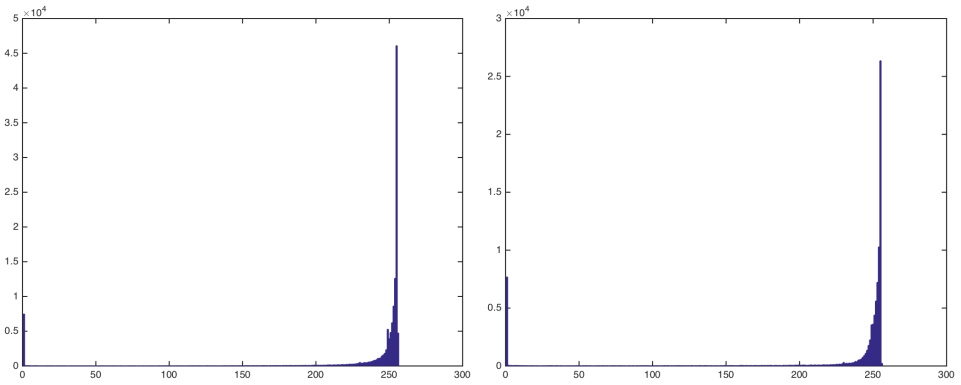


Fig. 4.1: *Histograms of weights supplied by the IRLS procedure in the robust estimation of the image dominant motion for two different experiments.*

The statistic $\mathcal{F}(m)$ allows us to decide whether model m is a more significant representation of the unknown true motion than the full model M over $\mathcal{I}_m$ which is the validity domain of model m in Ω. However, it will supply all the models m of that type. We need to take into account the dimension q_m of model m to further select the right one.

4.4.2 Fisher-based robust information criterion (FRIC)

Starting from (4.15), and penalizing the complexity of the model expressed by the number q_m of model parameters, we define:

$$\mathcal{S}_1(m) = \mathcal{F}(m)(q_M - q_m) + 2q_m. \tag{4.17}$$

Under the assumption of validity of model m, $\mathcal{F}(m)$ follows a Fisher distribution $F(q_M - q_m, |\mathcal{I}_m| - q_M)$. Then, the first term of the right member of (4.17) (approximately) follows a χ^2 distribution with $q_M - q_m$ degrees of freedom.

We can now write the test for selecting the best motion model $\tilde{m}$ in this robust model selection framework:

$$\tilde{m} = \arg\min_m \mathcal{S}_1(m). \tag{4.18}$$

The theoretical behavior of this test can be qualitatively described as follows. $\mathcal{S}_1(m)$ is supposed to decrease when evaluating in turn the first successive models in decreasing (or equivalently increasing) complexity order up to the optimal model m^*, and then to increase for the subsequent models. This is confirmed by Fig.4.2 which contains plots of $\mathcal{S}_1(m)$ values for several experiments.

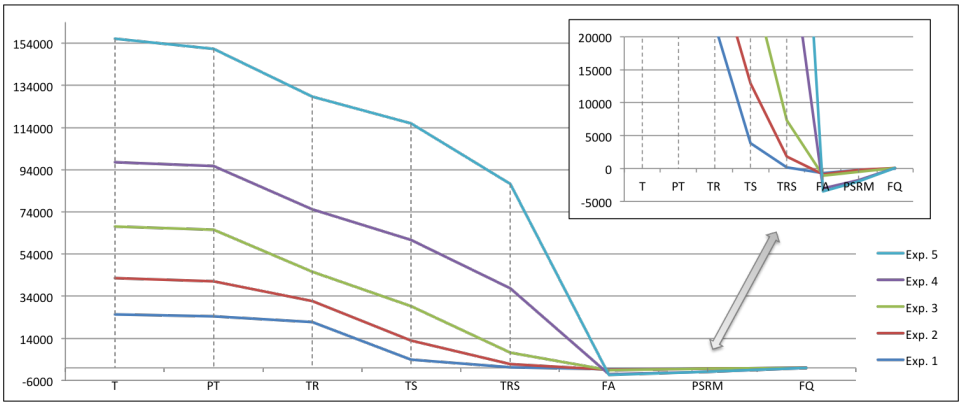


Fig. 4.2: *Plots of $\mathcal{S}_1(m)$ values corresponding to the set of tested motion models for several experiments. Tested motion models the models listed in Table 4.1, and ordered according to the number of parameters. The true model is FA.*

We design a second version of the Fisher-based robust model selection criterion, by incorporating the number of inliers in the model complexity penalization as in the BIC criterion, that is:

$$\mathcal{S}_2(m) = \mathcal{F}(m)(q_M - q_m) + 2\log(|\mathcal{I}_m|)q_m. \tag{4.19}$$

4.4.3 Interpretation of FRIC as robust C_p

We now provide another interpretation of the Fisher-based robust information criterion (FRIC) defined in (4.17). Let us first make $\hat{\sigma}^2_M(\mathcal{I}_m)$ appear in the expression of $\mathcal{S}_1(m)$ as follows:

$$\mathcal{S}_1(m) = \frac{(\mathrm{RSS}_m - \mathrm{RSS}^+_m)}{\hat{\sigma}^2_M(\mathcal{I}_m)} + 2q_m, \tag{4.20}$$

By exploiting (4.3) and (4.16), it can be further developed into:

$$\mathcal{S}_1(m) = \frac{1}{\hat{\sigma}^2_M(\mathcal{I}_m)} \sum_{p\in\mathcal{I}_m} r^2_{\hat{\theta}_m}(p) - |\mathcal{I}_m| + q_M + 2q_m. \tag{4.21}$$

If we neglect q_M which is a constant term for the test (4.18), expression (4.21) can be viewed as the Mallows' C_P criterion [91], computed over the inlier set attached to model m with $|\mathcal{I}_m|$ as the number of observations. Then, our test (4.18) could also be interpreted as a robust version of the Mallows' C_P criterion.

Let us point out that (4.21) explicitly involves the aforementioned trade-off between maximizing the size $|\mathcal{I}_m|$ of the inlier set and minimizing the complexity (i.e., the number q_m of parameters) of the selected motion model. In contrast, in existing robust

model selection criteria as RAIC or RBIC of the form:

$$RAIC = \sum_{p \in \Omega} \rho(r_{\hat{\theta}_m}(p)) + q_m, \tag{4.22}$$

$$RBIC = \sum_{p \in \Omega} \rho(r_{\hat{\theta}_m}(p)) + \log|\Omega| \; q_m, \tag{4.23}$$

the model selection is only implicitly influenced by the size of the inlier set attached to model m through the values of the robust function $\rho(.)$ at the outlier points. Hence, the impact depends on the asymptotic behavior of the robust function. The same holds for [126] where in addition the penalty term requires additional expensive computation to be evaluated.

4.5 Experimental results

4.5.1 Statistical objective evaluation

To assess the performance of the model selection criteria we carried out a comparative objective evaluation through synthetic experiments. We generated a series of image pairs by applying a velocity field to a real image, as demonstrated in Fig.4.3. The velocity field involves two parametric subfields chosen from the list given in Table 4.1. The first parametric motion subfield is the dominant motion, and the outliers, forming a rectangular region in the middle of the image, undergo the second one. Three groups of 3000 synthetic image pairs were generated, each group formed by different dominant and secondary motions. The first group involves a translation (T) motion model as dominant motion model and a full affine (FA) as secondary motion. The second set has a FA model as dominant motion and a planar surface rigid motion (PSRM) as secondary motion model. The last group has a PSRM model as the dominant one and a T model as the secondary one. Each group is divided in two sub-groups of 1500 image pairs each depending on the range used for the values of the parameters of the dominant motion, as summarized in Table 4.2. For each motion model used to create the image pairs, the value of its parameters is randomly selected in the interval given in Table 4.2.

We proceed to the selection of the dominant motion model in each experiment for each tested criterion as illustrated in Fig.4.4. In all the experiments and for all the four compared robust motion model selection criteria, we use the Talwar penalty function. Scores are given in percentage of the total number of the images in each experiment. Overall, criterion $\mathcal{S}_2$ outperforms $\mathcal{S}_1$ and $RBIC$. It successfully selects the true motion model in at least 74.4% of frames in a group, with a maximum of 83.4%. $\mathcal{S}_1$ also provides

Dominant motion model	a_1, a_4	a_2, a_3, a_5, a_6	a_7, a_8
$T1$	[-10,10]	-	-
$T2$	[-10,-1]∪[1,10]	-	-
$FA1$	[-10,10]	[-0.001,0.001]	-
$FA2$	[-10,-1]∪[1,10]	[-0.1,-0.001]∪[0.001,0.1]	-
$PSRM1$	[-5,5]	[-0.01,0.01]	[-0.001,0.001]
$PSRM2$	[-10,-1]∪[1,10]	[$-1e10^{-4}$,$-1e10^{-2}$]∪[$-1e10^{-4}$,$-1e10^{-2}$]	[$-1e10^{-5}$,$-1e10^{-4}$]∪[$1e10^{-5}$,$1e10^{-4}$]

Table 4.2: Range of parameter values for the different motion models.

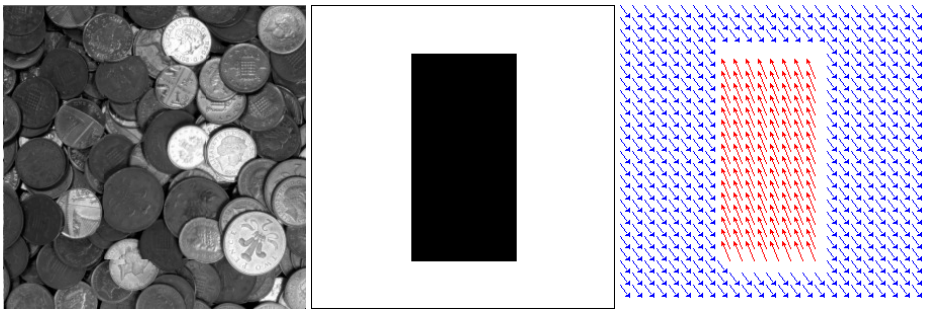


Fig. 4.3: *From left to right: the input image, the outlier subset (in black) in the middle of the image, typical velocity field applied to the input image formed by a dominant motion and a secondary one.*

good results, but it has a lowest success rate with a minimum of 61.2% and a maximum of 82.1%. In general, RBIC has a close but lower success rate than all the other criteria, with the lowest value at 38.6% for one of the experiments. RTIC, on the other hand, has a lower success rate in the T1 experiment. However, when the complexity of the dominant motion models used to generate the image pairs increases, RTIC produces better results, scoring over a 94% success rate in a couple of experiments. Overall, RTIC yields the best performance.

Tables 4.3 and 4.4 detail the scores obtained with $\mathcal{S}_2$ and RTIC for all the tested models and for the six subsets of experiments, respectively. As expected, selection errors are spread and concern mostly more complex models than the true one.

4.5.2 Results on real image sequences

In order to analyse the behaviour of the proposed criteria on real image sequences, we performed experiments on two types of video sequences. For these experiments, we adopt a subset of the models presented in Table 4.1, $\{T, TR, TS, PSRM, FQ\}$. FA was removed since it does not physically correspond to any given camera motion. PT was

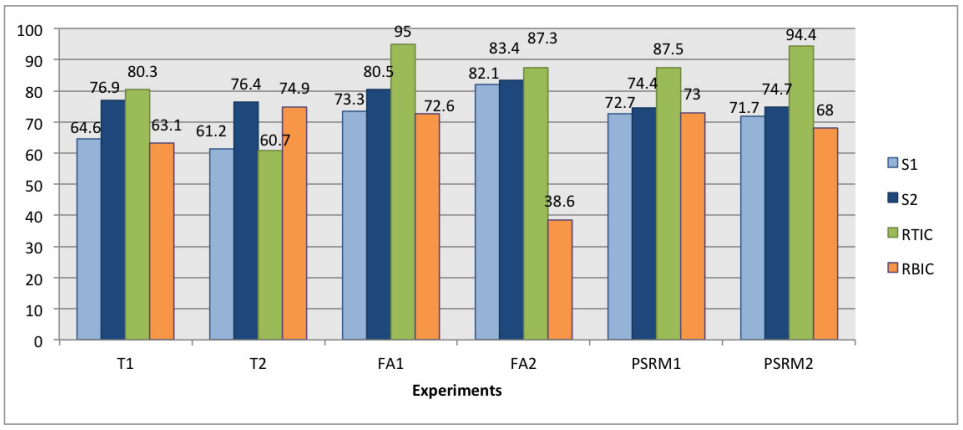


Fig. 4.4: *Results for each group of experiments obtained with the four model selection criteria (given in percentage of the total number of examples in each experiment).*

	True dominant motion models					
Tested models	**T**	**T**	**FA**	**FA**	**PSRM**	**PSRM**
T	**76.9**	**76.4**	0.1	1.5	0.1	0.1
TR	9.3	7.7	0.7	0.6	0.1	0.2
TS	1.4	11.8	0.3	1.1	0.1	0.1
TRS	0.1	0.6	2.1	0.2	0.3	0.2
FA	0.4	0.3	**80.5**	**83.4**	0.2	2.1
PT	2.1	0.7	0.1	3.9	0.9	2.4
PSRM	1.3	0.4	7.3	6.9	**74.4**	**74.7**
FQ	8.5	2.1	8.9	2.4	23.9	20.2

Table 4.3: *Scores obtained with criterion $\mathcal{S}_2$ for all the tested models and for the six experiments.*

	True dominant motion models					
Tested models	**T**	**T**	**FA**	**FA**	**PSRM**	**PSRM**
T	**80.3**	**60.7**	0	0	0	0
TR	11	15.9	0	0	0	0.1
TS	4	17.2	0	0	0	0
TRS	0.4	3.2	0.2	0	0	0
FA	0.6	0.5	**95**	**87.3**	0	0.2
PT	3.4	1	0	0.4	0	0
PSRM	0	0.7	1.1	9.1	**87.5**	**94.4**
FQ	0.3	0.8	3.7	3.2	12.5	5.3

Table 4.4: *Scores obtained with criterion RTIC for all the tested models and for the six experiments.*

removed since it only has two parameters, the same as T. The first set of real image sequences used in these tests were produced using a camera mounted on a Cartesian coordinate robot. The setup allows us to apply a determined motion model to the robot, in order to move the camera while capturing an image sequence. Ideally, the motion applied to the robot induces the dominant motion of the image sequence, therefore providing ground truth. Since we deal with robust motion selection able to handle the presence of outliers, we introduce a secondary motion using an additional single axis robot bearing a flat object and moving in the field of view of the camera. The complete setup is shown in Fig.4.5.

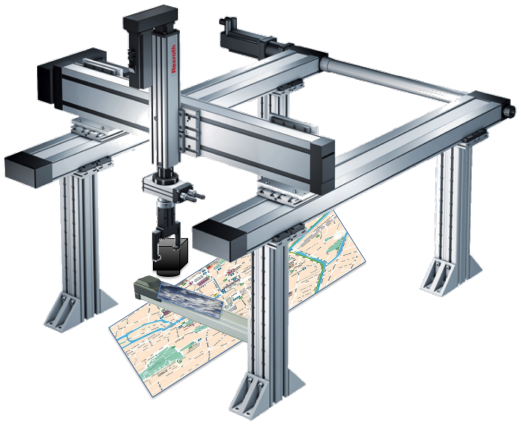


Fig. 4.5: *Robotic setup for the acquisition of image sequences. The scene is a planar surface formed by a poster.*

We report two of these experiments. In the first experiment, a rotation around the view axis is applied to the robot to produce an image sequence of 146 frames as illustrated in Fig.4.6. Since the rotation axis does not pass by the optical center, the expected dominant motion model is the combination of a translation and a in-plane rotation (TR).

Table 4.5 contains the model selection results provided by our criteria $\mathcal{S}_1$, $\mathcal{S}_2$ and RTIC, along with RBIC. $\mathcal{S}_2$ and RTIC select the assumed true motion model (TR) with a good percentage rate of 64.8% and 76% respectively. $\mathcal{S}_1$ selection is spread between TR and the full model FQ, while RBIC selects the full model in almost the whole sequence.

For the second experiment using the robotic setup, a translation along the camera optical axis is applied to the robot, producing a divergent motion in the image, but the focus of expansion is not at the origin. Then, the expected dominant motion model is TS. A sequence of 170 frames was acquired. Results are collected in Table 4.6. This experiment illustrates that the problem is not that simple and the criteria may fail. The

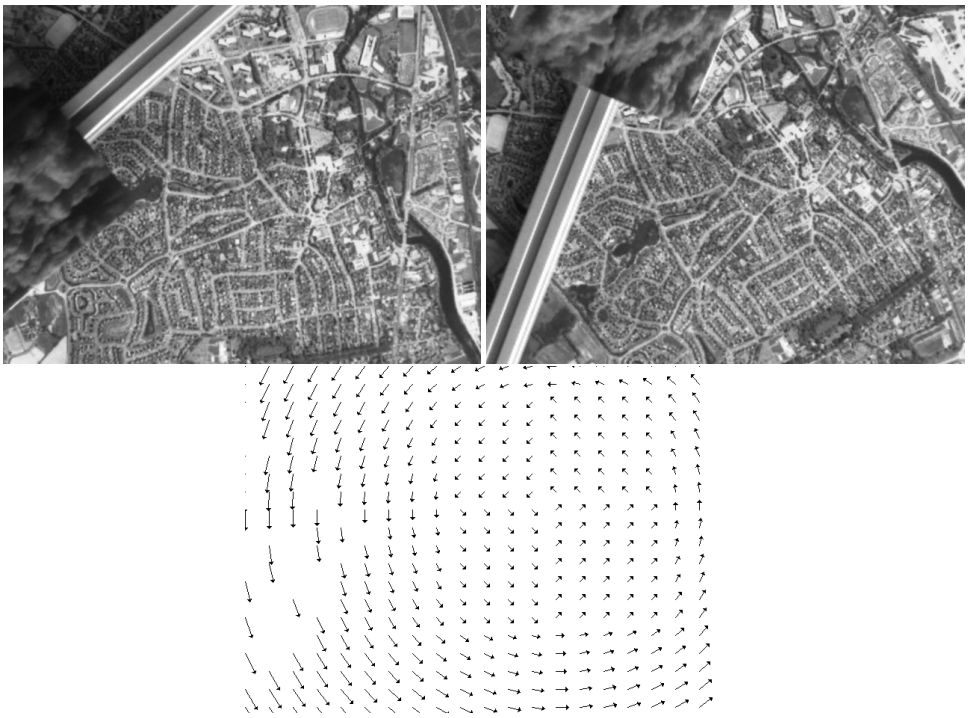


Fig. 4.6: *First and last frames of the first robot sequence, and the dominant flow between frames 0 and 1 computed with TR model.*

	T	TR	TS	PSRM	FQ
$\mathcal{S}_1$	0	46.4	0	7.8	45.8
$\mathcal{S}_2$	0	64.8	0	6.7	28.5
$RTIC$	0	76	0	0.5	23.5
$RBIC$	0.6	0	0	3.9	95.5

Table 4.5: *Selected motion models over the sequence of Fig.4.6 for the four compared criteria (in percentage of the total number of frames).*

three proposed criteria mostly select PSRM as dominant motion model. RTIC and $\mathcal{S}_1$ select FQ in second place and TS in third place, while $\mathcal{S}_2$ selects TS (the assumed true dominant motion model) in second place. Meanwhile, RBIC still selects the full model. However, the expected true model (TS) may possibly be questionable, since the robot expected motion may slightly depart from the specified command. The PSRM yet is also a relevant choice since the surface is planar and the robot motion is rigid.

	T	TR	TS	PSRM	FQ
$\mathcal{S}_1$	0	0	23.5	47.7	28.8
$\mathcal{S}_2$	0	0	28.2	49.4	22.4
$RTIC$	0	0.6	15.3	42.9	41.2
$RBIC$	0	0	4.1	3.5	92.4

Table 4.6: *Motion models selected by the compared criteria over the second robot sequence (in percentage of the total number of frames).*

We now report results on three real image sequences taken from the net. The first one depicts a field scene acquired from an airborne camera. The sequences contains 84 frames and the scene is almost planar. The outlier moving object is the reaping machine with the dust cloud behind it (Fig.4.7). It is difficult to infer the precise ground truth from the video alone, we do not know the type of camera motion. However, it can be assumed that the PRSM model should be the most relevant one.

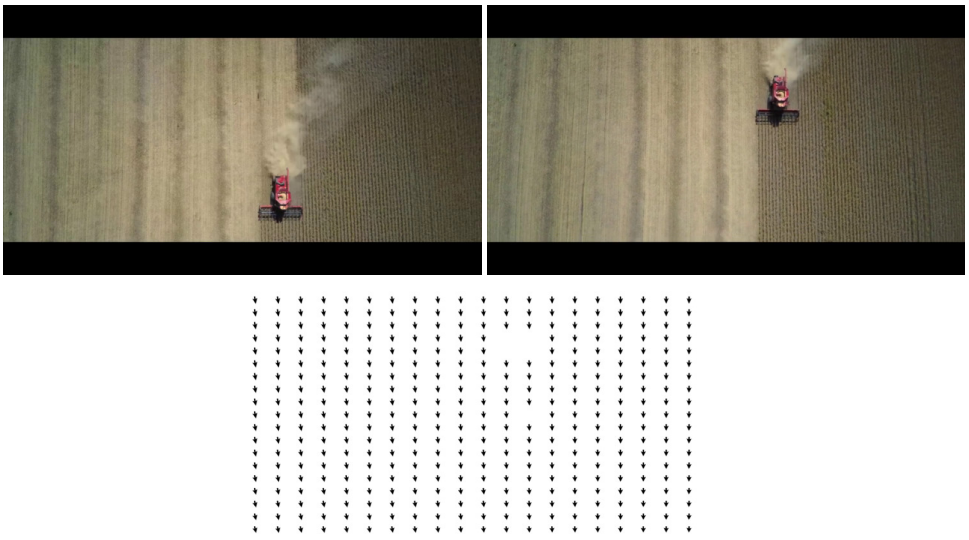


Fig. 4.7: *First and last frames of the field scene sequence, and the dominant flow between frames 0 and 1 computed with PSRM model.*

As observed in Table 4.7, the three proposed criteria $\mathcal{S}_1$, $\mathcal{S}_2$ and RTIC all select PSRM as the dominant motion of the sequence, both $\mathcal{S}_1$ and $\mathcal{S}_2$ with a rate of 73.8% and RTIC in 54.8% of the image pairs. RBIC selects the full model as dominant motion, with PSRM in second place.

	T	TR	TS	PSRM	FQ
$\mathcal{S}_1$	9.5	0	3.6	73.8	13.1
$\mathcal{S}_2$	9.5	0	3.6	73.8	13.1
$RTIC$	0	0	2.4	54.8	42.8
$RBIC$	0	1.2	13.1	34.5	51.2

Table 4.7: *Selected motion models over the field sequence of Fig.4.7 for the four compared criteria.*

The second real video consists of a sequence of 54 frames (Fig.4.8). Visually, the camera moves away from the scene, which leads to consider TS as the true dominant motion model. As in the previous sequence, the scene is almost planar and the vehicles present in it constitute the outliers to the dominant motion.

We report in Table 4.8 the selection scores of the compared criteria for the five tested

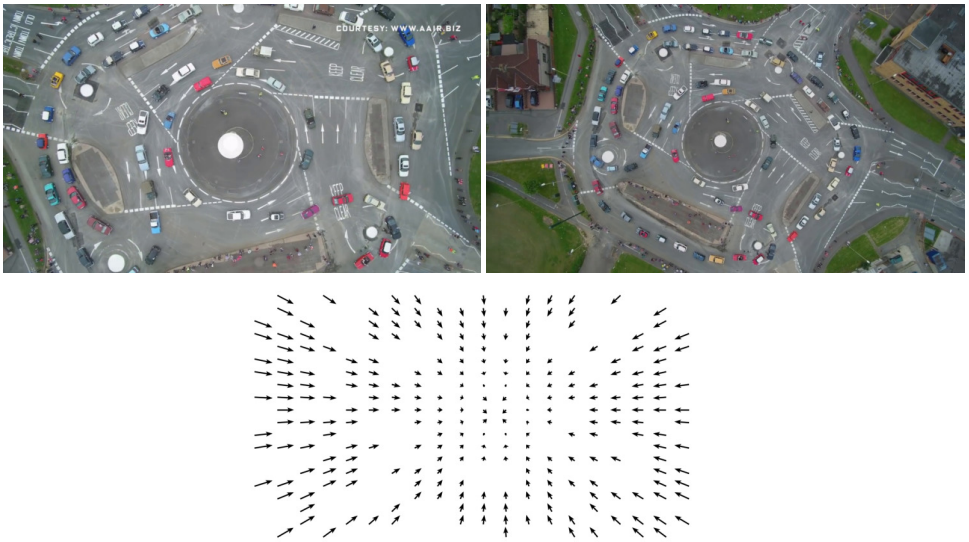


Fig. 4.8: *First and last frames of the roundabout sequence and the computed dominant motion at the time point corresponding to the first image, computed with TS model.*

motion models in the sequence of Fig.4.8. The right motion model TS is correctly selected by $\mathcal{S}_1$ and $\mathcal{S}_2$, but not by RTIC nor RBIC, both selecting the FQ model.

	T	TR	TS	PSRM	FQ
$RBIC$	0	0	0	16.7	83.3
$RTIC$	0	5.6	0	16.6	77.8
$\mathcal{S}_1$	16.6	0	72.2	5.6	5.6
$\mathcal{S}_2$	16.6	0	72.2	5.6	5.6

Table 4.8: *Motion models selected by the compared criteria over the roundabout sequence of Fig.4.8.*

The last real video example involves a sequence where a partly planar scene is recorded from an aerial camera (Fig.4.9). A passing train introduces outliers to the dominant motion, The camera motion is parallel to the ground with a slight rotation. We can assume that the TR motion model is the true one. Table 4.9 shows the results of this

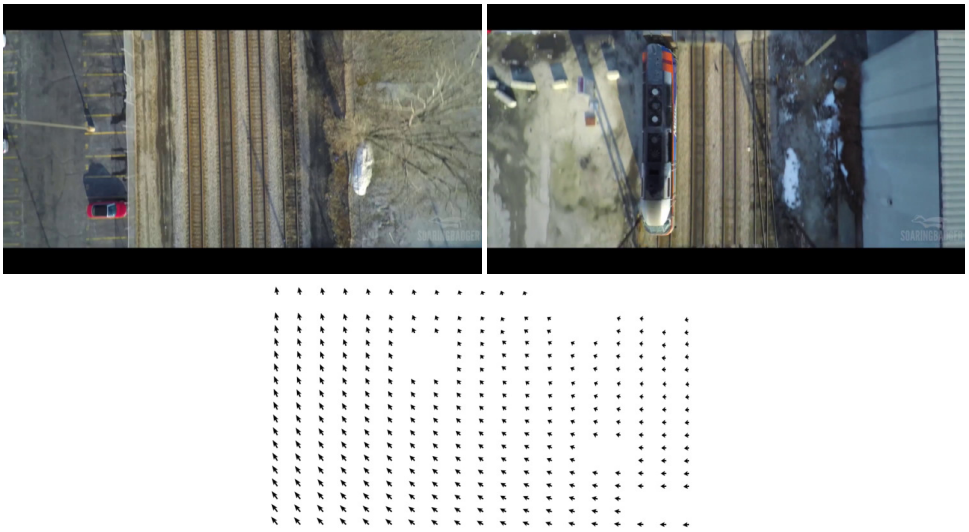


Fig. 4.9: *First and last frames of the train sequence and the dominant motion between the first and second frames computed with TR model.*

experiment. $\mathcal{S}_1$ and $\mathcal{S}_2$ select TR as dominant motion, both with a rate of 45.9%. They also select T and TS in almost half of the sequence, which are still reasonable choices. RTIC and RBIC incorrectly select the full model for most of the sequence.

4.6 Robust motion model selection for 2D CLEM registration

As mentioned in Chapter 1, there are many different techniques to acquire CLEM images. Deformation of the sample can be generated when using chemical fixation, during

	T	TR	TS	PSRM	FQ
$\mathcal{S}_1$	25	45.9	20.8	8.3	0
$\mathcal{S}_2$	25	45.9	20.8	8.3	0
$FRIC$	0	0	0	4.2	95.8
$RBIC$	0	8.3	8.3	0	83.4

Table 4.9: *Motion models selected by the compared criteria over the train sequence of Fig.4.9.*

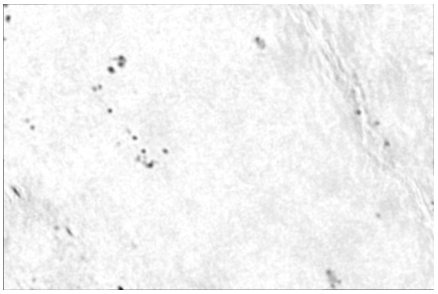
sectioning of the sample, when introducing markers for location, etc. Significant deformations can be avoided using more recent techniques, such as cryoEM, where the sample is frozen shortly after the live LM images have been acquired. In most cases, since the acquisition of LM and EM images is not performed using integrated setups, different kinds of geometrical transformations can be introduced between the LM and EM images, depending on the protocol used. The choice of the transformation to be estimated when registering CLEM images can have an impact on the accuracy of the registration. Even when registration has been performed, it is profitable to know if a rigid, an affine or a more complex transformation model is needed to improve the alignment of the images. In consequence, applying our robust motion model selection method could provide important information for the 2D CLEM registration problem.

When first exploring the options to perform registration of 2D CLEM images, we tested Motion2D to estimate the transformation between the pre-registered EM and LM images [157]. Although it delivered acceptable results, it was necessary to investigate options more adapted to the multimodality of CLEM images. As noted before, the differences between EM and LM are vast, including a significant discrepancy between the respective intensity of the images. This discrepancy can be appreciated in Figure 4.10, where the displaced frame difference (DFD) between the EM and LM images of experiment 1 after final registration is displayed. The DFD is defined as follows,

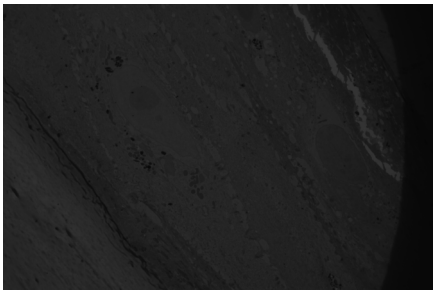
$$DFD(p,d) = I_2(p+d) - I_1(p) \tag{4.24}$$

where I_1 and I_2 are the intensity of the two images, $p = [x\,y]$ is a pixel of the image grid Ω and $d = (dx, dy)$ is the displacement vector.

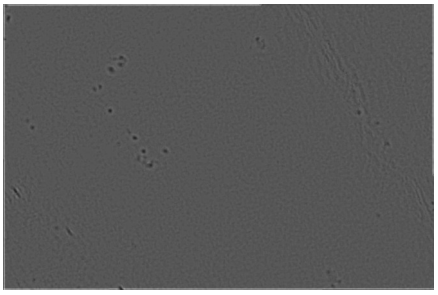
It can be observed in the histogram plot of Fig.4.10e that the difference between the intensities of LM (Fig.4.10a) and EM (Fig.4.10b) images is far too great to be overcome by an intensity-based registration method, such as the one implemented in Motion2D. However, the difference in intensities (Fig.4.10f) between the LoG representations of LM (Fig.4.10c) and EM (Fig.4.10b) images is manageable by intensity-based methods. Therefore, we decided to test our robust motion model selection on LoG-EM and LoG-LM



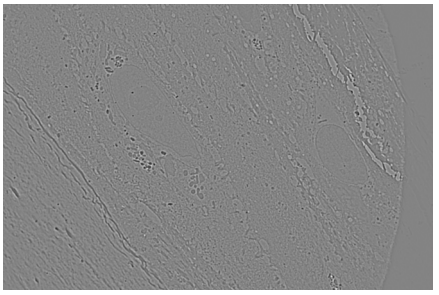
(a) Registered LM image



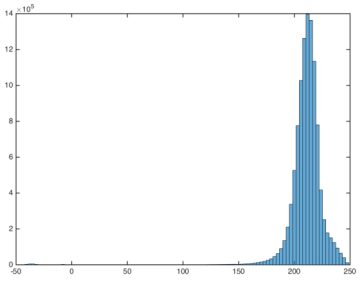
(b) Registered EM image



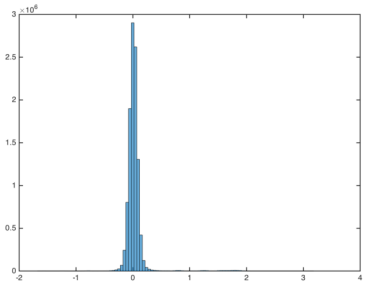
(c) LoG of registered LM image



(d) LoG of registered EM image



(e) DFD of registered LM and EM images



(f) DFD of registered LoG-LM and LoG-EM images

Fig. 4.10: *images and histograms of DFD for the CLEM experiment 1.*

patches registered with the mutual information (MI) method described in Chapter 2 to investigate if the registration can be further improved and which motion model is the most suitable to do so. We could also apply it in place of the MI-based affine registration refinement step, but previous testing demonstrated that the results were not accurate enough.

The first test was performed on the registered CLEM patches displayed in Figures 4.11a and 4.11b. To apply our method, the LoG representation of these images is used, as shown in Figures 4.11c and 4.11d. The overlay of the patches after the mutual information-based registration can be observed in Fig.4.11e. In this experiment, TR motion model was selected by our criterion $\mathcal{S}_1$ while the T model was selected by $\mathcal{S}_2$, RTIC and RBIC. The selection of these models is comprehensible since the initial overlay is quite accurate already, therefore no significant transformation is needed between the two datasets. Let us stress that the registration is the most improved with the TR model selected by criterion $\mathcal{S}_1$The registration errors of the initial registration, the refined registration using T model and the refined registration using TR model are detailed in Table 4.10, where we include the mean, standard deviation (SD), median, minimum (Min.) and maximum (Max.) of the errors computed from the corresponding overlays, in LM pixels.

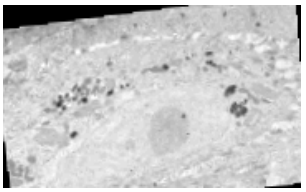
Errors	Initial registration	T	TR
Mean	0.22	0.24	0.20
SD	0.27	0.32	0.27
Median	0.17	0.20	0.15
Max.	0.67	0.71	0.50
Min.	0.00	0.00	0.00

Table 4.10: *Registration errors for experiment 1 after applying the robust motion model selection method with criteria $\mathcal{S}_2$, RTIC and RBIC for T and criterion $\mathcal{S}_1$ for TR.*

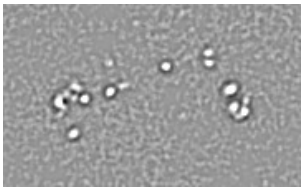
In experiment 2 (Fig.4.12) the MI-based alignment between EM and LM patches is good, although not as accurate as experiment 1. $\mathcal{S}_1$ and RTIC selected FQ as the motion model between EM and LM images, while $\mathcal{S}_2$ selected TR model and RBIC selected T model. The registration improves slightly using both models, as observed in the error Table 4.11, and in Figures 4.12f, 4.12g and 4.12h. In this experiment, TR is the one which further decreases the registration error, making $\mathcal{S}_2$ the most appropriate criterion.

Similar to experiment 1, in experiment 3, presented in Fig.4.13, the registration between LM and EM images is already accurate and not much improvement is expected. Both $\mathcal{S}_1$ and $\mathcal{S}_2$ select PSRM as motion model between the CLEM images, while RTIC

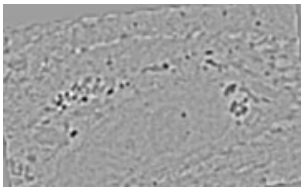
(a) Registered LM patch



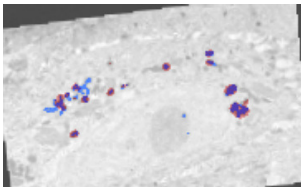
(b) Registered EM patch



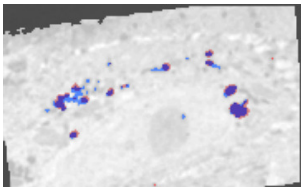
(c) LoG of registered LM patch



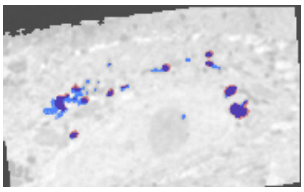
(d) LoG of registered EM patch



(e) Overlay of registered EM and LM patches



(f) Overlay after backwarping using the TR model, selected by $\mathcal{S}_1$

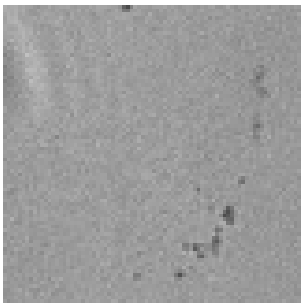


(g) Overlay after backwarping using the T model, selected by $\mathcal{S}_2$, RTIC and RBIC.

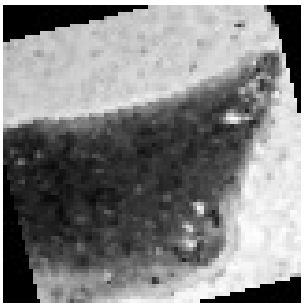
Fig. 4.11: *Experiment 1. From top to botton and left to right: Registered LM patch and EM patch. LoG representation of registered LM and EM patches. Overlay of registered LM and EM patches. Overlay after backwarping using the models T and TR selected respectively by $\mathcal{S}_1$, and by $\mathcal{S}_2$ or RTIC.*

Errors	Initial registration	TR	FQ	T
Mean	2.17	2.05	2.43	2.17
SD	1.36	1.20	1.40	1.52
Median	1.70	1.40	1.60	1.67
Max.	4.24	4.07	4.61	4.71
MIN	0.67	0.40	0.40	0.67

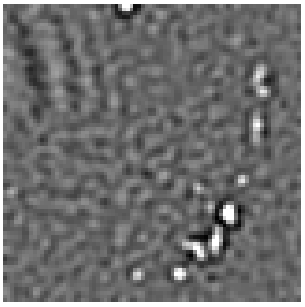
Table 4.11: *Registration errors for experiment 2 after applying the robust motion model selection method with criteria $\mathcal{S}_1$ and RTIC for FQ, criterion $\mathcal{S}_2$ for TR and criterion RBIC for T model.*



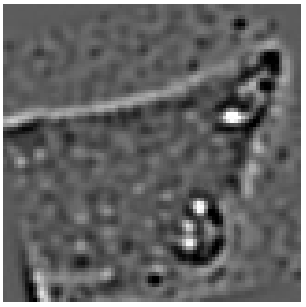
(a) Registered LM patch



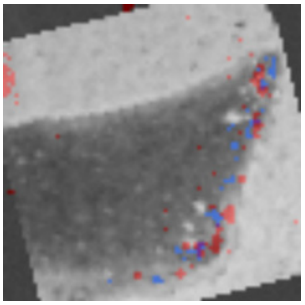
(b) Registered EM patch



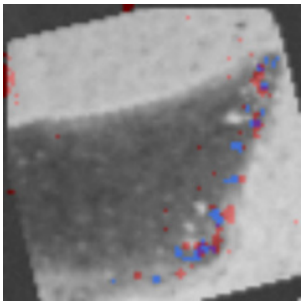
(c) LoG of registered LM patch



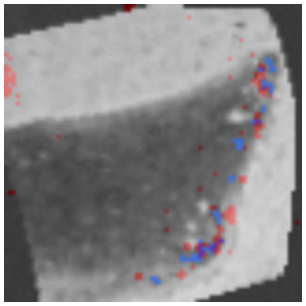
(d) LoG of registered EM patch



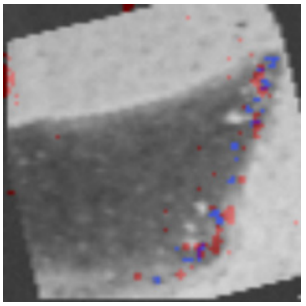
(e) Overlay of registered EM and LM patches



(f) Overlay after backwarping using the TR model, selected by $\mathcal{S}_2$



(g) Overlay after backwarping using the FQ model, selected by $\mathcal{S}_1$ and RTIC



(h) Overlay after backwarping using the T model, selected by RBIC

Fig. 4.12: *Experiment 2. From top to bottom and left to right: Registered LM patch and EM patch. LoG representation of registered LM and EM patches. Overlay of registered LM and EM patches. Overlay after backwarping using the models T, TR and FQ.*

selects FQ and RBIC selects TS. There is not significant visual improvement in the final registration between EM and LM, as observed in Figures 4.13e, 4.13f and 4.13g and in Table 4.12, where the error is computed in the EM image. However, we can infer from Table 4.12 that PSRM is more adapted than FQ, since the former decreases the error registration error. Then, $\mathcal{S}_1$ and $\mathcal{S}_2$ perform better.

Errors	Initial registration	PSRM	FQ	TS
Mean	6.6	6.46	7.49	7.3
SD	3.61	3.61	2.95	3.5
Median	6.7	5.99	6.71	6.68
Max.	10.81	10.37	11.99	10.95
MIN	2.99	2.99	3.24	2.99

Table 4.12: *Registration errors of experiment 3 after applying the robust motion model selection method with criteria $\mathcal{S}_1$ and $\mathcal{S}_1$ for PSRM, criterion RTIC for FQ and criterion RBIC for TS.*

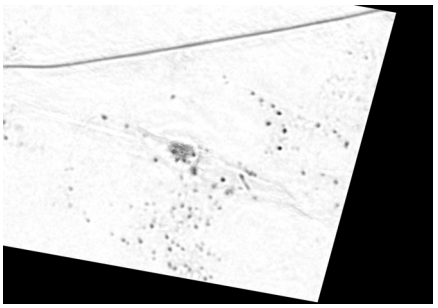
The registration in experiment 4 is good enough although it can be improved. In this particular case, the intensity values of EM and LM images are similar, therefore Motion2D can be applied directly on the original intensity images. After testing our robust motion model selection, all criteria select PSRM as the motion model which better suits the displacement between the EM and LM patches. We can notice in Figures 4.14c and 4.14d and in Table 4.13 that the alignment is slightly improved.

Errors	Initial registration	PSRM
Mean	3.76	2.53
SD	1.72	0.76
Median	4.06	2.50
Max.	5.59	3.6
MIN	1.0	1.58

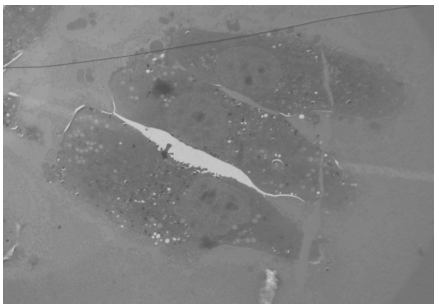
Table 4.13: *Registration errors of experiment 4 after applying the robust motion model selection method with criteria $\mathcal{S}_1$, $\mathcal{S}_1$, RTIC and RBIC for PSRM.*

4.7 Conclusion

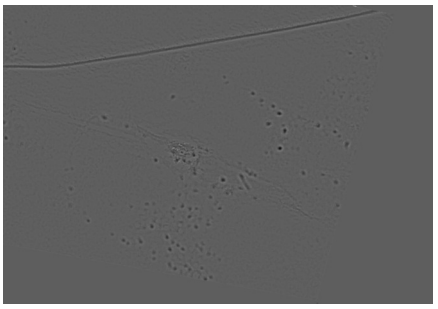
We have proposed two new robust motion model selection criteria. The first one is a robust version of the Takeuchi information criterion. The second one departs from the usual approaches by starting from the Fisher statistics. We propose two variants of the later. All three are easy to compute. The two criteria explicitly tackle the tradeoff between



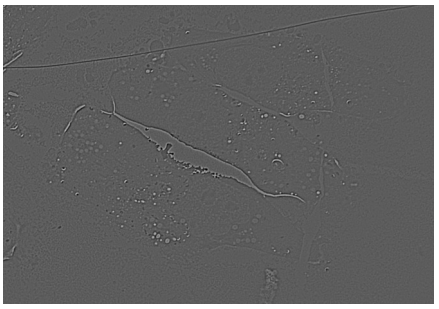
(a) Registered LM patch



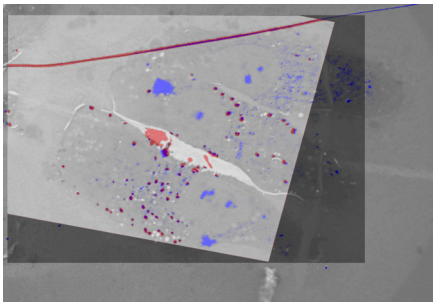
(b) Registered EM patch



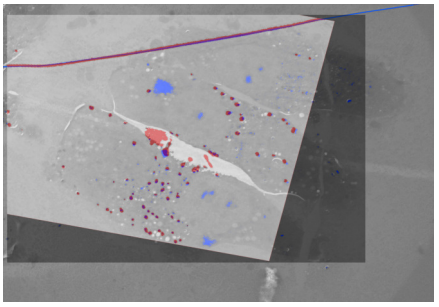
(c) LoG of registered LM patch



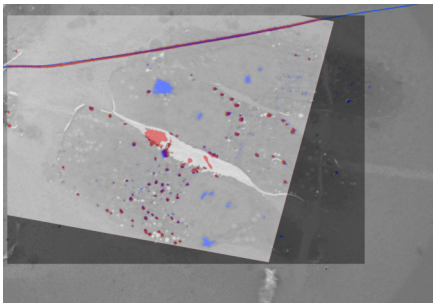
(d) LoG of registered EM patch



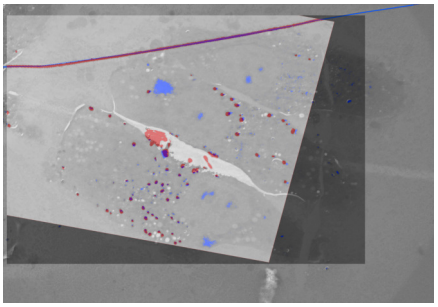
(e) Overlay of registered EM and LM patches



(f) Overlay after backwarping using the PSRM model, selected by $\mathcal{S}_1$ and $\mathcal{S}_2$

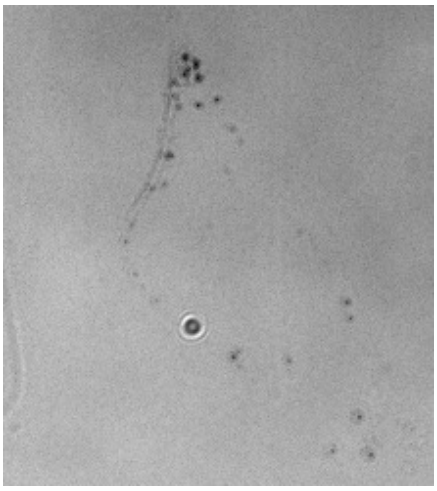


(g) Overlay after backwarping using the FQ model, selected by RTIC

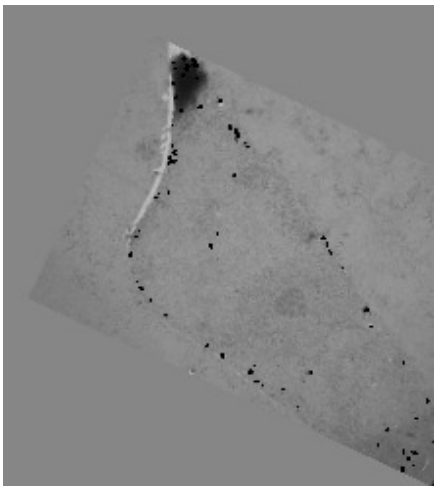


(h) Overlay after backwarping using the TS model, selected by RBIC

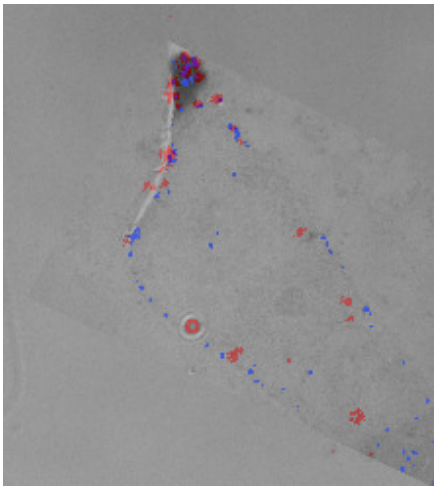
Fig. 4.13: *Experiment 3. From top to bottom and left to right: Registered LM patch and EM patch. LoG representation of registered LM and EM patches. Overlay of registered LM and EM patches. Overlay after backwarping using the models PSRM, FQ and TS.*



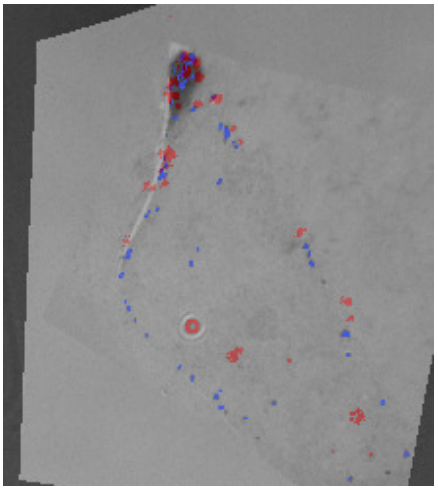
(a) Registered LM patch



(b) Registered EM patch



(c) Overlay of registered EM and LM patches



(d) Overlay after backwarping using the TR model, selected by $\mathcal{S}_1$, $\mathcal{S}_2$ and RTIC

Fig. 4.14: *Experiment 4. From top to bottom and left to right: Registered LM patch and EM patch. LoG representation of registered LM and EM patches. Overlay of registered LM and EM patches. Overlay after backwarping using the models PSRM.*

the size of the inlier set (to be maximized) and the complexity of the motion model (to be minimized). In addition, the FRIC criterion can be viewed as a proposition for a robust Mallows' C_P criterion. Experiments on synthetic and real image sequences, along with comparison with RBIC, demonstrate that our criteria provide superior performance. The application of our robust motion model selection method can be tricky in CLEM images due to their multimodality nature and the fact that the geometrical transformation to align them cannot be strictly compared to the global motion model between two successive images of a video sequence If we exploit the LoG representation of the two images first registered with the MI-based registration method to be closer to the usual video sequence configuration, our robust motion model criteria can still provide useful information about the remaining transformation between EM and LM images, which can help to improve the final overlay.

CHAPTER 5

3D CLEM

Correlative microscopy (CM) enables the study of cellular and subcellular elements in different and complementary ways by incorporating two different microscopy imaging techniques in the study of a single sample. Notably, the combination of light and electron microscopy offers a wide range of information for biological applications.

To profit from CLEM a reliable registration between the pair of images is essential. We have already proposed a general automatic registration method for 2D CLEM with promising results, and in consequence, we pursue to develop a new method for automatic 3D CLEM registration. Due to the large discrepancies in appearance, field-of-view, resolution and position expected to occur in most 3D CLEM acquisitions, a pre-alignment stage is required before any 3D fine registration stage. As for the 2D CLEM registration, we define an intensity-based method for both stages, which leverages a common representation of the two involved image modalities. We report experimental results on different real datasets of 3D correlative microscopy, demonstrating time efficiency and overlay accuracy.

5.1 Introduction

As previously mentioned, correlative microscopy has recently become an important and powerful tool in the bioimaging field. It combines two (or more) microscopy modalities, allowing the study of biological specimens in complementary ways and at different resolutions. Current tools used by biologists require manual intervention. In the case of 3D CLEM, this is an utterly complicated and burdensome task, from visually locating

elements of interest to correlate the CLEM image stacks to applying the necessary transformations by hand. As stated by biologists, this task can take up dozens of minutes for each pair of image stacks when having the advantage of using a specialized software such as ec-CLEM [51], which facilitates a semi-automatic registration of 3D CLEM images by manually picking matching pairs of points and computing a geometric transformation to register the stacks. When no such tools are available, the registration may take even more time.

Consequently, one key challenge in this area is to make the image overlay between modalities as automatic as possible for an easier and more efficient CM workflow, enabling the handling of large amount of images. This is particularly crucial for 3D acquisition. It will also favour the dissemination of 3D correlative microscopy as a valuable tool for biologists. Then, our goal is precisely to propose an automated registration method for 3D correlative microscopy.

The most frequent combination of correlative microscopy is between light microscopy (LM) and electron microscopy (EM), referred to as correlative light and electron microscopy (CLEM). As already outlined for 2D CLEM, fluorescence LM allows identification and tracking of labeled biomolecules, and provides information on their dynamics and interactions. However, the diffraction limits their resolution. Additionally, fluorescence LM cannot visualize unlabeled cellular structures. On the other hand, EM displays much higher resolution, and it can delineate cellular ultrastructure but at the cost of fixing the sample. Since there are several types of 3D LM and EM modalities, any combination results in a specific 3D CLEM method, also dependent on the biological problems under study. Other imaging techniques may accompany CLEM, such as microCT, or X-rays [47, 70, 138].

The technological and technical differences between EM and LM images are in many aspects: field of view, pixel resolution, image size, content, appearance, raising important and specific challenges for automated registration in 3D correlative microscopy. Furthermore, large gaps occur between the initial locations and orientations of the objects of interest in the respective LM and EM stacks in the data acquisition workflow. Then, geometrical and appearance discrepancies between LM and EM stacks hamper a direct computation of the 3D transformation mapping one image stack to the other.

3D multimodal image registration has a long story in medical imaging [90]. Recalling Chapter 2, there are two main kinds of approaches for registration [178]. The first one is the geometric approach, which relies on extracting specific points in each image, matching them, and computing the 3D registration transformation from the matched

pairs of points. This approach has the advantage of being able to handle large transformations. However, because of CM being multimodal and the differences in orientation, field of view and resolution, extracting points that can be correctly matched may not be the most dependable approach. The second kind of approach is intensity-based, which iteratively computes the parameters of a 3D transformation by minimizing a similarity criterion such as mutual information (MI), adopted in many multimodal applications. Intensity-based methods implicitly assume that an initial positioning is given not too far from the optimal solution. However, as we have already stated, 3D CM registration will probably start from a big misalignment. Therefore, 3D correlative microscopy raises specific issues, which makes both approaches not straightforwardly applicable for an automatic workflow.

Geometric registration approaches are the ones usually adopted in the existing semi-automatic 3D correlative microscopy workflows, but the selection and pairing of corresponding points are typically manually performed. In contrast, existing automatic segmentation methods cannot deliver repeatable point extraction in all 3D LM and EM images. Indeed, these methods may not extract a sufficient amount of reliable point pairs in the two image modalities, in particular due to the lack of structure and texture in fluorescence images. Intensity-based approaches for direct 3D registration computation are usually employed in medical imaging, since they can benefit from close initialization and isotropic volumes. This does not stand for correlative microscopy, which then requires a pre-alignment stage [75].

This chapter is organized as follows. First, we briefly survey related work in 3D CLEM. Then, we introduce the datasets used in our experiments and the corresponding image properties. Afterwards, we present the two stages of our method. We report experimental results, providing a comparison with manual interaction using eC-CLEM [51]. Finally, we present our concluding remarks.

5.2 Related work in 3D CLEM

In current 3D CLEM setups, registration is performed by a first global pre-alignment followed by a refined alignment, usually involving significant user interaction. The manual pre-alignment or matching of the EM and LM stacks is commonly performed by biologists, either during the acquisition step and/or during the post-acquisition registration step. During acquisition, the pre-alignment relies on visually identifying predefined marks on the sample holder, or exploiting fiducial markers introduced in the sample [38, 111]. Likewise, during the registration stage, the biologist has to select pairs

of points in the two images, whatever they are (fiducial markers, distinctive features on the sample, among others) [11, 27]. The landmark pairs are then used to compute the 3D geometric transformation between EM and LM stacks [15, 39, 131]. The matching step is a demanding and tedious task, especially if a large amount of data must be processed [87]. Such semi-automatic framework is adopted in [47] for registering 3D microCT and LM images, in [51] with the recently released eC-CLEM software for 3D correlative microscopy, or in [103] to overlay Scanning EM (SEM) and confocal LM volumes.

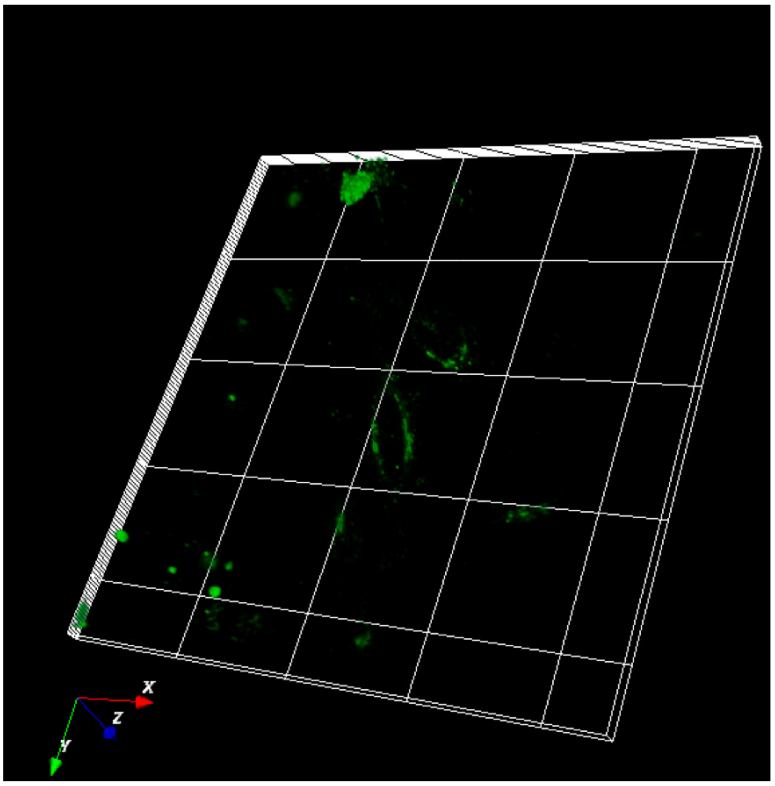
In the above works, the registration is addressed by pairing corresponding points, and the user has to iteratively re-apply this procedure (picking points, computing the transformation) until the final alignment is acceptable enough. A few works propose an automated refined registration (the pre-alignment being still manual). In [103], the cell membrane is segmented in the SEM stack and modified to a resolution similar to the one of LM. Then, EM and LM volumes are registered with an affine transformation and normalized correlation. [83] proposes an automated refined registration based on mutual information and rigid transformation. In [51], an extension to an automatic registration was also proposed, provided spots can be correctly detected in both images, but it was only tested on 2D real CLEM.

Our goal is to propose automated procedures for both the pre-alignment and the refined registration for 3D correlative microscopy. Our method is intensity-based for both stages, and involves the common representation based on the LoG transform of EM and LM images already used for 2D CLEM registration

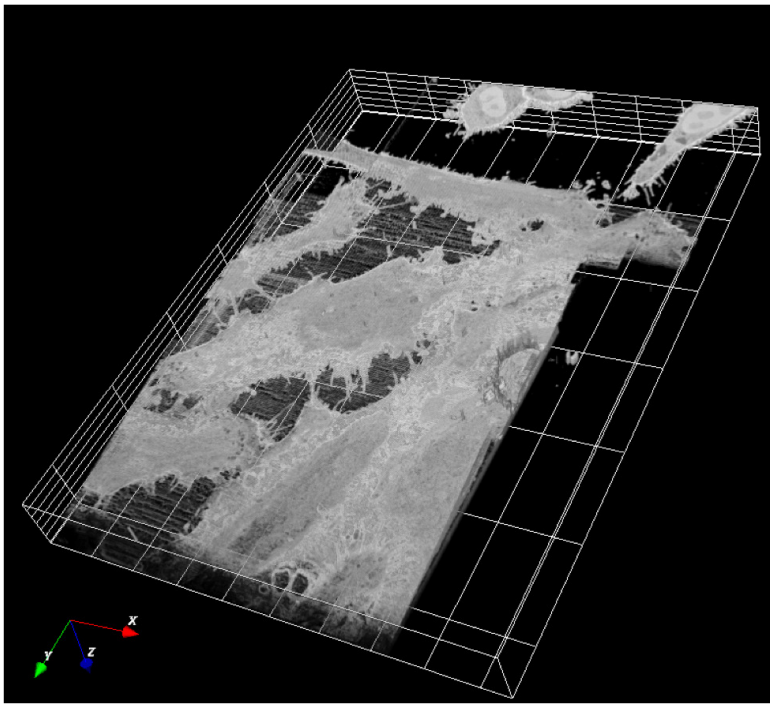
5.3 3D CLEM framework

To motivate our contribution, we first present the three 3D correlative microscopy sets we will work with, and we give information on the acquired images. The first set[1] is composed of a serial block face (SBF) SEM image stack and a confocal image stack. HeLa cells were imaged for LM in a 710 LSM Zeiss inverted confocal microscope; afterwards, they were fixed, stained and embedded for SBF-SEM imaging and acquired using a Sigma VP scanning electron microscope (Carl Zeiss). The LM stacks has two channels: transmitted light and one fluorescent channel, with volume size of $1024 \times 1024 \times 16$ and pixel size of $0.22\mu m$ in X and Y, and $0.6\mu m$ in Z, knowing that the XY-space corresponds to the lateral dimensions of the stack and the Z-axis to the vertical dimension of the stack. The size of the EM stack is $666 \times 903 \times 247$ and its pixel size is $0.14\mu m$ in X and Y, and $50nm$ in Z, which is rather low-magnification EM. A 3D rendering is displayed in Fig. 5.1.

[1] CLEM dataset provided by the Francis Crick Institute, UK.



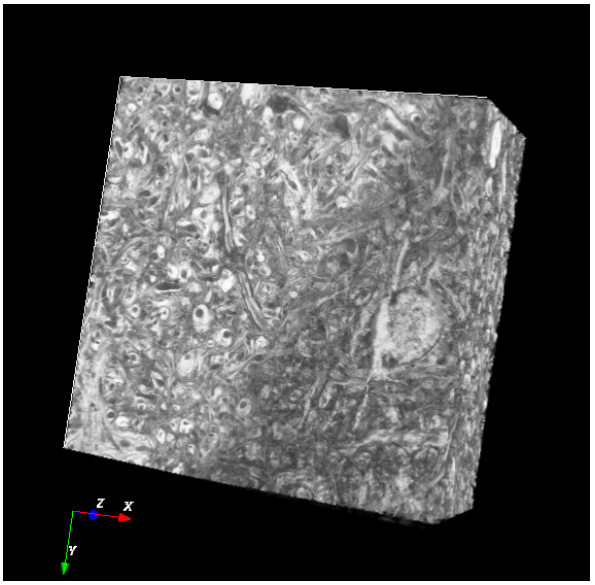
(a) Input fluroescence LM stack



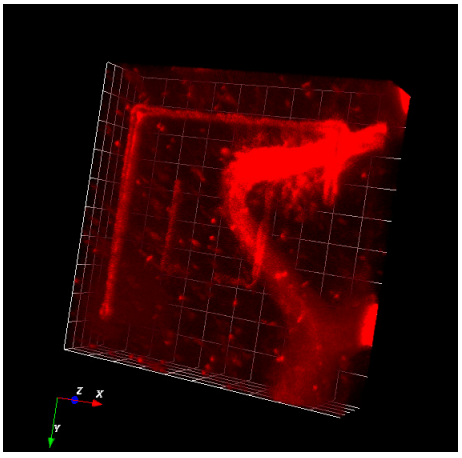
(b) Input EM stack

Fig. 5.1: *Experiment 1. Input LM and EM stacks.*

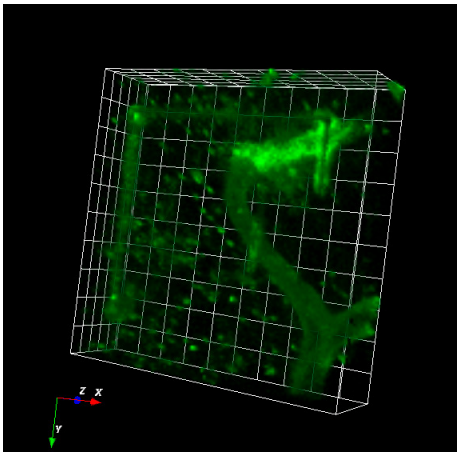
The second 3D CM[2] set depicts whole brain tissue. The EM stack of size $2048 \times 2048 \times 300$ has pixel size of $16nm$ in X and Y and $50nm$ in Z. The LM stack comprises two fluorescence channels, the red fluorescent protein (RFP) channel depicting mitochondrial structures and the green fluorescent protein (GFP) channel depicting axonal structures. The dimensions of these LM stacks are $1024 \times 1024 \times 27$ and their pixel size is $94nm$ in X and Y, and $1\mu m$ in Z. A 3D rendering of the stacks is displayed in Fig. 5.2.



(a) Input EM stack



(b) Input RFP LM stack

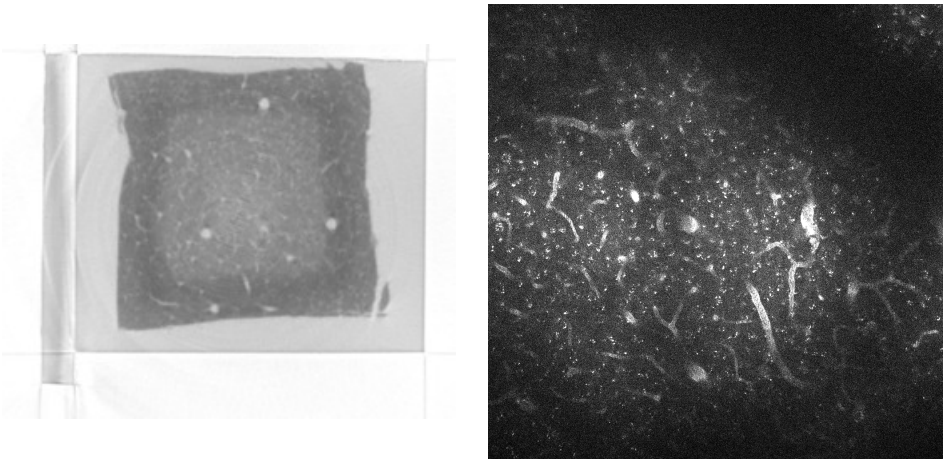


(c) Input GFP LM stack

Fig. 5.2: *Experiment 2. Input LM and EM stacks.*

[2]We thank Robert Lees from Ashby lab group, University of Bristol, UK for providing us the dataset and the associated explanations.

The third set[3] is of a different type and does not correspond to 3D CLEM. Instead, it includes a stack of microCT and a two-photon fluorescence microscopy (FM) stack. The microCT stack of size $1084 \times 896 \times 1094$ has pixel size of $1\mu m$ in X and Y and $1\mu m$ in Z, while the FM stack of size $1024 \times 1024 \times 107$ has pixel size of $0.59\mu m$ in X, $0.59\mu m$ in Y and $3.0\mu m$ in Z. A corresponding pair of slices from the FM and microCT stacks is shown in Fig.5.3.



(a) Slice 414 of microCT stack (b) Slice 81 of LM stack

Fig. 5.3: *Experiment 3. 3D microCT and fluorescence LM*

In the three datasets, the 3D stacks are anisotropic and have significant differences in terms of volume size, pixel resolution, content, appearance, and location of structures of interest between modalities. Nevertheless, in these examples, it turns out that the discrepancy between the two volumes in terms of shift along the Z-axis and of rotation in the XY plane are not that significant, but other discrepancies remain.

5.4 Our automated overlay method

5.4.1 Intensity-based XY pre-alignment

CLEM overlay may be performed in both ways, EM to LM stacks, or LM to EM stacks. To address the large gap between the initial positions of EM and LM stacks, we have defined an intensity-based prealignment method. We restrict the shift to the XY-space, since it is expected to be the larger component of misalignment (which is often the case

[3]Dataset provided by EMBL, Heidelberg, Germany.

in CLEM registration problems). To compute this initial XY shift, we accordingly reduce the dataset to two dimensions. We compute a projection along the Z-axis of the LM stack, using the maximum intensity projection (MIP), which yields a good distribution of the volume content. If the LM stack is of size $m \times n \times l$, LM-MIP is a 2D $m \times n$ image obtained by scanning LM intensities along the Z-axis and selecting the highest one among the l intensity values for each (x, y) location of the image grid. The same cannot be applied to the EM stack, since the resulting MIP image will be saturated due to the high resolution and great quantity of elements imagined by EM. Then, we simply select a few slices in the EM stack for the pre-alignment stage intersecting the predefined region of interest (ROI) delineated in the EM stack. The overlay is indeed focused on a region of interest (ROI) of the biological sample, which may involve a specific cell or a subcellular structure in a cell, and is specific to each biological project. The ROI is expected to be delineated either by the biologist (which will be in that case the only user interaction), or by an algorithm able to detect the content of interest.

In our case, we start with the EM stack. Given a 3D ROI delineated in the EM stack, we take the corresponding 2D ROIs in the selected EM slices, and we search for their corresponding patch in the LM-MIP image. The patch dimensions of the LM-MIP image are inferred from the EM-ROI ones, while taking into account the ratio between pixel resolutions available in the metadata attached to the acquisition. To facilitate the search, we require a comparable appearance for the two images. To do so, we use the Laplacian of Gaussian (LoG) transform we introduced in Chapter 2 for 2D CLEM registration. The LoG transform maps a 2D EM image and a 2D LM image to closer representation, decreasing the big difference in appearance between the images, as illustrated in Fig.5.5. The LoG transform enhances high spatial frequencies while suppressing local linear intensity variations. The LoG transform of an image I is expressed as:

$$\mathcal{L}_\sigma(I) = \Delta(G_\sigma * I), \tag{5.1}$$

where the Gaussian standard deviation σ acts as an adaptable scale linked to the size of the objects of interest in I. It is automatically selected by using the method described in [7] which exploits the Lindeberg's scale-space approach. We select the scale with the maximum number of blobs, which are local minima in the constructed scale-space domain. By adapting the LoG scale to the EM and LM image content, we lessen the difference in content scale of LM and EM images.

Once a common representation of EM and LM images is created, we have to match the 2D LoG-EM-ROI with its corresponding patch in the 2D LoG-LM-MIP image. We handle several 2D LoG-EM-ROIs, one by EM slice, to ensure robustness by making the

matching not dependent on one particular 2D EM slice. Then, a 2D patch-based exhaustive search is carried out over the LoG-LM-MIP. Due to the differences in orientation between the LM and EM images, it is necessary to adopt a similarity measure that is invariant to rotation and scaling. As already discussed in Chapter 2, we take histogram-based descriptors for the matching process. We have used two types of histograms: the LoG-values histogram and the Local Directional Pattern (LDP) [61]. We have also implemented two histogram distances as similarity measures: the cosine similarity and the histogram intersection distance to evaluate the similarity of the LoG-EM-ROI with each tested LoG-LM-MIP patch.

If we use n EM slices taken from the EM stack, we have n candidate patches sp for the ROI location in the 2D LM-MIP image, as illustrated in Fig. 5.4. The final ROI center location in the LM-MIP image is selected using a robust technique since outliers may occur, i.e., due to wrong matching. This robust technique comprises the following steps:

1. The median of the center location of all candidate patches is computed, $\text{median}(sp)$.
2. The median absolute deviation of all center locations, $\text{mad}(sp)$, is computed.
3. To discard outliers, the absolute distance of all patches centers with respect to the median, $\text{d}(sp, \text{median}(sp))$, is used as follows:

$$\text{if} \quad \text{d}(sp_i, \text{median}(sp)) \leq \text{mad}(sp) \quad \text{then} \quad \text{inliers} \leftarrow sp_i, \quad i \in \mathbb{N}, \qquad i \leq n \tag{5.2}$$

4. The location of the finally selected patch is computed as a weighted average of the center positions of the inlier patch candidates. The weights are defined as an inverse of the distance between the center of the patch and the median of the centers positions of inlier patches candidates:

$$w_i = \begin{cases} \frac{1}{\text{d}(sp_i, \text{median}(sp))^2} & \text{if } \text{d}(sp_i, \text{median}(sp)) \neq 0 \\ 1 & \text{if } \text{d}(sp_i, \text{median}(sp)) = 0 \end{cases} \tag{5.3}$$

5. The finally selected patch is defined as

$$sp_f = \frac{\sum_{i=1}^{n} w_i sp_i}{\sum_{i=1}^{n} w_i} \tag{5.4}$$

The difference between the center location of the finally selected patch in the LoG-LM-MIP image and the center of any 2D EM-ROI (since the 2D EM-ROIs are vertically aligned in the 3D EM-ROI stack) yields the 2D shift in the XY-space. Then, the full LM stack is shifted in the XY plane by applying the estimated 2D translation.

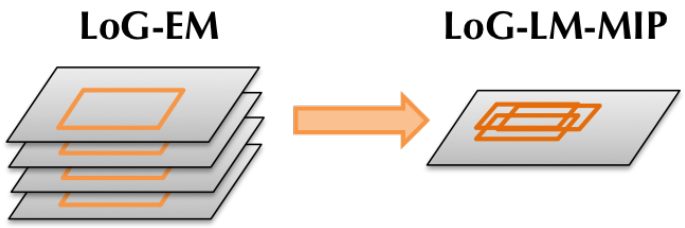


Fig. 5.4: *ROIs extracted from multiple LoG-EM slices to be located in the LoG-LM-MIP image*

To summarize, our pre-alignment method comprises four main steps:

1. Maximum intensity projection of the LM stack to a 2D image
2. LoG representation of 2D-LM-MIP image and 2D EM slices
3. Matching of the ROI of every LoG-EM slice with a patch selected in LoG-LM-MIP image
4. Pre-alignment of EM and LM stacks by the estimated XY-shift

We implemented a reproducibility test in order to validate the need for a pre-alignment stage in the 3D registration framework. This test can be found in Annex A.

5.4.2 3D ROI-based registration

Once the pre-alignment is achieved, we can estimate the 3D transformation which will allow to overlay the 3D LM stack onto the 3D EM stack. The 3D transformation is computed between the 3D EM-ROI and the pre-aligned 3D LM-ROI. The transformed LM stack is resampled to fit the same size in the Z-dimension as the EM stack using a bilinear interpolator. The 3D interpolated LM-ROI is then recovered using the same coordinates of the 3D EM-ROI.

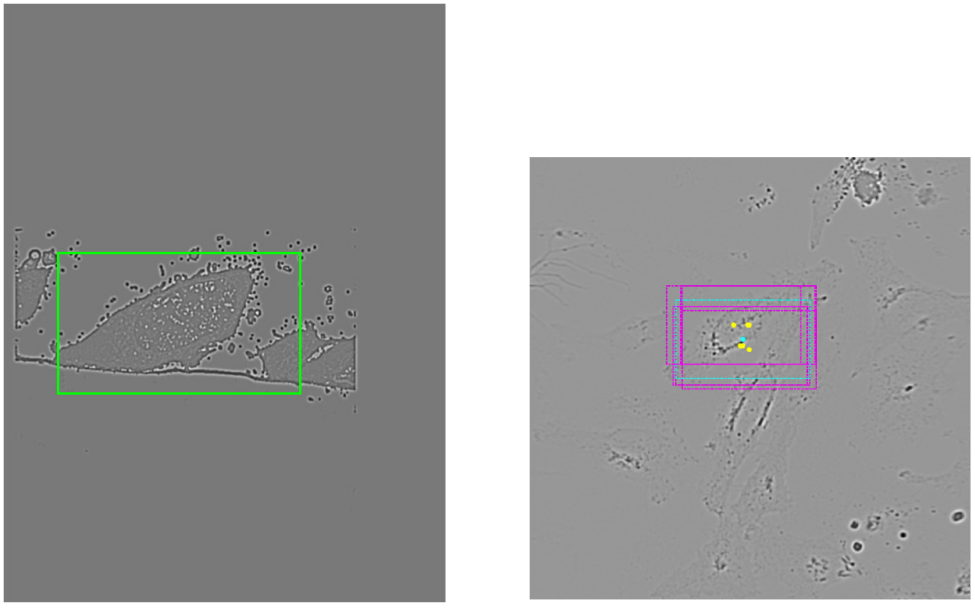
The contents of the two CM stacks are different enough so that a point-to-point intensity-difference similarity measure cannot provide satisfactory results. We resort again to mutual information (MI) as a similarity measure, due to its well-known capability of handling multimodal images. First, a 3D rigid transformation (composed of a rotation and a translation) is estimated. However, given the 3D CLEM workflow, it is not sufficient to fully account for the geometric relationship between the LM and EM stacks. The EM slices do not lie in the same plane as the LM slices, and distortions between

stacks may occur during the acquisition. Consequently, we further refine the registration with a 3D affine transform.

Our overall method does not involve any extraction of image features from the EM and LM stacks. It is merely intensity-based, which prevents from any possible segmentation errors and saves computation time. Furthermore, it does not require any parameter setting by the user. We use ITK libraries in C++ for the MI-based registration.

5.5 Experimental results

Experiment 1: EM and LM modalities. To pre-align the EM and LM stacks, we follow the 2D search process described in Section 5.4.1. A 2D projection of the LM stack is generated using MIP. EM slices 107-118 were selected to extract 2D ROIs, whose LoG representation (Fig.5.5a) is matched within the LoG-LM-MIP image following an exhaustive search. Using [7], we obtain an approximate scale for the LoG transform of EM and LM images, respectively estimated as $\sigma_{EM} = 4.3$ and $\sigma_{LM} = 2.5$.
A corresponding patch in the LoG-LM-MIP is found for each LoG-EM-ROI by the patch-



(a) LoG-EM slice (b) LoG-LM-MIP image

Fig. 5.5: *Experiment 1. a) Slice 107 of LoG-EM stack with the ROI delineated in green, b) LoG-LM-MIP image with patch candidates overprinted in magenta and the finally selected one in light blue.*

based search, with a 90% overlap between tested patches, using LoG-value histograms

along with both cosine and histogram intersection distances. Then, we robustly compute the finally selected patch from the weighted average of the centers of the patch candidates (Fig.5.5b). The computed 3D rigid registration corresponds to a 3D translation of $(-1.21, -0.5, 7.17)$, and a rotation of angles $-7.52°$, $-2.63°$ and $-1.46°$ around the X, Y and Z axis respectively. The final overlay of the registered EM and LM ROIs is shown in Figure 5.6. The accuracy can be appreciated by observing the fluorescence (in green) correctly surrounding the cell nucleus in the EM slice. By picking corresponding points in LM and EM stacks, we evaluated an average registration error of 6.5 EM pixels.

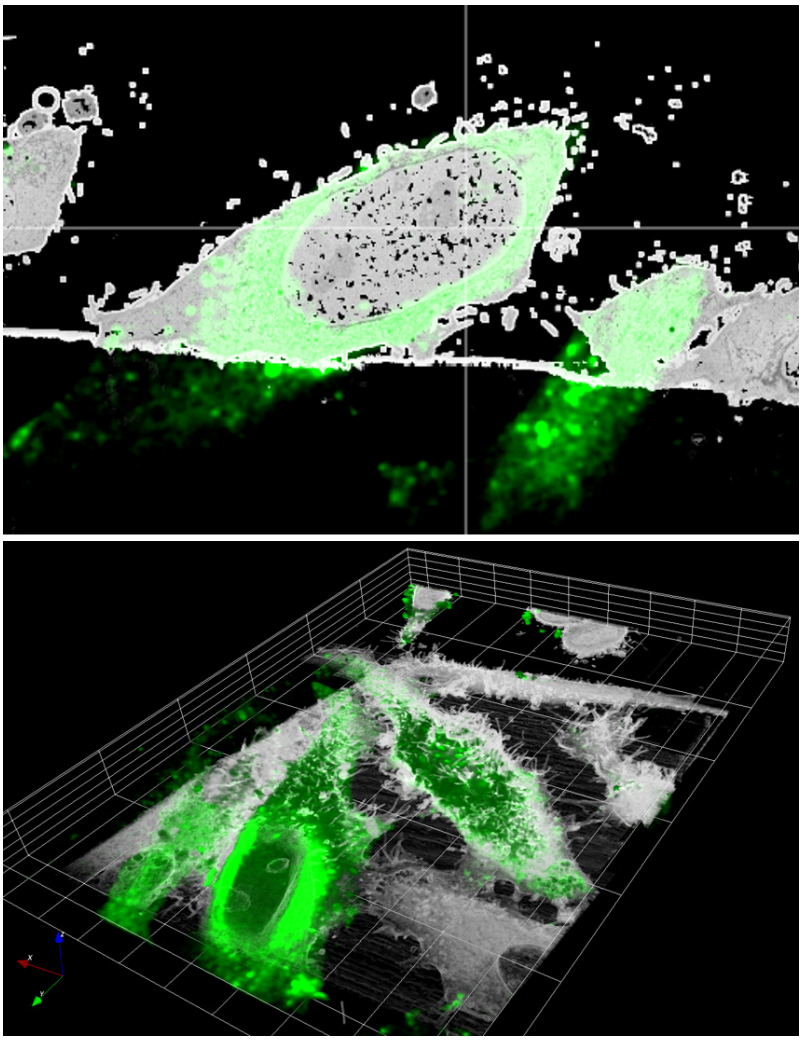
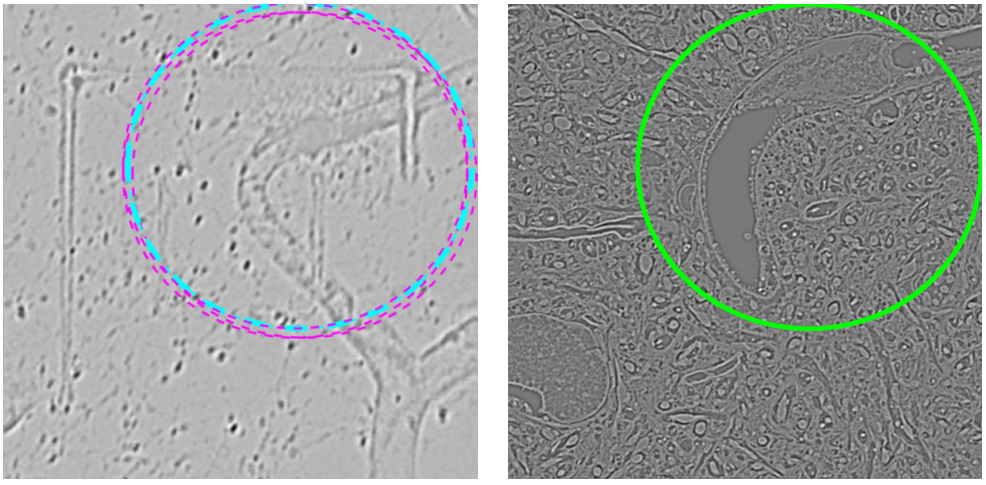


Fig. 5.6: *Final overlay for experiment 1. The fluorescence LM is overprinted in green. Top: Overlay visualized for one given slice. Bottom: Overlay visualized on the full 3D stack.*

To compute this registration, it takes 92.4 seconds for the pre-alignment and 13.5 seconds for the 3D registration. Comparatively, manual correlation by a trained biologist

on a known dataset using eC-CLEM took about 20 minutes, most of the time being dedicated to primary orientation. Once the initial rotation, translation and scaling are achieved (the first four points required in eC-CLEM), the fine correlation can be estimated to 5 minutes per points-pair until the desired resolution is reached. On an unknown dataset, initial correlation can take more than one hour.

Experiment 2: EM and LM modalities. In the second experiment, slices 100-120 from the EM stack are used to extract a set of 2D ROIs after generating their LoG representation. The location corresponding to each 2D ROI is searched in the LoG representation of the MIP of the LM stack. The LoG scales estimated for the EM slices and the LM-MIP are $\sigma_{EM} = 22.2$ and $\sigma_{LM} = 18.5$, respectively. Fig. 5.7 displays the LoG representation of the LM-MIP and the slice 60 of the EM slice, where the ROI is delineated in green.



(a) LoG-LM-MIP image (b) LoG-EM slice

Fig. 5.7: *Experiment 2. a) Slice 60 of LoG-EM stack with the ROI delineated in green, b) LoG-LM-MIP image with patch candidates overprinted in magenta and the finally selected one in light bue.*

An exhaustive search for each 2D ROI extracted from the LoG-EM slices is performed over the LoG-LM-MIP image using the same parameters and procedure as in the previous experiment. The resulting patch candidates are plotted in Fig. 5.7 in magenta, along with the weighted patch computed from the set of candidates, framed in light blue. After applying the XY shift corresponding to this pre-alignment step, a 3D rigid transformation followed by an affine transformation are computed between the 3D EM-ROI and the 3D LM selected patch (LM-SP) to create the final overlay displayed in Fig. 5.8. The registration accuracy can be visualized by observing the red fluorescence overlapping the

white structure of the EM image. 88.2 seconds are necessary for the pre-alignment step, while the 3D registration takes 70.56 seconds. The average error computed is 10.8 EM pixels.

Experiment 3: MicroCT and LM modalities. In this experiment, slices 15-25 are extracted from the microCT stack, and a 2D ROI is extracted from each of their LoG representations. Then, a MIP of the LM stack is computed. The estimated LoG scales are $\sigma_{\mu CT} = 22.2$ and $\sigma_{LM} = 2.1$. The full LM stack is pre-aligned by applying the shift given by the first stage. Then, we compute the 3D rigid transformation and the affine one. Computation time is 234 seconds for the pre-alignment stage and 15.4 seconds for the 3D registration. Visual evaluation of the overlay is given in Fig.5.9. It is normal that most of the elongated shapes are not overlaid, since they are present in only one of the two modalities. The average registration error is 10.6 pixels.

5.6 Conclusion

We have defined an original and efficient intensity-based method for automatically overlaying 3D image stacks of different modalities in correlative microscopy (CM). Whereas existing workflows require substantial user interaction, our method can cope with large appearance, resolution and position discrepancies in an automatic way. It proceeds in two stages: 2D pre-alignment and 3D fine registration. It does not involve any parameter setting, and then, is straightforward to use by biologists. Results on several combinations of 3D correlative microscopy show its genericity. Registration accuracy is satisfying for a fully automatic overlay of the two 3D image stacks. Future work will deal with still more general handling of the pre-alignment (fully 3D shift) and possibly non-rigid 3D registration, while continuing to validate the proposed method on other types of 3D CM datasets, when more datasets would be available.

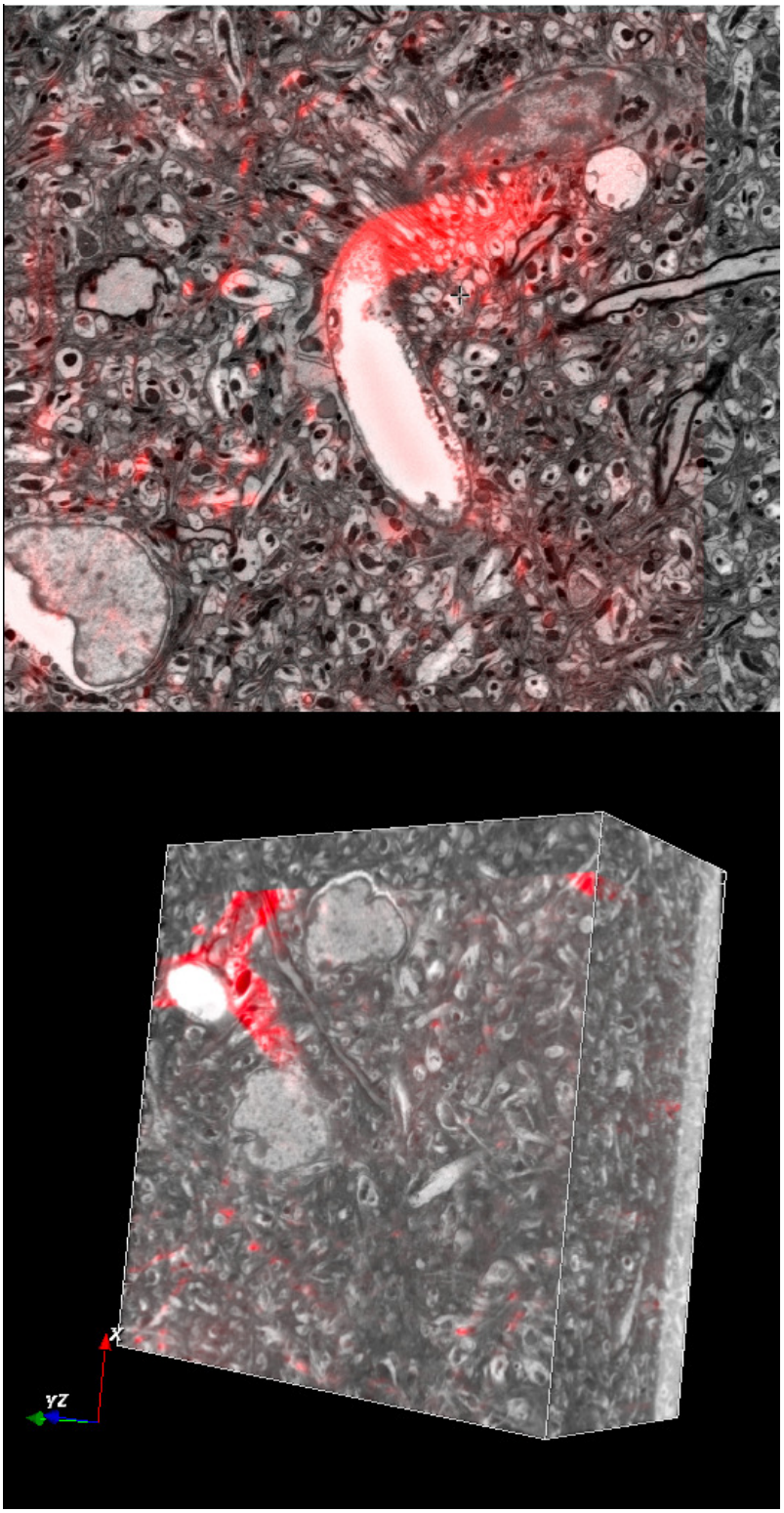


Fig. 5.8: *Final overlay for experiment 2. The fluorescence LM is overprinted in red. Top: Overlay visualized for one given slice. Bottom: Overlay visualized on the full 3D stack.*

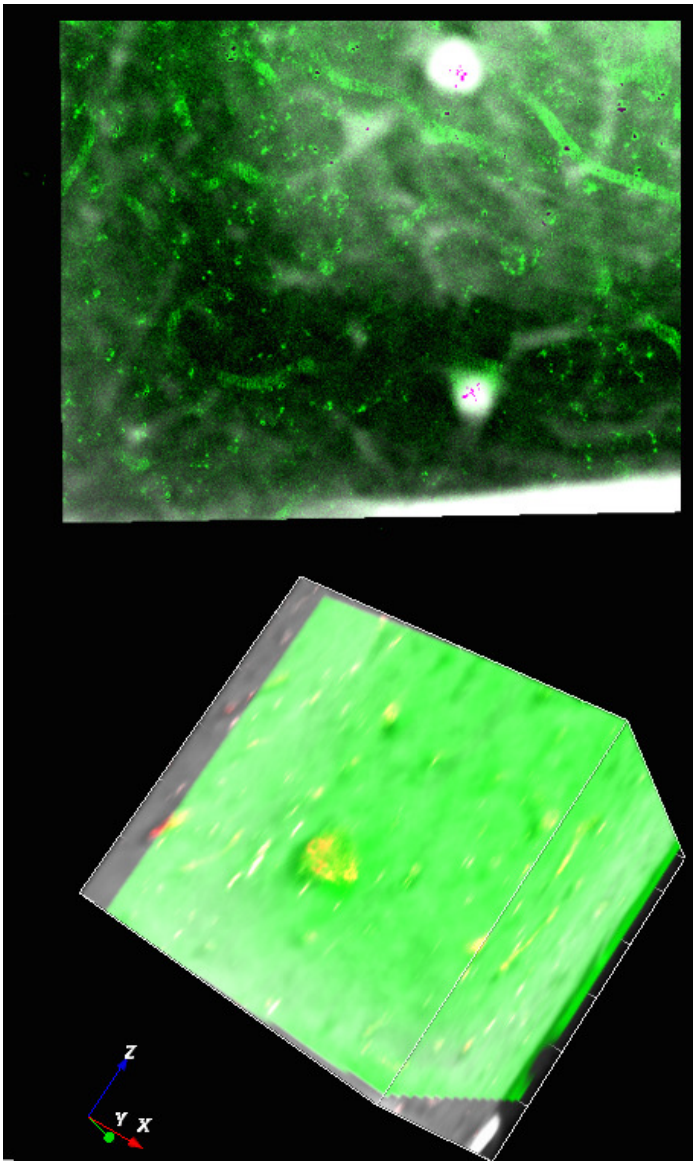


Fig. 5.9: *Final overlay for experiment 3. The fluorescence LM is overprinted in green. Top: Overlay visualized for one given slice. Bottom: Overlay visualized on the full 3D stack.*

CONCLUSIONS AND PERSPECTIVES

Contributions

Correlative microscopy is a recent methodology where two different microscopy imaging techniques are combined to generate new and important information about a biological sample. It has many different applications in the fields of cell biology, cancer research, marine biology and neuroscience. Different imaging modalities can be employed for CLEM studies, such as light microscopy (LM), electron microscopy (EM), micro-computed tomography (microCT), x-ray microscopy, ultrasound microscopy, etc. although the most commonly used are light microscopy and electron microscopy, denoted as correlative light electron microscopy (CLEM). Light microscopy techniques allow the observation of living samples and the labeling of cellular elements through the use of fluorescent markers. However, LM has relatively low resolution and poor structural information. Electron microscopy, on the other hand, has very high resolution, allowing the observation of ultrastructural details. However, a sample must be fixed to be imaged with EM, which effectively eliminates any living specimen. It is evident that the complementary nature of EM and LM techniques, and their combination have provided biologists with highly valuable information to answer many kinds of biological questions.

Multiple different protocols exist for CLEM acquisition, and they are usually defined depending on the subject of study. As output of the CLEM pipeline, the resulting images are registered, generally creating an overlay of the EM and LM images. Existing CLEM acquisition protocols may extend through several days.

We have mainly focused our research work on the last stage of the general CLEM workflow, the registration of EM and LM images. Currently, the greatest part of the existing investigations regarding CLEM images are focused on the biological part of the

problem, developing new CLEM setups and protocols, or adopting existing ones to different kinds of problems. To analyze the resulting images, it is necessary to align them, so that the information they contain can be visually and quantitatively examined. This implies that on image (or both images) is rotated, scaled, shifted, and sometimes deformed to match or correlate the shared imaged structures and overlay the images. In these works, the alignment or registration process is completely or partly carried out in an intensive interactive way using an image editing software. This is a difficult task, since visually locating the shared structures of interest and applying the necessary transformations can be burdensome and time-consuming and it could still produce unsatisfactory results. Additionally, it requires a certain level of expertise and previous knowledge of the current dataset. Very few automatic 2D CLEM registration approaches have been investigated so far, and the majority is designed for pre-registered images. Semi-automatic CLEM registration tools are also available, based on the manual selection of features of interest to compute the targeted transformation.

Consequently, we proposed methods for automatic 2D CLEM registration. Due to significant differences in terms of intensity values, pixel resolution, image size, field of view, orientation of the sample and content, automatic registration of EM and LM images is not a straightforward process. EM and LM images are expected to be initially largely misregistered. Different methods must be found to tackle each of the discrepancies aforementioned in a fast and effective way. Throughout our research project, we explored different image processing techniques to find which were more amenable to the CLEM registration problem. General image registration approaches can be classified in two ways: feature-based registration and intensity-based registration. Feature-based registration methods are usually able to handle large transformations, provided that enough features are extracted from the images and their pairing can be found. Intensity-based methods exploit information from the whole image to compute the transformation needed for registration by optimizing a similarity metric. It is profitable when reliable features cannot be extracted, but it assumes that the initial misregistration between the images is small.

After analyzing the results of the application of different feature detection approaches, we concluded that a feature-based registration method would likely be unable of producing satisfactory results, since the extracted features were often not repeatable in the two types of images, and when common to the two images, were too far from each other, making their correct automatic pairing difficult. corresponding features are not guaranteed to be extracted. Therefore, intensity-based methods were selected.

The implementation of intensity-based registration methods is very straightforward in single modality problems, where numerous similarity metrics exist for computing

the transformation. However, the implementation becomes more complicated when dealing with multimodal images. We adopted the mutual information scheme to create our CLEM registration framework. Beforehand, to perform the search step for pre-alignment, we decreased the appearance difference between EM and LM images in terms of intensity, by adopting a Laplacian of Gaussian (LoG) transform equipped with an automatic scale selection framework. We came up with a common representation of the input images making them easier to compare, while smoothing out noise and enhancing structures in the image.

Given an initial large misregistration is expected, directly computing a transformation to register the input EM and LM images would not generate satisfactory results for sure. Then, we designed a pre-alignment step to bring the EM and LM images far closer. Since a region of interest (ROI) is assumed to be delineated either on the LM or the EM image, we designed a patch-based exhaustive search method to locate the patch matching the ROI in the other image. The ROI is extracted from the LoG representation of the source image (either LM or EM) and is located in the LoG representation of the target image (EM or LM) by exploiting histogram-based similarity measures, which are invariant to scale and rotation. The search patch is defined from the geometry of the ROI and the resolution ratio between the EM and LM images. This step supplies a shift transformation. We tested this pre-registration step on seven CLEM image pairs, all containing very different structures, image intensities and fields of view. In all cases, a patch containing the ROI was correctly located, demonstrating the effectiveness of our patch-based exhaustive search approach.

Once the shift transformation is applied to the source image, the ROI is extracted from both target and source input images. Then, we implemented a mutual information (MI) based registration process. First, since the difference between the orientation of the input images is expected to be large, we computed a coarse rotation angle by an exhaustive method. Then, after applying the rotation, we compute rigid followed by affine transformations to obtain the final registration of the EM and LM images. By visually evaluating the overlays created from the registered EM and LM ROIs and computing registration errors, we observed that our 2D CLEM registration framework performs satisfactorily, in some cases with registration errors around 1 pixel in the LM image. Although it is expected that semi-automatic registration methods where the user manually picks corresponding features to register CLEM images have lesser registration errors, our method provides good results for a fully automatic approach.

Additionally, we explored feature extraction methods for spot-like structures in images. Since many of the elements imagined by fluorescence microscopy techniques are

shaped like spots of various sizes, we developed a multiscale spot detection and segmentation method. One of the most widely known methods for spot detection is the use of the LoG transform. By specifying the Gaussian scale, which is related to the size scale of elements of interest in the image, spots of such scale will be highlighted and a method for their segmentation can be implemented. Based on a *a contrario* approach, we designed a method for automatically selecting the of meaningful scales in the processed image. Then, by an adaptive thresholding, the spots associated to each selected scale can be extracted. Finally, we defined a coarse-to-fine approach for generating a final segmentation map containing all spots detected at the different scales. We successfully tested our method on both synthetic and real images. Afterwards, we defined a registration refinement step based on our multiscale spot detection method. By applying our method, we extracted significant spots from pre-registered CLEM images. Our objective was to improve the final registration by implementing a point-based registration method based on a random sample consensus (RANSAC) technique and nearest neighbor pairing. We validated it on five CLEM datasets with satisfactory results. Then, we can conclude that a feature-based registration can still improve the registration of 2D CLEM images, provided a correct extraction of features is available.

Knowing the motion model that better describes the global motion between two images can very helpful in many computer vision problems. We developed two robust motion model selection criteria. Based on the robust estimation of a set of parametric motion models between two frames in an image sequence, our two criteria can select the motion model that better fits the global motion between two successive images of the sequence. One criterion is a robust extension of the Takeuchi information criterion, while the other one is based on the Fisher statistic. We tested extensively our method on synthetic and real video sequences with very satisfactory results. Additionally, we exploited our criteria on registered 2D CLEM image pairs to infer information on which motion model would better fit the remaining transformation between the images. Since the images were already well registered, our method mostly selects simple motion models, such as translation and translation plus rotation, which is to be expected.

Finally, we extended our research work to 3D CLEM registration. Similar to 2D CLEM, very few investigations have been undertaken regarding automatic 3D CLEM registration, while most works focus on semi-automatic approaches based on user-assisted selection and matching of features. Manual alignment of 3D CLEM images is even more complicated than in the 2D case, taking significantly more time, requiring more expertise and more prone to registration errors. Discrepancies similar to 2D CLEM are to be expected between 3D EM and 3D LM images in terms of appearance, content, field of view, orientation, volume size and pixel resolution. We proposed a method analogous

to the approach developed for 2D CLEM, where we exploit the common LoG representation of the 3D CLEM images to compute a pre-registration step. A robust patch-based exhaustive search approach is implemented to locate the ROI in the target image by creating 2D representations of the image stacks. After the XY pre-alignment step, a 3D MI-based registration is performed by computing rigid and affine transformations. We tested our method on three 3D CLEM datasets with good results for an automatic registration approach.

Perspectives

Short-term perspectives

Occasionally, deformation can be introduced to the sample during the CLEM workflow due to the fixation, embedding or sectioning. We have limited of our method to rigid and affine transformations to register EM and LM images. However, further estimation of deformable methods could provide valuable results on cases where the sample is distorted. In addition, incorporating spatial information for topology preservation through the use of regularization terms in both 2D and 3D registration frameworks is likely to be profitable for improving the final registration step. Although we have good computation times for the 2D and 3D registration experiments, they can still be improved through optimization of the code.

Further testing on 2D and especially 3D CLEM datasets will help to assess the robustness or our method.

We have successfully tested our method in a simulated low-to-high magnification acquisition of EM images guided by the registered LM image. Additional testing of similar cases will add to the versatility of our method.

Developing a plug-in of our code our within a specialized biological image processing software such as ICY[4] could help spread the use of our method as a valuable tool for biologists.

[4] http://icy.bioimageanalysis.org/

Long-term perspectives

A new application of our method within the CLEM framework could be envisioned. Given that LM provides a wide field of view of the sample, our method could be utilized as a positioning system acquisition protocol. If multiple ROIs are delineated in the LM image, their correspondence is sought for in a low-magnification EM image of the same sample using our method. Then, high magnification EM images can be acquired for all the localized regions. This could improve and speed up the acquisition workflow within the CLEM framework.

Exploring the generation of synthetic 2D and 3D CLEM datasets for validation of our method and other automatic registration approaches is of the utmost importance. Additionally, exploring emerging technologies such as machine learning techniques may help to create interesting new strategies. However, to implement any machine learning method it is important to have access to far more annotated data.

Although we have successfully tested our 2D and 3D CLEM registration frameworks on real datasets, it is obvious that more testing needs to be carried out. Unfortunately, the availability of 3D CLEM dataset is low, even though CLEM setups and protocols have been popularized in recent years. Many datasets are confidential and openly sharing data is still a "sensitive" topic in the biological community. However, we expect that the diffusion of tools like the one created in this thesis project will help in propagating and creating new collaborations.

APPENDIX A

3D REPRODUCIBILITY TEST

Reproducibility test

Implementing a pre-registration stage in the CLEM registration framework seems like an obvious choice, specially for 2D CLEM images. However, when working with 3D CLEM, the need for a pre-registration stage may not be as clear, given that the ROI and ground truth location on the source and target 3D stacks appear to be closer than in the 2D CLEM case, at least in the datasets that we analyzed. To validate the importance of the pre-registration stage, we implemented a reproducibility test. The main objective is to analyze the impact of the XY shift given by the pre-alignment step on the 3D MI-based stage of the registration workflow. As previously explained, the 2D pre-alignment is defined in the XY plane since it is where the biggest discrepancy is expected. The 2D shift is computed from the distance between the x and y coordinates of the center of a predelineated ROI in the first image and the coordinates of the corresponding patch in the second image, obtained through a robust patch-based exhaustive search approach. After applying the 2D shift to the 3D LM stack, 3D rigid and affine transformations are estimated to register the LM and EM stacks, using mutual information as similarity metric.

A XY translation of various random magnitudes is applied over the finally registered 3D LM stack, generating a set of altered 3D LM stacks. Then, the consecutive 3D rigid and affine transformations are computed to register the altered 3D LM and EM stacks, in the same way as the registration computed for the unaltered 3D LM stack in the experiment presented in Section 5.5 of Chapter 5. Then, the error between the altered LM stacks and

the unaltered LM stack after registration is computed by measuring the average distance between the coordinates of the 8 vertices of the bounding box of the 3D LM patches after registration and the coordinates of the 3D EM-ROI. This process is illustrated in Fig. A.1.

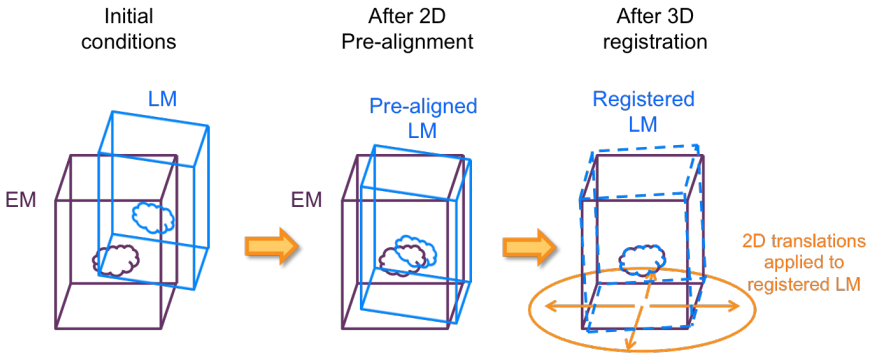


Fig. A.1: *Registration process for 3D CLEM. First, a 2D shift in the XY space is computed to prealign the 3D LM and EM stacks. Then, the 3D registration is estimated. For the reproducibility test, 2D translations are applied to the registered LM stack to create a series of altered LM stacks. Then, the 3D registration is re-estimated.*

Starting from the registered location, XY translation is drawn at random within $[-130,-2] \cup [2,130]$ for X and Y (independently). The drawn translation is rejected if its magnitude lies outside the disc defined by $\sqrt{tx^2+ty^2} \leq 130$, where tx and ty are the translations drawn for X and Y respectively. A set of 650 altered 3D LM patches were created by applying an XY translation of random magnitudes ranging from 2 to 130 pixels to the registered 3D LM of the first experiment presented in Chapter 5. In this experiment, the 2D shift computed between the center of the pre-delineated EM-ROI and the corresponding patch in the LM-MIP image is $(-205.9, 30.5)$ pixels in the XY space and it has a magnitude of 208.2 pixels, while the estimated 2D translation induced by the 3D registration is $(10.0, 50.1)$ pixels in the XY space, with magnitude of 50.99 pixels. The distance mentioned in the previous paragraph is computed between the bounding boxes of each 3D altered LM patch and 3D EM-ROI after estimating the 3D rigid and affine transformations. The resulting distance values are presented in Fig. A.2 in terms of the magnitude of the translation applied to generate the 3D altered LM stack.

It can be observed that the average distance (or error) between the vertices is mild for translations of magnitudes between 2 and approximately 98 pixels. The resulting registrations are classified as good by visually examining the overlay of the registered LM patches and 3D EM-ROI. Then, there is a clear transition around translations of magnitude 98, where large errors occur and the resulting registrations are not satisfactory. Two examples of the registration of altered LM patches and the EM-ROI are displayed in

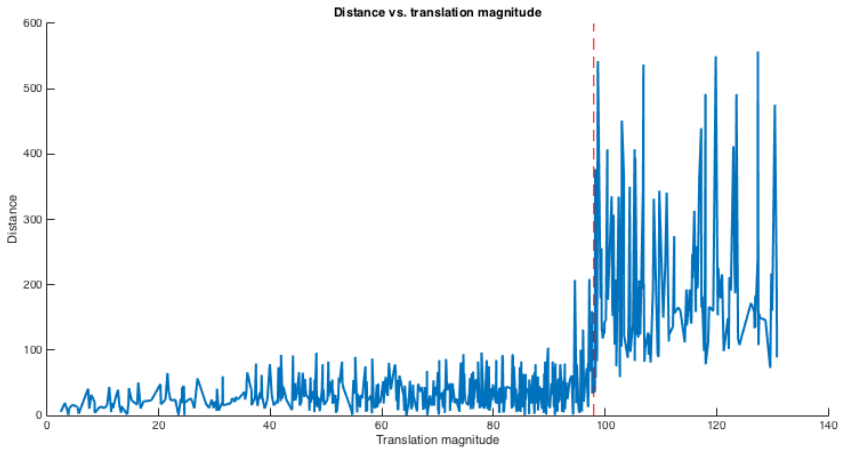
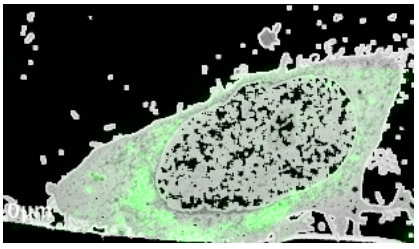
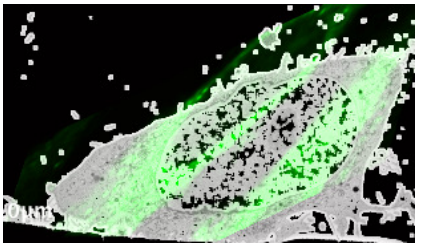


Fig. A.2: *Average distance between translated 3D LM patches after registration and 3D EM-ROI.*

Fig. A.3. The registration of the EM-ROI and an altered LM patch, translated by 37.9 pixels in the XY plane is displayed in Fig. A.3a, where the green fluorescent structures are correctly overlaid on the border of the EM structure. Additionally, the registration of the EM-ROI and a LM patch translated by 97,8 pixels is shown in Fig. A.3b, where the green LM fluorescent structures and the white EM structures are clearly misregistered.



(a) Overlay of altered LM patch and EM-ROI on a given slice. XY translation of magnitude of 37.9 pixels.



(b) Overlay of altered LM patch and EM-ROI on a given slice. XY translation of magnitude of 97.8 pixels.

Fig. A.3: *Overlay after registration of EM-ROI and altered LM patch. The registration can be assessed by observing the green fluorescent structures overlaid on the border of the EM structure.*

In this experiment, it is observed that the value of the magnitude of the XY translation after which the distance between the vertices significantly increases and the 3D registration is not correctly estimated is $\approx$ 98 pixels. This magnitude is less than 50% of the magnitude of the 2D shift obtained in the pre-alignment stage. Therefore, from this experiment it can be concluded that the 2D pre-registration step is really required.

It is a crucial part in the overall 3D CLEM registration framework. This experiment also demonstrates that the pre-alignment is not required to be very accurate, since the 3D pose can be retrieved even for not that small deviation from the estimated registration. In future developments on the 3D CLEM registration problem, a full 3D pre-registration stage should be explored to compensate the initial misalignment of 3D correlative microscopy stacks.

LIST OF PUBLICATIONS

- B.M. Toledo Acosta, X. Heiligenstein, G. Malandain, P. Bouthemy. Intensity-based matching and registration for 3D correlative microscopy with large discrepancies. In IEEE International Symposium on Biomedical Imaging (ISBI'18), Washington, April 2018.
- B.M. Toledo Acosta, A. Basset, P. Bouthemy, C. Kervrann. Multi-scale spot segmentation with selection of image scales. In IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'17), New Orleans, March 2017.
- B.M. Toledo Acosta, P. Bouthemy, C. Kervrann. A common image representation and a patch-based search for correlative light-electron-microscopy (CLEM) registration. In IEEE International Symposium on Biomedical Imaging (ISBI'16), Prague, April 2016.
- P. Bouthemy, B.M. Toledo Acosta, B. Delyon. Robust selection of parametric motion models in image sequences. In IEEE International Conference on Image Processing (ICIP'16), Phoenix, September 2016.

BIBLIOGRAPHY

[1] C. Agostinelli. Robust model selection in regression via weighted likelihood methodology. *Statistics and Probability Letters*, 56(3):289 – 300, 2002.

[2] A. V. Agronskaia, J. A. Valentijn, L. F. van Driel, C. T. Schneijdenberg, B. M. Humbel, P. M. van Bergen en Henegouwen, A. J. Verkleij, A. J. Koster, and H. C. Gerritsen. Integrated fluorescence and transmission electron microscopy. *Journal of Structural Biology*, 164(2):183–189, Nov. 2008.

[3] H. Akaike. A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6):716–723, Dec. 1974.

[4] J. R. Anderson, B. W. Jones, J. H. Yang, M. V. Shaw, C. B. Watt, P. Koshevoy, J. Spaltenstein, E. Jurrus, K. UV, R. T. Whitaker, D. Mastronarde, T. Tasdizen, and R. E. Marc. A computational framework for ultrastructural mapping of neural circuitry. *PLOS Biology*, 7(3), Mar. 2009.

[5] S. Ayer and H. S. Sawhney. Layered representation of motion video using robust maximum-likelihood estimation of mixture models and MDL encoding. In *Proceedings of IEEE International Conference on Computer Vision*, pages 777–784, June 1995.

[6] P. Bajcsy, A. Cardone, J. Chalfoun, M. Halter, D. Juba, M. Kociolek, M. Majurski, A. Peskin, C. Simon, M. Simon, A. Vandecreme, and M. Brady. Survey statistics of automated segmentations applied to optical imaging of mammalian cells. *BMC Bioinformatics*, 16:330, Oct. 2015.

[7] A. Basset, J. Boulanger, J. Salamero, P. Bouthemy, and C. Kervrann. Adaptive Spot Detection With Optimal Scale Selection in Fluorescence Microscopy Images. *IEEE Transactions on Image Processing*, 24(11):4512–4527, Nov. 2015.

[8] A. Basset, P. Bouthemy, and C. Kervrann. Recovery of motion patterns and dominant paths in videos of crowded scenes. In *IEEE International Conference on Image Processing*, pages 184–188, Oct. 2014.

[9] M. Basu. Gaussian-based edge-detection methods-a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 32(3):252–260, 2002.

[10] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool. Speeded-Up Robust Features (SURF). *Computer Vision and Image Understanding*, 110(3):346–359, 2008.

[11] M. S. Beckwith, K. S. Beckwith, P. Sikorski, N. T. Skogaker, T. H. Flo, and Ø. Halaas. Seeing a Mycobacterium-Infected Cell in Nanoscale 3d: Correlative Imaging by Light Microscopy and FIB/SEM Tomography. *PLOS ONE*, 10(9):1–19, 2015.

[12] P. J. Besl and N. D. McKay. A method for registration of 3-D shapes. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(2):239–256, Feb. 1992.

[13] E. Betzig, G. H. Patterson, R. Sougrat, O. W. Lindwasser, S. Olenych, J. S. Bonifacino, M. W. Davidson, J. Lippincott-Schwartz, and H. F. Hess. Imaging Intracellular Fluorescent Proteins at Nanometer Resolution. *Science*, 313(5793):1642–1645, 2006.

[14] M. J. 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.

[15] D. Bohorquez, F. Haque, S. Medicetty, and R. A. Liddle. Correlative Confocal and 3D Electron Microscopy of a Specific Sensory Cell. *Journal of Visualized Experiments*, (101):e52918, July 2015.

[16] P. Bouthemy, B. M. T. Acosta, and B. Delyon. Robust selection of parametric motion models in image sequences. In *IEEE International Conference on Image Processing*, pages 3743–3747, Sept. 2016.

[17] D. S. Bright and E. B. Steel. Two-dimensional top hat filter for extracting spots and spheres from digital images. *Journal of Microscopy*, 146(2):191–200, 1987.

[18] K. P. Burnham and D. R. Anderson. *Model selection and multimodel inference: a practical information-theoretic approach*. Springer, 2 edition, 2002.

[19] Y. S. Bykov, M. Cortese, J. A. G. Briggs, and R. Bartenschlager. Correlative light and electron microscopy methods for the study of virus–cell interactions. *FEBS Letters*, 590(13):1877–1895, 2016.

[20] T. Cao, C. Zach, S. Modla, D. Powell, K. Czymmek, and M. Niethammer. Multimodal registration for correlative microscopy using image analogies. *Medical Image Analysis*, 18(6):914–926, 2014.

[21] A. Cardona, S. Saalfeld, S. Preibisch, B. Schmid, A. Cheng, J. Pulokas, P. Tomancak, and V. Hartenstein. An integrated micro- and macroarchitectural analysis of the Drosophila brain by computer-assisted serial section electron microscopy. *PLOS Biology*, 8(10), 2010.

[22] S.-H. Cha. Comprehensive Survey on Distance/Similarity Measures between Probability Density Functions, 2007.

[23] N. Chacko, K. G. Chan, and M. Liebling. Intensity-based point-spread-function-aware registration for multi-view applications in optical microscopy. In *IEEE International Symposium on Biomedical Imaging*, pages 306–309, Apr. 2015.

[24] W. Cheung and G. Hamarneh. N-SIFT: N-DIMENSIONAL SCALE INVARIANT FEATURE TRANSFORM FOR MATCHING MEDICAL IMAGES. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 720–723, Apr. 2007.

[25] A. Collignon, F. Maes, D. Delaere, D. Vandermeulen, P. Suetens, and G. Marchal. Automated multi-modality image registration based on information theory. *In: Bizais*, 1995.

[26] K. Cortese, A. Diaspro, and C. Tacchetti. Advanced correlative light/electron microscopy: current methods and new developments using Tokuyasu cryosections. *Journal of Histochemistry and Cytochemistry*, 57(12):1103–1112, Dec. 2009.

[27] K. Cortese, G. Vicidomini, M. C. Gagliani, P. Boccacci, A. Diaspro, and C. Tacchetti. Chapter 6 - 3d HDO-CLEM: Cellular Compartment Analysis by Correlative Light-Electron Microscopy on Cryosection. In T. Müller-Reichert and P. Verkade, editors, *Correlative Light and Electron MIcroscopy*, volume 111 of *Methods in Cell Biology*, pages 95 – 115. Academic Press, 2012.

[28] D. Cremers and S. Soatto. Motion Competition: A Variational Approach to Piecewise Parametric Motion Segmentation. *International Journal of Computer Vision*, 62(3):249–265, May 2005.

[29] P. de Boer, J. P. Hoogenboom, and B. N. G. Giepmans. Correlated light and electron microscopy: ultrastructure lights up! *Nature Methods*, 12:503 EP –, May 2015. Review Article.

[30] A. Desolneux, L. Moisan, and J.-M. Morel. A Grouping Principle and Four Applications. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(4):508–513, Apr. 2003.

[31] R. M. Dickson, A. B. Cubitt, R. Y. Tsien, and W. Moerner. On/off blinking and switching behaviour of single molecules of green fluorescent protein. *Nature*, 388(6640):355, 1997.

[32] J. Dubochet, M. Adrian, J.-J. Chang, J.-C. Homo, J. Lepault, A. W. McDowall, and P. Schultz. Cryo-electron microscopy of vitrified specimens. *Quarterly reviews of biophysics*, 21(2):129–228, 1988.

[33] P. A. V. d. Elsen, J. B. A. Maintz, E. J. D. Pol, and M. A. Viergever. Automatic registration of CT and MR brain images using correlation of geometrical features. *IEEE Transactions on Medical Imaging*, 14(2):384–396, June 1995.

[34] M. A. Fischler and R. C. Bolles. Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography. *Communications of the ACM*, 24(6):381–395, June 1981.

[35] D. Fortun, P. Bouthemy, and C. Kervrann. Optical flow modeling and computation: A survey. *Computer Vision and Image Understanding*, 134:1 – 21, 2015.

[36] E. François and P. Bouthemy. Derivation of qualitative information in motion analysis. *Image and Vision Computing*, 8(4):279 – 288, 1990.

[37] J. Frank. Averaging of low exposure electron micrographs of non-periodic objects. *Ultramicroscopy*, 1(2):159–162, 1975.

[38] J. Franks, C. T. Wallace, M. Shibata, M. Suga, N. Erdman, D. B. Stolz, and S. C. Watkins. *Correlative Fluorescence and Electron Microscopy in 3D-Scanning Electron Microscope Perspective*, volume 80, pages 1–12. John Wiley & Sons, Inc., Apr. 2017.

[39] Y. Fukuda, N. Schrod, M. Schaffer, L. R. Feng, W. Baumeister, and V. Lucic. Coordinate transformation based cryo-correlative methods for electron tomography and focused ion beam milling. *Ultramicroscopy*, 143:15 – 23, 2014.

[40] N. Gheissari and A. Bab-Hadiashar. Motion analysis: model selection and motion segmentation. In *12th International Conference on Image Analysis and Processing, 2003.Proceedings.*, pages 442–447, Sept. 2003.

[41] N. Gheissari and A. Bab-Hadiashar. A Comparative Study of Model Selection Criteria for Computer Vision Applications. *Image and Vision Computing*, 26(12):1636–1649, Dec. 2008.

[42] A. Gil, O. M. Mozos, M. Ballesta, and O. Reinoso. A comparative evaluation of interest point detectors and local descriptors for visual SLAM. *Machine Vision and Applications*, 21(6):905–920, Oct. 2010.

[43] J. I. Goldstein, D. E. Newbury, J. R. Michael, N. W. Ritchie, J. H. J. Scott, and D. C. Joy. *Scanning electron microscopy and X-ray microanalysis*. Springer, 2017.

[44] B. Grosjean and L. Moisan. A-contrario Detectability of Spots in Textured Backgrounds. *Journal of Mathematical Imaging and Vision*, 33(3):313, Sept. 2008.

[45] L. Gueguen and M. Pesaresi. Multi scale Harris corner detector based on Differential Morphological Decomposition. *Pattern Recognition Letters*, 32(14):1714 – 1719, 2011.

[46] S. Halary, S. Duperron, and T. Boudier. Direct Image-Based Correlative Microscopy Technique for Coupling Identification and Structural Investigation of Bacterial Symbionts Associated with Metazoans. *Applied and Environmental Microbiology*, 77(12):4172–4179, June 2011.

[47] S. Handschuh, N. Baeumler, T. Schwaha, and B. Ruthensteiner. A correlative approach for combining microCT, light and transmission electron microscopy in a single 3d scenario. *Frontiers in Zoology*, 10(1):44, 2013.

[48] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of Fourth Alvey Vision Conference*, pages 147–151, 1988.

[49] Y. He, A. B. Hamza, and H. Krim. A generalized divergence measure for robust image registration. *IEEE Transactions on Signal Processing*, 51(5):1211–1220, May 2003.

[50] X. Heiligenstein, J. Heiligenstein, C. Delevoye, I. Hurbain, S. Bardin, P. Paul-Gilloteaux, L. Sengmanivong, G. Régnier, J. Salamero, C. Antony, and G. Raposo. The CryoCapsule: Simplifying Correlative Light to Electron Microscopy. *Traffic*, 15(6):700–716, 2014.

[51] X. Heiligenstein, P. Paul-Gilloteaux, M. Belle, G. Raposo, and J. Salamero. eC-CLEM: Flexible Multidimensional Registration Software for Correlative Microscopies with Refined Accuracy Mapping. *Microscopy and Microanalysis*, 23(S1):360–361, 2017.

[52] S. W. Hell and J. Wichmann. Breaking the diffraction resolution limit by stimulated emission: stimulated-emission-depletion fluorescence microscopy. *Optics letters*, 19(11):780–782, 1994.

[53] R. Henderson, J. M. Baldwin, T. Ceska, F. Zemlin, E. a. Beckmann, and K. Downing. Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *Journal of molecular biology*, 213(4):899–929, 1990.

[54] D. L. Hill, P. G. Batchelor, M. Holden, and D. J. Hawkes. Medical image registration. *Physics in Medicine and Biology*, 46(3):1–45, Mar. 2001.

[55] L. Hodgson, D. Nam, J. Mantell, A. Achim, and P. Verkade. Retracing in correlative light electron microscopy: where is my object of interest? *Methods in Cell Biology*, 124:1–21, 2014.

[56] L. Hogrebe, A. R. Paiva, E. Jurrus, C. Christensen, M. Bridge, L. Dai, R. Pfeiffer, P. R. Hof, B. Roysam, J. R. Korenberg, and T. Tasdizen. Serial section registration of axonal confocal microscopy datasets for long-range neural circuit reconstruction. *Journal of Neuroscience Methods*, 207(2):200–210, June 2012.

[57] 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.

[58] B. Huang, H. Babcock, and X. Zhuang. Breaking the Diffraction Barrier: Super-Resolution Imaging of Cells. *Cell*, 143(7):1047 – 1058, 2010.

[59] B. Huang, M. Bates, and X. Zhuang. Super resolution fluorescence microscopy. *Annual Review of Biochemistry*, 78:993–1016, 2009.

[60] P. J. Huber. *Robust Statistics*. Wiley, 1981.

[61] T. Jabid, M. H. Kabir, and O. Chae. Local Directional Pattern (LDP) for face recognition. In *Digest of Technical Papers International Conference on Consumer Electronics*, pages 329–330, Jan. 2010.

[62] M. Jain, J. C. v. Gemert, H. Jégou, P. Bouthemy, and C. G. M. Snoek. Tubelets: Unsupervised action proposals from spatiotemporal super-voxels. *CoRR*, abs/1607.02003, 2016.

[63] A. Jaiswal, W. J. Godinez, R. Eils, M. J. Lehmann, and K. Rohr. Tracking Virus Particles in Fluorescence Microscopy Images Using Multi-Scale Detection and Multi-Frame Association. *IEEE Transactions on Image Processing*, 24(11):4122–4136, Nov. 2015.

[64] N. Jiménez, E. Van Donselaar, D. De Winter, K. Vocking, A. Verkleij, and J. Post. Gridded Aclar: preparation methods and use for correlative light and electron microscopy of cell monolayers, by TEM and FIB–SEM. *Journal of Microscopy*, 237(2):208–220, 2010.

[65] A. C. Jirapatnakul, S. V. Fotin, A. P. Reeves, A. M. Biancardi, D. F. Yankelevitz, and C. I. Henschke. Automated nodule location and size estimation using a multi-scale

Laplacian of Gaussian filtering approach. In *International Conference of the IEEE Engineering in Medicine and Biology Society*, pages 1028–1031, Sept. 2009.

[66] H. J. Johnson, M. M. McCormick, and L. Ibanez. *Template:The ITK Software Guide Book 1: Introduction and Development Guidelines-Volume 1*. Kitware, Inc., 2015.

[67] K. Kanatani. Geometric Information Criterion for Model Selection. *International Journal of Computer Vision*, 26(3):171–189, Feb. 1998.

[68] K. Kanatani. Model selection for geometric inference. In *Proceedings of Asian Conference on Computer Vision*, 2002.

[69] M. A. Karreman, L. Mercier, N. L. Schieber, T. Shibue, Y. Schwab, and J. G. Goetz. Correlating Intravital Multi-Photon Microscopy to 3d Electron Microscopy of Invading Tumor Cells Using Anatomical Reference Points. *PLOS ONE*, 9(12):1–23, 2014.

[70] M. A. Karreman, B. Ruthensteiner, L. Mercier, N. L. Schieber, G. Solecki, F. Winkler, J. G. Goetz, and Y. Schwab. Chapter 13 - Find your way with X-Ray: Using microCT to correlate in vivo imaging with 3d electron microscopy. In T. Müller-Reichert and P. Verkade, editors, *Correlative Light and Electron Microscopy III*, volume 140 of *Methods in Cell Biology*, pages 277 – 301. Academic Press, 2017.

[71] K. Kobayashi, D. Cheng, M. Huynh, K. R. Ratinac, P. Thordarson, and F. Braet. Imaging fluorescently labeled complexes by means of multidimensional correlative light and transmission electron microscopy: practical considerations. *Methods in Cell Biology*, 111:1–20, 2012.

[72] I. Kolotuev, Y. Schwab, and M. Labouesse. A precise and rapid mapping protocol for correlative light and electron microscopy of small invertebrate organisms. *Biol. Cell*, 102(2):121–132, Feb. 2010.

[73] H. Kong, H. C. Akakin, and S. E. Sarma. A Generalized Laplacian of Gaussian Filter for Blob Detection and Its Applications. *IEEE Transactions on Cybernetics*, 43(6):1719–1733, Dec. 2013.

[74] S. Konishi and G. Kitagawa. Information Criteria and Statistical Modeling. *Biometrics*, 64(2):661–661, 2008.

[75] B. G. Kopek, M. G. Paez-Segala, G. Shtengel, K. A. Sochacki, M. G. Sun, Y. Wang, C. S. Xu, S. B. van Engelenburg, J. W. Taraska, L. L. Looger, and H. F. Hess. Diverse protocols for correlative super-resolution fluorescence imaging and electron microscopy of chemically fixed samples. *Nature Protocols*, 12:916 EP –, Apr. 2017.

[76] W. Kukulski, M. Schorb, S. Welsch, A. Picco, M. Kaksonen, and J. A. Briggs. Precise, correlated fluorescence microscopy and electron tomography of lowicryl sections using fluorescent fiducial markers. *Methods in Cell Biology*, 111:235–257, 2012.

[77] W. Kukulski, M. Schorb, S. Welsch, A. Picco, M. Kaksonen, and J. A. G. Briggs. Correlated fluorescence and 3D electron microscopy with high sensitivity and spatial precision. *The Journal of Cell Biology*, 192(1):111–119, 2011.

[78] Y. Leng. Light Microscopy. In *Materials Characterization*, pages 1–44. John Wiley and Sons (Asia) Pte Ltd, 2008.

[79] X. Lin, S. Ruan, T. Qiu, and D. Guo. Nonrigid medical image registration based on mesh deformation constraints. *Computational and Mathematical Methods in Medicine*, 2013:373082, 2013.

[80] T. Lindeberg. Scale-Space for Discrete Signals. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12(3):234–254, Mar. 1990.

[81] T. Lindeberg. *Scale-Space Theory in Computer Vision*. Kluwer Academic Publishers, Norwell, MA, USA, 1993.

[82] T. Lindeberg. Feature Detection with Automatic Scale Selection. *International Journal of Computer Vision*, 30(2):79–116, 1998.

[83] C. S. López, C. Bouchet-Marquis, C. P. Arthur, J. L. Riesterer, G. Heiss, G. Thibault, L. Pullan, S. Kwon, and J. W. Gray. Chapter 8 - A fully integrated, three-dimensional fluorescence to electron microscopy correlative workflow. In T. Müller-Reichert and P. Verkade, editors, *Correlative Light and Electron Microscopy III*, volume 140 of *Methods in Cell Biology*, pages 149 – 164. Academic Press, 2017.

[84] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2):91–110, 2004.

[85] M. S. Lucas, M. Gunthert, P. Gasser, F. Lucas, and R. Wepf. Bridging microscopes: 3d correlative light and scanning electron microscopy of complex biological structures. *Methods in Cell Biology*, 111:325–356, 2012.

[86] J. A. Machado. Robust Model Selection and M-Estimation. *Econometric Theory*, 9(3):478–493, 1993.

[87] B. Maco, M. Cantoni, A. Holtmaat, A. Kreshuk, F. A. Hamprecht, and G. Knott. Semiautomated correlative 3D electron microscopy of in vivo–imaged axons and dendrites. *Nature Protocols*, 9:1354–1366, 2014.

[88] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. Multimodality image registration by maximization of mutual information. *IEEE Transactions on Medical Imaging*, 16(2):187–198, Apr. 1997.

[89] J. B. A. Maintz, P. A. v. d. Elsen, and M. A. Viergever. Evaluation of ridge seeking operators for multimodality medical image matching. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(4):353–365, Apr. 1996.

[90] J. B. A. Maintz and M. A. Viergever. A survey of medical image registration. *Medical Image Analysis*, 2(1):1 – 36, 1998.

[91] C. L. Mallows. Some Comments on C_p. *Technometrics*, 15(4):661–675, 1973.

[92] S. Marsland, C. J. Twining, and C. J. Taylor. A Minimum Description Length Objective Function for Groupwise Non-rigid Image Registration. *Image and Vision Computing*, 26(3):333–346, Mar. 2008.

[93] J. Matas, O. Chum, M. Urban, and T. Pajdla. Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767, 2004.

[94] Y. Matsushita, E. Ofek, W. Ge, X. Tang, and H.-Y. Shum. Full-frame video stabilization with motion inpainting. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(7):1150–1163, July 2006.

[95] D. Mattes, D. R. Haynor, H. Vesselle, T. K. Lewellyn, and W. Eubank. Nonrigid multimodality image registration. In *Proceedings of Medical Imaging*, volume 4322, pages 4322 – 4322–12. SPIE, 2001.

[96] P. Meer. Robust techniques for computer vision. *Emerging topics in computer vision*, pages 107–190, 2004.

[97] C. Meisslitzer-Ruppitsch, C. Rohrl, J. Neumuller, M. Pavelka, and A. Ellinger. Photooxidation technology for correlated light and electron microscopy. *Journal of Microscopy*, 235(3):322–335, Sept. 2009.

[98] A. Merdes and J. De Mey. Kinetochore Microtubules, Analyzed by Correlated Light and Immunoelectron Microscopy. In *Microtubule Dynamics: Methods and Protocols*, pages 209–221. Humana Press, Totowa, NJ, 2011.

[99] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and Gool. A Comparison of Affine Region Detectors. *International Journal of Computer Vision*, 65(1-2):43–72, 2005.

[100] M. S. Miri, M. D. Abràmoff, Y. H. Kwon, and M. K. Garvin. Multimodal registration of SD-OCT volumes and fundus photographs using histograms of oriented gradients. *Biomedical Optics Express*, 7(12):5252–5267, Dec. 2016.

[101] S. Müller and A. H. Welsh. Outlier Robust Model Selection in Linear Regression. *Journal of the American Statistical Association*, 100(472):1297–1310, 2005.

[102] W. E. Moerner and L. Kador. Optical detection and spectroscopy of single molecules in a solid. *Physical review letters*, 62(21):2535, 1989.

[103] G. E. Murphy, K. Narayan, B. C. Lowekamp, L. M. Hartnell, J. A. W. Heymann, J. Fu, and S. Subramaniam. Correlative 3D imaging of whole mammalian cells with light and electron microscopy. *Journal of Structural Biology*, 176(3):268 – 278, 2011.

[104] D. Nam, J. Mantell, L. Hodgson, D. Bull, P. Verkade, and A. Achim. Feature-based registration for correlative light and electron microscopy images. In *IEEE International Conference on Image Processing*, pages 3567–3571, 2014.

[105] H. Nishiyama, M. Suga, T. Ogura, Y. Maruyama, M. Koizumi, K. Mio, S. Kitamura, and C. Sato. Atmospheric scanning electron microscope observes cells and tissues in open medium through silicon nitride film. *Journal of Structural Biology*, 169(3):438–449, Mar. 2010.

[106] J. M. Odobez and P. Bouthemy. Robust Multiresolution Estimation of Parametric Motion Models. *Journal of Visual Communication and Image Representation*, 6(4):348–365, 1995.

[107] F. P. Oliveira and J. M. Tavares. Medical image registration: a review. *Computer Methods in Biomechanics and Biomedical Engineering*, 17(2):73–93, 2014.

[108] J.-C. Olivo-Marin. Extraction of spots in biological images using multiscale products. *Pattern Recognition*, 35(9):1989 – 1996, 2002.

[109] B. A. Othman, C. Greenwood, A. F. Abuelela, A. A. Bharath, S. Chen, I. Theodorou, T. Douglas, M. Uchida, M. Ryan, J. S. Merzaban, and A. E. Porter. Correlative Light-Electron Microscopy Shows RGD-Targeted ZnO Nanoparticles Dissolve in the Intracellular Environment of Triple Negative Breast Cancer Cells and Cause Apoptosis with Intratumor Heterogeneity. *Advanced Healthcare Materials*, 5(11):1310–1325, 2016.

[110] K. Pan, A. Kokaram, J. Hillebrand, and M. Ramaswami. Gaussian mixture models for spots in microscopy using a new split/merge em algorithm. In *IEEE International Conference on Image Processing*, pages 3645–3648, Sept. 2010.

[111] C. J. Peddie, K. Blight, E. Wilson, C. Melia, J. Marrison, R. Carzaniga, M.-C. Domart, P. O'Toole, B. Larijani, and L. M. Collinson. Correlative and integrated light and electron microscopy of in-resin GFP fluorescence, used to localise diacylglycerol in mammalian cells. *Ultramicroscopy*, 143:3 – 14, 2014.

[112] J.-M. Pérez-Rúa, A. Basset, and P. Bouthemy. Detection and Localization of Anomalous Motion in Video Sequences from Local Histograms of Labeled Affine Flows. *Frontiers in ICT*, 4:10, 2017.

[113] A. Perfumo, A. Elsaesser, S. Littmann, R. A. Foster, M. M. Kuypers, C. S. Cockell, and G. Kminek. Epifluorescence, SEM, TEM and nanoSIMS image analysis of the cold phenotype of Clostridium psychrophilum at subzero temperatures. *FEMS Microbiology Ecology*, 90(3):869–882, 2014.

[114] M. Perkovic, M. Kunz, U. Endesfelder, S. Bunse, C. Wigge, Z. Yu, V.-V. Hodirnau, M. P. Scheffer, A. Seybert, S. Malkusch, E. M. Schuman, M. Heilemann, and A. S. Frangakis. Correlative Light- and Electron Microscopy with chemical tags. *Journal of Structural Biology*, 186(2):205–213, 2014.

[115] J. Piovano and T. Papadopoulo. Local Statistic Based Region Segmentation with Automatic Scale Selection. In *Proceedings of European Conference on Computer Vision*, pages 486–499. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[116] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Image registration by maximization of combined mutual information and gradient information. *IEEE Transactions on Medical Imaging*, 19(8):809–814, Aug. 2000.

[117] J. P. W. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, Aug. 2003.

[118] G. Qian and H. R. Künsch. On model selection via stochastic complexity in robust linear regression. *Journal of Statistical Planning and Inference*, 75(1):91 – 116, 1998.

[119] S. Ram, J. J. Rodríguez, and G. Bosco. Segmentation and detection of fluorescent 3d spots. *Cytometry Part A*, 81A(3):198–212, 2012.

[120] A. Rényi. On Measures of Entropy and Information. In *Proceedings of Berkeley Symposium on Mathematical Statistics and Probability*, pages 547–561, Berkeley, Calif., 1961. University of California Press.

[121] S. H. Rezatofighi, R. Hartley, and W. E. Hughes. A new approach for spot detection in total internal reflection fluorescence microscopy. In *IEEE International Symposium on Biomedical Imaging*, pages 860–863, May 2012.

[122] J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465 – 471, 1978.

[123] A. Roche, X. Pennec, G. Malandain, and N. Ayache. Rigid registration of 3-D ultrasound with MR images: a new approach combining intensity and gradient information. *IEEE Transactions on Medical Imaging*, 20(10):1038–1049, Oct. 2001.

[124] M. Rohde. *microscopy*. Elsevier, 2011.

[125] E. Ronchetti. Robust model selection in regression. *Statistics & Probability Letters*, 3(1):21 – 23, 1985.

[126] E. Ronchetti and R. G. Staudte. A Robust Version of Mallows's C P. *Journal of the American Statistical Association*, 89(426):550–559, 1994.

[127] E. Rosten and T. Drummond. Fusing points and lines for high performance tracking. In *Proceedings of IEEE International Conference on Computer Vision*, volume 2, Oct. 2005.

[128] G. G. Roussas and G. G. Roussas. *A course in mathematical statistics*. Academic Press, New York; London, 1997.

[129] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. G. Hill, M. O. Leach, and D. J. Hawkes. Nonrigid registration using free-form deformations: application to breast MR images. *IEEE Transactions on Medical Imaging*, 18(8):712–721, Aug. 1999.

[130] D. B. Russakoff, C. Tomasi, T. Rohlfing, and C. R. Maurer. *Image Similarity Using Mutual Information of Regions*, pages 596–607. Springer Berlin Heidelberg, Berlin, Heidelberg, 2004.

[131] M. R. G. Russell, T. R. Lerner, J. J. Burden, D. O. Nkwe, A. Pelchen-Matthews, M.-C. Domart, J. Durgan, A. Weston, M. L. Jones, C. J. Peddie, R. Carzaniga, O. Florey, M. Marsh, M. G. Gutierrez, and L. M. Collinson. 3D correlative light and electron microscopy of cultured cells using serial blockface scanning electron microscopy. *Journal of Cell Science*, 130(1):278–291, 2017.

[132] P. Ruusuvuori, T. Äijö, S. Chowdhury, C. Garmendia-Torres, J. Selinummi, M. Birbaumer, A. M. Dudley, L. Pelkmans, and O. Yli-Harja. Evaluation of methods for detection of fluorescence labeled subcellular objects in microscope images. *BMC Bioinformatics*, 11(1):1–17, 2010.

[133] D. Sage, F. R. Neumann, F. Hediger, S. M. Gasser, and M. Unser. Automatic tracking of individual fluorescence particles: application to the study of chromosome dynamics. *IEEE Transactions on Image Processing*, 14(9):1372–1383, Sept. 2005.

[134] S. Saleh. Robust AIC with High Breakdown Scale Estimate. *Journal of Applied Mathematics*, 2014:286414:1–286414:7, 2014.

[135] H. Samet and M. Tamminen. Efficient component labeling of images of arbitrary dimension represented by linear bintrees. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 10(4):579–586, July 1988.

[136] M. J. Sanderson, I. Smith, I. Parker, and M. D. Bootman. Fluorescence Microscopy. *Cold Spring Harbor Protocols*, 2014(10):pdb.top071795–pdb.top071795, Oct. 2014.

[137] A. Sartori, R. Gatz, F. Beck, A. Rigort, W. Baumeister, and J. M. Plitzko. Correlative microscopy: bridging the gap between fluorescence light microscopy and cryo-electron tomography. *Journal of Structural Biology*, 160(2):135–145, Nov. 2007.

[138] L. Schaad, R. Hlushchuk, S. Barré, R. Gianni-Barrera, D. Haberthür, A. Banfi, and V. Djonov. Correlative Imaging of the Murine Hind Limb Vasculature and Muscle Tissue by MicroCT and Light Microscopy. *Scientific Reports*, 7:41842 EP –, Feb. 2017.

[139] K. Schindler, D. Suter, and H. Wang. A model-selection framework for multibody structure-and-motion of image sequences. *International Journal of Computer Vision*, 79(2):159–177, Aug. 2008.

[140] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of Interest Point Detectors. *International Journal of Computer Vision*, 37(2):151–172, June 2000.

[141] M. Schorb and J. A. Briggs. Correlated cryo-fluorescence and cryo-electron microscopy with high spatial precision and improved sensitivity. *Ultramicroscopy*, 143:24–32, 2014.

[142] G. Schwarz. Estimating the Dimension of a Model. *The Annals of Statistics*, 6(2):461–464, 1978.

[143] T. Senst, V. Eiselein, and T. Sikora. Robust Local Optical Flow for Feature Tracking. *IEEE Transactions on Circuits and Systems for Video Technology*, 22(9):1377–1387, Sept. 2012.

[144] D. Shen and C. Davatzikos. HAMMER: hierarchical attribute matching mechanism for elastic registration. *IEEE Transactions on Medical Imaging*, 21(11):1421–1439, Nov. 2002.

[145] K. A. Sjollema, U. Schnell, J. Kuipers, R. Kalicharan, and B. N. Giepmans. Correlated light microscopy and electron microscopy. *Methods in Cell Biology*, 111:157–173, 2012.

[146] I. Smal, M. Loog, W. Niessen, and E. Meijering. Quantitative Comparison of Spot Detection Methods in Fluorescence Microscopy. *IEEE Transactions on Medical Imaging*, 29(2):282–301, Feb. 2010.

[147] I. Smal, W. Niessen, and E. Meijering. A new detection scheme for multiple object tracking in fluorescence microscopy by joint probabilistic data association filtering. In *IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 264–267, May 2008.

[148] P. Soille. *Morphological Image Analysis: Principles and Applications*. Springer-Verlag New York, Inc., Secaucus, NJ, USA, 2 edition, 2003.

[149] G. E. Sosinsky, B. N. Giepmans, T. J. Deerinck, G. M. Gaietta, and M. H. Ellisman. Markers for correlated light and electron microscopy. *Methods Cell Biol.*, 79:575–591, 2007.

[150] A. Sotiras, C. Davatzikos, and N. Paragios. Deformable Medical Image Registration: A Survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, July 2013.

[151] C. Spiegelhalter, J. Laporte, and Y. Schwab. Correlative Light and Electron Microscopy: From Live Cell Dynamic to 3d Ultrastructure. In J. Kuo, editor, *Electron Microscopy*, volume 1117 of *Methods in Molecular Biology*, pages 485–501. Humana Press, 2014.

[152] C. Studholme, E. Novotny, I. G. Zubal, and J. S. Duncan. Estimating Tissue Deformation between Functional Images Induced by Intracranial Electrode Implantation Using Anatomical {MRI}. *NeuroImage*, 13(4):561–576, 2001.

[153] H. A. Sturges. The Choice of a Class Interval. *Journal of the American Statistical Association*, 21(153):65–66, 1926.

[154] R. Szeliski. Image Alignment and Stitching: A Tutorial. *Found. Trends. Comput. Graph. Vis.*, 2(1):1–104, Jan. 2006.

[155] D. Thomann, D. R. Rines, P. K. Sorger, and G. Danuser. Automatic fluorescent tag detection in 3d with super-resolution: application to the analysis of chromosome movement. *Journal of Microscopy*, 208(1):49–64, 2002.

[156] B. M. Toledo-Acosta, A. Basset, P. Bouthemy, and C. Kervrann. Multi-scale spot segmentation with selection of image scales. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 1912–1916, Mar. 2017.

[157] B. M. Toledo-Acosta, P. Bouthemy, and C. Kervrann. A common image representation and a patch-based search for correlative light-electron-microscopy (CLEM)

registration. In *2016 IEEE 13th International Symposium on Biomedical Imaging (ISBI)*, pages 257–260, Apr. 2016.

[158] B. M. Toledo-Acosta, X. Heiligenstein, G. Malandain, and P. Bouthemy. Intensity-based matching and registration for 3D correlative microscopy with large discrepancies. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI)*, Apr. 2018.

[159] P. H. S. Torr. An assessment of information criteria for motion model selection. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 47–52, June 1997.

[160] B. Triggs. Detecting Keypoints with Stable Position, Orientation and Scale under Illumination Changes. In T. Pajdla and J. Matas, editors, *European Conference on Computer Vision*, volume 3024 of *Lecture Notes in Computer Science (LNCS)*, pages 100–113, Prague, Austria, May 2004. Springer-Verlag.

[161] S. Tu and L. Xu. A Theoretical Investigation of Several Model Selection Criteria for Dimensionality Reduction. *Pattern Recognition Letters*, 33(9):1117–1126, July 2012.

[162] C. van Rijnsoever, V. Oorschot, and J. Klumperman. Correlative light-electron microscopy (CLEM) combining live-cell imaging and immunolabeling of ultrathin cryosections. *Nature Methods*, 5(11):973–980, Nov. 2008.

[163] T. Veit, F. Cao, and P. Bouthemy. An a contrario Decision Framework for Region-Based Motion Detection. *International Journal of Computer Vision*, 68(2):163–178, June 2006.

[164] G. Vicidomini, M. C. Gagliani, M. Canfora, K. Cortese, F. Frosi, C. Santangelo, P. P. Di Fiore, P. Boccacci, A. Diaspro, and C. Tacchetti. High data output and automated 3d correlative light-electron microscopy method. *Traffic*, 9(11):1828–1838, Nov. 2008.

[165] L. Vincent. Morphological grayscale reconstruction in image analysis: applications and efficient algorithms. *IEEE Transactions on Image Processing*, 2(2):176–201, Apr. 1993.

[166] P. Viola and W. M. Wells. Alignment by Maximization of Mutual Information. *International Journal in Computer Vision*, 24(2):137–154, Sept. 1997.

[167] S. Watanabe, A. Punge, G. Hollopeter, K. I. Willig, R. J. Hobson, M. W. Davis, S. W. Hell, and E. M. Jorgensen. Protein localization in electron micrographs using fluorescence nanoscopy. *Nature Methods*, 8(1):80–84, Jan. 2011.

[168] M. Weber, M. Mickoleit, and J. Huisken. Chapter 11 - light sheet microscopy. In J. C. Waters and T. Wittman, editors, *Quantitative Imaging in Cell Biology*, volume 123 of *Methods in Cell Biology*, pages 193 – 215. Academic Press, 2014.

[169] H. Wechsler, Z. Duric, F. Li, and V. Cherkassky. Motion estimation using statistical learning theory. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(4):466–478, Apr. 2004.

[170] W. Wein, S. Brunke, A. Khamene, M. R. Callstrom, and N. Navab. Automatic CT-ultrasound registration for diagnostic imaging and image-guided intervention. *Medical Image Analysis*, 12(5):577 – 585, 2008.

[171] William, P. Viola, H. Atsumi, S. Nakajima, and R. Kikinis. Multi-modal volume registration by maximization of mutual information. *Medical Image Analysis*, 1(1):35–51, 1996.

[172] D. B. Williams and C. B. Carter. The Transmission Electron Microscope. In *Transmission Electron Microscopy: A Textbook for Materials Science*, pages 3–17. Springer US, Boston, MA, 1996.

[173] T. J. Woehl, S. Kashyap, E. Firlar, T. Perez-Gonzalez, D. Faivre, D. Trubitsyn, D. A. Bazylinski, and T. Prozorov. Correlative Electron and Fluorescence Microscopy of Magnetotactic Bacteria in Liquid: Toward In Vivo Imaging. *Microscopy and Microanalysis*, 21(S3):1499–1500, 2015.

[174] J. Yang and H. Li. Dense, accurate optical flow estimation with piecewise parametric model. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, pages 1019–1027, June 2015.

[175] J. Yang, J. P. Williams, Y. Sun, R. S. Blum, and C. Xu. A robust hybrid method for nonrigid image registration. *Pattern Recognition*, 44(4):764 – 776, 2011.

[176] A. Yilmaz, O. Javed, and M. Shah. Object Tracking: A Survey. *ACM Comput. Surv.*, 38(4), Dec. 2006.

[177] B. Zhang, M. J. Fadili, J. L. Starck, and J. C. Olivo-Marin. Multiscale Variance-Stabilizing Transform for Mixed-Poisson-Gaussian Processes and its Applications in Bioimaging. In *IEEE International Conference on Image Processing*, volume 6, Sept. 2007.

[178] B. Zitová and J. Flusser. Image registration methods: a survey. *Image and Vision Computing*, 21(11):977–1000, 2003.

Résumé

Cette thèse porte sur la définition d'un schéma de recalage automatique en microscopie corrélative 2D et 3D, en particulier pour des images de microscopie optique et électronique (CLEM). Au cours des dernières années, la CLEM est devenue un outil d'investigation important et puissant dans le domaine de la bio-imagerie. En utilisant la CLEM, des informations complémentaires peuvent être collectées à partir d'un échantillon biologique. La superposition des différentes images microscopiques est généralement réalisée à l'aide de techniques impliquant une assistance manuelle à plusieurs étapes, ce qui est exigeant et prend beaucoup de temps pour les biologistes. Pour faciliter et diffuser le procédé de CLEM, notre travail de thèse est axé sur la création de méthodes de recalage automatique qui soient fiables, faciles à utiliser et qui ne nécessitent pas d'ajustement de paramètres ou de connaissances complexes. Le recalage CLEM doit faire face à de nombreux problèmes dus aux différences entre les images de microscopie électronique et optique et leur mode d'acquisition, tant en termes de résolution du pixel, de taille des images, de contenu, de champ de vision et d'apparence. Nous avons conçu des méthodes basées sur l'intensité des images pour aligner les images CLEM en 2D et 3D. Elles comprennent plusieurs étapes : représentation commune des images LM et EM à l'aide de la transformation LoG, pré-alignement exploitant des mesures de similarité à partir d'histogrammes avec une recherche exhaustive, et un recalage fin basé sur l'information mutuelle. De plus, nous avons défini une méthode de sélection robuste de modèles de mouvement, et un méthode de détection multi-échelle de spots, que nous avons exploitées dans le recalage CLEM 2D. Notre schéma de recalage automatisé pour la CLEM a été testé avec succès sur plusieurs ensembles de données CLEM réelles 2D et 3D. Les résultats ont été validés par des biologistes, offrant une excellente perspective sur l'utilité de nos développements.

Mots clés - Recalage multimodal d'images biologiques, microscopie corrélative, microscopie optique, microscopie électronique.

Abstract

This thesis is concerned with the definition of an automated registration framework for 2D and 3D correlative microscopy images, in particular for correlative light and electron microscopy (CLEM) images. In recent years, CLEM has become an important and powerful tool in the bioimaging field. By using CLEM, complementary information can be collected from a biological sample. An overlay of the different microscopy images is commonly achieved using techniques involving manual assistance at several steps, which is demanding and time consuming for biologists. To facilitate and disseminate the CLEM process for biologists, the thesis work is focused on creating automatic registration methods that are reliable, easy to use and do not require parameter tuning or complex knowledge. CLEM registration has to deal with many issues due to the differences between electron microscopy and light microscopy images and their acquisition, both in terms of pixel resolution, image size, content, field of view and appearance. We have designed intensity-based methods to align CLEM images in 2D and 3D. They involved a common representation of the LM and EM images using the LoG transform, a pre-alignment step exploiting histogram-based similarities within an exhaustive search, and a fine mutual information-based registration. In addition, we have defined a robust motion model selection method, and a multiscale spot detection method which were exploited in the 2D CLEM registration. Our automated CLEM registration framework was successfully tested on several real 2D and 3D CLEM datasets and the results were validated by biologists, offering an excellent perspective in the usefulness of our methods.

Keywords - Multimodal biological image registration, correlative microscopy, light microscopy, electron microscopy.