Image Analysis and Registration Methods for Cargo and vehicles X-Ray Imaging

Abraham Marciano

THÈSE DE DOCTORAT

de l'Université de recherche Paris Sciences et Lettres
PSL Research University

Préparée à Université Paris-Dauphine

Méthodes d'Analyse et de Recalage d'Images Radiographiques de Fret et de Véhicules

École doctorale n°ED 543

ECOLE DOCTORALE DE DAUPHINE

Spécialité SCIENCES

COMPOSITION DU JURY :

M. Laurent D. COHEN
CEREMADE, Directeur de Thèse

Mme Isabelle BLOCH
Télécom ParisTech, Rapporteur

M. Laurent YOUNES
John Hopkins University, Rapporteur

M. Jamal ATIF
LAMSADE, Président du jury

M. Najib GADI
Smiths Detection, Membre du jury

M. Ger KOOMEN
Dutch Customs Laboratory, Membre du jury

Soutenue par **Abraham MARCIANO**
le **3 Juillet 2018**

Dirigée par **Laurent D. COHEN**

DAUPHINE UNIVERSITÉ PARIS | PSL RESEARCH UNIVERSITY PARIS

PSL RESEARCH UNIVERSITY PARIS

PSL Research University

Doctoral School **École Doctorale de l'Université Paris Dauphine**

University Department **Centre De Recherche en Mathématiques de la Décision, Université Paris-Dauphine, CNRS, UMR 7534**

Thesis defended by **Abraham Marciano**

Defended on **3rd July, 2018**

In order to become Doctor from PSL Research University

Academic Field **Sciences**

Speciality **Mathematics**

Image Analysis and Registration Methods for Cargo and Vehicles X-Ray Imaging

Thesis supervised by Laurent D. COHEN

Committee members

Referees	Isabelle BLOCH	Professor at Télécom ParisTech
	Laurent YOUNES	Professor at John Hopkins University
Examiners	Jamal ATIF	Professor at LAMSADE - Université Paris Dauphine PSL
	Najib GADI	Senior Researcher at Smiths Detection
	Ger KOOMEN	Senior Researcher at Dutch Customs Laboratory
Supervisor	Laurent D. COHEN	Professor at CEREMADE - Université Paris Dauphine PSL

This thesis has been prepared at

Centre De Recherche en Mathématiques de la Décision, Université Paris-Dauphine, CNRS, UMR 7534

Place du Maréchal De Lattre De Tassigny
75016 Paris
France

☎ +33 1 44 27 42 98
Web Site `http://ceremade.dauphine.fr`

Remerciements

Je tiens à remercier les Professeurs Isabelle Bloch et Laurent Younes pour avoir accepté le travail de rapporteurs du présent mémoire. Je remercie aussi le Dr Jamal Atif et Ger Koomen pour leur participation à mon jury de thèse.

Je souhaiterais tout particulièrement remercier mon directeur de thèse le Professeur Laurent D. Cohen pour ses encouragements, son aide et ses précieux conseils tout au long de mes travaux. Son approche m'a fortement marqué, notamment sa propension à traiter des problèmes complexes tout en favorisant la recherche de solutions simples. Je tiens aussi à exprimer ma gratitude envers mon responsable au sein de Smiths Detection, le Dr Najib Gadi qui m'a permis de faire évoluer ma recherche avec flexibilité et s'est toujours montré à l'écoute, conjuguant compétences techniques et humilité.

Mes remerciements vont aussi au Dr Irène Dorion ainsi qu'à toute l'équipe P&T de Smiths Detection: Oussama et Youssef qui m'ont considérablement apporté en machine learning, Pierre, Luis, Cindy, Serge, JOF, Bruno, David, Yann, Guillaume, Christophe, Estelle... Merci à Xavier pour ses remarques pertinentes, à Pavel pour nous avoir aidé à la collecte de données. Je n'oublierai jamais les acquisitions d'armes dans un froid sibérien aux côtés de Fred, avec qui j'ai passé d'excellents moments à manipuler mortiers, grenades et autres fusils d'assaut...

Je souhaiterais aussi témoigner ma reconnaissance à l'égard du CEREMADE, notamment à César Faivre, Isabelle Bellier et Marie Belle, pour m'avoir accompagné dans le cadre des différentes conférences et publications.

Il m'incombe évidemment de remercier Manu, mon ami de toujours, pour m'avoir aidé à clarifier un grand nombre de points techniques dans le cadre de mes travaux. Nos escapades de célibataires lors des conférences à Venise et en Normandie resteront mémorables! Bonne chance pour ta thèse!

Je dois exprimer toute ma gratitude à mes parents qui m'ont énormémént soutenu tout au long de mes études et m'ont toujours poussé à perséverer. Cette thèse constitue le fruit de mon travail mais aussi et avant tout de leur éducation. Un grand merci à mon frère et mes soeurs à qui je souhaite beaucoup de réussite dans la vie. J'embrasse très fort mes filles Ethel et Livna que j'adore.

Enfin, il m'est naturel d'exprimer ma reconnaissance envers mon épouse, Shirley, sans qui tout ce travail n'aurait jamais abouti. Merci pour tes encouragements perpétuels et pour avoir su gérer habilement mes absences répétées ainsi que les semaines parsemées de nuits blanches...

Image Analysis and Registration Methods for Cargo and Vehicles X-Ray Imaging

Abstract

Our societies, faced with an unprecedented level of security threat since the second world-war, must provide fast and adaptable solutions to cope with a new kind of menace. Illicit trade also, often correlated with criminal actions, is viewed as a defining stake by governments and agencies. Enforcement authorities are thus very demanding in terms of technological features, as they explicitly aim at automating inspection processes.

The main objective of our research is to develop computer-vision assisting tools designed for the detection of weapons and narcotics in cargo X-ray images. Image registration is the cornerstone of the proposed strategies throughout this thesis. Judging by recent applications in various domains (medical image analysis, remote sensing, surveillance, motion tracking...) it is still actively explored. In the present work, we introduce new methods combining both advanced classification and registration techniques for irregularity detection in X-ray cargo screening images (differences detected at abnormal locations).

Rather than employing machine-learning recognition techniques, our methods prove to be very efficient while targeting a very diverse type of threats from which no specific features can be extracted. Moreover, the proposed approaches significantly enhance the detection capabilities for law-enforcement officers, particularly in dense regions where both humans or trained learning models would probably fail.

In the first part, deep-learning based classification methods (pre-trained Convolutional Neural Networks) are employed to label a sample under inspection $\mathbf{R}$. Via an original adaptation of Active Shape Models, we are able to retrieve the nearest template $\mathbf{T}$ for registration. The resulting pair of images is then partitioned in sub-images and a piece-wise feature-based linear registration is performed. The resulting field of displacement is regularized via a diffusion process applied at sub-images borders. We show that for X-ray images, differences between $\mathbf{R}$ and $\mathbf{T}$ must be computed by a rationing approach, in compliance with Beer Lambert's attenuation law. By applying thresholding and morphological operations on the map of differences, irregularities are easily identifiable. Our method is described through a specific study case; the inspection of refrigeration units inside containers. Though, we demonstrate that our algorithm can be easily adapted to a wider range of applications in X-ray cargo vision.

The second part of this thesis addresses the particular registration problem involving top-view radiographic images of same-model vehicles. Starting from the observation that linear and standard non-linear methods do not provide an accurate solution, we propose to formulate a unidimensional energy-minimization scheme along with a column-wise constancy constraint on the displacement field $\mathbf{u}$.

We demonstrate that the suggested method does not account properly for cases where significant sidewards translation occurs between the cars in $\mathbf{R}$ and $\mathbf{T}$. Indeed, the radiated beam is pyramidal, leading to a stereoscopic effect between both images. We introduce a sigular way to address the registration issue by adding a 1-D unconstrained vertical correction to the previous scheme, now viewed as a rectification stage. Volume preserving issues are also studied to prevent the softening of the differences amplitudes.

Our work reviews state-of-the art methods in terms of classification and image registration. Various numerical solutions are also explored. The proposed algorithms are tested on a very large number of images, showing their necessity and performances both visually and numerically.

Keywords: image analysis, x-ray images, threat detection, image registration, classification, energy-minimization methods, variational methods

Centre De Recherche en Mathématiques de la Décision, Université Paris-Dauphine, CNRS, UMR 7534
Place du Maréchal De Lattre De Tassigny – 75016 Paris – France

Résumé

Les sociétés contemporaines font face à un niveau de menace sécuritaire sans précédent depuis la seconde guerre mondiale. La lutte contre le trafic illicite mobilise aussi l'ensemble des organes de police, visant à endiguer un mode de financement du crime organisé. Dans cet effort, les autorités s'engagent à employer des moyens de plus en plus modernes, cherchant notamment à automatiser les divers processus d'inspection.

L'objectif de cette étude est de développer des outils basés sur la vision par ordinateur afin d'assister les officiers des douanes dans la détection d'armes et de narcotiques. Les approches de recalage d'images constituent la pierre angulaire de la présente recherche. A en juger les récentes contributions applicatives (analyse d'images médicales, télédétection, surveillance satellite...) le recalage d'image fait encore l'objet d'investigations approfondies. Le travail présenté examine l'introduction de nouvelles approches combinant des techniques avancées de classification et de recalage pour l'identification d'irrégularités dans des images radiographiques de fret. Une anomalie est ici définie comme un objet détecté à un endroit censé demeurer vide.

Plutôt que de recourir à la reconnaissance par apprentissage, nos méthodes revêtent un intérêt particulier lorsque les objets ciblés présentent des caractéristiques visuelles variées. De plus, elles augmentent notablement la détectabilité d'objets cachés dans des zones denses où même des approches de reconnaissance automatique n'identifieraient pas d'irrégularité.

Une première approche, basée sur des méthodes de classification par apprentissage profond (réseaux de neurones à convolutions pré-entrainés), permet d'associer une image inspectée $\mathbf{R}$ à une catégorie correspondante. Via une utilisation originale de la méthode des modèles de formes actives, nous sommes en mesure de retrouver l'image de référence la plus proche $\mathbf{T}$ en vue du recalage. La paire de scans est alors divisée en sous-images et un recalage linéaire par parties, basé sur des attributs extraits, est réalisé. Le champ de déformation résultant subit aussi une régularisation par diffusion appliquée à la jonction des sous-images. Nous montrons que pour des images obtenues par rayons X, les différences entre $\mathbf{T}$ et $\mathbf{R}$ doivent être mesurées en calculant une division terme-à-terme, conformément à la loi d'atténuation de Beer Lambert. Via un seuillage appliqué à la carte des différences ainsi que des opérations morphologiques, les irrégularités peuvent facilement être soulignées. Notre méthode est illustrée au travers d'un cas d'étude précis : l'inspection automatique de modules frigorifiques de conteneurs. Toutefois, l'algorithme peut être aisément adapté à d'autres applications pour l'imagerie de fret par rayons X ou dans des domaines transverses.

La seconde partie de cette thèse traite du problème de recalage d'images radiographiques de véhicules de même modèle (vue de dessus). Après avoir constaté que les méthodes traditionnelles, linéaires et non-linéaires n'apportent pas de résultats convenables, nous proposons une nouvelle solution en formulant un schéma de minimisation d'énergie unidimensionnel jumelé à une contrainte de constance par colonne, appliquée au champ de déplacement $\mathbf{u}$.

Nous montrons ensuite que lors d'un décalage du véhicule entre les acquisitions $\mathbf{R}$ et $\mathbf{T}$, la méthode suggérée ne produit pas le résultat escompté en termes de détectabilité des différences. En réalité, le faisceau à rayons X est pyramidal, entrainant un effet stéréoscopique entre les deux images. De ce fait, nous présentons une nouvelle contribution en ajoutant une correction verticale 1-D supplémentaire au schéma de minimisation précédent, considéré alors comme étape de rectification. Les problématiques de préservation des volumes sont aussi étudiées afin de palier toute altération des amplitudes dans l'image des différences résultante.

Nos travaux détaillent l'état de l'art des méthodes de classification et de recalage. Plusieurs approches de résolution numérique sont aussi explorées. Les algorithmes sont testés sur d'importantes bases de données afin d'apprécier leurs performances d'un point de vue visuel et quantitatif.

Mots clés : analyse d'images, images radiographiques, détection de menaces, recalage d'images, classification, méthodes de minimisation d'énergie, méthodes variationnelles

Contents

Introduction (Français)

Sommaire du présent chapitre

Contexte Général

Le contexte géopolitique actuel érige la sécurité en tant qu'enjeu décisif pour les instances internationales. La mondialisation efface progressivement les frontières inter-étatiques, rendant *de facto* les contrôles plus délicats. Le déploiement de moyens humains et technologiques aux frontières terrestres, aériennes ou maritimes est donc inévitable pour une vérification approfondie des biens transportés ainsi que des personnes.

Différentes méthodes sont employées par les organes de police afin de contenir le crime organisé et les groupuscules terroristes. A moyen-terme, s'attaquer au trafic d'armes s'avère être une solution efficace, privant les criminels de toute capacité de nuire. À titre d'exemple, dans les Balkans, le Royaume-Uni ou l'Espagne, de récentes saisies ont permis de démanteler des réseaux très structurés de contrebande d'armes (voir Figure 1). Les arsenaux découverts contenaient non seulement des armes de poing et des fusils d'assaut mais aussi de véritables machines de guerre tels que des lance-roquettes sol-air menaçant l'aviation civile [33].

(a)

(b)

Figure 1 – (a) Saisie importante d'armes en Espagne par Europol en mars 2017 [81] ; (b) 1350 kg de résine de cannabis retrouvés dans les parois latérales d'un camion tracteur (Douanes marocaines, [128]).

Dans une optique similaire, le trafic de stupéfiants représente une importante source de profit pour les réseaux mafieux. Au-delà des problèmes de santé publique qu'il pose, l'impact sur le plan sécuritaire est notable dans la mesure où les mouvances terroristes y trouvent souvent un mode de financement. Alors que le trafic concernait traditionnellement la cocaïne, le cannabis ou les opiacés, les autorités douanières font état de nouvelles variétés synthétiques sur le marché. Pour se faire une idée générale de l'ampleur du commerce illicite de drogues, en 2016, plus d'1,5 millions de kilos de narcotiques ont été saisis dans le monde [128].

D'autres trafics, à l'image de la contrebande de cigarettes, d'oeuvres d'art ou historiques sont aussi assignés au coeur des activités douanières.

Des investissements considérables ont été dédiés à l'élaboration de technologies innovantes afin d'assister les officiers de l'armée, de la police ou des douanes. Au fil des années, une vaste gamme d'outils a émergé sur les sites sensibles (aéroports, points de contrôle routiers...) à l'image des détecteurs de trace d'explosifs et de narcotiques ou encore des scanners de bagages par rayons X.

Afin de lutter contre des trafics de plus grande échelle, des machines d'imagerie à rayons X ont été spécialement conçues pour l'inspection non invasive de fret. La société britannique Smiths Detection (anciennement Heimann Systems), au sein de laquelle cette thèse a conjointement été réalisée, fait partie des leaders du marché des systèmes à haute énergie. Heimann Systems fut dès 1970 l'un des pionniers de la commercialisation de systèmes d'inspection de bagages par rayons X. Les premières machines fixes dédiées au controle de fret ont été mise au point dans les années 90. Progressivement, l'émergence de concurrents internationaux à l'image de Rapiscan Systems (Etats-Unis) ou Nuctech (Chine) a accéléré la compétitivité technologique et contribué à diversifier les gammes de produits.

L'Imagerie à Rayons X de Fret

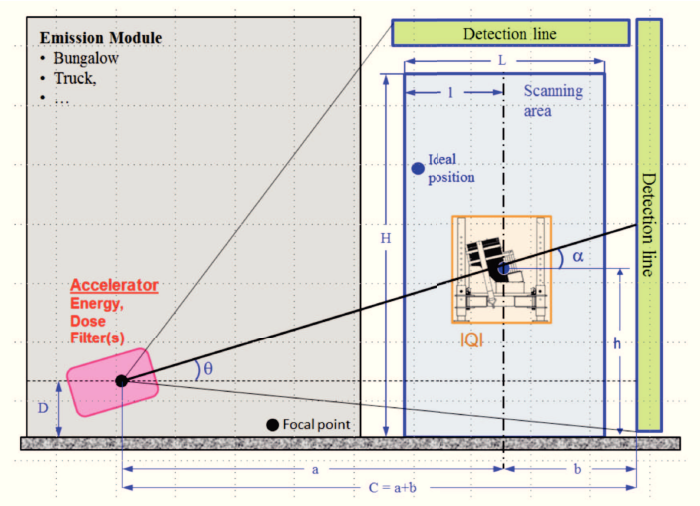


Figure 2 – Croquis du principe d'imagerie de fret par rayons X.

Sans pour autant examiner en profondeur la physique du fonctionnement des systèmes, nous allons expliquer les fondements de l'imagerie par rayons X pour le fret. Les photons de rayons X sont produits via les chocs d'électrons à haute énergie dans un accélérateur linéaire à particules. Le faisceau résultant est alors filtré pour ajuster sa distribution en énergie.

Lorsque le faisceau émis heurte alors un objet, une partie des photons incidents est absorbée. Sinon, des interactions avec l'environnement entrainent un phénomène de diffusion, souvent assimilé à un bruit gaussien. Après avoir traversé le fret, le faisceau résiduel est mesuré pas des capteurs, permettant de reconstruire une image de l'atténuation du signal en deux dimensions (Figure 2).

L'atténuation du faisceau à rayons X est modélisée par la loi de Beer Lambert. Pour un matériau ayant un coefficient d'absorption μ (en m^{-1}) avec une épaisseur traversée x (m), l'intensité I (nombre de photons transmis par unité de volume après franchissement du matériau) décroit exponentiellement par rapport à l'amplitude du faisceau incident I_0 :

$$I = I_0 \exp(-\mu x) \tag{1}$$

L'expression générale de la loi d'absorption constitue une propriété fondamentale, utilisée tout au long de notre travail de recherche. Comme détaillé sur la Figure 3, lorsque le faisceau traverse N matériaux différents, caractérisés par μ_i et x_i, la loi devient :

$$I = I_0 \exp(-\sum_{i=1}^{N} \mu_i x_i) \tag{2}$$

Le signal brut alors obtenu est ensuite traité via différentes transformations pour être stocké dans une image codée sur 16 bits.

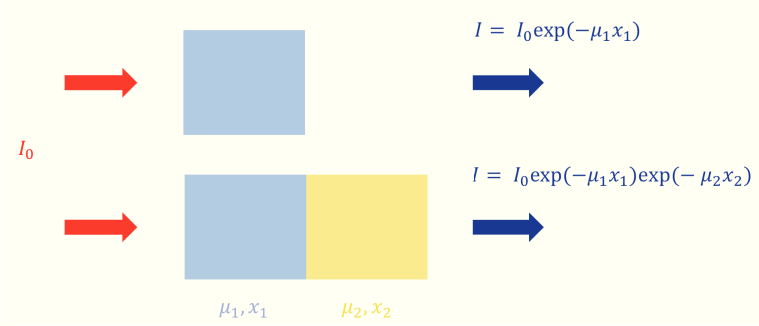


Figure 3 – Illustration de la loi d'atténuation de Beer-Lambert

Pour l'inspection de fret, des accélérateurs à très hautes énergies sont employés (rayonnement continu de fréinage) avec une énergie pouvant aller de 4 à plus de 10 MeV. En effet, les rayonnements à forte énergie (courte longueur d'onde) sont moins absorbés par la matière, et donc plus pénétrants. Les rayons X à haute énergie sont appelés "rayons durs" avec des longueurs d'ondes inférieures au nanomètre. Ces systèmes radiographiques permettent une pénétration de plus de 30 à 40 cm d'acier pour des véhicules se déplaçant jusqu'à une vitesse de 13 km/h. Il est à noter que certaines machines disposent d'une option "multi-énergie", avec un *low mode* et *high mode*, permettant la discrimination des matériaux à partir du numéro atomique Z calculé. Ainsi, les matériaux organiques se distinguent aisément des corps non-organiques.

Les Systèmes d'Imagerie de Fret

Figure 4 – Système d'imagerie HCVS

Le marché propose des types de systèmes d'imagerie à rayons X assez variés. Avec différentes géométries (longues/courtes), énergies, modes de scan et nombre de vues, les machines peuvent être déployées dans divers environnements et traiter des tâches d'inspection spécifiques. Certains modèles sont statiques alors que d'autres, transportés sur un camion, sont mobiles. Le but premier de la technologie est d'inspecter des conteneurs ou autres véhicules afin de vérifier les manifestes de douane, et d'identifier la présence de menaces (explosifs, armes, éléments radioactifs) ou de marchandise illicite (stupéfiants, cigarettes non déclarées, ...).

Les machines à haute énergie produites par Smiths Detection sont désignées par l'acronyme HCV pour Heimann Cargo Vision. Les principaux systèmes fixes sont :

- le système HCVS (*Stationary*), conçu pour des inspections de semi-remorques et conteneurs au niveau de ports ou de points de contrôle frontaliers. C'est un système avec une énergie

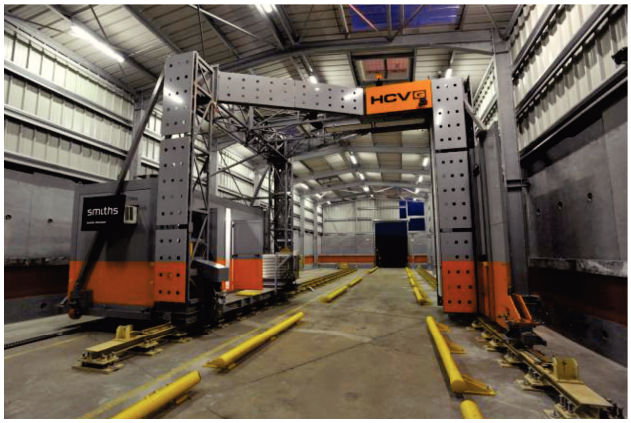


Figure 5 – Système d'imagerie HCVG

Figure 6 – Système d'imagerie HCVP

de 9 MeV permettant la discrimination des matériaux organiques avec une coloration basée sur le numéro atomique Z. Le HCVS donne une visualisation de côté et de dessus avec une très bonne résolution (Figure 4).

- Le HCVG (*Gantry*) est une machine d'inspection avec un portique mobile se translatant le long du véhicule statique. Une vue de côté est produite via un accélérateur délivrant des énergies de 4 à 6 MeV par une technologie dite *pulse-to-pulse* (Figure 5).
- Le HCVP (*Portal*) est un système à portique dit "drive-through", i.e. au travers duquel les véhicules pénètrent à vitesse réduite. Il a des propriétés assez similaires au HCVG (Figure 6).
- Le système HCVL (*Light* pour véhicules légers) sert à inspecter des véhicules légers comme des voitures, vans, minibus ou mobile-homes. Il renvoie des scans avec une vue de dessus. Le véhicule inspecté est tracté par un convoyeur afin d'être scanné à une vitesse constante. Le HCVL est généralement déployé au niveau de check-points frontaliers, à l'entrée de sites sensibles ou au sein de terminaux portuaires (Figure 7).

Deux types de systèmes mobiles sont actuellement mis sur le marché par Smiths Detection :

Figure 7 – Système d'imagerie HCVL

- le HCVM (*Mobile*, voir Figure 8) est disponible sous la forme de camion ou de remorque tractable. Son accélérateur a une énergie allant de 4 à 6 MeV, permettant la fonction de discrimination des matériaux (technologie *pulse-to-pulse*). Les capteurs sont placés au niveau d'un bras déployable qui permet deux modes de scan : un mode portique ("gantry") ou "pass through". L'image résultante est une vue de côté du chargement.
- HCVMe (*easy*) est l'équivalent du HCVM, plus compacte avec une énergie plus faible. Le système est représenté sur la Figure 9.

Figure 9 – Système d'imagerie HCVMe

Au cours de notre étude, nous avons été amenés à travailler sur des images d'HCVS, HCVM et HCVL pour l'élaboration d'algorithmes. En effet, la majeure partie des données (en milliers d'images) dont nous disposions proviennent de ces trois dispositifs de balayage. Néanmoins, le champ d'application de nos méthodes peut être facilement étendu à d'autres systèmes.

Le Rôle des Techniques de Traitement d'Image

Les techniques de traitement d'image et de vision par ordinateur sont omniprésentes dans l'industrie de la défense, contribuant à de multiples applications : du suivi par vidéo-surveillance en

(a) mode compacte pour transport

(b) bras déployé pour scanner

Figure 8 – Système d'imagerie HCVM

passant par les satellites espions aux missiles à têtes chercheuses.

Dans le domaine d'inspection de fret par rayons X, le traitement d'image occupe une place prépondérante dans la formation des images, le débruitage, l'amélioration visuelle ou encore les problématiques de segmentation (par exemple pour la discrimination de matériaux). Notre étude, quant à elle, traite davantage de l'exploitation des données de l'image résultante que de sa formation même.

Notre ère se caractérise par un afflux massif de données, du "Big Data" qui doit être traité dans des temps de plus en plus courts. Parallèlement, les pays souffrent de déficits publics engendrant des baisses d'effectif à tous les niveaux, notamment chez les officiers de douane (Figure 10). Pour ces diverses raisons, la transition rapide vers une automatisation des procédés d'inspection est décisive, a fortiori au regard de la crise sécuritaire actuelle.

L'industrie dans son ensemble a été grandement impactée par ce phénomène. Il lui incombe de résoudre de nouvelles problématiques, souvent au moyen d'approches ayant déjà fait leurs preuves dans des domaines transverses. De fait, les marchandises illicites et les armes peuvent être détectées de deux façons, à partir d'une image obtenue par rayons X :

- En adoptant des méthodes de reconnaissance par apprentissage (par exemple en entrainant un réseau de neurones à convolutions). Cela présuppose disposer d'un ensemble d'entrainement composé de menaces recherchées et d'images saines. Les objets à détecter doivent aussi présenter des caractéristiques visuelles communes à l'image telles que la texture.
- Une méthode alternative consiste à rechercher des irrégularités indifférenciées plutôt qu'une certaine catégorie d'objets. Par "irrégularité", nous définissons ici tout objet détecté dans

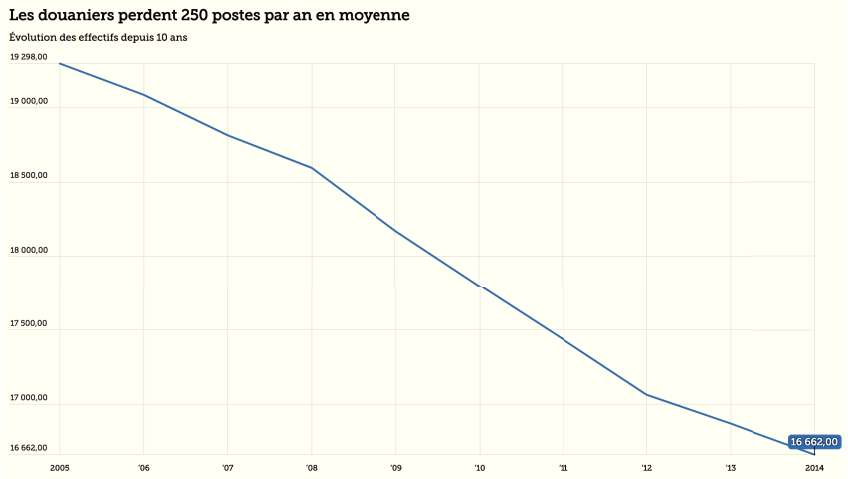


Figure 10 – Réduction des effectifs parmi les fonctionnaires des douanes en France au cours des dix dernières années ([112]).

une région censée rester vide. Dans le cadre de véhicules légers, cela s'applique notamment à la partie avant, où aucun chargement ne doit être décelé (hormis le moteur et autres composants mécaniques). L'idée est de comparer le scan à analyser (ou une région spécifique d'intérêt) avec une image de référence analogue en les recalant puis en calculant leurs différences. Comme il sera démontré dans cette thèse, les drogues ne possèdent pas d'attributs spécifiques pouvant permettre d'envisager une identification par apprentissage. L'une des façons d'aborder le problème serait d'analyser des cachettes traditionnelles via la comparaison d'images. Par ailleurs, de nombreux cas mentionnés au fil de la thèse montrent que les techniques proposées permettent la visualisation d'anomalies dans des zones très denses, où la vision humaine ou des techniques de reconnaissance par apprentissage échoueraient probablement.

Les concepts théoriques des méthodes de recalage d'images seront précisément étudiés dans cette thèse. De façon générale, le recalage sera défini comme la recherche de l'alignement optimal entre une image inspectée $\mathbf{R}$ et un modèle de référence $\mathbf{T}$, pourvu que la transformation appliquée soit plausible. La détection d'anomalies étant notre objectif premier, nous maintiendrons l'image à analyser $\mathbf{R}$ statique et le champ de déplacement sera appliqué à $\mathbf{T}$ (afin d'éviter toute déformation d'éventuelles menaces).

En raison de la confidentialité imposée au domaine ainsi que de l'émergence relativement récente des systèmes d'imagerie à haute énergie, les travaux de recherche concernant le sujet se font rares. Les articles de [74], traitant de l'inspection de scans de fret doivent être considérés, mais ils ne donnent toutefois qu'un aperçu très général de techniques assez connues. Par conséquent, la plupart des approches employées dans cette thèse sont inspirées de la recherche en imagerie médicale, où elles ont d'ores et déjà fait l'objet d'une validation rigoureuse.

Plan de la Thèse et Contributions

Cette thèse vise essentiellement à introduire des méthodes mathématiques pour résoudre des problématiques de recalage d'images appliquées au contexte de contrôle de fret par radiographie.

Étant donné un scan d'entrée **R** ou une région d'intérêt spécifique, l'idée est de combiner deux étapes successives :

1. Retrouver un modèle correspondant **T** en adaptant des techniques avancées de classification.
2. Trouver un schéma de recalage approprié et ajustable aux déformations observées ainsi qu'aux contraintes mécaniques imposées par les systèmes d'imagerie par rayons X. À partir de la paire d'images résultante, un alignement optimal peut être réalisé afin de calculer une image des différences et enfin détecter les principales irrégularités.

Notre travail se scinde en deux parties. Alors que la première introduit en détail l'ensemble des blocs de construction employés, la seconde vise à définir les algorithmes conçus, leurs implémentations et performances.

Partie 1

Le premier chapitre passe en revue la théorie des modèles de formes actives (ASM - *Active Shape Models*) [23] : l'annotation, l'entrainement ainsi que l'ajustement d'ensembles de points. Les ASM ont été introduits pour la segmentation d'objets dont la forme est apprise à partir d'une base d'images annotée manuellement, formant une vérité terrain. L'idée que nous proposons ici est de détourner l'emploi usuel des ASM par leur représentation réduite des formes, pour retrouver parmi notre jeu de données, l'image **T** la plus proche de **R** préalablement classifiée. La similarité est ici définie par une fidélité au niveau des formes les plus saillantes, présentes dans **R** et **T** (décrites par les formes annotées).

Ainsi, après annotation, chaque vecteur de forme défini par : $\mathbf{x} = (x_1, y_1, ..., x_n, y_n)^T$ est approché via la méthode des PCA (*Principal Component Analysis*) par :

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{P}\mathbf{b} \tag{3}$$

où les t vecteurs propres correspondant aux plus grandes valeurs propres λ_i de la matrice de covariance sont empilés dans $\mathbf{P}$. $\mathbf{b}$ est un vecteur de dimension t contenant les poids/paramètres du modèle. Par conséquent, le modèle peut être appliqué à chaque image d'entrainement d'une classe m donnée, pour obtenir q vecteurs $\mathbf{b}_i$. En ajustant les ASM à une image d'entrée $\mathbf{I}_{new}$, $\mathbf{b}_{new}$ est calculé et l'index de l'image la plus proche en termes de formes est donnée par :

$$i^* = \underset{i \in [\![1,q]\!]}{\arg\max} \|\mathbf{b}_i - \mathbf{b}_{new}\| \tag{4}$$

Par ailleurs, nous adaptons aussi les ASM de façon originale pour formuler un coût combinant à la fois des données de forme et d'aspect calculées après application d'un modèle sur **R**. Ce coût permet d'affiner les résultats de classification via une double vérification, souvent nécessaire lorsque peu de données d'entrainement sont disponibles.

Dans le second chapitre, les approches de représentation et de classification des données sont décrites. Les descripteurs denses ou parcimonieux tels que HoG, SIFT ou SURF sont présentés. Un aperçu global des techniques d'apprentissage en profondeur est fourni, avec un accent particulier sur les réseaux de neurones à convolutions (CNN - *Convolutional Neural Networks*). Les réseaux pré-entrainés utilisés pour l'extraction d'attributs sont longuement détaillés ; ceux-ci ayant un rôle de premier plan tout au long de notre recherche. La classification dite non-supervisée de type K-means est brièvement introduite, servant notamment à accélérer la construction de bases de données à partir d'images non labélisées. De même, les méthodes supervisées sont abordées,

en particulier via les machines à vecteur de supports (SVM), et le choix du schéma multiclasse *un-contre-un* est aussi justifié.

En ce qui concerne les méthodes de recalage, elles sont passées en revue dans le chapitre 3. Tout d'abord, des méthodes linéaires basées sur des attributs extraits sont décrites. La deuxième section du chapitre est consacrée à l'état de l'art des méthodes de recalage non linéaires. La formulation du problème de recalage non rigide dans le cadre variationnel constitue la pierre angulaire du corps principal de cette thèse. Notons le champ de déplacement $\mathbf{u} : \Omega \to \mathbb{R}^d$ (pour d dimensions), le problème d'optimisation est écrit comme suit :

$$\begin{aligned} &\text{trouver } \mathbf{u} \text{ minimisant} \\ &\mathcal{J}[\mathbf{u}] = \mathcal{D}[\mathbf{T}(\mathbf{u}), \mathbf{R}] + \alpha \mathcal{S}[\mathbf{u}] \end{aligned} \tag{5}$$

où $\mathcal{D}$ représente le terme d'attache aux données et $\mathcal{S}$ le composant de régularisation. Les approches paramétriques et non paramétriques sont décrites ainsi que différents termes de similarité et de lissage. Aussi, les problèmes de préservation du volume sont traités par l'application de contraintes spécifiques. L'aspect de la résolution numérique est également discuté au travers de méthodes de descente du premier ou second ordre.

Enfin, notre approche linéaire par morceaux est introduite, visant à considérer les déformations de manière plus locale, sans pour autant faire appel à des paradigmes non-linéaires complexes, plus coûteux en termes de calculs. Elle consiste à partitionner $\mathbf{R}$ et $\mathbf{T}$ sur la base des formes ASM ajustées à ces images. Les points-clés SURF sont extraits et mis en correspondance entre les s_d paires de sous-images $\{(\mathbf{T}_i, \mathbf{R}_i)\}_{1 \leqslant i \leqslant s_d}$. Les champs de déplacement denses $\mathbf{u}_i$ sont calculés pour être appliqués à chaque sous-image du template déformable $\mathbf{T}$. Les discontinuités aux limites des sous-images sont traitées en utilisant un schéma de régularisation par diffusion isotropique.

Le chapitre 4 décrit le processus de détection des différences à la suite du recalage de $\mathbf{R}$ et $\mathbf{T}$. Nous démontrons, à l'aide de la loi d'atténuation de Beer Lambert, que l'utilisation d'un ratio au lieu d'un simple schéma de différences améliore sensiblement la capacité de détection des objets cachés. En effet, tout élément ajouté j contribue à diminuer l'intensité du faisceau d'un facteur additionnel $\exp(-\mu_j x_j)$ appliqué à I, obtenu en l'absence de j. Ainsi, le calcul de l'absorption causée par j revient à effectuer une simple division. Les différences $\mathbf{D}$ sont donc calculées à partir du template déformé $\mathbf{T}_{\mathbf{u}^*}$ et l'image statique à inspecter $\mathbf{R}$ de la façon suivante :

$$\mathbf{D}(\mathbf{x}) = \begin{cases} 1 - \frac{\mathbf{R}(\mathbf{x})}{\mathbf{T}_{\mathbf{u}^*}(\mathbf{x})} \text{ si } \mathbf{T}_{\mathbf{u}^*}(\mathbf{x}) \geqslant \mathbf{R}(\mathbf{x}) \text{ et } \mathbf{T}_{\mathbf{u}^*}(\mathbf{x}) \neq 0 \\ 0 \text{ sinon} \end{cases} \tag{6}$$

Des résultats qualitatifs sur des images radiographiques confirment ce choix. Ils montrent que, grâce à notre approche, des irrégularités peuvent être détectées dans des endroits extrêmement denses. Un bref aperçu des méthodes de seuillage et des opérateurs morphologiques est également fourni.

Partie 2

Dans la seconde partie de la thèse, nous proposons des solutions à deux problèmes de recalage de types différents. Le premier concerne l'automatisation de l'analyse d'irrégularités dans les conteneurs. Bien que la méthode proposée se concentre essentiellement sur les modules de réfrigération, elle peut également être étendue à d'autres régions. Le deuxième cas d'étude porte sur la problématique de recalage de scans de véhicules d'un même modèle, obtenus par HCVL. Pour

les deux applications, les méthodes proposées ont été testées sur des centaines voire des milliers d'images.

Le chapitre IRU chapter introduit la nécessité d'une inspection automatique des unités de réfrigération (UR) dans les conteneurs ainsi que les compartiments refroidisseurs d'air des semi-remorques (voir Figure 11). La procédure de construction de la base de données est décrite ainsi que la classification SVM effectuée sur les descripteurs d'entités CNN-PCA extraits de **R**. Cependant, au sein de chaque catégorie d'UR, de nombreuses variations peuvent être observées à travers les différentes images, compliquant ainsi la sélection automatique du template **T**. C'est la problématique dite de "variabilité intra-classe". Des expériences ont été menées en utilisant des points-clés mis en correspondance entre **R** et **T**. D'autres approches utilisent des vecteurs HoG ou CNN-PCA pour trouver l'échantillon le plus proche de **R** parmi l'ensemble de données. Néanmoins, ces techniques ne tiennent pas compte des déformations locales et résultent souvent en un **T** trop différent de **R** pour envisager un recalage ultérieur.

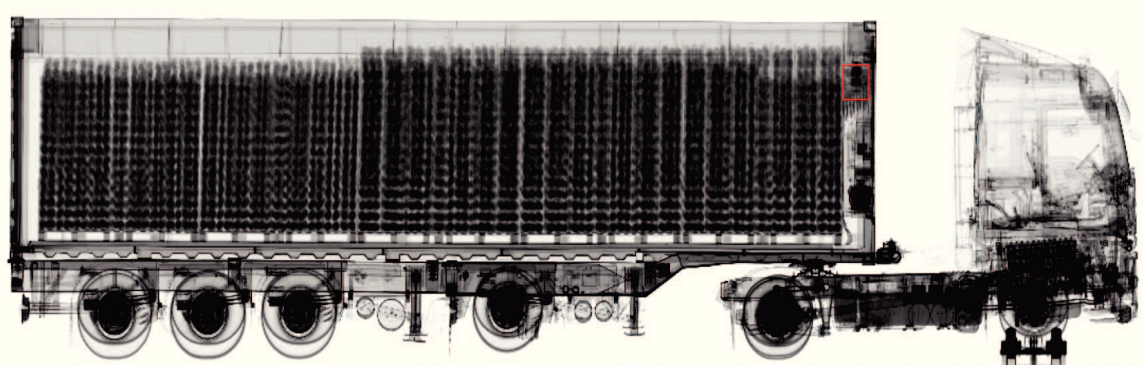
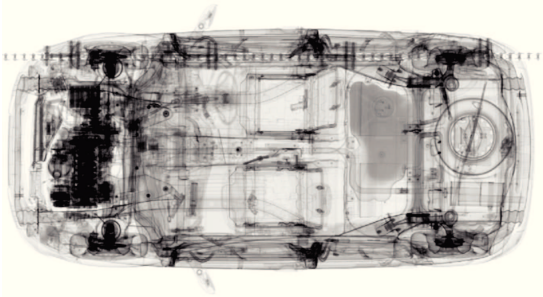


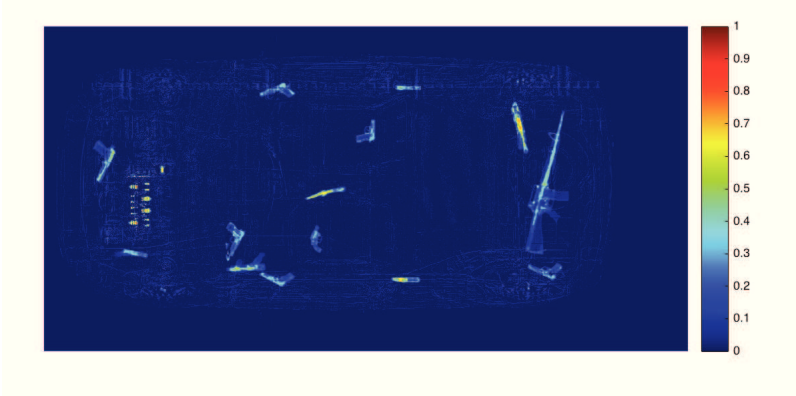
Figure 11 – Détection d'irrégularités dans l'unité de réfrigération d'un conteneur ISO. L'inspection manuelle avait alors abouti à la saisie d'une cargaison de cocaïne.

Par voie de conséquence, nous proposons une intégration nouvelle des ASM pour l'extraction du template le plus proche, permettant l'appariement précis des formes saillantes. Grâce à cette méthode, les formes d'intérêt, susceptibles de supporter des déformations locales, sont directement ciblées par l'application de l'ASM. Le bénéfice de la méthode est assez significatif : il n'est pas nécessaire d'analyser la nature des déformations observées et de concevoir un schéma de recalage personnalisé. En revanche, sachant que **T** a été sélectionné pour sa similitude à **R**, nous avons suggéré de combiner une simple approche linéaire par morceaux avec une régularisation par diffusion aux limites des sous-images. Un seuillage élémentaire suivi d'opérations morphologiques permet alors la détection d'irrégularités. Plusieurs tests, effectués sur les systèmes HCVM et HCVS, donnent une précision de détection élevée.

Dans le chapitre 6, le recalage d'images radiographiques de véhicules légers d'un même modèle est abordé (scans HCVL, vue de dessus, voir Figure 12). Par définition, une voiture est un corps rigide, de telle sorte que de simples approches linéaires devraient suffire pour résoudre le problème d'alignement. Néanmoins, les images obtenues présentent fréquemment de fortes déformations non linéaires, affectant le résultat de recalage lorsqu'un schéma rigide est adopté. Les méthodes classiques non paramétriques et paramétriques (respectivement B-spline et démons) sont testées mais ne permettent pas d'obtenir un alignement précis et transgressent sensiblement la rigidité intrinsèque du véhicule représenté.



(a)



(b)

Figure 12 – (a) Une image de voiture inspectée où plusieurs pistolets et fusils d'assaut ont été intentionnellement disposés ; (b) images des différences après recalage avec une référence vide. Plusieurs armes à feu sont facilement repérables.

Notre idée consiste à examiner les origines de distorsions non linéaires observées. De fait, en pénétrant le système HCVL, le véhicule est traîné par un convoyeur via des cylindres poussant sur ses roues arrière. Bien que durant le scan, la vitesse est censée demeurer constante (12 m/min), les chocs entre les cylindres et les roues peuvent provoquer une accélération suivie d'un ralentissement de la voiture. Ainsi, ce phénomène conduirait à un suréchantillonnage de certaines régions du véhicule scanné, entrainant ainsi de fortes déformations. Deux hypothèses simplificatrices ont pu être énoncées :

- Le déplacement recherché $\mathbf{u} = (u_x, u_y)$ est constant par colonne. En effet, chaque coupe latérale du véhicule est acquise à une fréquence d'échantillonnage constante et correspond à une colonne de l'image. Il vient donc : $u_x(x, y) = u_x(x)$.
- La déformation observée ne se produit que selon une direction horizontale (dans le sens du mouvement de la voiture), de telle sorte que la composante verticale de $\mathbf{u}$ est nulle : $u_y(x, y) = 0$.

Ces hypothèses sont intégrées au schéma d'optimisation (5) et permettent de passer d'un problème bi-dimensionnel à un paradigme unidimensionnel, résolu par un algorithme de descente du premier ordre. L'approche que nous présentons fait usage d'un recalage 1-D par diffusion pour pénaliser les transformations non plausibles.

$$\begin{aligned} &\text{trouver } u_x(x) \text{ minimisant} \\ &J\left[u_x(x)\right] = \frac{1}{2}\left\|\mathbf{T}(u_x(x)) - \mathbf{R}\right\|_2^2 + \frac{\alpha}{2}\left\|\nabla_x u_x(x)\right\|_2^2 \end{aligned} \tag{7}$$

La minimisation est effectuée par la méthode du gradient associée à la condition d'Armijo et une approche multi-résolution. Nous montrons que notre approche surpasse les méthodes classiques testées en termes d'alignement des images et respect des contraintes de rigidité.

Au-delà des résultats obtenus, d'autres expériences ont montré que notre modèle n'est plus valide lorsque l'un des deux véhicules scannés est déplacé latéralement par rapport à l'autre (ou un même véhicule entre deux acquisitions). De fait, Le faisceau incident est pyramidal et conduit à un effet stéréoscopique : les objets plus proches de la source subissent une déformation plus significative entre les deux images du véhicule translaté.

En d'autres termes, la composante verticale de $\mathbf{u}$ doit être intégrée dans le processus d'optimisation. Cependant, les corrections verticales et horizontales sont profondément différentes de par leur nature et les deux problèmes doivent être traités séparément. Nous proposons de rectifier tout d'abord $\mathbf{T}$ et $\mathbf{R}$ afin d'aligner leurs colonnes (premier recalage horizontal 1-D). Et dans un second temps, pour chaque colonne, nous suggérons d'optimiser les emplacements des pixels par une correction verticale.

L'enjeu de la préservation des volumes est largement discuté et l'ajout d'une contrainte souple d'incompressibilité conduit à un schéma de recalage élastique (avec la transformation $\phi = u_y + id$) :

$$\begin{aligned} &\text{trouver } u_y(x,y) \text{ minimisant} \\ &J\left[u_y\right] = \frac{1}{2}\left\|\mathbf{T}(u_y) - \mathbf{R}\right\|_2^2 + \frac{\alpha}{2}\left\|\nabla u_y\right\|_2^2 + \frac{\beta}{2}\left\|\det(\nabla u_y + id) - 1\right\|_2^2 \end{aligned} \tag{8}$$

L'approximation l-BFGS de la méthode de Gauss-Newton est adoptée pour résoudre le problème d'optimisation. Divers exemples de recalage d'images présentés démontrent la nécessité de notre méthode ainsi que l'amélioration notable des capacités de détection qu'elle offre.

Introduction

Outline of the current chapter

Context

In the current geopolitical context, security has increasingly imposed itself as a top priority for governments and worldwide agencies. In our contemporary, globalized world, international borders are less obvious, which indeed turns counter-terrorism alongside illicit trade eradication into defining stakes. It implies the deployment of human means associated with advanced technologies at sea, air and land borders for more thorough controls of passengers and trans-shipment goods.

Enforcement agencies strive to tackle organized crime and terrorist activities through various means. A mid-term solution is to counter effectively the smuggling of lethal weapons, and whereby deprive criminals from any capacity to harm. In Europe for instance, different contraband networks from the Balkans, UK and Spain have been recently unveiled (see Figure 13). Spectacular seizures have shown that not only handguns but also genuine war weapons such as ground-air rocket launchers were part of the contraband arsenal [33].

(a) (b)

Figure 13 – (a) Huge weapons seizure in Spain by Europol in March 2017 [81]; (b) 1350kg of cannabis resin found in the side walls of a tractor truck (Moroccan Customs, [128]).

In a similar perspective, the illicit narcotics trade is known to be the major source of profit for organized criminal syndicates. Besides causing public health concerns, it tightly relates to security issues, as it very often serves to fund larger-scale crimes as terrorism. Traditionally, the drug market includes cannabis, cocaine and opiates. Yet, Customs authorities are growingly faced with newer types of synthetic banned narcotics and stimulants smuggling. To get a general picture, in 2016 more than 1.5 million kg of narcotics were seized as they were being transported illicitly between and within countries [128].

Apart from these critical illicit-trade concerns, cigarettes, artworks or cultural objects smuggling are also at the very heart of Customs authorities enforcement actions.

Massive investments have been dedicated to find innovative solutions in order to assist the military or custom officers in their tasks. Over time, a very wide range of products has emerged at sensitive sites, airports, harbors and border-crossings. It is the case of explosives and narcotics trace detectors as well as X-ray scanning machines for luggage inspection.

Still, to address larger-scale illegal trade and homeland security threats, X-ray scanning facilities were especially designed for cargo non-invasive inspection. The Smiths Detection company (formerly Heimann Systems), within which the present work has been carried out, is one of the current leaders on the specialized high energy scanning machines market. Since 1970, Heiman Systems is renowned for being one of the first manufacturers of X-ray luggage inspection machinery. The first stationary systems, dedicated to the control of freight were developed in the 1990s. Gradually, the emergence of international companies such as Rapiscan Systems (USA) or Nuctech (China) boosted the technological competitiveness and led to a fast diversification of the range of products.

Cargo X-Ray Scanning

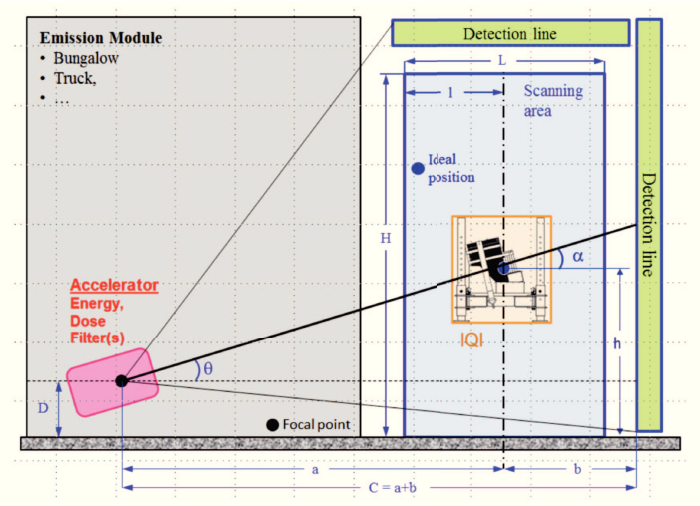


Figure 14 – Sketch of cargo X-ray scanning.

Without getting into the detailed physics behind scanning systems, we will explain the fundamentals of cargo X-ray imaging. X-ray photons are produced by high-energy electrons hitting an object through a linear particle accelerator. The resulting beam undergoes a filtering process in order to refine its energy distribution.

Then, when the emitted filtered beam penetrates an obstacle, a fraction of incident photons is being absorbed. Alternatively it interacts with the environment, and gives rise to a scattering phenomenon often associated with a Gaussian noise. After crossing the cargo, the residual beam is measured by sensors placed opposite to the accelerator, to produce a two-dimensional image (Figure 14).

The X-ray beam attenuation is modelized by the Beer-Lambert's law. Given a material of linear attenuation coefficient μ (in m^{-1} and of thickness x (m), the intensity I (number of photons transmitted after getting through the material, per volume unit) decreases exponentially with respect to the incident beam intensity I_0:

$$I = I_0 \exp(-\mu x) \tag{9}$$

The general expression of the absorption law is a fundamental property employed all along our research study. When the beam goes through N different materials characterized by μ_i and x_i, the law is written:

$$I = I_0 \exp(-\sum_{i=1}^{N} \mu_i x_i) \tag{10}$$

The resulting rough signal is then processed via diverse transforms and finally stored in a 16 bits/pixel image format.

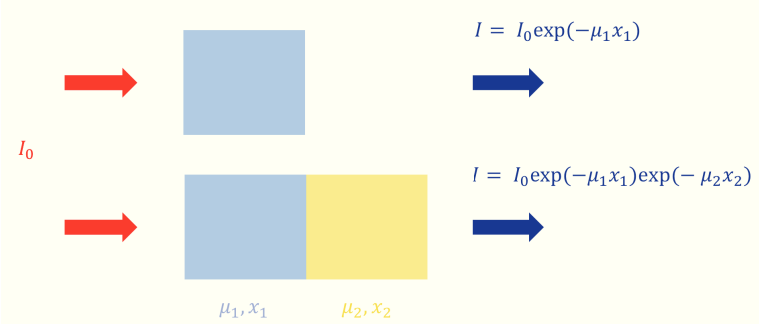


Figure 15 – Illustration of Beer-Lambert's attenuation law.

Concerning freight inspection, very high energy accelerators are used (by Bremsstrahlung) with an energy ranging from 4 to more than 10 MeV. Indeed, high energy (short wavelength) radiation is less absorbed by the material, and therefore performs a further penetration. High energy X-rays are called "hard rays" with wavelengths below one nanometer. These X-ray systems can allow for a penetration of more than 30 to 40 cm of steel for vehicles moving up to a speed of 13 km / h. It should be noted that some machines have a "multi-energy" option, with a *low mode* and *high mode*, providing the discrimination information between materials on the basis of the computed atomic numbers Z. Thus, organic materials are easily distinguished from non-organic bodies.

Cargo Imaging Machines

Figure 16 – HCVS imaging system

Various types of scanning systems are usually proposed on the market. With different possible geometries, energy ranges, number of views and scanning modalities, the machines can be deployed in multiple types of environments and address specific inspection tasks. While a few models are stationary, other scanning facilities are mobile systems, carried on a truck facility. The main purpose of this technology is to inspect containers and vehicles by checking manifests and looking for threats (explosives, narcotics, weapons of mass destruction) and contraband (narcotics, undeclared cigarettes...).

The high-energy machines produced by Smiths Detection are referred to by the acronym HCV for Heimann Cargo Vision. The main stationary systems are:

- The HCVS (*Stationary*) screening system is designed for security checks at ports and border crossings (heavy traffic locations) by screening whole trucks and containers. It is

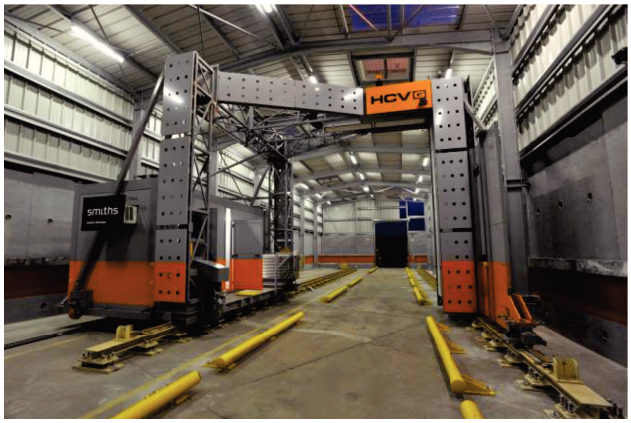


Figure 17 – HCVG scanning machine

Figure 18 – HCVP scanning machine

a 9 MeV energy machine, delivering organic and inorganic material discrimination and colorization based on atomic number. It provides both side and top-view scans with a very high resolution (Figure 16).

- The HCVG is a *gantry* inspection system on which a single side-view scan is yielded, with an interlaced energy from 4 to 6 MeV via a *pulse-to-pulse*. The gantry is translated along the motionless inspected subject during scanning. A description is provided on Figure 17.

- The HCVP (*Portal* is a drive-through portal system with identical energy specifications to the HCVG. The main difference lies in that the system remains fixed during scanning, as trucks move through the portal at low speed (Figure 18).

- The HCVL (*Light*) is designed to inspect lighter vehicles including cars, vans, minibuses and mobile homes. It outputs top-view images. Available in conveyor configuration, it is dispatched at border-crossings, sea port terminals, airports and other sensitive sites (Figure 19).

Smiths Detection manufactures two types of mobile X-ray scanning systems:

Figure 19 – HCVL scanning machine

- The HCVM (*Mobile*, see Figure 20) is available as a truck or towable trailer and uses accelerators ranging from 4 MeV to 6 MeV allowing for material discrimination. Both accelerators and sensors are located in a deployable arm that enables two scanning modes: gantry and pass-through. It produces a single side-view of inspected vehicles.
- HCVMe (*easy*) is the equivalent, more compact and lower energy mobile scanning system. It is represented on Figure 21.

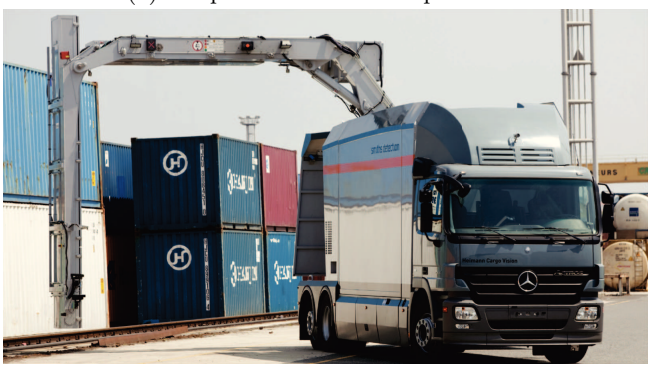
Figure 21 – HCVMe X-ray machine

In the course of our study, the focus was given on features applied on HCVS, HCVM and HCVL screening machines. Indeed, the major part of the available data (in thousands of images) were obtained from these three systems. Yet, most of our methods can be cautiously extended to other systems.

Role of Image Processing Techniques

Image processing and computer vision methods are utterly ubiquitous in defense industries and provide wide-ranging applications: from CCTV tracking, remote sensing satellite surveillance to heat-seeking missiles...

(a) compact mode for transportation



(b) scanning arm deployed

Figure 20 – HCVM system

In the field of cargo X-ray screening, image processing is unavoidable, especially for the very formation of the resulting samples, denoising, enhancement, or even segmentation problems (*e.g* for material discrimination). However, our main objective throughout this study was to address other types of issues.

Actually, in our so-called "Big-Data" era, huge amounts of images have to be processed in shorter periods of time. At the same time, public deficits lead to very sharp workforce cuts, and security officers are not been left unaffected as reported on Figure 22. Thereby, the switch towards threat targeting automation constitutes an imperative step, considering the current security situation.

The industry, as a whole, has significantly been impacted by this move and must address new problems, by often using approaches already proven in cross-cutting areas.

In fact, illicit goods and security threats can be automatically detected by two types of methods:

- Adopting recognition machine-learning (ML) techniques (*e.g* by training a convolutional neural network). It supposes providing a training set of images containing targeted threats. It also assumes that specific textures or other feature signatures can be extracted from the type of objects being searched.

- An alternative approach consists in looking for irregularities instead of identifiable threats. A discrepancy is defined here as any item placed at a spot where no foreign element is meant to be set. In vehicles, this applies to car fronts for instance. The idea is to compare the (or a part of the) current scan under inspection with a reference sample through image

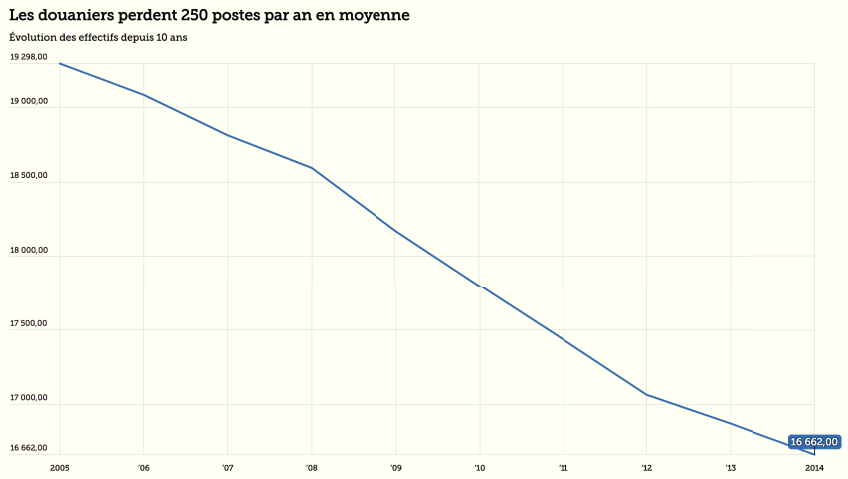


Figure 22 – Downsizing among customs officials in France over the past ten years ([112]).

registration techniques. As we will bring-forward in this thesis, narcotics do not display specific features that would enable machine-learning recognition. Instead, one way to detect them is to analyze regular concealment locations in vehicles and containers via image comparison. Moreover, several cases reported all along our study demonstrate that the proposed techniques make irregularity visualization possible on regions where both human vision and machine learning approaches are very prone to yield poor detection results.

The concepts of image registration will be explored in a very precise manner throughout this thesis. From a general point of view, it is defined as the search for an optimal alignment between an inspected image $\mathbf{R}$ and a retrieved reference template $\mathbf{T}$, provided the applied transformation is plausible. Since detecting irregularities is our primary objective, to avoid possible threats distortion, $\mathbf{R}$ will always remain static and $\mathbf{T}$ will stand for the deformable image.

Nevertheless, because of the confidential nature of the field of research, along with the relatively recent emergence of cargo screening technologies, a very few literature is available on the subject. The works of [74] must be cited, yet they only review very general and well-known approaches on X-ray cargo inspection. As a consequence, most of the explored methods throughout this thesis find their roots in medical image analysis, where they gained a lot of popularity and were already subject to comprehensive studies.

Thesis Outline and Our Contributions

This thesis addresses particular image registration issues applied to the context of cargo X-ray screening. Given an entire input sample $\mathbf{R}$ or a specific region of interest (ROI), the core idea lies in combining two successive stages:

1. Retrieving a corresponding template $\mathbf{T}$ by using state-of-the-art classification approaches.
2. Finding a smart registration scheme that suits the observed deformations and mechanical constraints imposed by the scanning systems. On the basis of the resulting pair of images, alignment can thus be achieved to compute a difference map and detect major changes.

Our works articulate into two main parts. While the first introduces in detail all utilized building-block methods, the second aims at providing the algorithm pipelines, their implementation and performances.

Part 1

The first chapter introduces the Active Shape Models (ASM) [23] theory in which shape annotation, training and fitting steps are reported. The technique was initially introduced for the segmentation of objects whose shape is learned from a manually annotated image database, forming a ground truth. We indeed stray from the classical method by using ASM reduced representations to retrieve, among our dataset, the closest template $\mathbf{T}$ - in terms of salient shapes - given a labelled input image $\mathbf{R}$. Thus, following annotation, each shape vector defined by $\mathbf{x} = (x_1, y_1, ..., x_n, y_n)^T$ is approximated by PCA (*Principal Component Analysis*):

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{P}\mathbf{b} \tag{11}$$

where the t eigenvectors related to the largest eigenvalues λ_i of the covariance matrix are stacked in $\mathbf{P}$. $\mathbf{b}$ is a t dimensional vector containing the parameters/weights of the model. Hence, the ASM can be applied to each image of a given class m from the training set, yielding q vectors $\mathbf{b}_i$. While fitting the ASM on the input image (under inspection) $\mathbf{I}_{new}$, $\mathbf{b}_{new}$ is computed and the closest image in terms of shape is given by:

$$i^* = \underset{i \in [\![1,q]\!]}{\arg\max} \|\mathbf{b}_i - \mathbf{b}_{new}\| \tag{12}$$

Besides, we also adapt ASM in a singular way by the formulation of a combined cost based on shape and appearance data, computed after fitting the model on $\mathbf{R}$. It actually aims at applying a double-check process to refine classification results; which turns out to be crucial in borderline cases (limited amount of data, image similarities between labels).

In chapter 2, we intend to present both representation and classification approaches employed, at different levels, all along this thesis. Standard dense and sparse feature descriptors such as HoG, SIFT and SURF are reported. A global overview of deep learning techniques, with a significant focus on Convolutional Neural Networks (CNN) is provided. Particularly, pre-trained CNN used for features extraction are described, as they take a front seat role throughout our research study. Unsupervised classification, which allowed to speed-up database construction processes is also brought forward, through the K-means algorithm. On the other hand, supervised classification, namely multi-class support-vector-machine (SVM) is outlined and the choice for a *one-to-one* scheme is also motivated.

As far as registration methods are concerned, they are reviewed in chapter 3. First, feature-based linear methods are described. The second section of the chapter is dedicated to the state-of-the-art of non-linear registration methods. The formulation of the non-rigid registration problem within the variational framework provides the cornerstone for the major part of this thesis. Let's denote the field of displacement $\mathbf{u} : \Omega \to \mathbb{R}^d$ (in d dimensions), the optimization problem is written as follows:

$$\begin{aligned} &\text{find } \mathbf{u} \text{ minimizing} \\ &\mathcal{J}[\mathbf{u}] = \mathcal{D}[\mathbf{T}(\mathbf{u}), \mathbf{R}] + \alpha \mathcal{S}[\mathbf{u}] \end{aligned} \tag{13}$$

with $\mathcal{D}$ standing for the data-fitting term and $\mathcal{S}$ the regularization component. Both parametric and non-parametric approaches are outlined as well as different similarity and smoothing terms. Volume-preserving issues are handled by the application of specific constraints. The numerical resolution aspect is also discussed through first an second-order descent methods.

Finally, our piece-wise linear approach is provided, aiming at handling deformations in a more local manner, without resorting to sophisticated non-linear paradigms (more computationally demanding). It consists in partitioning both $\mathbf{R}$ and $\mathbf{T}$ on the basis of ASM fitted shapes. SURF keypoints are extracted and matched between the s_d pairs of sub-images $\{(\mathbf{T}_i, \mathbf{R}_i)\}_{1 \leqslant i \leqslant s_d}$. Dense displacement fields $\mathbf{u}_i$ are computed to be applied on each sub-image of the moving template $\mathbf{T}$. Discontinuities at sub-images borders are addressed by using an isotropic direction diffusion scheme.

Chapter 4 describes the difference detection process following the registration of $\mathbf{R}$ and $\mathbf{T}$. We demonstrate, via Beer Lambert's attenuation law, that using a ratio instead of a simple difference scheme enhances considerably the detection capability of concealed objects. Indeed, any added object j contributes in decreasing the beam intensity by an additional factor $\exp(-\mu_j x_j)$ applied on I, obtained in absence of j. Thus, computing the absorption caused by j amounts to performing a simple division. The differences $\mathbf{D}$ are thus calculated from the deformed template $\mathbf{T}_{\mathbf{u}^*}$ and the static sample to be inspected $\mathbf{R}$ in the following way:

$$\mathbf{D}(\mathbf{x}) = \begin{cases} 1 - \frac{\mathbf{R}(\mathbf{x})}{\mathbf{T}_{\mathbf{u}^*}(\mathbf{x})} \text{ if } \mathbf{T}_{\mathbf{u}^*}(\mathbf{x}) \geqslant \mathbf{R}(\mathbf{x}) \text{ et } \mathbf{T}_{\mathbf{u}^*}(\mathbf{x}) \neq 0 \\ 0 \text{ otherwise} \end{cases} \tag{14}$$

Visual results along these lines are provided on X-ray images, and show that, through our approach, irregularities can be detected in very dense locations. A brief outline of thresholding methods paired with morphological operators is also supplied.

Part 2

In the second part of our thesis, we propose new solutions to two different types of registration problems. The first issue concerns the analysis automation of so-called rip-offs concealments in containers. Although the suggested method focuses essentially on refrigeration modules, it can be extended to other regions as well. The second study case deals with the registration problem of HCVL top-view scans of same-model vehicles. For both applications, our methods have been tested on hundreds to thousands of images.

Chapter 5 introduces the need for an automatic inspection of refrigeration units in containers and air-cooler compartments for semi-trailers (see 23). The database construction procedure is described as well as the SVM classification performed on CNN-PCA feature descriptors extracted from the input $\mathbf{R}$. Yet, for each label, many variations can be observed across the training set, turning the template $\mathbf{T}$ selection into a challenging task. **This issue is referred to as the "intra-class variability" problem**. Experiments have been conducted using matching keypoints between $\mathbf{R}$ and $\mathbf{T}$. Other approaches employed HoG or CNN-PCA feature vectors to find the closest sample to $\mathbf{R}$ across the dataset. Yet, all these techniques did not account well for local changes, and inputs were often mismatched with regards to a further registration task.

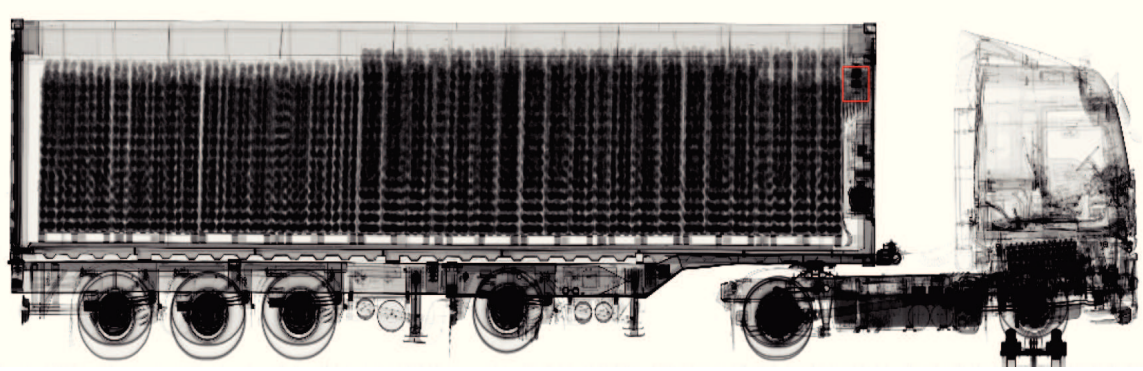


Figure 23 – Irregularity detection in the refrigeration unit of an ISO container. A further manual inspection resulted in the seizure of cocaine.

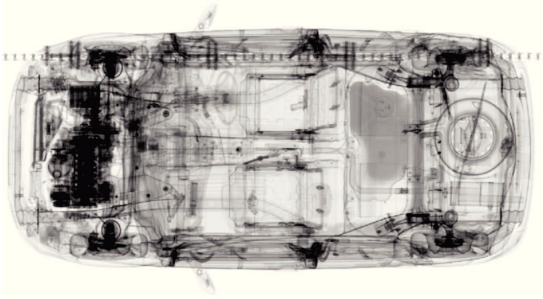
As a result, we propose a new embodiment of ASMs for closest-template retrieval, allowing for a better accuracy for salient shapes matching. Through this method, shapes of interest, very likely to bear local deformations, are directly targeted by the application of ASM. The benefit of the method is quite significant: one does not need to analyze the nature of the observed deformations to design a customized registration scheme. Instead, given that $\mathbf{T}$ was chosen to be as similar as possible to $\mathbf{R}$, we suggested to employ our simple and fast piece-wise linear approach paired with a diffusion regularization at sub-images borders. A simple thresholding followed by morphological operations outputs the irregularity detection result. Several tests, performed on HCVM and HCVS X-ray scans provide a high detection accuracy.

In Chapter 6, the registration of top-view scans of same-model light vehicles is addressed (HCVL scans, see Figure 24). By definition, a car is a rigid body such that linear-registration approaches should be sufficient to solve the alignment problem. Nevertheless, output images present frequently strong non-linear deformations, affecting the registration result while employing rigid schemes. Non-parametric and parametric standard methods (resp. B-spline and demons) are tested but do not achieve an accurate alignment and strongly violate the intrinsic stiffness of the car.

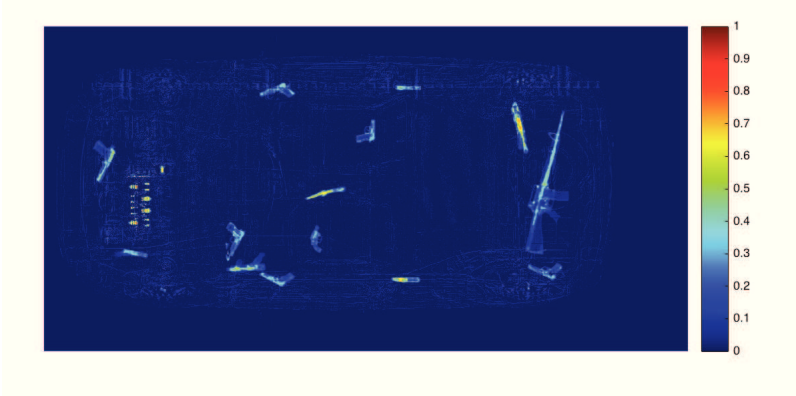
Our idea consists in the examination of the reported non-linear distortions origins. In fact, while entering the HCVL scanning system, a vehicle is trailed by a conveyor as rollers push on its back wheels. Although the speed during scanning is meant to remain steady (12 m/min), shocks between the rollers and the rear wheels may cause an acceleration followed by a slowdown of the car. Hence, this phenomenon leads to an oversampling on some regions of the scanned vehicle and entail strong deformations. Two simplifying assumptions were thereby stated:

- the sought displacement $\mathbf{u} = (u_x, u_y)$ is column-wise constant since each lateral cut of the car is acquired at a constant sampling rate (and corresponds to each column). It comes: $u_x(x, y) = u_x(x)$.
- the displacement occurs only on the horizontal direction, such that the vertical component of $\mathbf{u}$ is null: $u_y(x, y) = 0$.

We integrate these assumptions into the variational framework (13) and hence move from a 2-D scheme to a simple 1-D problem (15) solved by a first-order descent algorithm. In our approach, we used a 1-D diffusion registration to penalize unwanted transformations.



(a)



(b)

Figure 24 – (a) Inspected car image where several guns and rifles were intentionally arranged; (b) map of differences following registration with an empty reference. Several firearms are easily underlined.

$$\begin{aligned} &\text{find } u_x(x) \text{ minimizing} \\ &J\left[u_x(x)\right] = \frac{1}{2}\left\|\mathbf{T}(u_x(x)) - \mathbf{R}\right\|_2^2 + \frac{\alpha}{2}\left\|\nabla_x u_x(x)\right\|_2^2 \end{aligned} \tag{15}$$

Minimization is performed by gradient descent in combination with backtracking line-search and coarse-to-fine methods. We show that our approach outperforms the tested standard methods in terms of data fitting and the upholding of stiffness constraints.

Beyond the obtained results, further experiments showed that our model does not hold when one of the scanned vehicles is shifted side-wards, with respect to the other one. The incident beam is indeed pyramidal, and leads to a stereoscopic effect: objects that are closer to the source undergo a stronger deformation whenever the car is translated laterally.

In other words, the vertical component of $\mathbf{u}$ must be integrated in the optimization process. Still, vertical and horizontal corrections are thoroughly different by nature and both problems must be addressed in a separate fashion. We propose to first rectify $\mathbf{T}$ and $\mathbf{R}$ to align their columns (first 1-D horizontal registration). In a second stage, for each column, we suggest to optimize the pixels locations by a vertical correction.

The volume-preserving issue is widely discussed and the use of an additional incompressibility soft constraint leads to an elastic registration scheme, phrased as follows:

$$\begin{aligned} &\text{find } u_y(x,y) \text{ minimizing} \\ &J\left[u_y\right] = \frac{1}{2}\left\|\mathbf{T}(u_y) - \mathbf{R}\right\|_2^2 + \frac{\alpha}{2}\left\|\nabla u_y\right\|_2^2 + \frac{\beta}{2}\left\|\det(\nabla u_y + id) - 1\right\|_2^2 \end{aligned} \tag{16}$$

The l-BFGS approximation of the Gauss-Newton method is adopted to solve the optimization problem. Various reported registration examples demonstrate the necessity for our method as well as the substantive detection-capabilities enhancement it offers.

Part I

Mathematical Methods

Chapter 1

Closest Shape Detection using ASM

Outline of the current chapter

1.1 Introduction

Developed by Tim Cootes and Chris Taylor in the mid-90's, the *Active Shape Models* methods (ASM) [26], [24], [23] are still an active field of research in computer vision. They are particularly studied for segmentation tasks in medical imaging or face tracking (*e.g* identity and facial expression recognition).

Essentially, the technique consists in the construction of statistical models from the shape of objects, represented by a set of points. The Point Distribution Model (PDM) is extracted through a training process over an annotated dataset of shapes. As a consequence, shapes can be handled in much lower dimensions. Given a new sample, the technique aims at matching the trained statistical shape to the image. Here, the focus is given on 2-D images only; yet, the method can easily generalize to 3-D as reported in [84] for instance..

The main advantage over active contours and front propagation methods is related to their difficulty to integrate shape priors into the energy minimization scheme. Though, in some situations, a quick dataset examination shows that the main variations of a shape are known.

Constraining the model to fit in with the variations found in the training set, would be worthwhile in this case. To overcome these limitations, several methods have been proposed in the past. Yuille *et al.* [132] for instance developed hand-crafted parametric models, but their method is utterly application dependent in that it exclusively covered human eye shapes and face features. Fourier series shape models were also proposed by Staib and Duncan [104] but do not generalize to open boundaries.

Over the years, ASM methods have evolved and led to various improving techniques. The famous AAM approach (*Active Appearance Models*) [25] is analogous to the ASM and aims at fitting a statistical model to a new image. Still, in this method, the model is composed of both object shape and appearance information. It is widely discussed in [105]. Until recently, the CLM (*Constrained Local Models*) [29] scheme has drawn a lot of attention, especially due to its efficiency to handle facial expression recognition tasks (Figure 1.1). In this paradigm either, shape and appearance are considered. First, a set of "response images" are generated to perform a local exhaustive search. In a second phase, a shape model is used to optimize the global response. For further information, refer to some of the latest works on the subject: for instance Asthana *et al.* [6] Linder *et al.* [65] or Martins *et al.* [72]. For a complete overview, the works of [84] can be consulted.

Figure 1.1 – CLM fitting on faces images [6]

More advanced methods have been elaborated to address large amounts of variability in particular shapes. This is especially true for face tracking or medical applications where objects

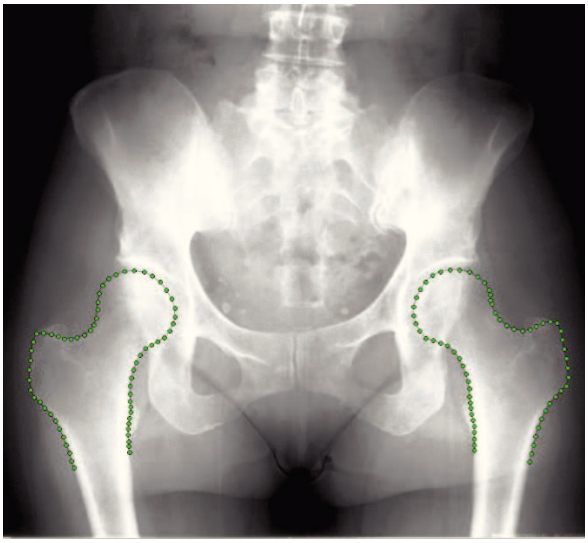


Figure 1.2 – Bone segmentation using the BoneFinderTM technology on 2-D images [66]

of interest may strongly deform in a nonrigid way (soft tissues). In our field of research, we are less prone to be confronted with these kinds of variability. A typical scanned vehicle, despite undergoing some distortion, still maintains an intrinsic global rigidity. Accordingly, we preferred to adopt the ASM approach, for its simplicity and low computational cost.

In most applications, statistical models are used for segmentation (*e.g.* lung or bones segmentation [119], see Figure 1.2) and recognition (face tracking, expression recognition or identification). **In our work, we propose an new approach to address an "intra-class" variability issue**. Let us consider a database gathering different categories of images; within each class, the images may display significant shape variations. Assume also that the classification task is already performed by some technique (see Chapter 2). Given a new X-ray scan, the purpose is to identify, within the corresponding group, the image carrying the closest shape of interest.

In this chapter, we will first briefly review the ASM approach through the description of the annotation process, the model construction as well as the fitting strategy. Then, we will present our particular way to adapt ASM for the "closest shape identification" problem. Finally,on the basis of the fitted shape, we introduce a simple formulation of a cost measure, with the objective of double-checking classification outputs.

1.2 Data Annotation

In ASM, shapes are represented by a set of n landmark points. Formally, for 2-D images, they are stacked in shape vectors $\mathbf{x}$ of dimension $2n$:

$$\mathbf{x} = (x_1, y_1, ..., x_n, y_n)^T \tag{1.1}$$

These cue points are manually annotated on the images and are usually located along the edges of shapes of interest. Annotation is achieved by picking landmark points at corners or

"T" junctions between boundaries in combination with spline interpolation to provide a denser description of the shape [24].

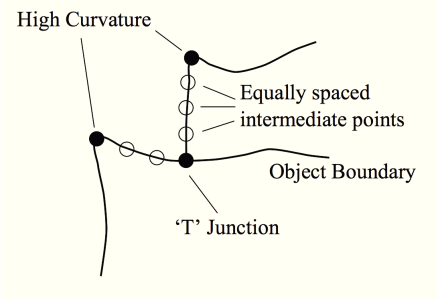


Figure 1.3 – Landmark annotation illustration on "T" junction [24]

More precisely, we designed a tool to annotate pairs of cue points with a few intermediate random instances depicting the local curve of the shape. From these samples, equidistant points are generated between the manually labelled landmarks through spline interpolation. Annotation descriptions are provided on Figure 1.3 and Figure 1.5.

As detailed in the next section, the connectivity between shape points must also be recorded. In fact, this information is crucial to define the shape structure and compute, for each landmark, gray-level profiles in the normal direction ([119], [24]). In the scope of this work, only closed shapes are considered.

Given s training examples, s vectors $\mathbf{x}_j$, $j \in [\![1, s]\!]$ are extracted, Procrustes Analysis ([42], [84]) is employed to normalize and align the shapes for a uniform representation across the training set. By employing an identical coordinate frame, the constructed statistical model is thus independent of the size, position and orientation of the objects (Figure 1.4). See Figure 1.6 for an example related to the type of shape annotated in Figure 1.5.

Note that in our study, a dozen of manually labelled images are sufficient to yield a coarse model, which, applied on the data-set remainder can achieve automatic annotation. This way, the model can be refined automatically in terms of robustness to new inputs. Considered that annotating a training sample takes about one minute (for both landmarks and intermediate points), the iterative refinement strategy proves to be a significant time saver.

1.3 Statistical Model Construction

Following the data labelling step on a set of training images, the core idea is to build a statistical model by applying Principal Component Analysis (PCA) [129] to the shape vectors $\mathbf{x}$. The mean shape is first computed,

$$\bar{\mathbf{x}} = \frac{1}{s} \sum_{i=1}^{s} \mathbf{x_i} \tag{1.2}$$

along with the covariance matrix:

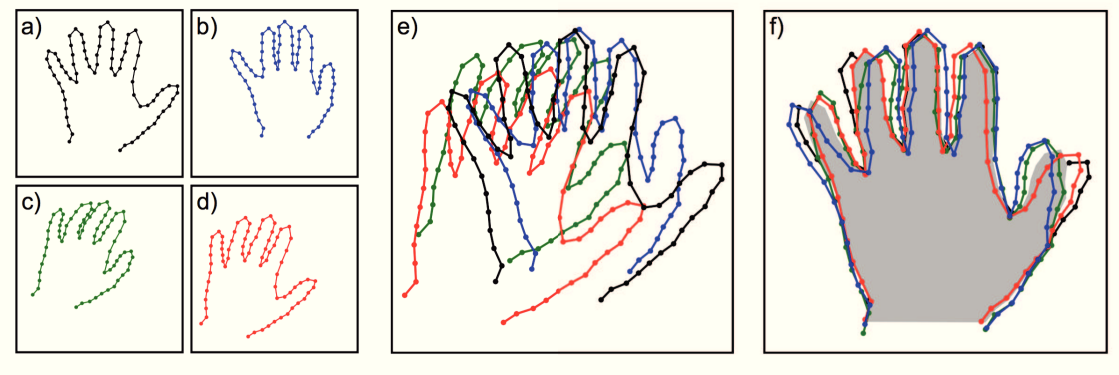


Figure 1.4 – Generalized Procrustes analysis example on hand shapes; a-d) Four training shapes; e) the shapes are initially not aligned well; f) Shapes alignment following Procrustes analysis; gray region illustrates mean shape (from [84]).

$$\mathbf{S} = \frac{1}{s-1}\sum_{i=1}^{s}(\bar{\mathbf{x}} - \mathbf{x_i})\,(\bar{\mathbf{x}} - \mathbf{x_i})^T \tag{1.3}$$

Thanks to a Single Value Decomposition (SVD) [41], both eigenvectors and eigenvalues of the covariance matrix $\mathbf{S}$ are computed. The t eigenvectors related to the largest eigenvalues λ_i are stacked in the matrix $\mathbf{P} = (\mathbf{p}_1|\mathbf{p}_2|...|\mathbf{p}_t)$. Actually, each eigenvalue λ_i corresponds to the variance of the data in the direction $\mathbf{p}_i$. Thus, the value of t is chosen by picking the t largest eigenvalues that account for a certain proportion r_v of the total variance. t is referred to as the number of *modes* and relates to the main shape deformations across the training image dataset.

$$\sum_{i=1}^{t}\lambda_i \geqslant r_v \sum_{i=1}^{2n}\lambda_i \tag{1.4}$$

r_v is usually set between 95% and 98 % ([24], [119]) . Now we can approximate any shape $\mathbf{x}$ by:

$$\mathbf{x} \approx \bar{\mathbf{x}} + \mathbf{Pb} \tag{1.5}$$

with $\mathbf{b}$ a t dimensional vector containing the model parameters/weights given by:

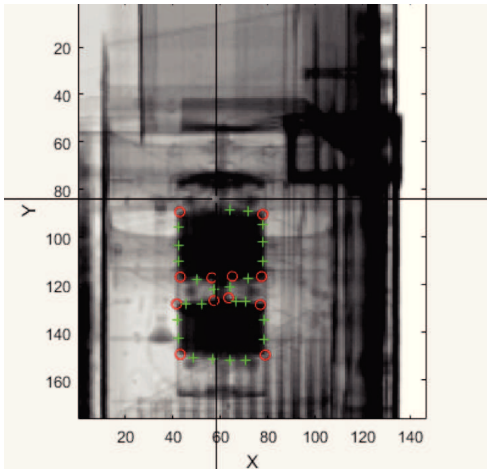
$$\mathbf{b} = \mathbf{P}^T(\mathbf{x} - \bar{\mathbf{x}}) \tag{1.6}$$

Indeed, the $\mathbf{b}$ vector gives a lower dimensional representation of the shapes ($t \ll 2n$). Specifically, it measures how much the layout of the points deviates from the average shape.

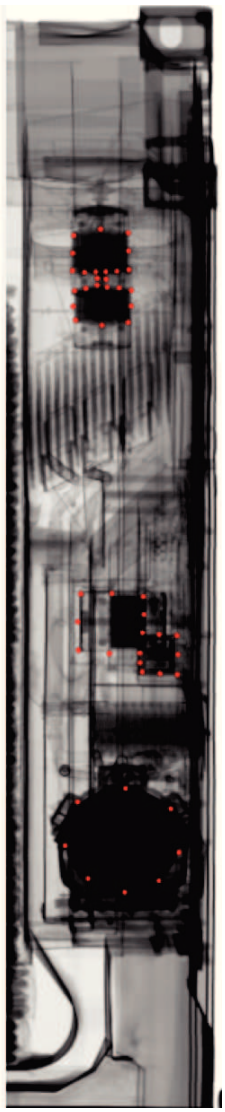
1.4 Gray-level Appearance and Model Fitting

1.4.1 Classical Method

The training set is also employed to build a statistical gray-level model. At each landmark, we compute the normalized derivatives of the intensity values along the normal to the shape ([27]).



(a)



(b)

Figure 1.5 – (a) Ongoing landmark manual annotation example on the refrigeration unit of an ISO container. Red circles and green crosses respectively designate annotated landmarks and intermediate points; (b) Fully annotated shape on the refrigeration module image.

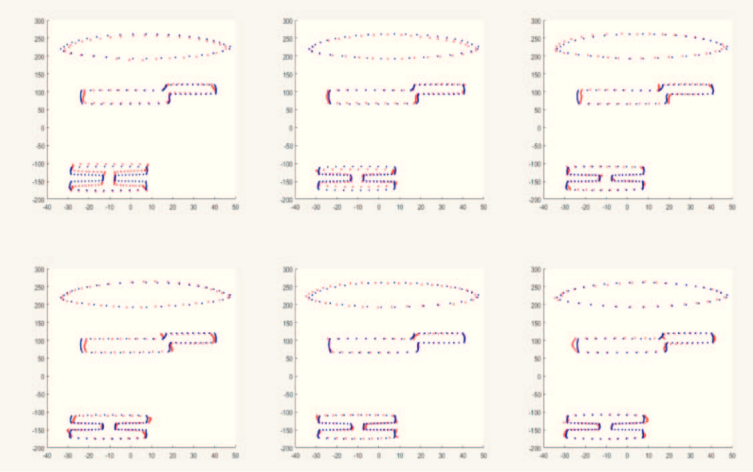


Figure 1.6 – Representation of six different modes of annotated shapes from Figure 1.5 following Procrustes Analysis. The mean shape is described in blue, while the modes are represented in red.

k pixels are sampled on either side of every cue points to yield vectors $\mathbf{g}_i$ of length $2k+1$. The profiles generation process is described on Figure 1.8,

For a specific model point indexed $i \in [\![1, n]\!]$, we consider the mean profile $\bar{\mathbf{g}}$ across the s training instances, as well as the covariance S_g. While applying ASM, the mean shape $\bar{\mathbf{g}}$ first undergoes the inverse alignment transform (Procrustes) and is set on the input image with initialization conditions. For the shape matching process Cootes and Taylor [26] propose to minimize the Mahalanobis distance between a new profile $\mathbf{g}_l$ and the profile model:

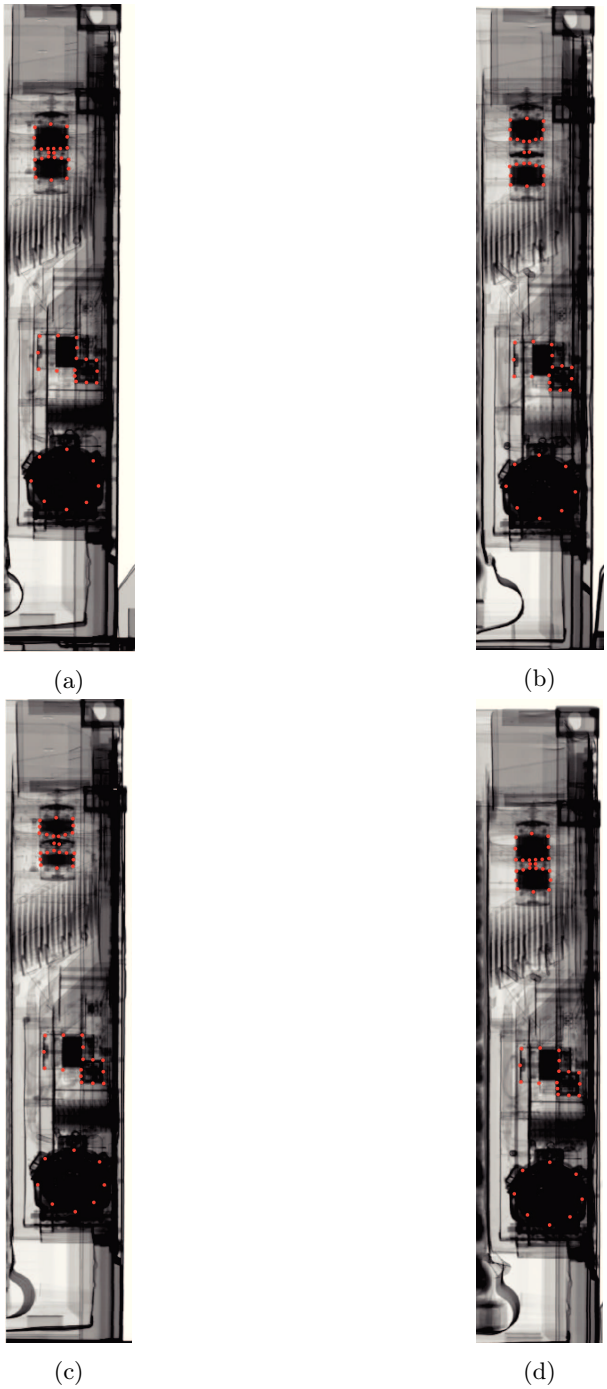
$$f(\mathbf{g}_l) = (\mathbf{g}_l - \bar{\mathbf{g}})^T S_g^{-1} (\mathbf{g}_l - \bar{\mathbf{g}}) \tag{1.7}$$

At a given iteration and for each landmark, a number n_s of profiles are computed along the perpendicular direction to the shape. The model point is then moved to the center of the profile minimizing the energy (1.7).

Besides, the plausibility of the resulting shapes is ensured by a constraint directly formulated on the parameters vector $\mathbf{b}$. Its values are generally restricted within the range $\pm m\sqrt{\lambda_i}$ with $m \in \{2, 3\}$ ([26], [24], [119]).

1.4.2 Coarse-to-Fine Scheme

Generally, a multi-resolution framework is adopted to improve the robustness and efficiency of the algorithm ([24], [23]). In both training and matching stages, profiles and appearance models are computed at different scales. The shape is first fitted to the input image at a lower scale to obtain a coarse segmentation approximation on global image structures. This estimate is then used as an accurate initialization for the finer resolution where refinement is performed.
Also, this strategy proves to be useful for shape initialization: the set of points is actually matched at a coarse resolution first, making poor starting estimates acceptable.



(a) (b)

(c) (d)

Figure 1.7 – Shape annotations of different samples from the same category of objects. Significant shape variations are raised between the training samples, especially at the top part.

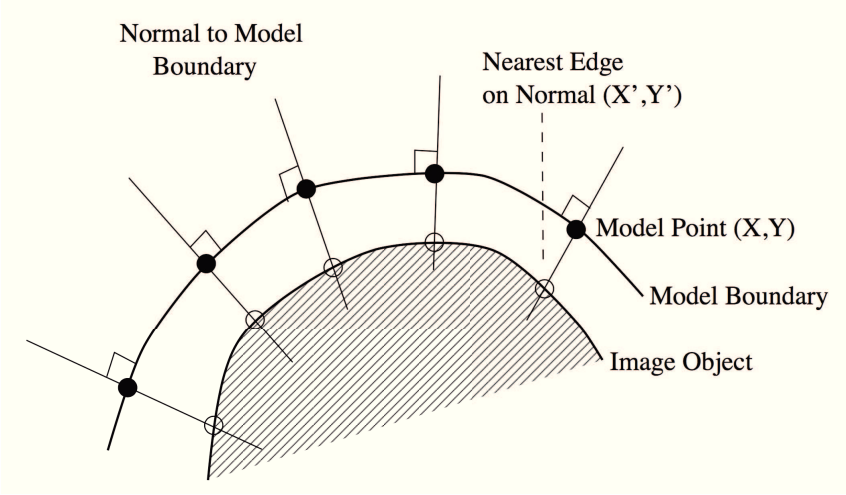


Figure 1.8 – Normal gray-level profiles description ([24]).

1.4.3 Slightly Different Approach

Yet, by choosing the Mahalanobis distance (1.7) as objective function, we assume a normal distribution of appearance profiles ([24], [119]). Finding the minimum of $f(\mathbf{g}_l)$ is then equivalent to maximizing the probability that $\mathbf{g}_l$ comes from a multivariate Gaussian distribution.
Though, this might be erroneous in cases where objects surrounding the shapes of interest vary throughout the training set ([119]), which is very likely in our field of study.

We therefore turn to a slightly different approach for the model fitting step. Alternative methods overcome the issue by extracting image features and using a kNN classifier ([119]). In our works, an appearance model is constructed by calculating the PCA representation of the gray-level profiles for every single landmark. The profiles along the contour normals boil down to the intensity values rather than the normalized derivatives vectors in [24]. Let us now consider, for each separate model point, the profiles mean denoted by $\bar{\mathbf{x}}_g$. Any appearance profile $\mathbf{x}_g$ can now be approximated as follows:

$$\mathbf{x}_g \approx \bar{\mathbf{x}}_g + \mathbf{P}_g \mathbf{b}_g \tag{1.8}$$

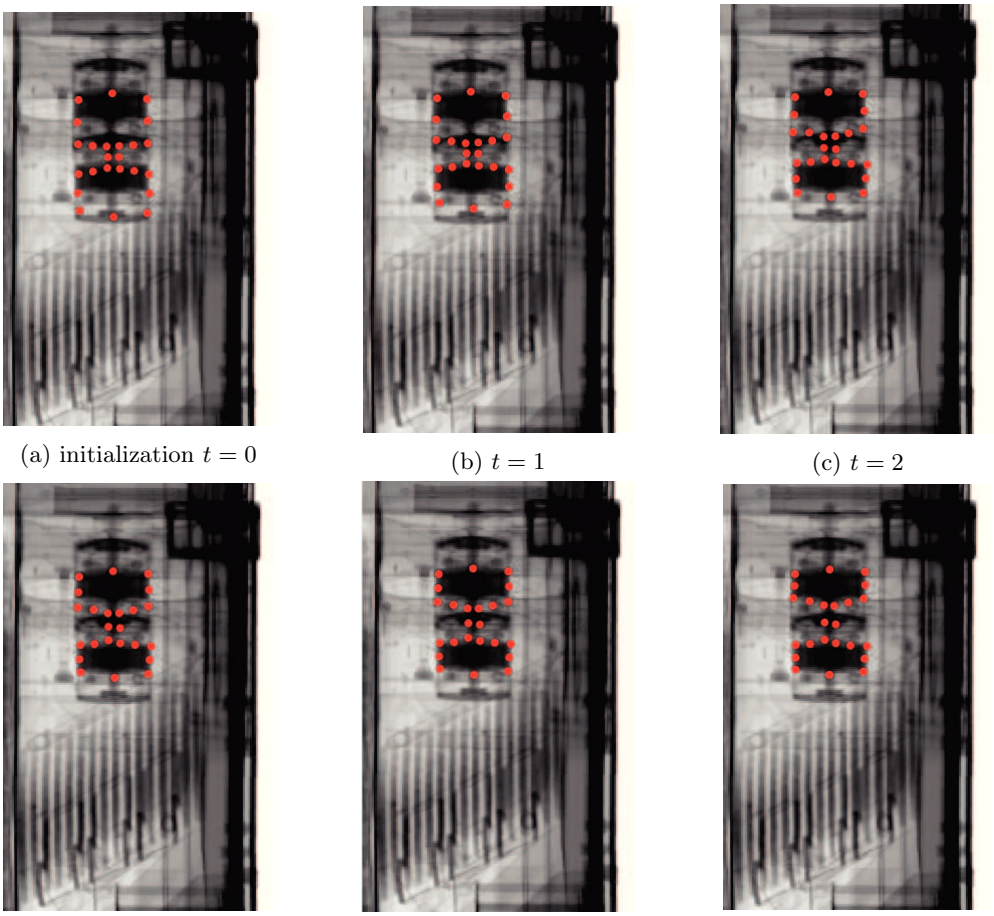
While fitting the model to a new image, n_s different profiles of length $2k+1$ are computed for every single point. Hence, we obtain:

$$\mathbf{b}_g = \mathbf{P}_g^T (\mathbf{x}_g - \bar{\mathbf{x}}_g) \tag{1.9}$$

The optimal position of a landmark minimizes the deviation from the mean profile. In other words, the new position is given by the center of the profile that gives the minimal $\mathcal{L}_2$ norm: $\|\mathbf{b}_g\|$. An example describing a few steps of the shape fitting process is displayed on Figure 1.9.

After convergence, a vector $\mathbf{v}$ stacking the final minimal $\|\mathbf{b}_g\|$ for each landmark is obtained. For each indexed landmark $i \in [\![1, n]\!]$, by noting $\left\|\mathbf{b}_g^{i\,*}\right\|$ the minimum cost, we get:

$$\mathbf{v} = \left(\left\|\mathbf{b}_g^{1\,*}\right\|, \left\|\mathbf{b}_g^{2\,*}\right\|, ..., \left\|\mathbf{b}_g^{n\,*}\right\|\right) \tag{1.10}$$



(a) initialization $t = 0$ (b) $t = 1$ (c) $t = 2$

(d) $t = 3$ (e) $t = 4$ (f) $t = 5$

Figure 1.9 – Description of the shape fitting process. A few iterations are needed to converge the set of points to the desired shape (takes less than a second).

Somehow, this method may seem quite similar to the AAM paradigm [25]. It is indeed a simplified version of the AAM where a PCA is also computed on a joint representation of the shape and appearance information. For further explanations, refer to [23].

Empirically, using this method gives better results on our dataset in comparison with the Mahalanobis cost function (1.7) based on the first derivative profiles.

1.5 Our Shape Recognition Strategy

1.5.1 Closest Shape Retrieval

As mentioned earlier, our main interest in using ASM is to address the shape variability issue within a considered category of images (see Figure 1.7 for an illustration). A new input sample is initially matched to a corresponding class. Among the set of instances defining this category, the objective is to find the closest image in terms of shape. As a consequence, difference analysis

by registering both images would be eventually made possible.

Let us consider a large set of images partitioned into L categories $(C_1, C_2, ..., C_L)$. On each subset C_l, $l \in [\![1, L]\!]$, images are annotated through n_l points, an ASM is trained whilst obtaining specific shape and appearance models. In a further step, the shape model is fitted to all training images and the shape vectors $\mathbf{b}$ are extracted from every single instance (*offline* process).

When a new image I_{new} is analyzed, it is first classified to a corresponding matching class C_M composed of q images. Subsequently, the ASM is fitted to I_{new} via shape and appearance models specifically trained on C_M. The weights vector $\mathbf{b}_{new}$ is then computed and compared to the bank of vectors $(\mathbf{b}_1^M, \mathbf{b}_2^M, ..., \mathbf{b}_q^M)$ previously fetched in C_M.

The closest image to I_{new} is then given by the minimization of the euclidean distance:

$$i^* = \underset{i \in [\![1,q]\!]}{\arg\min} \|\mathbf{b}_i - \mathbf{b}_{new}\| \tag{1.11}$$

In Figure 1.10, a detailed description of the closest shape recognition strategy is given.

1.5.2 Classification

One could easily think about resorting to the ASM for classification purposes. Given a set of shape models of different classes $(C_1, C_2, ..., C_L)$, the idea would be to apply this "bank of models" on a new instance I_{new}. The shape model resulting in the lowest $\|\mathbf{b}\|$ and $\|\mathbf{b}_g\|$ would indicate the category I_{new} belongs to.

This procedure was first adopted and then quickly abandoned. Since all shapes must be fitted to yield a classification decision, it was significantly time-consuming (especially for highly partitioned data-sets, more details are provided in Chapter 5). Hence, other classification strategies are employed, as developed in the next chapter.

Still, ASM may be used as a tool to "double-check" classification results. This applies in situations where small training sets are available, on which state-of-the art classifiers are prone to lead to poorer performances.

The final appearance descriptor $\mathbf{v}$ brought in (5.8) is obtained after convergence; as well as the corresponding shape vector of weights $\mathbf{b}$. We define a scalar $\hat{b} = \frac{\|\mathbf{b}\|}{t}$, with t the number of modes, *i.e* $\mathbf{b}$'s dimension.

These features are stacked in the following vector:

$$\mathbf{f} = \begin{pmatrix} \bar{\mathbf{v}} \\ \hat{b} \end{pmatrix} \tag{1.12}$$

with $\bar{\mathbf{v}}$ referring to the mean of $\mathbf{v}$. This vector controls whether the current fitted set of points significantly strives from the appearance and shape model. Whenever $\|\mathbf{f}\| > \upsilon$, the classification is considered as erroneous, where υ stands for a pre-set threshold.

1.6 Conclusion

In this chapter, we introduced the use of ASM techniques applied on cargo X-ray image processing problems. The mathematics of ASM are provided for both training and fitting stages. As mentioned, we used ASM for two distinct objectives. Given an input sample, we proposed our

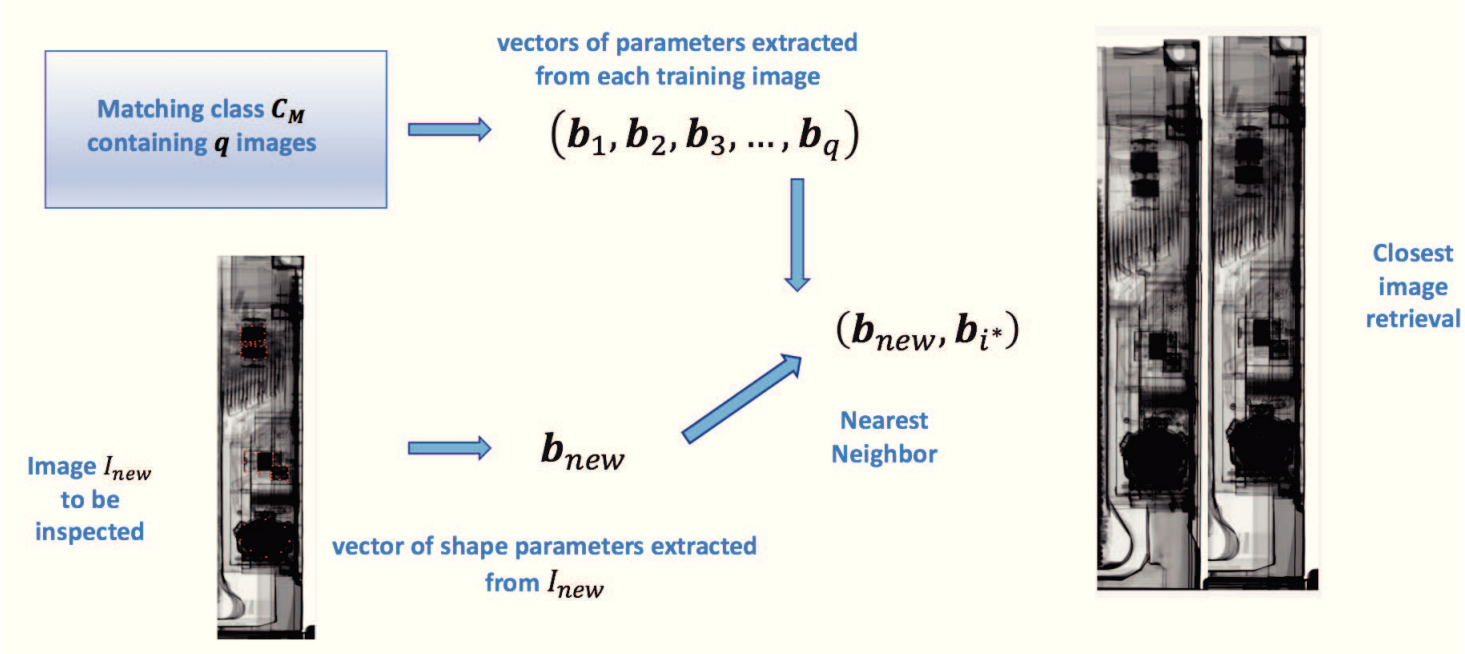


Figure 1.10 – Closest shape recognition scheme.

.

ASM-based strategy to find the closest image in terms of shape. Moreover, we introduced a cost combining both appearance and shape data at convergence. As a follow-up to traditional classification, it is indeed a smart way to use ASM for a double-check process.

At the end of the day, during training, shape and appearance models are built. At the same time, a table stacking all training X-ray images, their corresponding fitted shape points coordinates and b vectors, is also obtained. While shape and appearance data are employed for shape matching, training data is dedicated to the closest image retrieval application.

For more thorough details about speed, convergence issues or shape initialization the reader is urged to refer to Chapter 5 where the method is directly outlined within its applicative context.

Chapter 2

Descriptors and Classification Methods

Outline of the current chapter

2.1 Introduction

Representing an image by its pixel-wise gray-levels or RGB intensities is far from being optimal as far as classification issues and template-matching applications are concerned. Indeed, for a few decades, a substantial part of computer vision research has been dedicated to designing techniques to represent images in reduced dimensions.

Well-known dense and sparse descriptors have been developed, ranging from HoG, to SIFT or SURF features that seek to characterize a specific appearance and shape signature of given objects. However, in the dynamics of deep-learning methods and their major achievements, "hand-crafted" approaches are progressively being upstaged in favor of automatically-learned image descriptors.

In this chapter, we will briefly summarize the different descriptors used within the framework

of our research. The supervised-classification methods employed are also detailed. This subject intends to draw up a list of the the different image representation and classification tools employed in this work rather than reporting a contribution of our own.

2.2 Traditional Descriptors

2.2.1 Histogram of Gradients

The Histogram of Gradients, commonly referred to as HoG is a features descriptor that relies on first-order derivatives of images (magnitude and orientation). Introduced by Dalal and Triggs in 2005 [31], it stems from the hypothesis that local edge direction may be sufficient to characterize the local appearance of an object. Roughly, the approach consists in calculating occurrences of gradient orientations in small spatial regions ("cells"). This "dense" technique has proven to be quite efficient for object detection tasks (*e.g.* pedestrians detection in [108] or see Figure 2.2). It is composed of three stages:

1. **Gradients computation**: both horizontal and vertical gradients are calculated by applying the 1-D centered point discrete derivative mask $[-1, 0, 1]$.

2. **Histograms construction**: within a cell, each pixel votes for a given gradient magnitude and direction. Typically, the histograms are binned by 8 or 9 channels for each cell (Figure 2.1). Also, either signed or un-signed orientations (*resp.* ranges of 180° or 360°) may be used.

3. **Descriptor blocks**: local illumination and contrast variations are addressed by a normalization scheme over an agglomeration of connected cells forming a larger block. In this regard, different normalization schemes are studied in [31] (L_2-norm, L_1-norm, L_2-hys, L_1-sqrt).

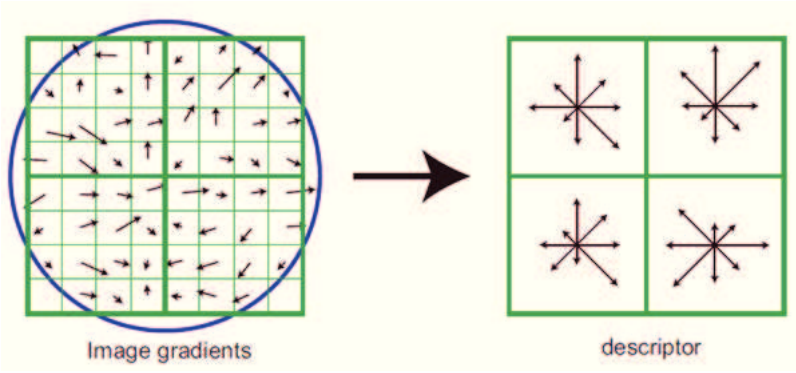
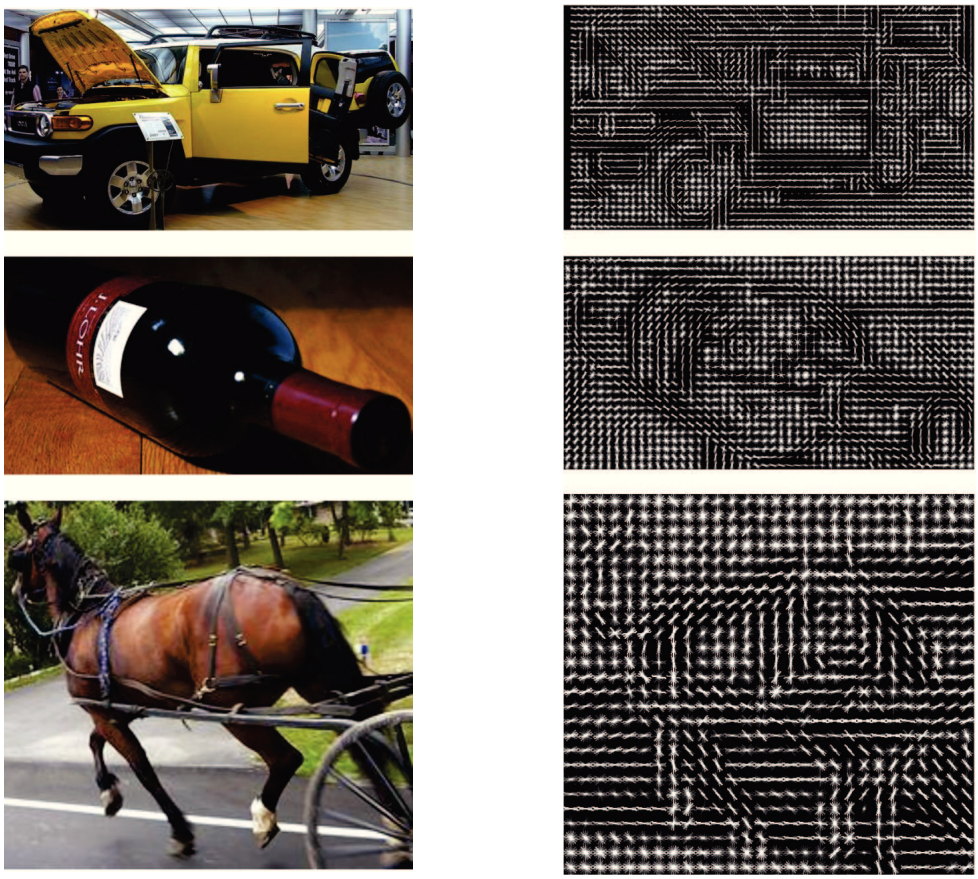


Figure 2.1 – Histograms of gradients extracted from different cells (8 channels)

.

Actually, HoG features descriptor is defined as the concatenation of the elements of the normalized cell histograms from each block. [31] distinguishes between two types of blocks: R-HoG (rectangular blocks) and C-HoG (circular blocks). We will only employ the R-HoG approach where cells per block and pixels per cell are specified as input parameters. The number



(a) Real images (b) HoG

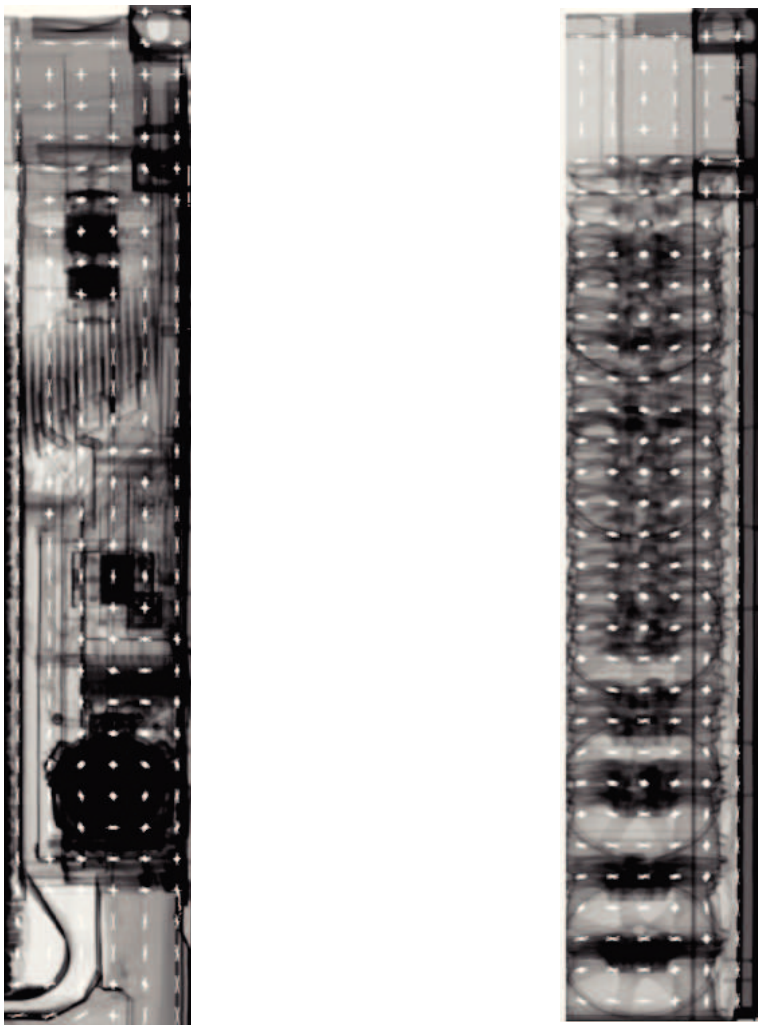
Figure 2.2 – Examples of extracted oriented histograms displayed for each defined cell on real images ([126]).

of overlapping cells between adjacent blocks may also be considered (we usually select at least half of the block size to ensure adequate contrast normalization).

Although the authors of [31] recommend to use an optimal set of parameters of their own for HoG extraction, we empirically fine-tuned these values regarding the specificity of our task described in chapter 5. In fact, for X-ray images, contrast and illumination changes are less prone to happen. Furthermore, the dimension of the HoG features vector constitutes a defining constraint; precisely while training a supervised classifier (SVM) on a limited amount of data. Note that a HoG-PCA approach has also been adopted to this end.

For these reasons, cells of 22×22 pixels were chosen (instead of 8×8 as advised by [31]) with blocks constructed from 2×2 cells. In section 5.3.3, HoG features are used to discriminate refrigeration units from regular freight (as in Figure 2.3).

Eventually, HoG descriptors have been extended to video in [32]. A 3-D adaptation was also initially proposed by Krückhans in [59]. Alternative faster methods using HoG descriptors combined with Adaboost algorithm were proposed in the famous works of [135] that are beyond the scope of the present study.



(a) (b)

Figure 2.3 – (a) HoG visualization on a HCVS image displaying a refrigeration unit; (b) HoG visualization on a HCVS image displaying regular freight.

2.2.2 SIFT

The Scale-Invariant Feature Transform (SIFT) proposed by David G. Lowe in [67] consists in detecting and describing sparse *keypoints* of an image. These interest points are defined as salient regions that can be detected despite viewpoint, illumination, scale or orientation changes. To this end, a particular emphasis on repeatability has been introduced in the method. It also shows high distinctiveness properties. Indeed, each keypoint is described by a 128-element SIFT-descriptor vector such that similar points have very close descriptors whereas different points are characterized by a significant distance between their vectors of features.
These keypoints and their descriptors are of high interest in the present work where matching problems must be considered to perform registration.
As mentioned above, the technique is composed of two stages:

Points of interest identification

Based upon the result of [64] , Gaussian kernel is used to create the scale-space. For computational considerations, the Difference of Gaussians (DoG) is preferred to the Laplacian of Gaussians (LoG). Keypoints are localized by taking the maxima/minima of the DoG at different scales. In further detail:

1. Two Gaussian masks, at scales σ and $k\sigma$ are considered: $G(x, y, \sigma)$ and $G(x, y, k\sigma)$.
2. Given $I(x, y)$ an input image, the DoG is obtained by computing:

$$D(x, y, \sigma) = (G(x, y, \sigma) - G(x, y, k\sigma)) * I(x, y)) = L(x, y, \sigma) - L(x, y, k\sigma) \tag{2.1}$$

3. Each pixel is compared to its 26 immediate cross-scale neighbors in terms of $D(x, y, \sigma)$ values (see Figure 2.4). Any pixel found to be an extremum is selected as a candidate keypoint.

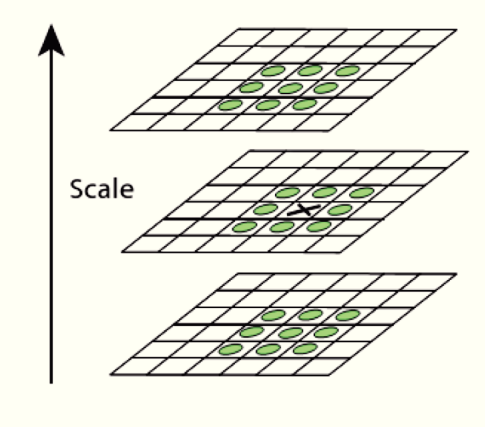


Figure 2.4 – Cross-scale neighbors of a given pixel.

.

4. Eventually, sub-pixel localization, low contrast points filtering and edge response elimination are employed to yield a more accurate keypoint detection.

Points of interest description

For each keypoint, a descriptor vector is obtained in two steps:

1. Orientation assignment: The gradient magnitude and orientation θ is computed for each blurred image $L(x, y, \sigma)$. In addition, for every point of interest located at (x, y), a window of size 1.5σ centered in (x, y) is defined. The window is rotated according to the computed $\theta(x, y)$.

2. Descriptor building: we divide the rotated window into $4 \times 4 = 16$ sub-regions. On each cell, an 8-bin histogram of gradients is created. The 16 histograms are concatenated into a $16 \times 8 = 128$ elements vector which is ultimately normalized (see Figure 2.5).

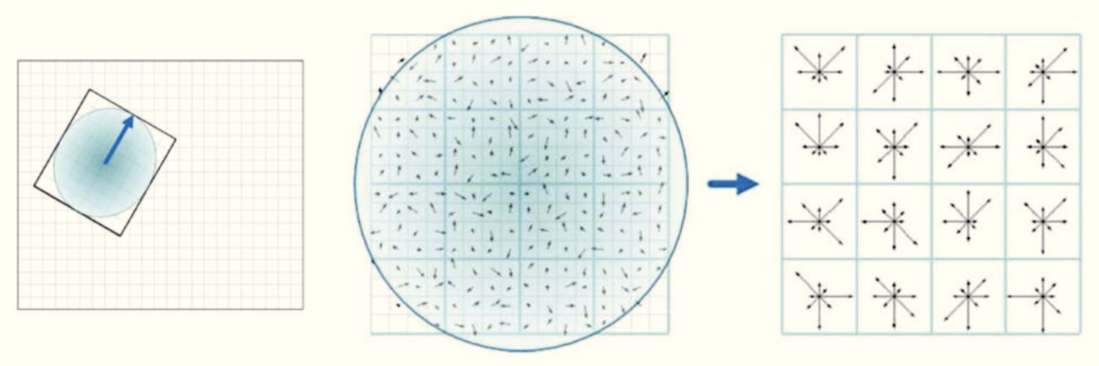


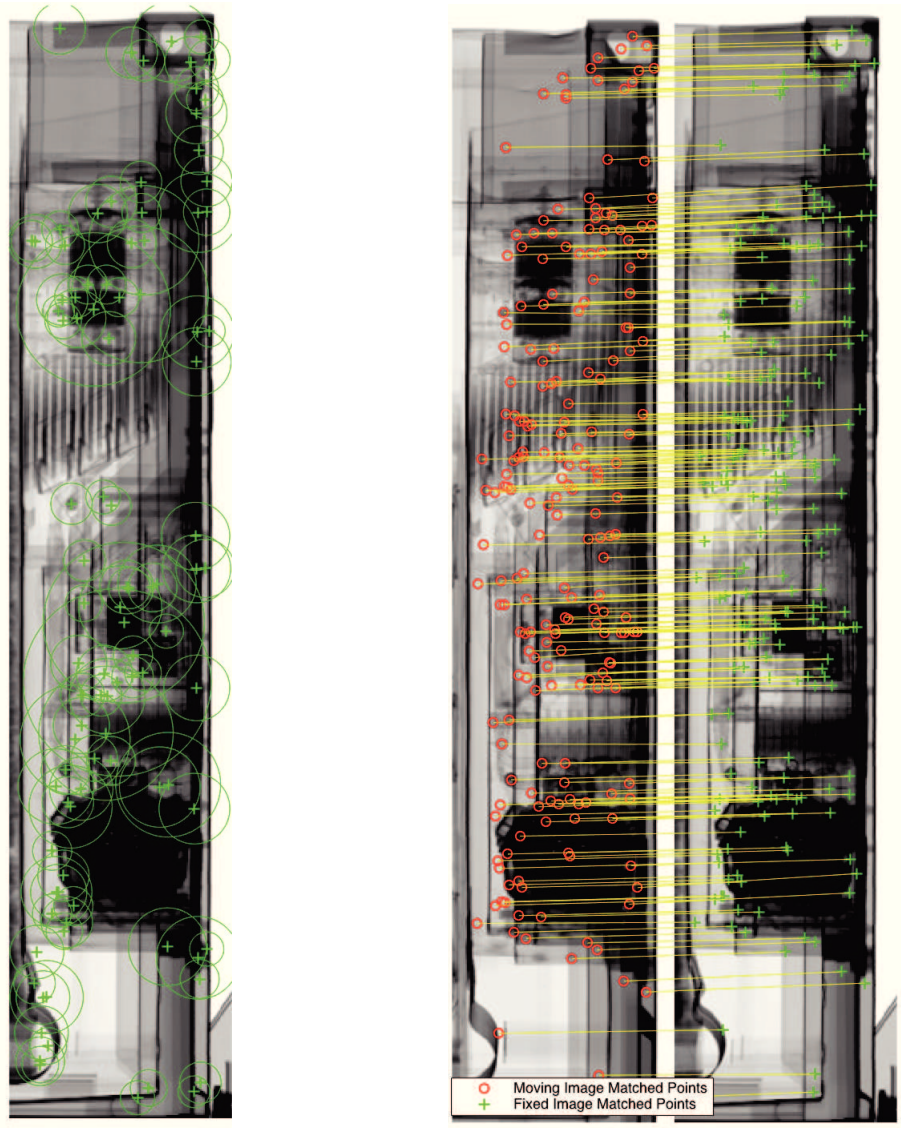
Figure 2.5 – Building a SIFT descriptor, from left to right: keypoint scale and rotation, image gradient and final keypoint descriptor (from [100]).

As a consequence, SIFT descriptor is invariant to scale and rotation thanks to the chosen window size and rotation update by the gradient orientation in step 1. It is also robust to illumination changes due to the normalization performed in the second step.
Note that HoG descriptors are quite similar to SIFT. They differ in that they are usually computed on a dense grid (without scale or orientation alignment), yielding a global image descriptor whereas SIFT features are computed on sparse salient keypoints.

The most famous improvement of SIFT known as Speeded Up Robust Features (SURF) was presented in [11]. It aims at improving speed and robustness of both detection and description stages. Instead of using DoG, Bay et *al.* resort to square-shaped filters to approximate Gaussian filters in combination with the integral image of I. Points of interest are found by computing the determinant of the hessian matrix for each location (x, y, σ) of $L(x, y, \sigma)$ and searching for local extrema.

Haar wavelet responses are used for the orientation assignment step as well as the descriptor construction. See [11] for more thorough details on the SURF approach.
The authors conclude that their method achieves a substantial dimension and complexity reduction of the descriptors without affecting discrimination and point-matching robustness/accuracy.

The viewpoint invariance property of these descriptors is actually crucial for many applications in large-scale X-ray imaging. Both scanning machines or vehicles are prone to different



(a) (b)

Figure 2.6 – (a) SURF keypoints detection on HCVS refrigeration unit scan (100 strongest points displayed); (b) SURF feature points matching between two different scans, following MLESAC filtering ([116]).

types of displacement, *e.g.* such as rotations around the z axis or translations. We will show later that small translations could lead to significant viewpoint changes in case of conic X-ray beam (top-view scanning modality, see Chapter 6). One could also be interested in matching images acquired by different systems and geometries. Likewise, more complex deformations may also occur in the resulting images, hence robust detectors and descriptors are essential for matching or recognition tasks. SURF features are employed in chapter 5 for our piece-wise linear registration scheme involving refrigeration units scans (in Figure 2.6, a SURF-based global registration illustration is provided). In chapter 6, SURF features are used for a rigid pre-registration step between top-view scans of vehicles.

Alternative famous "hand-crafted" methods have been developed: refer to FAST ([92]) or ORB ([93]) for instance.

2.3 Pretrained Convolutional Neural Networks

This section briefly reviews the principles of Convolutional Neural Networks (CNNs) and describes how their architectures could be smartly utilized for global feature descriptors extraction. It does not claim to present a comprehensive overview on ConvNets, but rather a general outline.

2.3.1 Convolutional Neural Networks - CNNs

As remarked in the introduction to this chapter, traditional machine learning techniques usually employed a dual-stage approach for classification tasks. Hand-engineered features were first extracted to represent the input image into lower dimensions and designate salient elements. Thus, either global descriptors such as Local Binary Patterns (LBP, [79]), HoG [31] or local descriptors such as SIFT [67], SURF [11] or ORB [93] were paired to a Bag-of-visual-Words (BoW) representation [30]. Throughout a second step, the extracted features served as input to trainable classifiers. Yet, at the same time, classification performances relied utterly on the quality of the designed descriptors and the learning algorithms [103].

With the progressive emergence of deep-learning methods, handcrafted feature descriptors are being replaced by learned filters that automatically update to yield an optimal classifier. The fact that no prior knowledge or human effort are required for feature design constitutes a major advantage.

Historical background

As for regular neural networks, CNNs (or ConvNets) architectures were motivated by the early works of two neurobiologists: Hubel and Wiesel in the 1960s [53]. Their main discovery showed that in early stages of the primary visual cortex, specific directed edges were given a strong response by neurons while more complex patterns were addressed in a later stage. They also identified two types of cells in the visual cortex: simple cells with local receptive fields reacting to straight edges in certain orientations, and more complex cells with larger receptive fields, insensitive to input distortion or translation. Both types were ordered in a hierarchical fashion. On the basis of these observations, Fukushima introduced a multi-layered model called the neocognitron in 1980 [39]. In spite of not resorting to back-propagation methods, it constitutes the genuine cornerstone to ConvNets models [60].

Yann Lecun *et al.* proposed in 1990 the first, fully automatic CNN for hand-written numbers recognition on bank checks or zip codes [62]. It was later extended to character recognition for document analysis in [61]: the 5 layers ConvNet LeNet-5 (Figure 2.7).

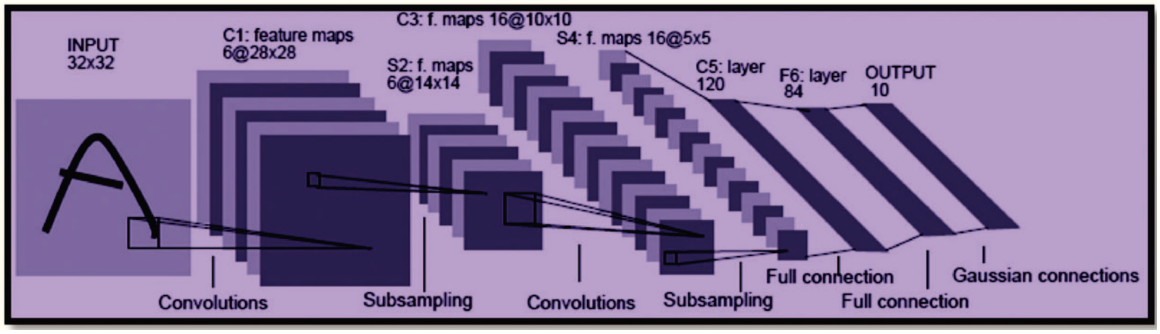


Figure 2.7 – Pipeline of LeNet-5 convolutional neural network (from [61]).

Research on CNNs was significantly diminished in the first decade of the 21^{th} century until its resurgence: the paper of Krizhevsky *et al.* ([58]) in 2012 is considered as a major breakthrough in the field. They reported an error rate brought down by half on a 1000-labels classification task (ImageNet LSVRC-2010 contest). In fact, this work literally initiated the recent surge to CNNs, fueled by optimized GPU computations ([22]), improved algorithms and wider access to huge labelled data-sets [86], [60].

CNN architecture general overview

Generally speaking, a ConvNet architecture receives a multiple array image (typically RGB) I which is processed through a series of hidden stages to report a classification output. I is commonly referred to as an input volume. In line with natural signals properties, four main types of layers are to be pointed out:

1. Convolutional layer: In these layers, units (or neurons) are arranged in feature maps. Each neuron is being connected to a patch of the feature maps from the previous layer through the application of a set of weights (filter bank) [60]. The use of convolutional filters controlled by a filter size (receptive field hyperparameter) ensures the spatial local connectivity between neurons of adjacent layers [103]. Note also that the same local filter bank is applied on the different locations of the image, it is known as the weight-sharing property [86], [103]. Besides, spatial arrangement is controlled by three hyperparameters: depth (number of filters), stride (convolution filters overlapping) as well as zero-padding [58].

2. Nonlinear activation layer: This layer allows the extraction of non-linear features rather than obtaining "less-expressive" linear systems ([103]). Typically, sigmoid or hyperbolic tangent were used. As reported by [58], employing the Rectified Linear Units (ReLUs) (2.2) on convolutional layers outputs shows a neat advantage in terms of learning speed (see Figure 2.9). Novel nonlinear activations *e.g.* RReLU or SReLU are widely explored in current research [86]. Given an input x, the ReLU activation function is defined by:

$$f(x) = max(0, x) \tag{2.2}$$

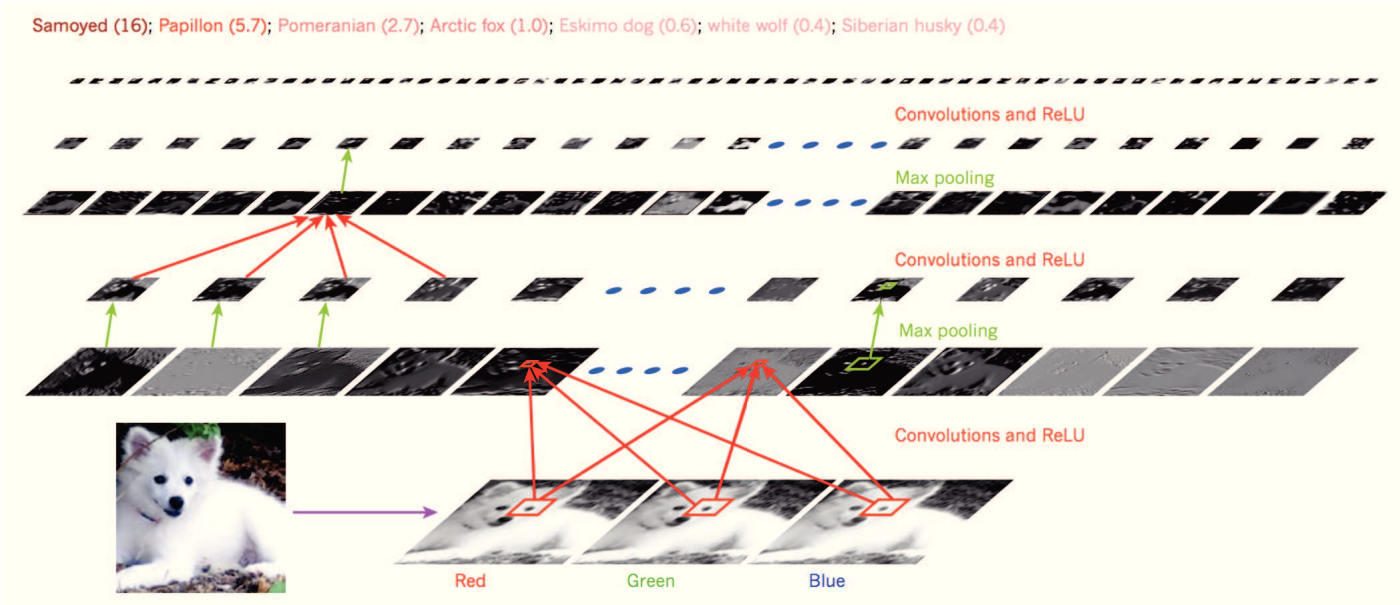


Figure 2.8 – Typical Convolutional Network architecture applied to the RGB image of a Samoyed dog. Each rectangular image is a feature map corresponding to the output for one of the learned features, detected at each of the image positions. On the top, the different possible labels with their respective scores. The image is classified as a samoyed since it gets the highest output score [60].

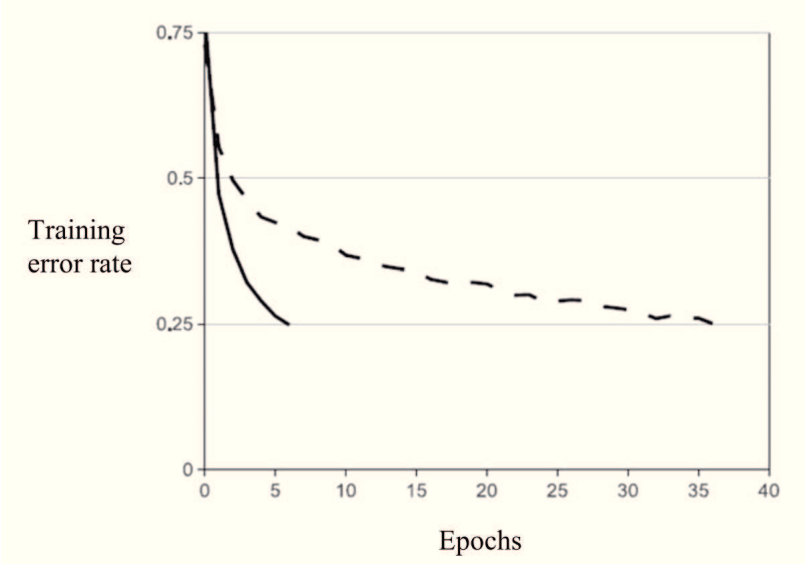


Figure 2.9 – Training times of ReLUs versus tanh activations ([58]).

3. Pooling layer: Aiming at spatial invariance to image distortion and translation (as observed in the visual cortex by [53]), the resolution of feature maps is reduced by pooling layers. Concomitantly, it entails lowering the computational cost for the weights training in deeper layers. While average pooling aggregation was initially employed [62], max pooling is very popular [86], [103]. Within a receptive field, semantically similar values are merged into the maximum value and propagated to the next layer (see Figure 2.8). Given a receptive field $\Re_{ij}$ located around the position (i, j), Y_{kij} is the pooling output corresponding to the k^{th} feature map with x_{kpq} input values:

$$Y_{kij} = \max_{(p,q)\in\Re_{ij}} x_{kpq} \tag{2.3}$$

4. Fully-connected layer: Following several stacked convolutional and pooling layers, fully-connected layers are added on top to perform high-level reasoning. Neurons are then connected to all activations in the previous layer. The softmax operator is commonly used for classification problems [86].

This architecture actually detects global features such as oriented edges in lower-levels. In deeper levels, the layers produce more complex features since they combine simpler responses from lower levels. They are dedicated to learn high-level visually semantic content in the image [60], [133]. This specific feature is depicted on Figure 2.10.

Learning algorithm

During learning, the weights are actually adjusted by the famous backpropagation algorithm based on the combination of the chain rule alongside a Stochastic Gradient Descent method (SGD) [61], [60],[103], [86].

The gradient of a formulated loss function is computed in a backward pass in order to update the weights such that errors are minimized (in the opposite direction to the gradient). Compre-

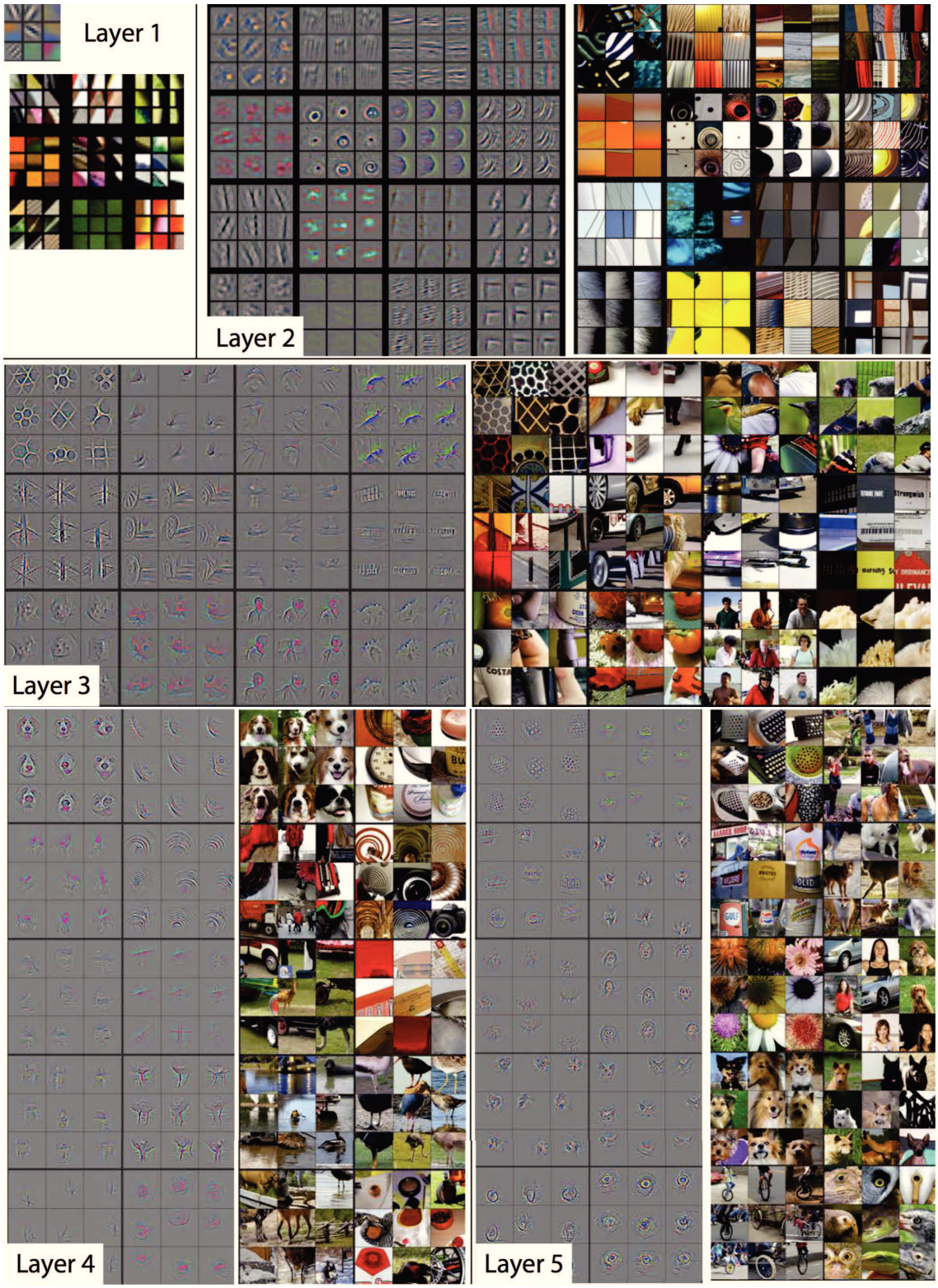


Figure 2.10 – Visualization of bottom to top levels activations. As described, responses become more and more abstract as layers get deeper (from [133]).

hensive explanations on the backpropagation gradient descent algorithm are provided in [60]. In order to address overfitting issues, ingenious regularization methods have been developed such as dropout, dropconnect or stochastic pooling. These details are beyond the scope of the present work. For further information, the reader is urged to refer to the papers of [86] or [103].

CNN-derived approaches

Besides classification tasks where a huge labelled dataset is used in several forward and backward stages (epochs) to train a ConvNet, different variations have gained a lot of popularity. Without concern for completeness, among well-known methods, the Region-based Convolutional Neural Networks (R-CNNs) [40] performs object detection. Its improved version, developed in Microsoft Research labs: Faster R-CNN [89] introduces a Region Proposal Network (RPN) that enables nearly cost-free region proposal. See also the impressive speed performances provided by the You Only Look Once (YOLO) framework [87], [88]. For an illustration, see Figure 2.11.

Figure 2.11 – Very diverse types of classes recognized by the YOLO9000 framework ([88]).

Generative Adversarial Networks (GANs) [43] introduced by Ian J. Goodfellow *et al.* have

recently drawn the attention of the computer vision community.
In this approach, the generative model G competes against an adversary A: for each category, G synthesizes artificial negative images while a discriminative model A learns to determine whether a sample hails from the model distribution or from real data. Both models are therefore refined until it is almost impossible to distinguish between the "counterfeits" and genuine positive samples. See [97] for the most up-to-date achievements on the subject.
The method constitutes a major breakthrough in the sense that, regarding binary classification, negative labelled samples are synthesized, and are thereby no longer required to train the network.

2.3.2 Pretrained CNN Features Extractor

A characteristic of deep learning is that it builds representations of the data. These representations tend to have a universal value, or at least to be applicable to an array of problems that transcends the specific application a model was trained for. This is fortunate as training complex models requires weeks of works on one or more GPUs; these models can then be "frozen" and re-employed for a number of additional tasks, with no or minimal additional work. Moreover, the collection of large training and testing image sets is a challenging task. Hence, resorting to pretrained ConvNet architectures represents a smart alternative to overcome these issues. Two methods are to be pointed out [103]:

- Transfer learning: Given a model, already trained towards image classification, the idea is to use its trained weights as an initialization for the new task and perform SGD again. The number n of layers required to be re-learnt depends mostly on the likeness between both current and pretrained networks tasks. Meaning that if both tasks display semantically close categories, then only a few layers will need to be trained anew. The first layers are traditionally kept unchanged whereas deeper layers, meant to learn visually semantic attributes, need to be re-learnt. Note that n is also estimated by taking the size of the training set into consideration: the more the data, the higher n can be.
  In addition, the method enables to pick a learning rate much lower than what was used in the original model. For an indepth review of the method, also known as fine-tuning, the reader may refer to [130].

- Extraction of features: as remarked earlier, the later layers in CNNs are crucial to catch high-level attributes of the visual content of an image. Hence, one could employ the activations available on the second-to-last fully connected layer of the network (i.e. before the final softmax classifier). These outputs stand for the feature vector supplied to a machine learning classifier which further learns to classify it. Paradigms using these features are deemed to report a better classification accuracy than those employing hand-crafted features such as SIFT, SURF or HoG ([34], [101], [50]).

  The idea, primarily proposed in [34] led the authors to adapt the technique to various visual recognition tasks, such as domain adaptation, fine-grained part-based recognition, and large scale scene recognition, all showing positive results. In a similar perspective, [9] puts forward the extraction of fully-connected layers outputs (referred to as "neural codes") for image retrieval. Figure 2.12 depicts the scheme used in their study. Razavian *et al.* extracted generic feature descriptors from the OverFeat architecture in [101]. After conducting various experiments, they prove the soundness of the method in several tasks. In 2015, [50] prefers fetching the outputs across all the network layers and stack them into a feature vector defined as a hypercolumn.

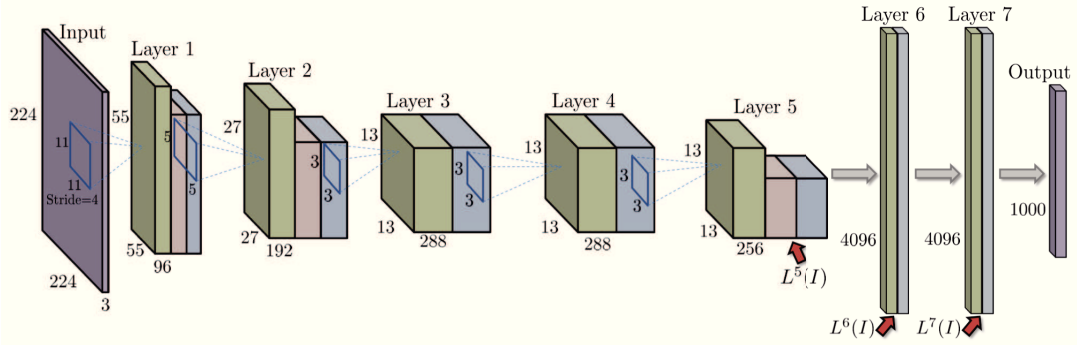


Figure 2.12 – The convolutional neural network architecture used for the experiments conducted in [9]. Purple nodes correspond to both input (RGB image of size 224×224) and output (1000 class labels). Green units correspond to outputs of convolutions, red units correspond to the outputs of max pooling, and blue units correspond to the ReLU activations. Layers 6, 7, and 8 (the output) are fully connected to the preceding layers. In this scheme, L^7 corresponds to the layer whose output is used for feature descriptor extraction (from [9]).

As far as our applications are concerned (see thorough explanations in chapter 5), data extraction yielded too limited amounts of samples to consider regular CNN training (despite relying on data augmentation - bootstrapping). The *imagenet-vgg-f* from [18] was used as pretrained model for features extraction. This architecture reports high performances on the ILSVRC 2012 challenge dataset. It is a 9 layers network composed of 5 convolutional layers paired with ReLU and pooling units followed by 3 fully-connected layers (ReLU + pooling) and the final softmax layer. This CNN takes RGB 224×224 input images so that our grayscale images are usually required to be resized and stacked into three channels.

Following the suggestions of [9] or [101], an additional principal component analysis compression was generally applied to reduce the dimension of the output feature descriptors.

2.4 K-Means

Throughout our work, we are often brought to deal with very large unlabelled image sets we seek to group into their respective categories. One of the simplest and most famous unsupervised classification approach is the K-means partitional clustering method, first introduced in [68] (also known as Lloyd's algorithm).

Given a set of N observations (feature vectors in our case) $\{\mathbf{x}_1, \mathbf{x}_2, ...\mathbf{x}_N\}$ of a random variable $\mathbf{x} \in \mathbb{R}^d$, the idea is to find a partition of these into K clusters. Each cluster is associated with a centroid $\mu_k \in \mathbb{R}^d, k = 1, ..., K$. The idea is to assign each data point $\mathbf{x}_n, k = 1, ..., N$ to the cluster k with the closest centroid μ_k by minimizing the intra-class variance.

We use the 1-of-K coding scheme notation as described in [14] by introducing binary indicators $r_{nk} \in \{0, 1\}$. If $\mathbf{x}_n$ is associated with the kth cluster, then $r_{nk} = 1$. Otherwise, for $j \neq k$ we will have $r_{nj} = 1$. The objective function is then expressed as the *Sum of Squared Differences* (*SSD*) between each data point being assigned to its cluster center μ_k :

$$J = \sum_{n=1}^{N} \sum_{k=1}^{K} r_{nk} \left\| \mathbf{x}_k - \mu_k \right\|^2 \tag{2.4}$$

The number of clusters K must be specified as an input. The algorithm is an iterative method composed of the following steps (Figure 2.13):

1. **Initialization**: cluster centroids are initialized in a random fashion.

2. **Clustering**: data points are affected to the closest centroid, thus forming clusters of observations.

$$r_{nk} = \begin{cases} 1 \text{ if } \arg\min_j \left\| \mathbf{x}_n - \mu_j \right\|^2 \\ 0 \text{ otherwise.} \end{cases} \tag{2.5}$$

3. **Centroids update**: the average of the observations within each cluster is computed to get new centroid locations.

$$\mu_k = \frac{\sum_{n=1}^{N} r_{nk} \mathbf{x}_k}{\sum_{n=1}^{N} r_{nk}} \tag{2.6}$$

4. Repeat 2. and 3. until cluster assignments remain the same or when the maximum number of iterations is reached.

Several improvements of K-means have been developed in order to address a couple of issues. One could name for instance *spherical* K-means or K-medoids [14], [56] that allow alternatives to the euclidean distance (*e.g* cosine, correlation ...). Concerning our task presented in section 5.4, we chose to work with the K-means ++ algorithm introduced by Arthur and Vassilvitskii in [4]. In short, they contribute in improving the solution optimality via a smarter initialization randomized seeding scheme. The results reported in their work show faster and more accurate clustering outputs. Indeed, their method is $O(\log k)$-competitive with the optimal clustering as proven in [4]. Likewise, for higher robustness, a few runs are usually performed and the result reporting the lowest sum of intra-class *SSD* is returned.

Besides, considerable literature has focused on automatically finding the number of clusters K [56]. Though, our use of clustering techniques is confined to data labelling with regard to a further supervised learning stage. As a matter of fact, K-means performs badly in complex data settings [56], hence, it cannot be used for a fully automatic annotation. In other words, K-means is only adopted to assist a manual labelling of the samples. Indeed, K is first estimated according to a preliminary rough appreciation of the operator (in the light of visual observations). A further labelling refinement is manually performed to get classes as homogeneous as possible. Very often, an extra K-means clustering is performed within the obtained categories with $K = 2$ or $K = 3$ to yield a more rigorous partition of the dataset (see section 5.4).

Especially, in line with each task specificity, observations $\mathbf{x}_n$ relate either to HoG or CNN-extracted feature vectors (global descriptors) jointly with a PCA pre-processing step for dimensionality reduction [110].

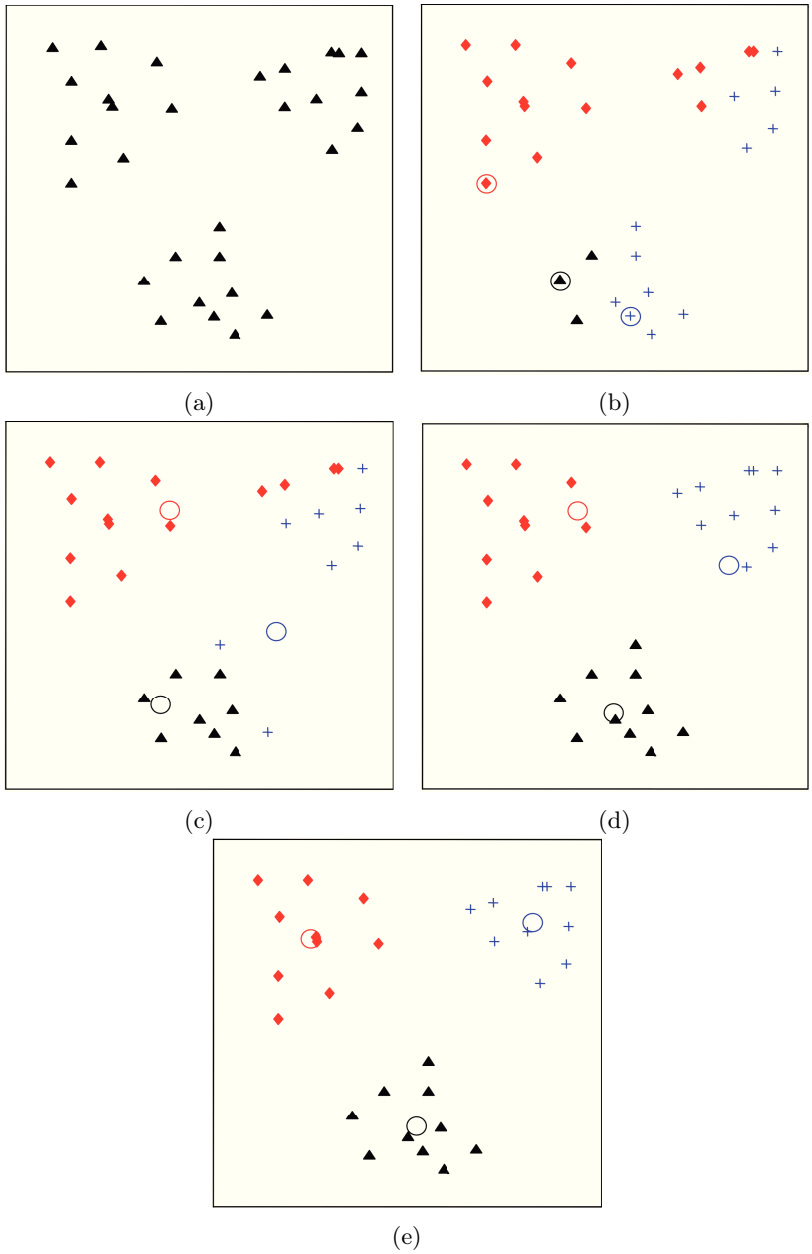


Figure 2.13 – Different stages of K-means algorithm. (a) input data with three obvious clusters; (b) initialization of three cluster centroids and data assignment ; (c) and (d) intermediate iterations describing labels and centroids updates; (e) final clustering output obtained by K-means algorithm at convergence ([56])

.

2.5 Support Vector Machines

The following section briefly reviews the main principles of the Support Vector Machines (SVM) classification method. Built on the seminal works on statistical learning theory of Vapnik *et al.* [120] in the sixties, the modern SVM version actually dates back to the nineties [121]. It became very popular because of its high performances in a very large range of fields.

The approach is indeed ubiquitous among the techniques developed in the framework of the present study. Its traditional version is dedicated to binary classification problems, where possible labels can either be -1 or 1 (or 0 and 1). In this chapter, we will also include a description and explanation of multi-labels extensions of the SVM method. For an indepth review on SVM, the reader is urged to study the comprehensive textbooks of [99] or [106].

2.5.1 Linear Support Vector Machines

Linear decision boundary

Let us first consider the most straightforward situation, i.e when the classifier is trained on separable data. Suppose that our training set consists of $\{\mathbf{x}_i, y_i\}$, $y_i \in \{-1, 1\}$ the assigned labels and the d-dimension feature vectors $\mathbf{x}_i \in \mathbb{R}^d$ for $i = 1, ..., l$, with l relating to the number of training inputs. A binary linear classifier would basically aim at finding a hyperplane that separates positive from negative samples.

Formally, given $\mathbf{w}$ a vector of weights, and b a scalar, an input feature vector $x \in \mathbb{R}^d$ is classified according to the sign of:

$$h(\mathbf{x}) = \mathbf{w}^T\mathbf{x} + b \tag{2.7}$$

Thus, the decision boundary is defined by a hyperplane such that $h(\mathbf{x}) = 0$. Hence, the separability conditions are formulated in the following way:

$$y_i(\mathbf{w}^T\mathbf{x} + b) \geqslant 0,\ i = 1, \ldots, l. \tag{2.8}$$

Maximal margin

Actually, a margin is associated to the separation line. It is defined by the distance between the separation line and the nearest positive and negative samples, referred to as "support vectors". Indeed, the decision boundary is chosen so that the margin is maximal. This choice is motivated by the *VC*-theory claiming that the largest capacity is achieved for a maximal margin [16]. Finding the optimal decision line amounts to solve a quadratic optimization problem using well-known algorithms.

By calculating the projection of a training sample $\mathbf{x}_i$ on the hyperplane, the linear decision boundary is then given by $\mathbf{w}$ and b satisfying:

$$\arg\max_{\mathbf{w},b} \left[\frac{1}{\|\mathbf{w}\|} \min_i (y_i(\mathbf{w}^T\mathbf{x} + b)) \right] \tag{2.9}$$

$\mathbf{w}$ and b are normalized such that the support vectors $\mathbf{x}^+_{SV}$ and $\mathbf{x}^-_{SV}$ are defined by $h(\mathbf{x}^\pm_{SV}) = \pm 1$. The separability condition (2.8) becomes:

$$y_i(\mathbf{w}^T\mathbf{x} + b) \geqslant 1,\ i = 1, \ldots, l. \tag{2.10}$$

Hence, (2.9) turns to the following quadratic minimization problem:

$$\begin{aligned} \underset{\mathbf{w},b}{\text{minimize}} \quad & \frac{1}{2}\|\mathbf{w}\|^2 \\ \text{subject to} \quad & y_i(\mathbf{w}^T\mathbf{x}+b) \geqslant 1,\ i=1,\ldots,l. \end{aligned} \tag{2.11}$$

This problem is solved using Lagrange multipliers α_i paired with Kuhn-Tucker conditions (KKT):

$$\begin{cases} \sum_{i=1}^{l} \alpha_i y_i \mathbf{x}_i = \mathbf{w}^* \\ \sum_{i=1}^{l} \alpha_i y_i = 0 \end{cases} \tag{2.12}$$

The dual formulation (2.11) is thereby expressed as follows:

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \quad & L(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{k,j} \alpha_k \alpha_j y_k y_j \mathbf{x}_k^T \mathbf{x}_j \\ \text{subject to} \quad & \alpha_i \geqslant 0,\ \sum_{i=1}^{l} \alpha_i y_i = 0. \end{aligned} \tag{2.13}$$

As a consequence, with a test feature vector sample $\mathbf{x}$, the solution hyperplane is given by replacing the optimal $\mathbf{w}^*$ by the first KKT condition in (2.7), formulated through the solutions α_i^* of the dual problem (2.13):

$$h(\mathbf{x}) = \sum_{i=1}^{l} \alpha_i^* y_i \mathbf{x}^T \mathbf{x}_i + b \tag{2.14}$$

2.5.2 Non-Linear Support Vector Machines

Kernel trick

However, typical datasets are generally not separable in a linear fashion. A widespread solution is to transform the representation space $\mathbb{R}^d$ into a higher dimensional space $\mathbb{R}^q$ in which linear classification is feasible (q can even be infinite). Actually, picking specific transform functions ϕ would lead to very high computational costs while calculating the dot product in:

$$\begin{aligned} \underset{\alpha}{\text{maximize}} \quad & L(\alpha) = \sum_{i=1}^{l} \alpha_i - \frac{1}{2}\sum_{k,j} \alpha_k \alpha_j y_k y_j \phi(\mathbf{x}_k)^T \phi(\mathbf{x}_j) \\ \text{subject to} \quad & \alpha_i \geqslant 0,\ \sum_{i=1}^{l} \alpha_i y_i = 0. \end{aligned} \tag{2.15}$$

Instead, a kernel function $K(\mathbf{x}_k, \mathbf{x}_j) = \phi(\mathbf{x}_k).\phi(\mathbf{x}_j)$ is used. The function is meant to be a symmetric positive semi-definite function [16]. In practice, ϕ is not required to be explicitly picked since the kernel function is directly chosen. See Figure 2.14 for an illustration of the kernel trick. A few popular kernel functions can be pointed out:

- Linear: $K(\mathbf{x}_j, \mathbf{x}) = \langle \mathbf{x}_j, \mathbf{x} \rangle$
- Polynomial (rth degree): $K(\mathbf{x}_j, \mathbf{x}) = (1 + \langle \mathbf{x}_j, \mathbf{x} \rangle)^r$
- Radial basis functions (RBF): $K(\mathbf{x}_j, \mathbf{x}) = \exp(-\gamma \|\mathbf{x}_i - \mathbf{x}\|^2)$

- Sigmoid : $K(\mathbf{x}_j, \mathbf{x}) = \tanh(\alpha_1 \langle \mathbf{x}_j, \mathbf{x} \rangle + \alpha_2)$

Several studies confirm the fame of RBF kernel functions (*e.g* [74] or see [51]), since it especially handles well situations where the relation between labels and attributes is nonlinear. We will therefore resort to this kernel throughout our works.

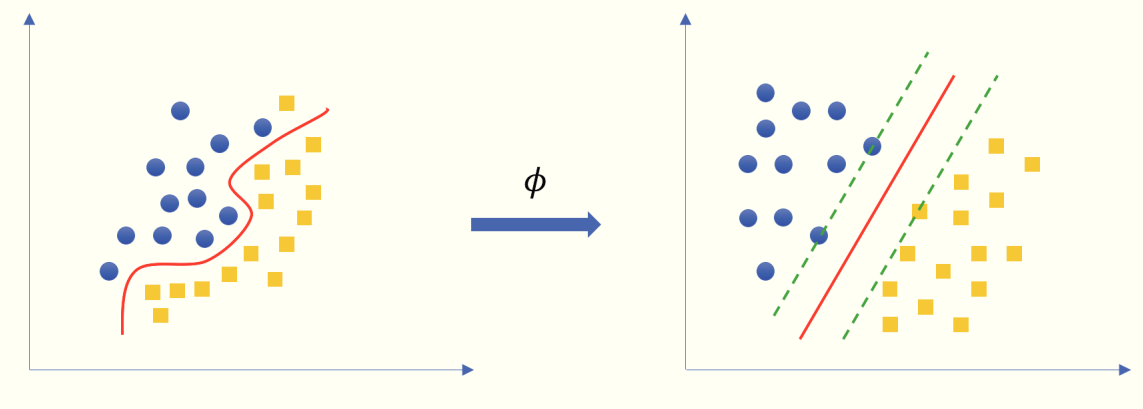


Figure 2.14 – Description of ϕ transforming the feature space to a higher dimensional space where linear classification is achievable

Soft margin

Nonetheless, the kernel trick is usually not sufficient for the classifier to generalize well enough. A regularization term is therefore introduced in the objective function (2.11) through slack variables that relax the constraint on the vectors. It actually enables to tolerate a few misclassified samples when searching for an optimal hyperplane.

(2.10) becomes:

$$y_i(\mathbf{w}^T \mathbf{x} + b) \geqslant 1 - \xi_i, \ \xi_i > 0, \ i = 1, \ldots, l. \tag{2.16}$$

and (2.11) can be written as:

$$\underset{\mathbf{w}, b, \xi_i}{\text{minimize}} \quad \frac{1}{2} \|\mathbf{w}\|^2 + C \sum_{i=1}^{l} \xi_i, \ C > 0 \tag{2.17}$$

where C is a regularization parameter balancing between the margin size and the number of tolerated misclassified samples. The choice of C plays a major role with respect to the classification accuracy; the reader is urged to consult [51] for solutions on this topic.

2.5.3 Multi-class Support Vector Machines

Let us consider M different categories $\{C_1, C_2, ..., C_M\}$. Two main multi-class extension methods are to be pointed out:

- the *one-versus-all* approach: M binary classifiers are trained while assigning successively 1 to the considered class and -1 to the rest of the samples. During testing, the classifier reporting the largest margin is selected, yielding straightforwardly the classification output.
- the *one-versus-one* technique: in this case, $\frac{M(M-1)}{2}$ binary classifiers are built between all pairs of categories (see Figure 2.15). While classifying an input sample $\mathbf{x}$, the corresponding class prediction is made by a voting scheme among all decision results.

Alternative multi-class approaches were put forward. The Directed Acyclic Graph SVM (DAGSVM) introduced by [82] uses the same training as one-against-one with a directed acyclic graph is employed to speed up testing and select the optimal classifier.

In [28], Crammer and Singer proposed a different technique by solving a single optimization problem, with a few slack variables. This way, both memory and training time are reduced. However, in their comparative study, [52] reported that the approach faces speed issues.

Throughout our study on refrigeration units analysis (section 5.5), we used the *one-versus-one* within the Error Correcting Output Codes *ECOC* generalized framework [127] where the multi-class problem is decomposed into a predefined set of binary problems. It is typically employed in combination with cross-validation to overcome overfitting issues. For a comprehensive discussion on the subject, the reader may refer to the recent works of [127].

2.6 Conclusion

In this chapter, both features description and classification tools employed hereby for X-ray image analysis are presented. HoG and CNN-extracted feature vectors are retained as global image descriptors while SURF are used as local descriptors toward matching applications. In addition, an overview of the main principles of ConvNets have also been presented.

The reported K-means method is adopted as an assistant tool for the manual labelling of large indiscriminated image sets. It is an essential tool in the sense that it significantly speeds up the database building process. Once classified into their respective categories, images can then be partitioned into training and testing sets and serve as input to a supervised learning classifier.

In this regard, the SVM supervised classification paradigm was finally outlined. Since we are interested in multi-label classification, its *ECOC* extension scheme was also described.

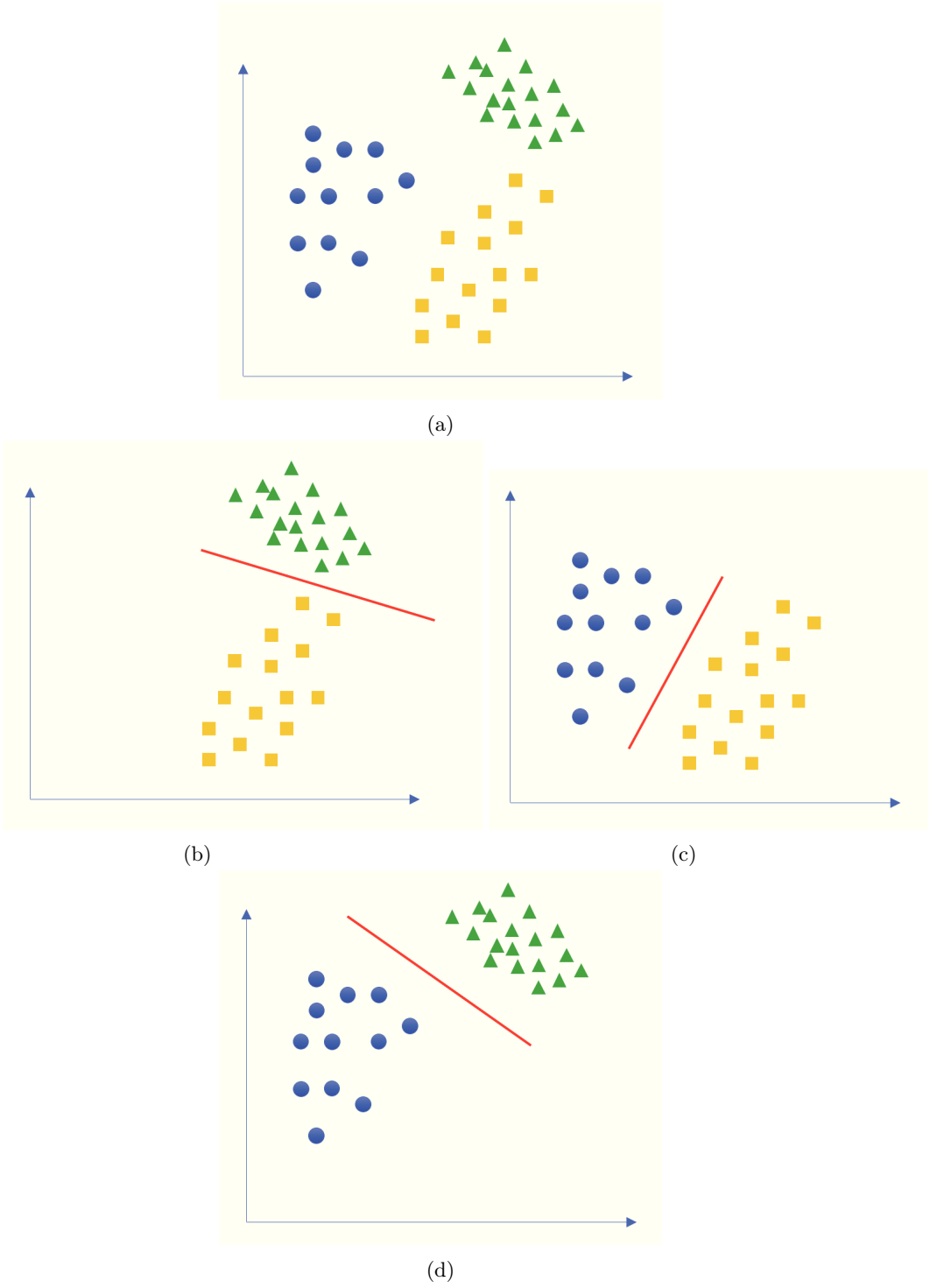


Figure 2.15 – *One-versus-one* multi-class SVM approach. Three binary classifiers are built between all pairs of classes.

Chapter 3

Image Registration Methods

Outline of the current chapter

3.1 Introduction

Image registration definitely belongs to one of the most prominent problems at the heart of computer vision research. Over the years, applications ranging from medical image analysis, remote sensing, robotics to geophysics and various implementations for the recreational sector (*e.g* image stitching, augmented reality [110]) have been elaborated. Meanwhile, it has increasingly drawn

attention of security and defense industries, particularly aiming at threat detection automation: surveillance, motion tracking, automatic target recognition, etc.

In a straightforward move, we were called upon to consider image registration methods to design functionalities for cargo X-ray imaging. With a view to automating "anomaly" detection in freight, commercial or civilian vehicles, change detection techniques may prove to be very efficient. The "anomaly" or irregularity, is defined as the presence of any foreign object inserted in a place that must remain empty (not necessarily to an identifiable threat). In our field, this applies to doors or refrigeration units in containers, car front-hoods and alike.

This feature turns out to be very helpful for custom officers. Indeed, in the current security context paired with sharp workforce cuts, operators are frequently prone to losses of awareness. This especially applies in situations of extreme tiredness. Still, another notable advantage lies in its capability to enable visualization of objects beyond human visual spectrum. In fact, the technique would achieve detection in very dense regions where even machine learning recognition algorithms would probably fail.

As defined by Modersitzki *et al.* [76], [75], image registration is commonly referred to as the searching process of a transformation that aligns, in an optimal but "plausible" way, a moving template image $\mathbf{T}$ with a static target sample $\mathbf{R}$. To be more precise, given an X-ray scan to be inspected $\mathbf{R}$, and assuming that a corresponding empty template $\mathbf{T}$ can be automatically retrieved, the underlying idea is to register both images to detect significant differences that could relate to irregularities.

In this chapter, we will give a general outline of registration methods. Both linear and non-linear approaches are described. **We introduce a particular piece-wise embodiment of rigid methods, paired with a displacement regularization strategy, to deal with local deformations in a more efficient manner** (section 3.5). Within the second category, a focus will be given on the variational registration framework where parametric and non-parametric methods are to be distinguished. The numerical schemes employed to solve the minimization problems for non-rigid registration are also discussed.

Note that deforming potential threats in the image is not desirable for further recognition tasks. We will therefore consider the current inspection scan as the static image $\mathbf{R}$ and the empty reference image from our database as the moving template $\mathbf{T}$ (such that $\mathbf{R}$ does not undergo any linear or non-linear warping).

3.2 Linear Registration

We assume that $\mathbf{T}$ and $\mathbf{R}$ are two given images and we are looking for the "best" transform Θ such that the deformed template $\Theta(\mathbf{T})$ is "similar" to the static image $\mathbf{R}$. The concept of "best" transform usually relates to the minimizer of some energy that includes a measure of similarity between images.

The linear or so-called "rigid" registration approach aims at finding a parametric planar transformation Θ to be applied on $\mathbf{T}$. Traditionally, this scheme is *feature-based*, i.e sparse keypoints descriptors are extracted in $\mathbf{R}$ and $\mathbf{T}$, and Θ is computed on the basis of their filtered correspondences [110].

3.2.1 Linear Transforms

As reported on Figures 3.1 and Figure 3.2, a few linear transforms can be expressed via 2×3 or 3×3 matrices. Given a 2D point $X = [X_x, X_y]^T$, its image X' is computed by the application of a parametrized transform matrix.

Translation

For the translation case, we have, while using $\bar{X} = [X_x, X_y, 1]$ (augmented coordinates) for a more compact notation:

$$X = \begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \end{bmatrix} \bar{X} = [Id\ \mathbf{t}]\, \bar{X} \tag{3.1}$$

Euclidean

Referred to as the rigid transform (strictly speaking, yet the term "rigid" has been abusively extended to all linear methods) it combines a rotation and a translation, thus preserving lengths of transformed shapes.

$$X = \begin{bmatrix} \cos(\theta) & -\sin(\theta) & t_x \\ \sin(\theta) & \cos(\theta) & t_y \end{bmatrix} \bar{X} = [R_\theta\ \mathbf{t}]\, \bar{X} \tag{3.2}$$

where R_θ is an orthonormal rotation matrix.

Similarity

The transformed is expressed as follows, with a scale factor applied on the rotation matrix:

$$X = \begin{bmatrix} a & -b & t_x \\ b & a & t_y \end{bmatrix} \bar{X} \tag{3.3}$$

It preserves angles between lines.

Affine

With this transform, parallelism between lines is maintained. It is expressed by a 2×3 abitrary parametrized matrix:

$$X = \begin{bmatrix} a_{00} & a_{01} & a_{02} \\ a_{10} & a_{11} & a_{12} \end{bmatrix} \bar{X} \tag{3.4}$$

Projective

The projective or perspective transform is applied via a 3×3 matrix H in homogeneous coordinates as described in Figure 3.2. It ensures the preservation of straight lines.

3.2.2 Feature Points Alignment

While performing linear registration, we use SIFT or SURF feature points and descriptors as developed in section 2.2.2. The keypoint matching process is performed through an exhaustive nearest-neighbor search. For each keypoint X_i in $\mathbf{T}$, a corresponding feature point X_i' in $\mathbf{R}$ is

found such that an *SSD* distance between feature descriptors is minimized. kd-trees could also be used to this purpose (see [110]).

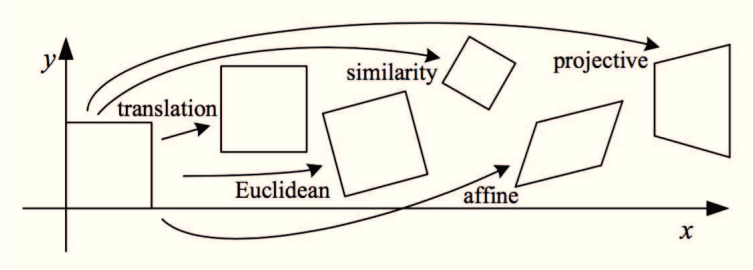


Figure 3.1 – Most common types of planar transforms ([110]). Matrices were set such that $\mathbf{p} = 0$ yields the identity.

In a second stage, the geometric consistency of the matching features between both images is analyzed to check whether the corresponding displacements can be interpreted as part of a 2D planar parametric transform. Let us consider a set of matched feature points $\{(X_i, X_i')\}$, we look for the motion parameters $\mathbf{p}$ defining Θ such that:

$$X' = \Theta(X, \mathbf{p}) \tag{3.5}$$

One could obviously solve the following least-squares problem to find $\mathbf{p}$

$$\begin{aligned} &\text{find } \mathbf{p} \text{ minimizing} \\ &E(\mathbf{p}) = \sum_i \|\Theta(X_i, \mathbf{p}) - X_i'\|^2 \end{aligned} \tag{3.6}$$

This scheme is simplified for affine, similarity or translation transforms where the amount of motion is linearly related to the unknown parameter $\mathbf{p}$ (linear regression). For more complex transforms such as Euclidean 2D transform, a non-linear regression problem must be solved by iterative methods such as the Levenberg-Marquardt algorithm (LMA) [110].

Though, in order to cope with measurements outliers, more robust procedures are preferred. The principle behind the M-estimator is to apply a robust penalty ϱ on the residuals:

$$\begin{aligned} &\text{find } \mathbf{p} \text{ minimizing} \\ &E_{\text{M-estimator}}(\mathbf{p}) = \sum_i \varrho(\|\Theta(X_i, \mathbf{p}) - X_i'\|) \end{aligned} \tag{3.7}$$

Still, the reported results show that the method often converges to a local minimum [110]. We therefore turn to the Random Sample Consensus technique known as RANSAC, an iterative method based on a voting scheme to filter matching points outliers. In short, the technique can be divided into two steps:

1. a sample subset of k matching points is randomly picked, and a parameter estimate $\hat{\mathbf{p}}$ is computed.

Transform	Matrix	Parameters p
translation	$\begin{bmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \end{bmatrix}$	(t_x, t_y)
Euclidean	$\begin{bmatrix} c_\theta & -s_\theta & t_x \\ s_\theta & c_\theta & t_y \end{bmatrix}$	(t_x, t_y, θ)
similarity	$\begin{bmatrix} 1+a & -b & t_x \\ b & 1+a & t_y \end{bmatrix}$	(t_x, t_y, a, b)
affine	$\begin{bmatrix} 1+a_{00} & a_{01} & t_x \\ a_{10} & 1+a_{11} & t_y \end{bmatrix}$	$(t_x, t_y, a_{00}, a_{01}, a_{10}, a_{11})$
projective	$\begin{bmatrix} 1+h_{00} & h_{01} & h_{02} \\ h_{10} & 1+h_{11} & h_{12} \\ h_{20} & h_{21} & 1 \end{bmatrix}$	$(h_{00}, h_{01}, \ldots, h_{21})$

Figure 3.2 – Parametrization of most common types of planar transforms ([110]).

2. the algorithm calculates the residuals according to the model defined by $\hat{\mathbf{p}}$. The inliers, defined by some error threshold: $\|\Theta(X_i, \hat{\mathbf{p}}) - X_i'\| \leqslant \varepsilon$ are counted and define the *consensus set*.

The process is repeated S times and the parametric model with the greater number of inliers is selected [110].

More particularly, an improved version of RANSAC: MLESAC filtering [116] was adopted for the computation of the linear transform to be applied on **T** towards image registration. Instead of maximizing the *consensus set* cardinality, the solution giving the maximum likelihood is chosen. It has also been shown to give superior results to those of RANSAC. See [116] for a comprehensive description of the method.

In our RU (refrigeration unit) irregularity detection application (chapter 5), we look for an affine transform to register **T** and **R**. Empirically, it allows to deal with various distortions in a more plausible way than projective transform does. Concerning the analysis of top-view scans of vehicles (chapter 6), a similarity transform is employed for the pre-registration step, as we wish to investigate the very nature of the displacement. **In the present thesis, the term "rigid" is therefore extended to either similarity or affine transforms**.

Still, by definition, linear registration outputs a global transform which cannot address local deformation issues. Two illustrations from the registration problems we strive to solve in this work are provided on Figure 3.3 and Figure 3.4. **Note that green and pink colors depict respectively the differences originating from the moving template T and target image R under inspection**.

Intuitively, a straightforward solution would be to partition **R** and **T** into sub-images on

which a linear feature-based registration is locally performed. In 3.5, we introduce a nice way to do so.

3.3 Non-Linear Methods in Variational Framework

In many registration problems, the non-linear distortions in images cannot be solved by simple schemes like the above-described techniques. Non-rigid methods offer more flexibility to address more sophisticated registration problems.

3.3.1 Mathematical Problem Setup

Let's consider a static image $\mathbf{R}$ and a moving template $\mathbf{T}$; we look for a transformation applied on $\mathbf{T}$ such that both images align as closely as possible [76], [75]. The general formulation is given by the definition of $\mathbf{R}$ and $\mathbf{T}$ as d-dimensional images represented by the mappings:

$$\mathbf{R}, \mathbf{T} : \Omega \to \mathbb{R}, \Omega \in [0,1]^d \text{ in normalized coordinates.}$$

The transformation ϕ is expressed via the displacement field $\mathbf{u} : \Omega \to \mathbb{R}^d$ with:

$$\phi = id + \mathbf{u}. \tag{3.8}$$

Hence, given $\mathbf{x} \in \Omega$, the transformed image is written $\mathbf{T}(\phi(\mathbf{x})) = \mathbf{T}(\mathbf{x} + \mathbf{u}(\mathbf{x}))$ where $\mathbf{u}$ is applied on the position $\mathbf{x}$ of $\mathbf{T}$. In other words, $\mathbf{T}(\mathbf{x})$ is the intensity of $\mathbf{T}$ at $\mathbf{x}$ and $\mathbf{T}(\mathbf{x} + \mathbf{u}(\mathbf{x}))$ designates the value of $\mathbf{T}$ at the translated position $\mathbf{x} + \mathbf{u}(\mathbf{x})$.

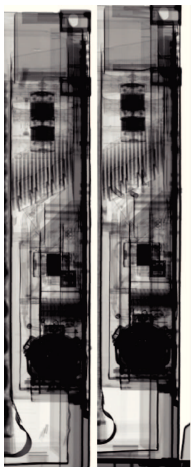
Since we wish to minimize the distance between the deformed template $\mathbf{T} \circ \phi$ and the target $\mathbf{R}$, the unified optimization framework for intensity-based registration is given by the following joint functional scheme:

$$\begin{aligned} &\text{find } \mathbf{u} \text{ minimizing} \\ &\mathcal{J}[\mathbf{u}] = \mathcal{D}[\mathbf{T}(\mathbf{u}), \mathbf{R}] + \alpha \mathcal{S}[\mathbf{u}] \end{aligned} \tag{3.9}$$

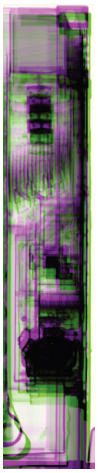
where $\mathcal{D}$ is referred to as the distance measure between $\mathbf{T} \circ \phi$, the transformed image, and $\mathbf{R}$. It commonly designates the data-fitting term defining the *external force* [37] to obtain an ideal alignment.

In order to tackle the ill-posedness of the optimization problem ([76], [75]), a regularization term $\mathcal{S}$ is introduced (relating to the *internal force*, [37]). It is designed to keep the displacement smooth during deformation by penalizing unwanted or implausible solutions. The regularization strength is controlled by the smoothing parameter α : increasing its value emphasizes the smoothness of the sought solution $\mathbf{u}$ and vice-versa. Note that a too strong value of α may push the transformation towards the identity. The choice of this parameter has been intensively studied, *e.g.* in [125] and [49], but remains utterly dependent on the considered application and the appreciation of the operator [44], [102], [46].

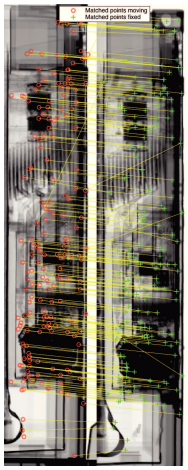
Similarly, both data-fitting and smoothing terms are selected according to the specificity of the registration task. The most popular similarity measure is the *Sum of Squared Differences* (*SSD*):



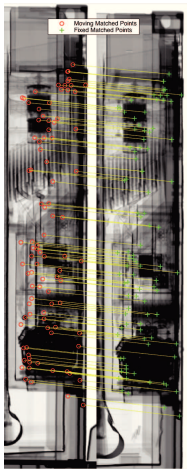
(a)



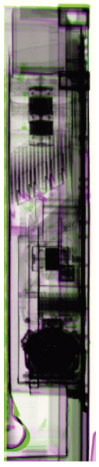
(b)



(c)

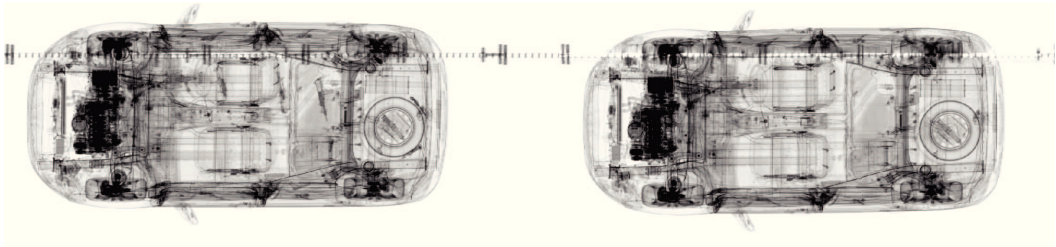


(d)

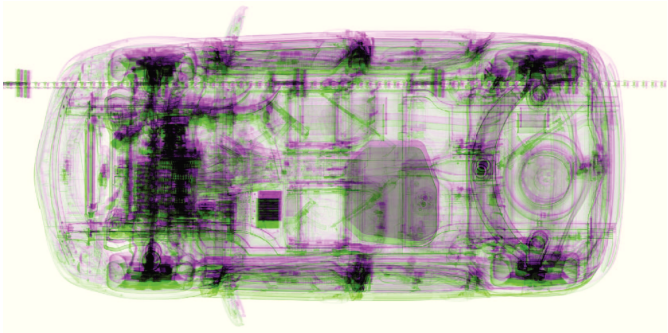


(e)

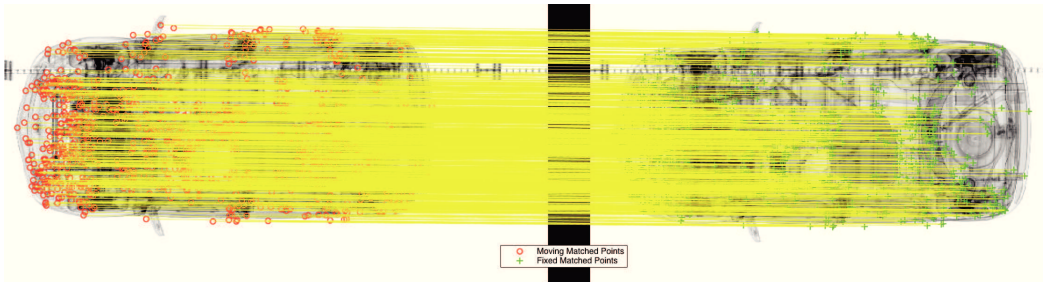
Figure 3.3 – (a) HCVS refrigeration unit images **R** and **T** from the same class; (b) Overlay before rigid registration; (c) SURF features extraction and matching; (d) MLESAC filtering to get an affine transform, (e) Overlay following rigid registration.



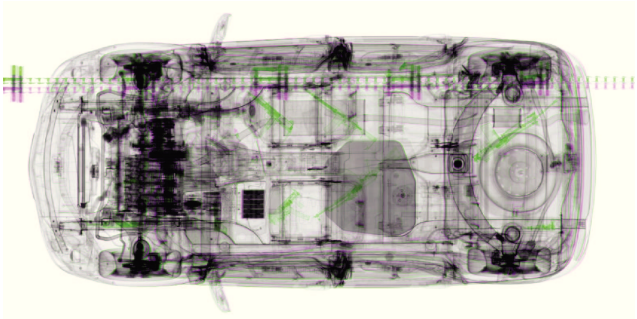
(a)



(b)



(c)



(d)

Figure 3.4 – (a) HCVL top-view images **R** and **T** of the same car; (b) Overlay before rigid registration; (c) SURF features extraction and MLESAC filtering of correspondences; (d) Overlay following rigid registration.

$$\mathcal{D}^{SSD}\left[\mathbf{T}(\mathbf{u}),\mathbf{R}\right]=\frac{1}{2}\int_{\Omega}\left[\mathbf{T}\left(\mathbf{x}+\mathbf{u}\left(\mathbf{x}\right)\right)-\mathbf{R}(\mathbf{x})\right]^{2}d\mathbf{x} \tag{3.10}$$

The *Normalized Cross Correlation* (*NCC*) is also widespread for matching problems where images sampled with the same modality are involved (*monomodal* registration):

$$\mathcal{D}^{NCC}\left[\mathbf{T}(\mathbf{u}),\mathbf{R}\right]=1-\frac{\langle\mathbf{T}(\mathbf{u}),\mathbf{R}\rangle^{2}}{\left\|\mathbf{T}(\mathbf{u})\right\|^{2}\left\|\mathbf{R}\right\|^{2}} \tag{3.11}$$

where

$$\langle\mathbf{T}(\mathbf{u}),\mathbf{R}\rangle=\int_{\Omega}\mathbf{T}\left(\mathbf{x}+\mathbf{u}\left(\mathbf{x}\right)\right)\mathbf{R}(x)d\mathbf{x}$$

and

$$\left\|\mathbf{R}\right\|=\sqrt{\langle\mathbf{R},\mathbf{R}\rangle}$$

Information-theoretic approaches based on Shannon's entropy are employed for *multimodal* registration where images are acquired by different procedures, *e.g.* while combining magnetic resonance (MRI) and computed tomography (CT) outputs of a patient (or PET and CT such as in Figure 3.5). The underlying idea behind *Mutual Information – MI* is to measure the mutual information between $\mathbf{T}$ and $\mathbf{R}$, which amounts to computing a normalized entropy of the joint density. Especially, it consists of measuring the co-occurrence of intensities in both images. This similarity measure was suggested in the seminal works of Collignon [69] and Viola [124]. With H denoting the entropy measure:

$$\mathcal{D}^{MI}\left[\mathbf{T}(\mathbf{u}),\mathbf{R}\right]=H(\mathbf{T}(\mathbf{u}))+H(\mathbf{R})-H(\mathbf{T}(\mathbf{u}),\mathbf{R}) \tag{3.12}$$

Both marginal and joint entropies can be estimated by either the joint histogram of $\mathbf{T}(\mathbf{u})$ and $\mathbf{R}$ or by resorting to kernel density estimators such as Parzen windowing (see [75] for an exhaustive explanation). A normalized extension of *MI* has also been proposed by Studholme [107] in the context of three-dimensional medical imaging:

$$\mathcal{D}^{MI}\left[\mathbf{T}(\mathbf{u}),\mathbf{R}\right]=\frac{H(\mathbf{T}(\mathbf{u}))+H(\mathbf{R})}{H(\mathbf{T}(\mathbf{u}),\mathbf{R})} \tag{3.13}$$

Admitting *MI* may be a measure of choice for future projects in large-scale X-ray security imaging, where the combination of information acquired with different energies and geometries could be desirable, it is still beyond the scope of this thesis work.

Over the years, more sophisticated measures have also been explored such as the so-called *NGF - Normalized Gradient Fields*. It aims at considering gradients as a crucial quantity for optimal alignment computation between $\mathbf{T}$ and $\mathbf{R}$ [76]. $\hat{\mathbf{n}}$ denotes the normalized gradient field:

$$\hat{\mathbf{n}}\left[\mathbf{T},\kappa\right]=\frac{\nabla\mathbf{T}}{\sqrt{\left|\nabla\mathbf{T}\right|^{2}+\kappa^{2}}} \tag{3.14}$$

κ stands for a threshold defining the edges to be taken into account ($|\nabla\mathbf{T}|>\kappa$) or rather those considered as noise ($|\nabla\mathbf{T}|<\kappa$).

The similarity measure is thus given by:

$$\mathcal{D}^{NGF}\left[\mathbf{T}(\mathbf{u}),\mathbf{R}\right]=\int_{\Omega}1-\left(\hat{\mathbf{n}}\left[\mathbf{T}(\mathbf{x}+\mathbf{u}(\mathbf{x}))\right]^{T}\hat{\mathbf{n}}\left[\mathbf{R}(\mathbf{x})\right]\right)^{2}d\mathbf{x} \tag{3.15}$$

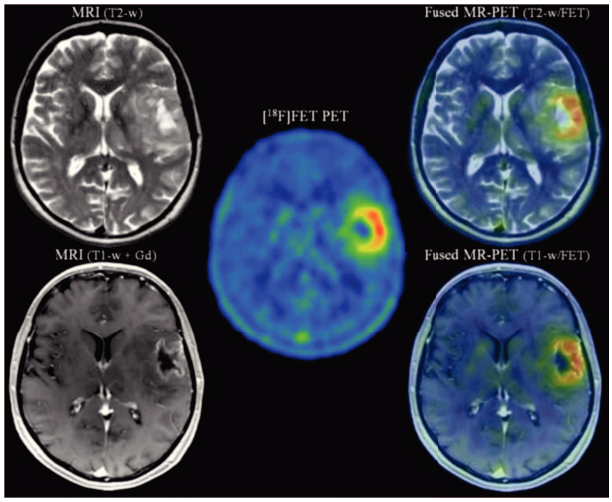


Figure 3.5 – Multimodal registration example for CT and PET images fusion ([77]).

Note that the idea of employing gradient-based information for dense matching has been explored earlier by Benayoun [13] within the context of optical flow computation.

The reader may refer to [102], [75], or [46] for further details about similarity measures.

In various registration tasks, the choice of the smoothing term is crucial [76], [102]. It allows the embedding of physical priors or possible acquisition constraints within the optimization scheme. The next sections will provide a larger overview of the most widespread regularization terms.

Note that the selection of both terms is conditioned upon the existence of a Gâteaux derivative ([76], [75], [46], [45]).

Registration can be landmark-based, i.e. the image is transformed via the matching of sparse elements (feature points or manually annotated control landmarks) combined with *TPS* (*Thin-Plate-Spline*) interpolation to yield dense correspondences ([76], [75], [102]).

Although these methods have proven their efficiency in many applications, particularly for medical image analysis, they show less flexibility in comparison with the unified variational framework introduced in the minimization problem (3.9) (see [37]).

Aiming at addressing specific physical models of deformation, it would therefore be more convenient to handle a model that offers more adaptability. In addition, despite the recent achievements on key points automatic generation (deep-learning based, as described in section 2.3.2), unmanned landmark detection remains a challenging task.

For these reasons, we chose to focus our work on intensity-based, image-based or so-called iconic methods. These approaches generally split into two categories ([76], [75], [102]): parametric and non-parametric schemes.

3.3.2 Parametric Approaches

Regarding iconic parametric image-registration techniques (a.k.a *PIR*), the displacement field $\mathbf{u}$ is parametrized. The optimization process thereby consists in finding the parameters of $\mathbf{u}$

minimizing $\mathcal{J}$.

In B-spline *Free-Form-Deformations* (*FFD*) approaches (short for basis-spline), the displacement is defined as a linear combination of a small set of basis functions. See [75] or [102] for a further description of *PIR* approaches. *FFD* constitute one of the most widespread types of transformations for medical image registration [102]. Bardinet Cohen and Ayache [10] fit a superquadric on unstructured 3D points through a free-form-deformation to describe the cardiac left ventricle. In [95] Rueckert *et al.* proposed the B-spline registration paradigm to handle the local breast motion in MRI images. More recently, they released a generalized framework with up-to-date experiments in [94].

For instance, let us consider the cubic-spline parametrization of the deformation $\mathbf{u}$:

$$\mathbf{u}(\mathbf{x}) = \sum_i \rho_i \beta_i(\mathbf{x}) \tag{3.16}$$

while ρ_i stands for a set of scaling factors and β is a piecewise cubic polynomial composed of the set of basis functions β_i.

In some cases, the smoothing terms are chosen such that $\mathcal{S}[\mathbf{u}] = 0$ with a parametrization meant to implicitly integrate regularization constraints. Though, Tychonov regularization may often be used, especially when the transformation space is high-dimensional [75]. Rueckert *et al.* recommend to use the two-dimensional energy of a thin-plate of metal as a penalty to render the problem well-posed [94] (for 2-D images). It is phrased as follows:

$$\mathcal{S}^{tp} = \int\int_\Omega \left[\left(\frac{\partial^2 \mathbf{u}}{\partial x^2} \right)^2 + \left(\frac{\partial^2 \mathbf{u}}{\partial y^2} \right)^2 + 2 \left(\frac{\partial^2 \mathbf{u}}{\partial x \partial y} \right)^2 \right] dxdy \tag{3.17}$$

As a matter of fact, parametric registration is adopted for *Free-Form-Deformations*, that is to say, for objects which do not own intrinsic rigidity properties. Although [94] expressed a penalty function for this purpose, it only imposes rigidity constraints in a local manner. Regarding X-ray imaging of vehicles, such characteristics are inherent to the systems and the types of scanned objects; B-spline or analogous *PIR* methods are thus very likely to yield poor registration outputs. Results along these lines will be provided in the second part of the present work (see section 6.4.1).

3.3.3 Non-Parametric Approaches

Non-parametric methods (*NPIR*) are very popular for offering further adaptability to formulate specific-task constraints. Here, the smoothing term $\mathcal{S}$ plays a major role while the displacement $\mathbf{u}$ is no longer parametrized. The regularization penalty is intended to constrain the displacement to physically-meaningful types of movements. Most common techniques employ L_2-normed terms. They are usually divided into four categories.

Diffusion registration

The diffusion smoother is often associated with techniques deriving from optical-flow applications [35]. Introduced by Fisher et al. in [36] it aims at quantifying the gradient of the deformation $\mathbf{u}$:

$$\mathcal{S}^{diff}[\mathbf{u}] = \frac{1}{2} \sum_{l=1}^{d} \int_\Omega \|\nabla u_l\|^2 \, d\mathbf{x} \tag{3.18}$$

Following the results of [36], the diffusion registration scheme is a method of choice for high-resolution problems. Indeed, with N standing for the number of voxels, an $\mathcal{O}(N)$ complexity can be reached. The method was applied on time-series three-dimensional MRI images (breast cancer) or for an online correction of the brain shift during a surgery operation.

Elastic registration

The elastic regularizer measures the elastic potential of the displacement $\mathbf{u}$ to be applied. The idea was first proposed in the seminal works of Broit [15] and extensively discussed in [113]. $\mathcal{S}^{elastic}$ is expressed as follows:

$$\mathcal{S}^{elastic}[\mathbf{u}] = \frac{1}{2}\int_{\Omega}\left[\frac{\lambda+\mu}{2}\left\|\nabla.\mathbf{u}\right\|^2 + \frac{\mu}{2}\sum_{l=1}^{d}\left\|\nabla u_l\right\|^2\right]d\mathbf{x} \tag{3.19}$$

When combined with the *SSD* similarity measure, this smoother defines the so-called elastic registration model. Its force-field (*internal force*), obtained from the corresponding Gâteaux derivative, yields the famous Navier-Lamé operator. Thus, μ and λ relate to the Lamé constants, controlling the smoother strength with respect to the characteristics of the underlying imaged body. This method has also proven its efficiency in medical image registration of organs deforming in an elastic fashion (histological sections of human brain) [35].

Fluid registration

The elastic model is local by nature, and can therefore hardly compensate for large deformations. The viscous fluid model, more flexible, was applied to image registration by Christensen [21]. The idea is to calculate the elastic potential of the velocity field $\mathbf{v}$ rather than the deformation $\mathbf{u}$. The partial derivative operator yields also the Navier-Lamé operator, applied to $\mathbf{v}$.

$$\mathcal{S}^{fluid}[\mathbf{u}] = \mathcal{S}^{elastic}[\mathbf{v}] \tag{3.20}$$

Curvature registration

In [35], Fisher and Modersitzki suggested a regularization technique based on the deformation curvature calculation. Its main advantage lies in the skipping of the rigid pre-registration step, which is imperative for the above-mentioned penalties. It is phrased as follows:

$$\mathcal{S}^{curv}[\mathbf{u}] = \frac{1}{2}\sum_{l=1}^{d}\int_{\Omega}\left(\Delta u_l\right)^2 d\mathbf{x} \tag{3.21}$$

For further details the reader may refer to [35].

Demons registration

The very popular demons method, introduced by Thirion [115] is also inspired from optical-flow techniques. The velocity $\mathbf{v}$ of the displacement field $\mathbf{u}$ is computed while applying a Gaussian smoothing at each iteration. Hence, this method falls within a diffusion-like regularization, as claimed by [35]. It is especially useful for high dimensional and computationally demanding non-linear problems [75].

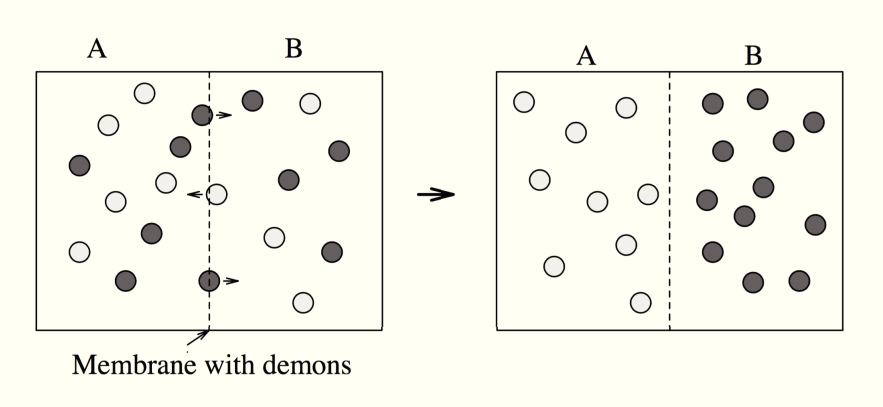


Figure 3.6 – Description of Maxwell's demons with a mixed gas [115].

Behind Thirion's concept, the underlying idea is to model an attractive force by an intuitive analogy to Maxwell's demons. The boundaries (or edges) in the target image $\mathbf{R}$ are *semi-permeable*, thus letting the moving sample $\mathbf{T}$ diffuse through these very "membranes". A depiction of the approach is provided on Figures 3.6 and 3.7. At each iteration, the update $\Delta\mathbf{u}$ or velocity $\mathbf{v}$ is given by:

$$\begin{cases} \mathbf{v} = \frac{(\mathbf{T}(\mathbf{u})-\mathbf{R})\nabla\mathbf{T}(\mathbf{u})}{\nabla\mathbf{T}(\mathbf{u})^2+(\mathbf{R}-\mathbf{T}(\mathbf{u}))^2} \\ \\ \mathbf{v} = 0 \text{ if } \nabla\mathbf{T}(\mathbf{u})^2 + (\mathbf{R}-\mathbf{T}(\mathbf{u}))^2 < \epsilon_d \end{cases} \tag{3.22}$$

with ϵ_d an arbitrarily small real number. As specified, each update is completed by regularizing the deformation field via a Gaussian kernel. Since heat propagation is viewed as Gaussian smoothing, this scheme is tightly associated with diffusion registration.

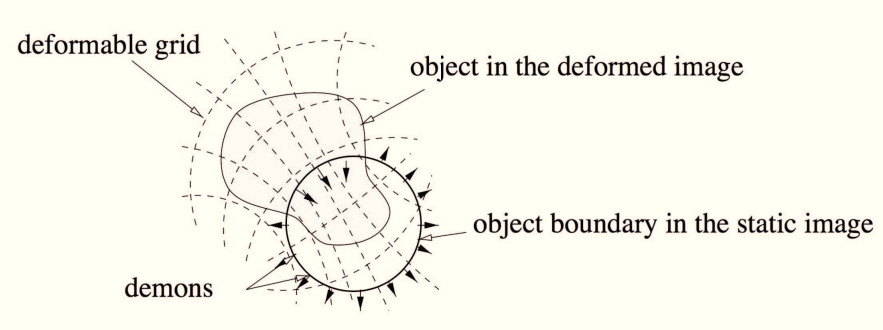


Figure 3.7 – Demons scheme: a deformed image is diffusing through the contours of the objects in the static image by "demons" modelizing attractive forces [115].

3.3.4 Volume Preservation - Diffeomorphic Registration

Topology-preserving registration is also a very active field of research, particularly in medical applications where most organs of interest maintain a fixed topology over time [83]. As we wish to preserve the topology of a transformed space $\phi : \Omega \to \Omega$, two properties are required:

- Distinct points should have distinct generated images. ϕ should therefore be injective (no folds or repeated values).
- no holes should be created, each point of the domain should present a pre-image.

As we look for a bijection, homeomorphisms can be considered to this regard:

Definition 3.3.1. A **homeomorphism** of Ω is a continuous bijection $\phi : \Omega \to \Omega$ such that its inverse ϕ^{-1} is also continuous.

Nevertheless, additional regularity constraints must often be applied to the deformation field; for this reason, diffeomorphisms are often preferred.

Definition 3.3.2. A **diffeomorphism** of Ω is a continuously differentiable homeomorphism $\phi : \Omega \to \Omega$ such that its inverse ϕ^{-1} is also continuously differentiable.

In fluid mechanics, incompressibility imposes that the determinant of the transformation Jacobian remains equal to 1 (see e.g. [102], [46]). On this basis, Rohlfing [91] and Christensen [19] simply add a soft constraint to the unified variational scheme (3.9) to guarantee the conservation of volumes:

$$\begin{aligned} \mathcal{C}^{soft}[\mathbf{u}] &= \|\det(\nabla\phi) - 1\|^2 \\ &= \|\det(Id + \nabla\mathbf{u}) - 1\|^2 \end{aligned} \tag{3.23}$$

Its logarithmic version [94] has also been proposed to ensure volume preservation. Similarly, various analogous smoothers are described in [91], [19] and [94]. Instead, Haber and Modersitzki opt for an inequality/equality hard constraint (resp. [45] and [46]). More recently, Rueckert et al. compare soft and hard-constraint techniques in [94].

Alternative approaches concentrate directly on the diffeomorphic framework. ϕ is then explicitly constrained to be a continuously differentiable bijection having a smooth inverse. In the context of implicit template deformation [83] shows that diffeomorphisms can naturally be handled by function composition whereas adding them has actually no geometrical meaning. This concept proves to be more efficient where the *small deformation model* of Ashburner *et al.* [5] is no longer valid (in case of large deformations). This issue has also been discussed in [122]. Similarly to [131], the deformation is modeled via a time-dependent velocity vector field. Further details will be given in subsection 3.4.4.

One should also mention the study of [20] and the *LDDMM* framework (*Large Deformation Diffeomorphic Metric Mappings*) introduced by Beg *et al.* [12]. The reader may also refer to the *log-Euclidean* registration proposed by [3] and improved by [5] or [122]. In this framework, the transformation ϕ is defined as the exponential of the displacement: $\phi = \exp(\mathbf{u})$. Yet, as observed by [83], some diffeomorphisms cannot be expressed by an exponential; and, despite allowing for accurate registration results in practice, the computed transformations are often non plausible. See also the recent contributions of Mang *et al.* about mass and volume preservation constraints for a comprehensive state of the art overview ([70], [71]).

Other methods suggest to apply a post-processing step over the displacement field following registration. In [48], Poisson's equation is solved to maintain a so-called solenoidal displacement field $\mathbf{u}$ and ensure the preservation of volumes.

For our car registration task (chapter 6), the search for diffeomorphic transformations is indeed crucial. Actually, a non volume-preserving function ϕ would push $\mathbf{T}$ to "mimic" the

added objects in $\mathbf{R}$ while deforming, entailing strong alleviation at the difference detection stage. Various experiments will be presented in the second part of this work to illustrate the relevance of incompressibility constraints.

Similarly, rigidity constraints (hard and soft versions) have been formulated in different publications. Though, they are beyond the scope of this paper (see [38] for instance).

3.4 Numerical Resolution for Non-Linear Registration

3.4.1 General Overview

First-order descent methods can be handled to find an optimizer $\mathbf{u}^*$ to the registration problem [35]. In [76], [75], Modersitzki suggests a general optimization framework to solve (3.9). An approximation of the Gauss-Newton method is employed [78]: the gradient $\nabla_{\mathbf{u}}\mathcal{J}$ and Hessian $\mathcal{H}$ of $\mathcal{J}$ are computed to yield the descent direction $\mathcal{H}\left[\mathbf{u}\right]^{-1}\nabla_{\mathbf{u}}\mathcal{J}\left[\mathbf{u}\right]$ at each iteration.

The variations of the displacement between consecutive iterations are estimated by $\|\nabla\mathbf{u}\|$ in order to stop the minimization process whenever it falls below a pre-set threshold ϵ. See [76], [75] for further details about the stopping criterion.

Alternative numerical techniques have also been presented for computationally-demanding problems. In [75], l-BFGS is described. Trust-region methods are quite popular as well since they can achieve fast quadratic convergence [78].

For volume-preserving constrained problems, Modersitzki *et al.* resort to the framework of Sequential Quadratic Programming methods (SQP) [46] or a variant of the log-barrier method [45]. Diffeomorphic image registration as described in [131] and [12] typically uses an implicit regularization (by solving a Poisson PDE) while gradient-descent methods are used to find an optimal solution. In their latest works, Mang et al. [71] proposed second-order numerical methods to produce diffeomorphic mappings.

Modersitzki generally opts for a discretize-then-optimize approach. In several applications, the choice of discretization paradigms is crucial. Staggered, nodal or cell-centered grids may be used. See [75] and [76] for further details about interpolation or discretization issues. Matrix-free approaches are also proposed to speed-up the computations in [75] and more recently in [96]. Still, for an in-depth, up-to-date description and analysis of registration methods, the reader is urged to refer to [102].

In this section, we will more precisely present the numerical optimization techniques which were employed to solve our specific registration tasks.

3.4.2 Gradient Descent

Due to various simplification hypotheses, we were naturally brought to employ first-order descent schemes. In order to approach a solution by gradient descent, the Gâteaux derivative of the objective function must be computed (for further details on the directional derivative calculation, see [76] for instance). It also amounts to computing the Euler-Lagrange equation. By definition, the Gâteaux derivative of $\mathcal{J}$ is expressed as:

$$d_{\mathbf{u},\mathbf{q}}\mathcal{J}\left[\mathbf{u}\right] := \lim_{h\to 0}\frac{1}{h}(\mathcal{J}\left[\mathbf{u}+h\mathbf{q}\right]-\mathcal{J}\left[\mathbf{u}\right]) \tag{3.24}$$

Still, it can also be phrased in the following way:

$$d_{\mathbf{u},\mathbf{q}}\mathcal{J}[\mathbf{u}] = \int_{\Omega} \langle \nabla_{\mathbf{u}}\mathcal{J}[\mathbf{u}], \mathbf{q}\rangle_{\mathbb{R}^d}\, d\mathbf{x} \tag{3.25}$$

Hence, as soon as the descent direction $\nabla_{\mathbf{u}}\mathcal{J}[\mathbf{u}]$ is computed, the following update scheme is obtained for each iteration (with some initial and boundary conditions):

$$\mathbf{u}^{k+1} = \mathbf{u}^k - \rho\nabla_{\mathbf{u}}\mathcal{J}\left[\mathbf{u}^k\right] \text{ with } k \in \mathbb{N} \tag{3.26}$$

while ρ is set to a fixed size meanwhile. The procedure is repeated until a stopping criteria is satisfied. For clarification purposes, we will employ the *SSD* as cost function:

$$\mathcal{J}[\mathbf{u}] = \mathcal{D}^{SSD}[\mathbf{T}(\mathbf{u}), \mathbf{R}] = \frac{1}{2}\int_{\Omega} [\mathbf{T}(\mathbf{u}(\mathbf{x}) + \mathbf{x}) - \mathbf{R}(\mathbf{x})]^2\, d\mathbf{x}$$

As far as our example is concerned, we obtain the following result ($\mathbf{u}(\mathbf{x})$ is written as $\mathbf{u}$ to make notations more readable):

$$\begin{aligned} d_{\mathbf{u},\mathbf{q}}\mathcal{J}[\mathbf{u}] &= \lim_{h\to 0}\frac{1}{h}(\mathcal{J}[\mathbf{u} + h\mathbf{q}] - \mathcal{J}[\mathbf{u}]) \\ &= \lim_{h\to 0}\frac{1}{2h}\int_{\Omega} [\mathbf{T}(\mathbf{x} + \mathbf{u} + h\mathbf{q}) - \mathbf{R}(\mathbf{x})]^2 - [\mathbf{T}(\mathbf{x} + \mathbf{u}) - \mathbf{R}(\mathbf{x})]^2\, d\mathbf{x} \\ &= \lim_{h\to 0}\frac{1}{2h}\int_{\Omega} [\mathbf{T}(\mathbf{x} + \mathbf{u} + h\mathbf{q}) - \mathbf{T}(\mathbf{x} + \mathbf{u})]\,[\mathbf{T}(\mathbf{x} + \mathbf{u} + h\mathbf{q}) + \mathbf{T}(\mathbf{x} + \mathbf{u}) - 2\mathbf{R}(\mathbf{x})]\, d\mathbf{x} \end{aligned} \tag{3.27}$$

Thus, using Taylor expansion of $\mathbf{T}(\mathbf{x} + \mathbf{u} + h\mathbf{q})$, we get:

$$\begin{aligned} d_{\mathbf{u},\mathbf{q}}\mathcal{J}[\mathbf{u}] &= \lim_{h\to 0}\frac{1}{2h}\int_{\Omega} \left[\nabla_{\mathbf{u}}\mathbf{T}^T(\mathbf{x} + \mathbf{u})h\mathbf{q} + h\mathcal{O}(\mathbf{q})\right][\mathbf{T}(\mathbf{x} + \mathbf{u} + h\mathbf{q}) + \mathbf{T}(\mathbf{x} + \mathbf{u}) - 2\mathbf{R}(\mathbf{x})]\, d\mathbf{x} \\ &= \int_{\Omega} \langle \nabla_{\mathbf{u}}\mathbf{T}(\mathbf{x} + \mathbf{u})\,[\mathbf{T}(\mathbf{x} + \mathbf{u}) - \mathbf{R}(\mathbf{x})], \mathbf{q}\rangle_{\mathbb{R}^d}\, d\mathbf{x} \end{aligned} \tag{3.28}$$

As a result, it comes that:

$$\nabla_{\mathbf{u}}\mathcal{J}\left[\mathbf{u}^k\right] = \nabla_{\mathbf{u}}\mathbf{T}(\mathbf{x} + \mathbf{u})\,[\mathbf{T}(\mathbf{x} + \mathbf{u}) - \mathbf{R}(\mathbf{x})]$$

3.4.3 Quasi-Newton Method - *l*-BFGS

In computationally demanding problems, second order descent methods are often chosen. Newton algorithm is a widespread solution for unconstrained optimization schemes [75], [78]. The Hessian matrix must then be explicitly computed "from scratch", for each update of $\mathbf{u}$. The iterative scheme is then given by:

$$\mathbf{u}^{k+1} = \mathbf{u}^k - \rho\mathcal{H}^{-1}\nabla_{\mathbf{u}}\mathcal{J}\left[\mathbf{u}^k\right] \tag{3.29}$$

Yet, calculating the Hessian inverse may turn to a high complexity problem while handling large images. A Gauss-Newton approximation of $\mathcal{H}$ could be computed at each iteration as proposed in [75] in order to significantly reduce the computational cost. It is achieved by a positive

semi-definite matrix (for a convex objective function) obtained by a second-order Taylor expansion. Still, this approximation is known to be inefficient for very high-dimensional minimization problems. Alternative optimization paradigms have been proposed to this effect, in which the Hessian (or its inverse) is iteratively approximated through gradient information instead of a memory-expensive explicit computation. One of the most common methods is known as limited Broyden-Fletcher-Goldfarb-Shanno (l-BFGS). Other techniques such as Trust Region Methods, DFP (Davidon-Fletcher-Powell) or Pearson can achieve similar results. The reader could refer to [78] (chapters 7 and 8) for a comprehensive overview of Newton and quasi-Newton methods.

In our works, we also adopted the l-BFGS method which approximates the inverse of the Hessian using an initial guess $\mathcal{H}_0$ and a sequence composed of descent directions and gradients [78]. More specifically, the original BFGS technique consisted in yielding an approximation of the Hessian inverse on the basis of the full history of gradients. The improvement of l-BFGS is that only a few (most recent) vectors must be retained (generally $m < 10$), which leads to a much cheaper algorithm in terms of memory.

The Broyden family of operators satisfy the secant equation:

$$\begin{aligned} &\mathcal{H}_k s_k = z_k \\ &z_k = \nabla_{\mathbf{u}} \mathcal{J}\left[\mathbf{u}^{k+1}\right] - \nabla_{\mathbf{u}} \mathcal{J}\left[\mathbf{u}^{k}\right] \end{aligned} \tag{3.30}$$

Since we aim at approximating the inverse of the Hessian, we will phrase the l-BFGS scheme with $\hat{\mathcal{H}} = \mathcal{H}^{-1}$. The descent direction is defined as:

$$\Delta\xi_k = -\hat{\mathcal{H}}_k \nabla_{\mathbf{u}} \mathcal{J}\left[\mathbf{u}^{k}\right] \tag{3.31}$$

The inverse of the Hessian is approximated by:

$$\hat{\mathcal{H}}_{k+1} = (Id - \delta_k s_k z_k^T)\hat{\mathcal{H}}_k(Id - \delta_k z_k s_k^T) + \delta_k s_k s_k^T \tag{3.32}$$

where:

$$s_k = \mathbf{u}^{k+1} - \mathbf{u}^{k}$$

and

$$\delta_k = \frac{1}{z_k^T s_k}$$

Starting with an initial estimate $\hat{\mathcal{H}}_k^0$, the "two-loop recursion" technique can be applied [78]:
For a given k, and with m BFGS vectors, a sequence of vectors $\{g_{k-m}, g_{k-m+1}..., g_k\}$ is defined such that:

$$\begin{aligned} &g_k = -\nabla_{\mathbf{u}} \mathcal{J}\left[\mathbf{u}^{k}\right] \\ &g_i = (Id - \delta_k z_k s_k^T) g_{i+1} \end{aligned} \tag{3.33}$$

By denoting $\alpha_i = \delta_i s_i^T g_{i+1}$, g_i is calculated from g_{i+1}:

$$g_i = g_{i+1} - \alpha_i z_i \tag{3.34}$$

At the same time, a second sequence of vectors $\{y_{k-m}, y_{k-m+1}..., y_k\}$ is created with $y_i = \hat{\mathcal{H}}_i g_i$. The vectors are indeed constructed in a recursive way: $y_{k-m} = \hat{\mathcal{H}}_k^0 g_{k-m}$. We define

$\beta_i = \delta_i s_i^T y_i$ and compute:

$$y_{i+1} = y_i + (\alpha_i - \beta_i) s_i \tag{3.35}$$

We then get y_k, which corresponds to the approximation of the descent direction at the k^{th} iteration $\Delta\xi_k$. Note that BFGS vectors are only retained with regard to the Hessian approximation if $z_k^T s_k > 0$. For an in-depth explanation of l-BFGS, refer to [78].

3.4.4 Flow of Diffeomorphisms

In the light of [131] (Chapter 11), we provide an example of a diffeomorphism calculation on the *SSD* similarity measure. It is phrased here via the transformation ϕ.

$$\mathcal{J}[\mathbf{u}] = \mathcal{D}^{SSD}[\mathbf{T}(\phi), \mathbf{R}] = \frac{1}{2}\int_\Omega [\mathbf{T}(\phi(\mathbf{x})) - \mathbf{R}(\mathbf{x})]^2 \, d\mathbf{x} \tag{3.36}$$

With L a smoothing operator acting on the $h \circ \phi$ function, a natural scalar-product on diffeomorphism groups is introduced (for a mathematical justification, see [131] in section 12.3):

$$\langle h, k \rangle_\phi = \int_\Omega \left\langle L(h \circ \phi), k \circ \phi^{-1} \right\rangle d\mathbf{x} \tag{3.37}$$

The central idea is to express the Gâteaux derivative of the objective function $\mathcal{J}$ as the scalar-product $\langle \vartheta_\phi, w \rangle_\phi$:

$$d_{\phi,w}\mathcal{J}[\phi] = \langle \vartheta_\phi, w \rangle_\phi \tag{3.38}$$

From (3.28), we have:

$$d_{\phi,w}\mathcal{J}[\phi] = \int_\Omega \langle \nabla_\phi \mathbf{T}(\phi(\mathbf{x})) [\mathbf{T}(\phi(\mathbf{x})) - \mathbf{R}(\mathbf{x})], w(\mathbf{x}) \rangle_{\mathbb{R}^d} \, d\mathbf{x} \tag{3.39}$$

By performing the change of variables: $y = \phi(\mathbf{x})$, we get:

$$d_{\phi,w}\mathcal{J}[\phi] = \int_\Omega \left\langle \nabla_y \mathbf{T}(y) \left[\mathbf{T}(y) - \mathbf{R}(\phi^{-1}(y))\right], w(\phi^{-1}(y)) \right\rangle_{\mathbb{R}^d} \left|\det\left(\nabla\phi^{-1}(y)\right)\right| dy \tag{3.40}$$

Hence, (3.38) is verified provided that $\vartheta_\phi \circ \phi^{-1}$ is solution of:

$$Lg(y) = \nabla_y \mathbf{T}(y) \left[\mathbf{T}(y) - \mathbf{R}(\phi^{-1}(y))\right] \left|\det\left(\nabla\phi^{-1}(y)\right)\right| \tag{3.41}$$

Accordingly, the scheme is composed of three steps:

1. Compute $\theta(y) = \nabla_y \mathbf{T}(y) \left[\mathbf{T}(y) - \mathbf{R}(\phi^{-1}(y))\right] \left|\det\left(\nabla\phi^{-1}(y)\right)\right|$)
2. Solve $Lg(y) = \theta(y)$. L generally consists of the Laplacian operator such that it turns to solve a Poisson equation. This step can be viewed as an implicit regularization.
3. Update: $\phi_{k+1} = \phi_k + \rho g_k \circ \phi_k$

In fact, the last step represents the discretized version of the following PDE:

$$\frac{\partial\phi}{\partial t} = g(t, \phi(t, \mathbf{x})) \tag{3.42}$$

Thus, for $\rho \to 0$, we get that that ϕ is a flow of diffeomorphisms ([131]). More details and examples are provided in Part 2 for the search of unidimensional diffeomorphisms.

3.4.5 Armijo's Backtracking Line-Search

In the above-mentioned methods, a fixed step size ρ was considered. Actually, in steepest gradient algorithm, backtracking line search techniques are employed to yield an optimal step size $\rho^{(k)}$ at each iteration k [78]. The same holds for l-BFGS and analogous second-order numerical solutions. Given a descent direction denoted by $\Delta\xi_k$, the core idea is to find a step size value $\rho^{(k)}$ such that $\mathcal{J}\left[\mathbf{u}^{k+1}\right]$ is smaller "enough" in comparison with $\mathcal{J}\left[\mathbf{u}^{k}\right]$. Beginning from an initialization of $\rho^{(k)}$, as long as the decrease condition of $\mathcal{J}$ is not fulfilled, the step size is multiplied by some pre-selected parameter $\tau \in]0,1[$. Armijo's scheme is described in Algorithm 1.

Result : $\rho^{(k)}$, $\mathbf{u}^{k+1}$, $\mathcal{J}\left[\mathbf{u}^{k+1}\right]$
initialization $\rho^{(k)} = \rho_i$;
while $\mathcal{J}\left[\mathbf{u}^{k+1}\right] > \mathcal{J}\left[\mathbf{u}^{k}\right] - \gamma\rho^{(k)} \left\|\Delta\xi_k\right\|^2$ **do**
  $\mathbf{u}^{k+1} = \mathbf{u}^{k} - \rho^{(k)}\Delta\xi_k$
  evaluate $\mathcal{J}\left[\mathbf{u}^{k+1}\right] = \mathcal{D}\left[\mathbf{T}(\mathbf{u}^{k+1}), \mathbf{R}\right] + \alpha\mathcal{S}\left[\mathbf{u}^{k+1}\right]$
  $\rho^{(k)} \leftarrow \tau\rho^{(k)}$ with $\tau \in]0,1[$;
end

Algorithme 1 : Armijo's backtracking line search.

Typically, we set $\rho_i = 1$, $\gamma = 10^{-4}$ and fix a number of possible step size updates $N_{iter} = 10$ It defines a genuine stopping criteria: if the decrease criteria is not fulfilled after N_{iter} trials, the minimization process is stopped and we can assume that a minima of the objective function has been reached. One could also make use of the more general Wolfe conditions, see [78].

3.4.6 Multi-Resolution Scheme

The minimization process is generally achieved jointly with a multiscale strategy. A Gaussian pyramid is generated from both original images $\mathbf{R}$ and $\mathbf{T}$ with a factor $\frac{1}{2}$ between each level. The re-assignment of pixel values at lower scales is performed by bicubic interpolation. Thus, each output pixel value is a weighted average of pixels in the nearest 4-by-4 neighborhood.

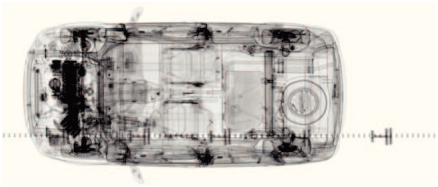
The optimization algorithm is first run on the coarsest level l_c such that an optimal displacement field $\mathbf{u}^{l_c}$ for this specific scale is computed. As a consequence, the upsampled version of $\mathbf{u}^{l_c}$ serves as a strong initialization estimate for the finer upper level $l_c - 1$. As for images downsampling, we resort to bicubic interpolation to upsample the obtained field of displacement $\mathbf{u}^{l}$ for each level $l \leqslant l_c$. The advantages of the technique are twofold [75]:

- Optimization is much less expensive on coarser levels so that a few iterations may be sufficient to compute a minimizer.
- At coarser scales, thin details are not observable, making optimization more robust to spurious and noisy details in the image. As a direct result, issues of convergence at local minima can be overcome.

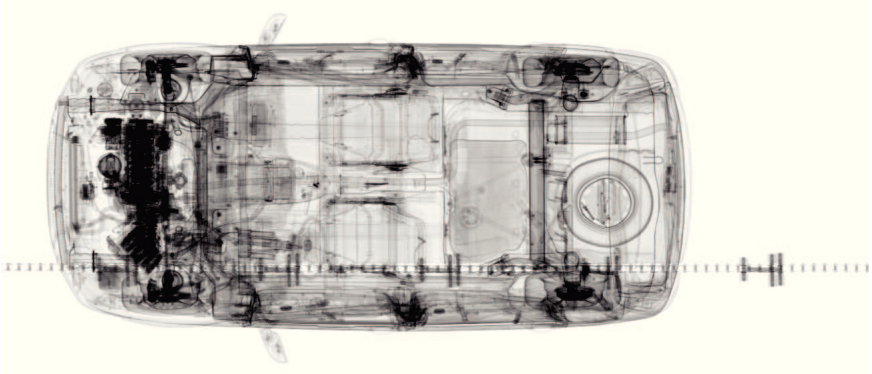
For the different registration schemes developed all along our works, the number of levels was generally set to $l_c = 4$. More complex techniques may be used, as in [63], where a multi-resolution approach based on a Fourier pyramid is put forward.

(a) level 3

(b) level 2



(c) level 1



(d) level 0

Figure 3.8 – Description of a 3-levels Gaussian pyramid generated from a top-view HCVL scan of an inspected vehicle.

3.4.7 Discretization

The central difference scheme is used to approximate the gradient. It actually stems from the sum of the so-called *backward* and *forward* differences. Given h_h the fixed grid spacing (all intervals are of equal size) and a function f, we get:

$$\left(\frac{\partial f}{\partial x}\right)_i = \frac{f_{i+1} - f_{i-1}}{2h_h} + \mathcal{O}(h_h)^2 \tag{3.43}$$

Similarly, a three-points stencil finite difference method is employed to calculate second derivatives:

$$\begin{aligned}\left(\frac{\partial^2 f}{\partial x^2}\right)_i &= \left[\frac{\partial}{\partial x}\left(\frac{\partial f}{\partial x}\right)\right]_i \\ &= \lim_{h_h \to 0} \frac{\left(\frac{\partial f}{\partial x}\right)_{i+1/2} - \left(\frac{\partial f}{\partial x}\right)_{i-1/2}}{h_h} \\ &= \frac{\frac{f_{i+1}-f_i}{h_h} - \frac{f_i - f_{i-1}}{h_h}}{h_h} + \mathcal{O}(h_h) \\ &= \frac{f_{i+1} - 2f_i + f_{i-1}}{h_h^2} + \mathcal{O}(h_h)\end{aligned} \tag{3.44}$$

Regarding a bivariate function $f(x, y)$, which is for instance, the case of the displacement field $\mathbf{u}$ with its two components u_x and u_y, we get the following results:

$$\left(\frac{\partial^2 f}{\partial x \partial y}\right)_{i,j} \approx \frac{\left(\frac{\partial f}{\partial y}\right)_{i+1,j} - \left(\frac{\partial f}{\partial y}\right)_{i-1,j}}{2h_h} \tag{3.45}$$

Besides, with h_v the vertical fixed grid spacing, we have:

$$\left(\frac{\partial f}{\partial y}\right)_{i+1,j} \approx \frac{f_{i+1,j+1} - f_{i+1,j-1}}{2h_v} \tag{3.46}$$

and:

$$\left(\frac{\partial f}{\partial y}\right)_{i-1,j} \approx \frac{f_{i-1,j+1} - f_{i-1,j-1}}{2h_v} \tag{3.47}$$

By combining (3.47) and (3.46), we get:

$$\left(\frac{\partial^2 f}{\partial x \partial y}\right)_{i,j} \approx \frac{f_{i+1,j+1} - f_{i+1,j-1} - f_{i-1,j+1} + f_{i-1,j-1}}{4h_h h_v} \tag{3.48}$$

We choose $h_h = h_v = 1$. For an in-depth review of finite differences methods, the reader is urged to look at the works of [7]. In [75], different discretization schemes are proposed such as nodal, staggered or cell-centered grids, while they must be carefully chosen with respect to the type of smoother to be applied. For the purposes of the present work, cell-centered grid was adopted, providing sound results.

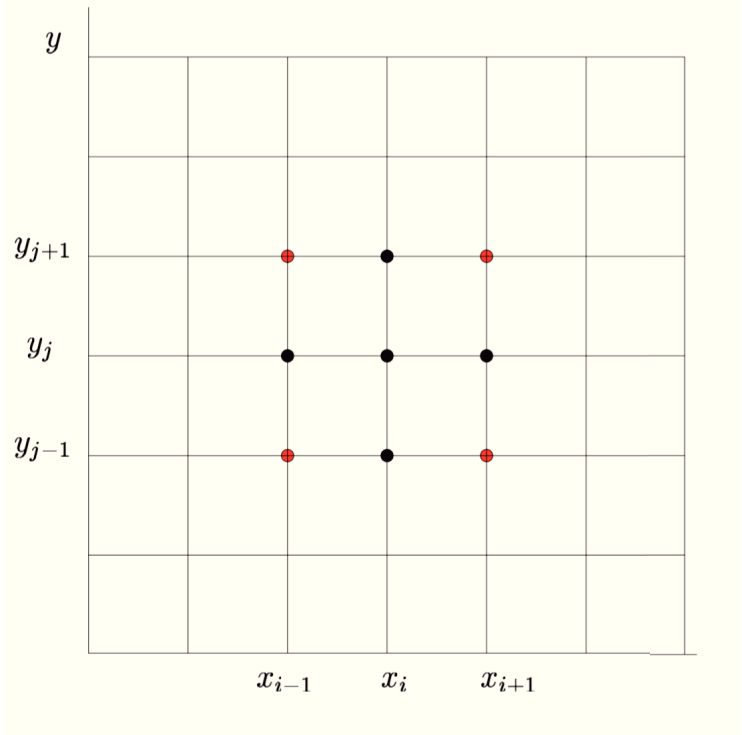
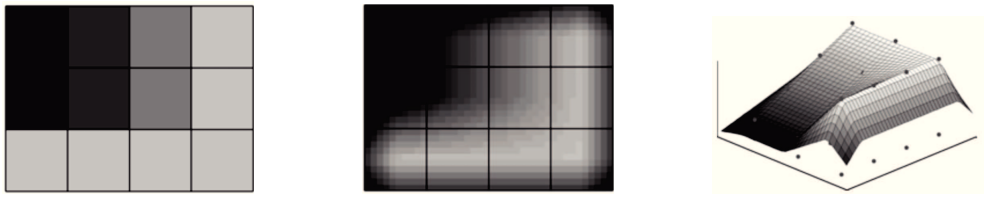


Figure 3.9 – Three-points stencil scheme.

3.4.8 Interpolation

When applying a transform ϕ on the deformable image $\mathbf{T}$, interpolation methods are indeed required to obtain a continuous result. The most straightforward interpolant is the *nearest-neighbor* scheme. Though, it causes discontinuities in the image, which may conspicuously affect the optimization process. Linear interpolation could be a satisfying choice since it is differentiable almost everywhere. Still, at the grid points, this property does not hold anymore, which may turn to a major issue in schemes where derivatives have to be calculated. Hence, regarding our specific applications, a smoother interpolant must be adopted. Consequently, *cubic*, also known as *spline* interpolation is handled. A visual comparison of both *linear* and *cubic* methods is provided in Figures 3.10 and 3.11.

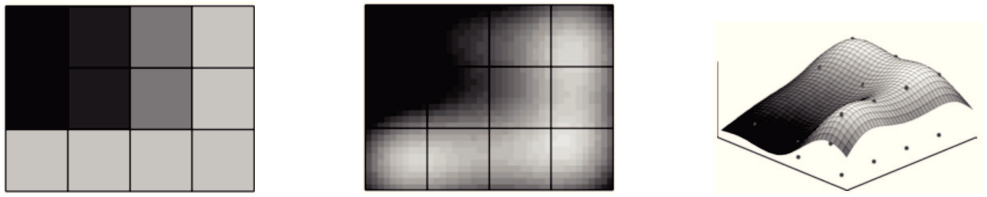


(a) coarse grid (b) finer grid (c) 3D view

Figure 3.10 – Example of linear interpolation ([75]).

Several alternatives are also suggested in the literature. To learn more about interpolation

issues, see a general overview in [76] or [75] and the works of [2] and [114] for a thorough discussion on the subject.



(a) coarse grid (b) finer grid (c) 3D view

Figure 3.11 – Same example as 3.10 with *spline* interpolation ([75]). The output appears much smoother.

3.5 Our Piecewise-Linear Procedure

In this section, we propose a new way to perform a **piece-wise rigid registration** by partitioning both fixed and moving samples into sub-images on the basis of ASM fitted data. Our method can be used to deal with local deformations without resorting to complex non-linear paradigms, particularly computationally expensive. A dense field of displacement vectors is inferred from the matching of extracted feature points, on each partition of the moving image. Our method introduces the application of a regularization scheme to get a smooth field of displacement at the vicinity of sub-images boundaries.

3.5.1 Sub-images

Several methods have been explored to deal with local deformations rather than using a single global transform as described in the previous section. A first idea was to divide the moving image **T** into several parts (generally two or three), and compute the linear transform to be applied on each of them. This procedure has been investigated using both vertical and horizontal partitioning. An alternative approach consisted in using a regular grid and extracting three possible types of rigid transforms: similarity, affine and projective. On each cell of the grid, the transform resulting in the lowest *SSD* between fixed and warped samples was selected.

A similar method has also been studied where several transforms of the same type (affine for instance) are extracted. In an iterative fashion, MLESAC is used to compute a first transform. Then, inliers are filtered out and outliers are kept to compute a new transform of the same type by applying MLESAC again. The procedure is repeated three or four times and a set of transforms is obtained. Corresponding SSD errors are also computed and the final result is given by the local linear warping that minimize the registration error in each cell.

The drawbacks of these last schemes are two-fold: given that a regular grid is used, any object may be split and subject to different transforms in a non-continuous fashion. In addition, detecting foreign objects is the main purpose of the procedure. Therefore, using a "naive" *SSD* minimization criteria may entail to decrease the visual detectability of a potential threat.

Though, as developed in the next part (see Chapter 5), a simpler registration scheme may be employed considering that the reference template **T** is found via ASM and is meant to present a very similar structure when compared to **R**.

Accordingly, in order to compute consistent local parametric transforms, $\mathbf{T}$ and $\mathbf{R}$ are both analogously partitioned. Let us assume that ASM is applied on $\mathbf{R}$ in order to retrieve the image displaying the closest annotated shape within our dataset $\mathbf{T}$. As explained in section 1.5.1, active shape models are fitted to the training images of each class to compute parameters vector $\mathbf{b}$. During this *offline* procedure, the set of points after convergence $\mathbf{x^T} = (x_1, y_1, ..., x_n, y_n)^T$ is also stored in a matrix. Hence, shapes $\mathbf{x^T}$ and $\mathbf{x^R}$ (fitted on $\mathbf{T}$ and $\mathbf{R}$) are available and can be useful to perform a similar partition in both images. In addition, this subdivision is "semantically meaningful" considering that ASM were built by annotating uniformly distributed and most salient objects of the region of interest (*e.g* refrigeration units).

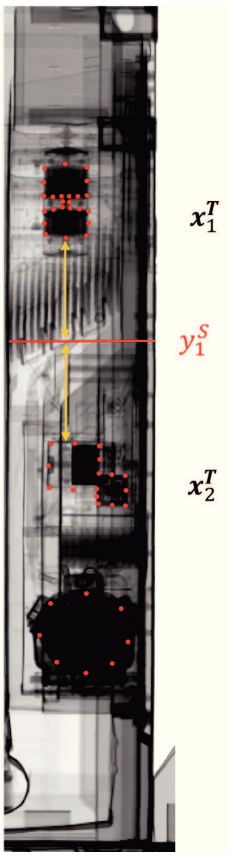


Figure 3.12 – Image partition based on fitted shape points

Let us consider an image $\mathbf{T}$ with its corresponding fitted shape $\mathbf{x^T}$. The c_s connected shapes composing $\mathbf{x^T}$ are defined such that $\mathbf{x^T} = \{\mathbf{x_1^T}, \mathbf{x_2^T}, ...\mathbf{x_c^T}\}$. Given that shapes are annotated in counterclockwise direction from top to bottom, sub-images boundaries could be chosen at equal distance between the maximal position in $\mathbf{x_i^T}$ and the minimal position of $\mathbf{x_{i+1}^T}$ for instance. With s_d the number of wanted subdivisions, the linear boundaries y_i^s, $i \in [\![1, s_d - 1]\!]$:

$$y_i^s = \frac{1}{2}(\max \mathbf{x_i^T} + \min \mathbf{x_{i+1}^T}) \tag{3.49}$$

Indeed, as depicted on Figure 3.12, one or two connected shapes are generally used to create

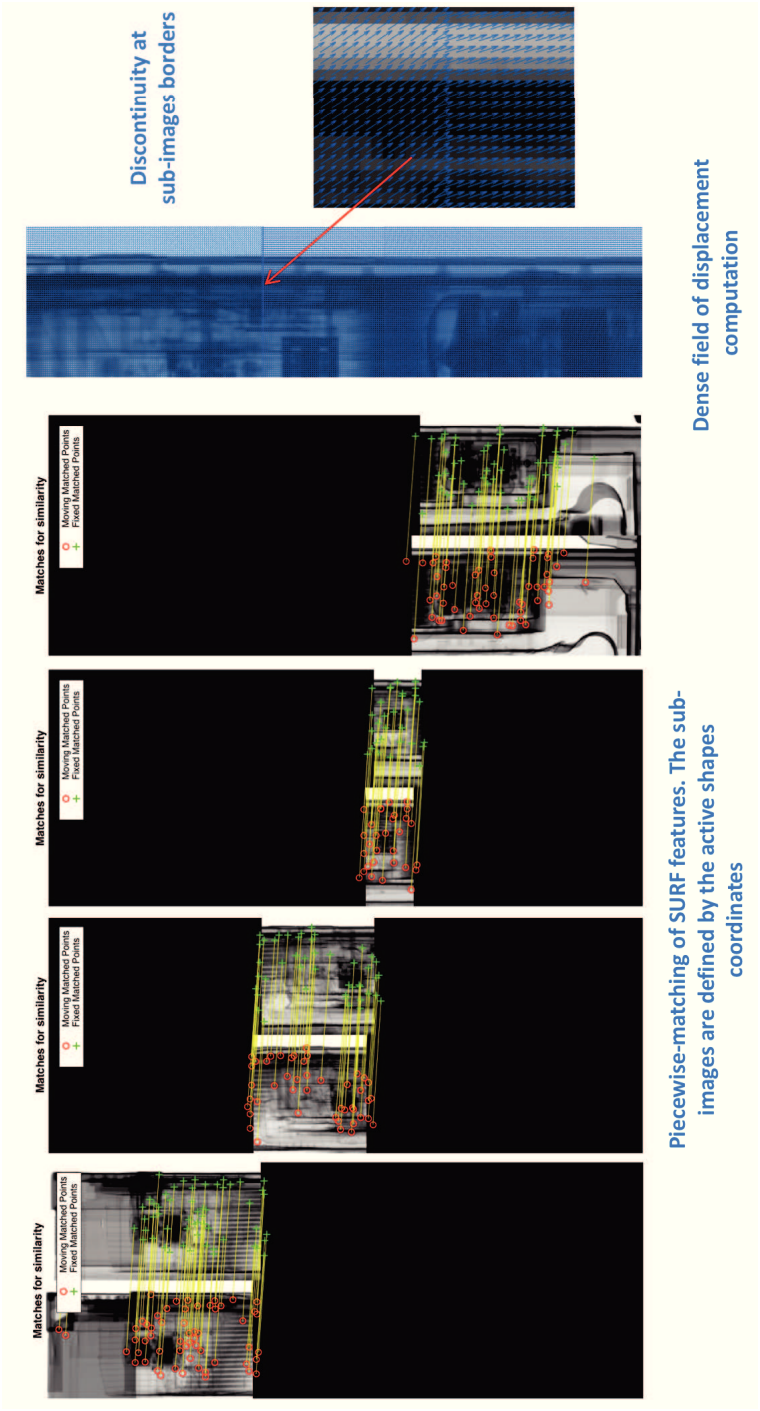


Figure 3.13 – Dense field of displacement extraction and discontinuity description at sub-image borders

a sub-image ($s_d \leqslant c_s$). Thus, we obtain a set of corresponding partitions $\{(\mathbf{T}_i, \mathbf{R}_i)\}_{1 \leqslant i \leqslant s_d}$. For each pair of sub-images, SURF features are extracted and matched through MLESAC filtering, given a wanted type of parametric transform. The parameters of a global transform matrix are thereby calculated, leading to the computation of a dense vector field of displacement $\mathbf{u}_i$. Yet, while stacking the different vector fields into matrices of vertical and horizontal components of displacement, one can observe sharp discontinuities at y_i^s boundaries. The piece-wise linear method (feature-based) along with the discontinuity phenomenon are depicted on Figure 3.13.

3.5.2 Displacement field regularization

This problem is solved by regularizing the vector field in the vicinity of the boundaries y_i^s that separate sub-images. The isotropic direction diffusion scheme developed by Tang, Sapiro and Caselles in [111] was retained to this regard.

From vectors to directions

Given the field of displacement $\mathbf{u} : \mathbb{R}^2 \to \mathbb{R}^d$, a first step is to consider directions instead of standard vectors. More formally, for a d-dimensional problem, an $\mathbb{R}^d$ direction is defined on an image of coordinates $(x, y) \in \mathbb{R}^2$ by a normalized vector:

$$\hat{\mathbf{u}}(x, y) = \frac{\mathbf{u}(x, y)}{\|\mathbf{u}(x, y)\|} \tag{3.50}$$

With $\hat{u}_k : \mathbb{R}^2 \to \mathbb{R}, 1 \leqslant k \leqslant d$ the components of $\hat{\mathbf{u}}$ such that its euclidean norm is equal to 1:

$$\|\hat{\mathbf{u}}\| = \sqrt{\sum_{k=1}^{d} \hat{u}_k^2} = 1 \tag{3.51}$$

The notation can also be simplified by defining $\hat{\mathbf{u}} : \mathbb{R}^2 \to \mathbb{S}^{d-1}$ where $\mathbb{S}^{d-1}$ stands for the unit ball in $\mathbb{R}^d$ (in 2-D, it relates to the unit sphere $\mathbb{S}^1$).

The underlying idea in [111] is to extend diffusion paradigms of grayscale values to images representing normalized vectors (directions). Hence, diffusion is directly applied on the direction space whilst maintaining the unit norm constraint. During the diffusion process, a family of direction images $\hat{\mathbf{u}}(x, y, t) : \mathbb{R}^2 \times [0, \tau) \to \mathbb{S}^{d-1}$ is computed by solving a system of coupled partial differential equations (PDE's).

General problem formulation

The method is widely inspired from the theory on harmonic maps in crystals [111] where the optimization problem is formulated as follows:

$$\min_{\hat{\mathbf{u}}:\mathbb{R}^2 \to \mathbb{S}^{d-1}} \quad \int\int_{\Omega} \|\nabla \hat{\mathbf{u}}\|^p \, dxdy \tag{3.52}$$

where $p \geqslant 1$ and Ω designates the image domain. The euclidean norm of the vectorial image gradient is given by:

$$\|\nabla\hat{\mathbf{u}}\| := \sqrt{\sum_{k=1}^{d}\left(\frac{\partial\hat{u}_k}{\partial x}\right)^2 + \left(\frac{\partial\hat{u}_k}{\partial y}\right)^2} \tag{3.53}$$

Isotropic diffusion

Regarding our X-ray images, the displacement field is composed of two-dimensional vectors, thus $d = 2$. In addition, we resort to the isotropic diffusion scheme such that $p = 1$ is picked. To make notations clearer, $\mathbf{u}$ and $\hat{\mathbf{u}}$ are redefined as follows:

$$\mathbf{u} = (u_x, u_y) \tag{3.54}$$

$$\hat{\mathbf{u}} = (\hat{u}_x, \hat{u}_y) \tag{3.55}$$

In the isotropic case, the gradient descent flow of (3.52) is expressed by the set of coupled PDE's:

$$\frac{\partial\hat{u}_{x,y}}{\partial t} = \Delta\hat{u}_{x,y} + \hat{u}_{x,y}\,\|\nabla\hat{\mathbf{u}}\|^2 \tag{3.56}$$

where $\Delta\hat{u}_x$ stands for the x-component Laplacian given by:

$$\Delta\hat{u}_x = \frac{\partial^2\hat{u}_x}{\partial x^2} + \frac{\partial^2\hat{u}_x}{\partial y^2} \tag{3.57}$$

The same holds for the vertical component $\hat{u}_y$. From equation (3.56), a solution is iteratively approached:

$$\begin{cases} \hat{u}_{x,y}(t+1) = \hat{u}_{x,y}(t) + \Delta t\left(\Delta\hat{u}_{x,y}(t) + \hat{u}_{x,y}(t)\,\|\nabla\hat{\mathbf{u}}\|^2\right) \\ \text{s.t } \|\hat{\mathbf{u}}(t)\| = 1 \\ \text{with initial conditions: } \hat{\mathbf{u}}(t=0) = \frac{\mathbf{u}}{\|\mathbf{u}\|} \end{cases} \tag{3.58}$$

Actually, in order to maintain the diffusion at the vicinity of the boundaries, a weight is applied such that other regions do not undergo the smoothing process. Since images are partitioned vertically, we use a unidimensional (non-binary) Gaussian mask for this purpose (see Figure 3.14). Equation 3.58 becomes thus:

$$\begin{cases} \hat{u}_{x,y}(t+1) = \hat{u}_{x,y}(t) + \Delta t\sum_{i=1}^{s_d-1} G(y, y_i, \sigma)\left[\Delta\hat{u}_{x,y}(t) + \hat{u}_{x,y}(t)\,\|\nabla\hat{\mathbf{u}}\|^2\right] \\ \text{s.t } \|\hat{\mathbf{u}}(t)\| = 1 \\ \text{ and with initial conditions: } \hat{\mathbf{u}}(t=0) = \frac{\mathbf{u}}{\|\mathbf{u}\|} \end{cases} \tag{3.59}$$

where $G(y, y_i, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(y-y_i)^2}{2\sigma^2}\right), i \in [\![1, s_d - 1]\!]$.

Figure 3.14 – Mask composed of Gaussians as described in equation 3.59. It is used to confine the diffusion process to sub-images boundaries.

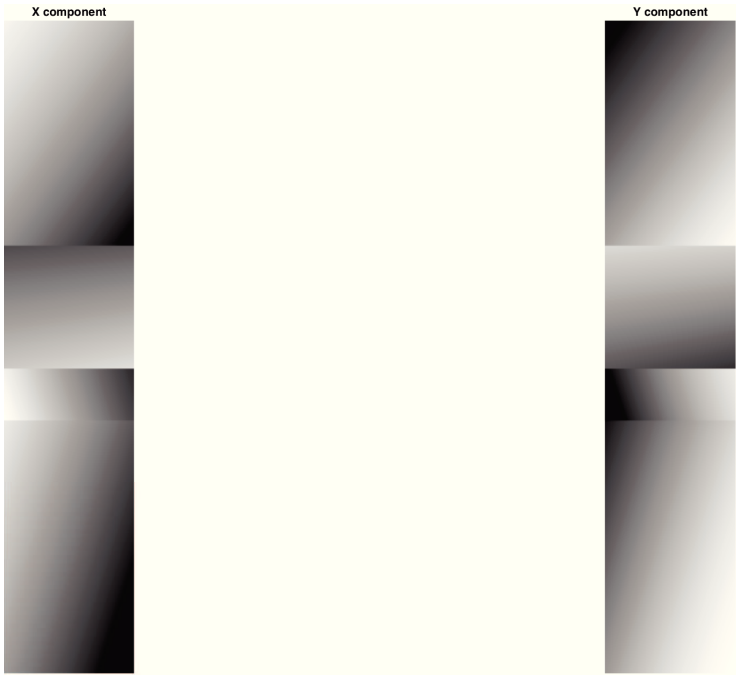
The descent step size Δt is kept constant to 0.1. σ designates the number of rows impacted by the diffusion above and below each boundary y_i, it is set to $\sigma = 20$. Moreover, the number of iterations N_{it} for which a stable solution $\mathbf{u}_f$ is considered to be reached, is fixed to a few dozens ($N_{it} = 50$ is generally chosen).

In the course of the diffusion process, we deal with normalized vectors (referred to as directions). The final result $\mathbf{u}_f(x, y)$ can therefore be obtained through the multiplication of the direction vector by the initial norm of the displacement:

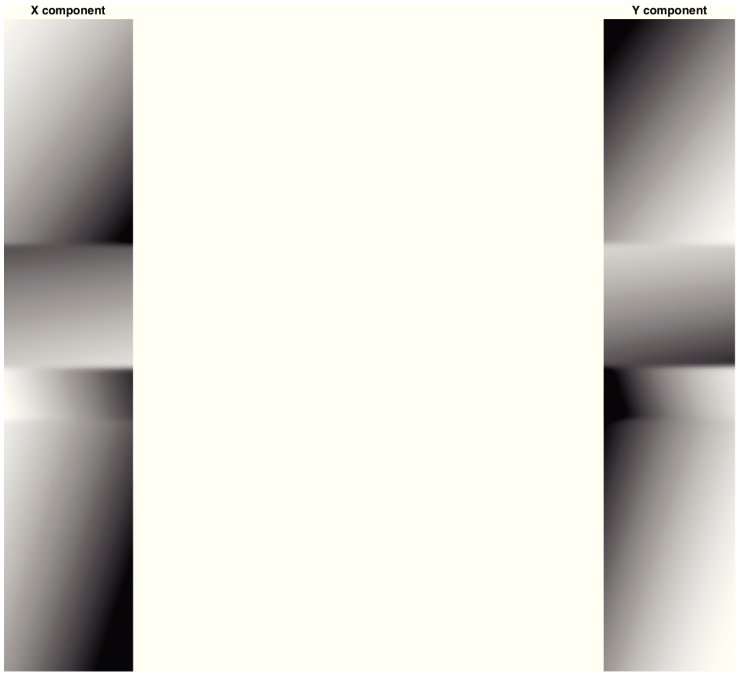
$$\mathbf{u}_f(x, y) = \hat{\mathbf{u}}(x, y, t_f) \, \|\mathbf{u}(x, y)\| \tag{3.60}$$

In other words, the vector amplitude is assigned back to each component with a new proportion, obtained from the diffusion mechanism. In Figure 3.15, the evolution of the displacement field throughout the diffusion process is described on the basis of Figure 3.13 example. The regularization is maintained to the vicinity of the sub-images boundaries while $\mathbf{u}$ is kept unchanged in remote areas. The smoothing impact on the deformed image is directly illustrated on Figure 3.16.

Note that alternative approaches such as the one proposed by Tschumperlé *et al.* [117] could have been adopted as well.



(a) t = 0



(b) t = 50

Figure 3.15 – Visual description of the smoothing process on the components of normalized displacement vectors at different stages. The preservation of the vector field can be noticed in remote areas with respect to sub-images boundaries.

(a) **T** before diffusion

(b) **T** after diffusion

(c) zoom before diffusion

(d) zoom after diffusion

Figure 3.16 – Visual effect of the smoothing process on discontinuities in the deformed image **T**.

3.6 Conclusion

In this chapter, the main purpose was to introduce the different registration methods and frameworks we resorted to, all along our work. Linear registration was presented, as well as our new approach combining a piecewise-rigid paradigm with a diffusion smoothing at sub-images borders.

Non-linear techniques were put forward, specifically, through the variational framework. The principles behind parametric and non-parametric schemes were clearly outlined. Also, a general overview of data fitting and regularization terms is provided, side-by-side with numerical resolution methods. Finally, implementation details concerning discretization strategies and interpolants are supplied.

Chapter 4

Change Detection

Outline of the current chapter

4.1 Introduction

In the present chapter, we will describe the methods adopted to compute a map of differences between the transformed image denoted $\mathbf{T}_{\mathbf{u}^*}$ and the target scan $\mathbf{R}$ ($\mathbf{u}^*$ being an optimizer of the minimization scheme (3.9)). Through the map of differences, the intention is to provide a description of possible changes to be remarked. We introduce two similarity measures adapted to the X-ray imaging context and the specificity of our irregularity detection applications. Image differencing (simple differences) is thus compared to our image rationing scheme (using the ratio of images), with respect to the visual detectability of an irregularity. We will demonstrate, on both visual and physical aspects, that the proposed image rationing approach leads to more accurate results than a simpler image differencing technique.

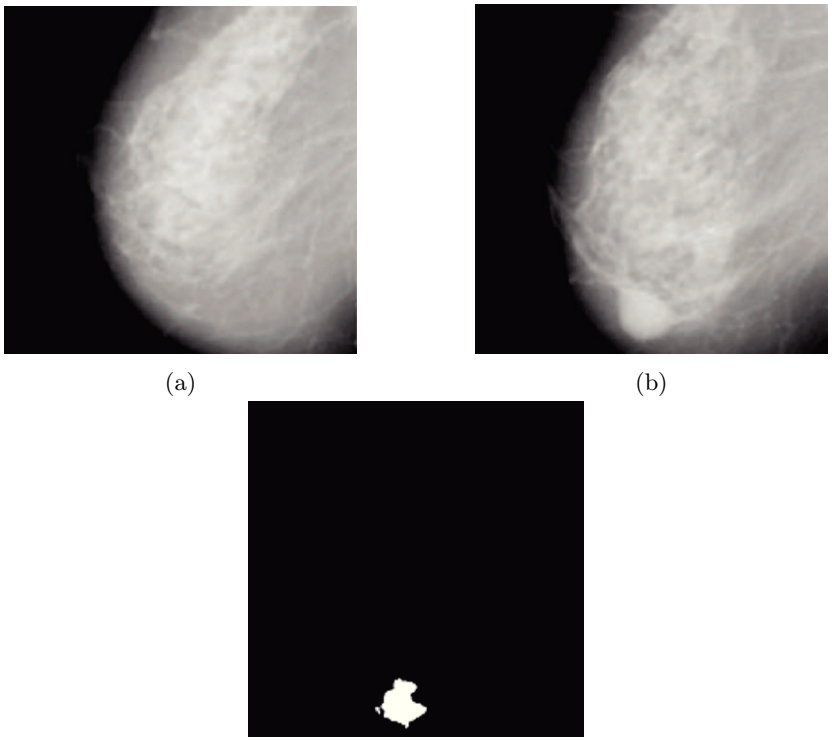
Many applications require a full automation of the irregularity detection process. In such cases, the map of differences is segmented in a binary fashion, filtered through morphological

operations and a bounding box locating the anomaly spots can be set on the inspected image **R**.

The employed binary segmentation algorithm along with morphological operators are briefly presented in the following sections, and illustrated through an applicative example.

4.2 Difference Measures

Beyond image registration algorithms, change detection can be considered as a field of research by itself. It is intensively explored in medical image analysis, where CT or MRI scans of a subject, acquired at distinct times of a treatment are registered and undergo a difference detection process (*e.g* see Figure 4.1). Among other things, the technique enables to diagnose tumors and monitor their evolution over time [95], [98], [123].



(a) (b)

(c)

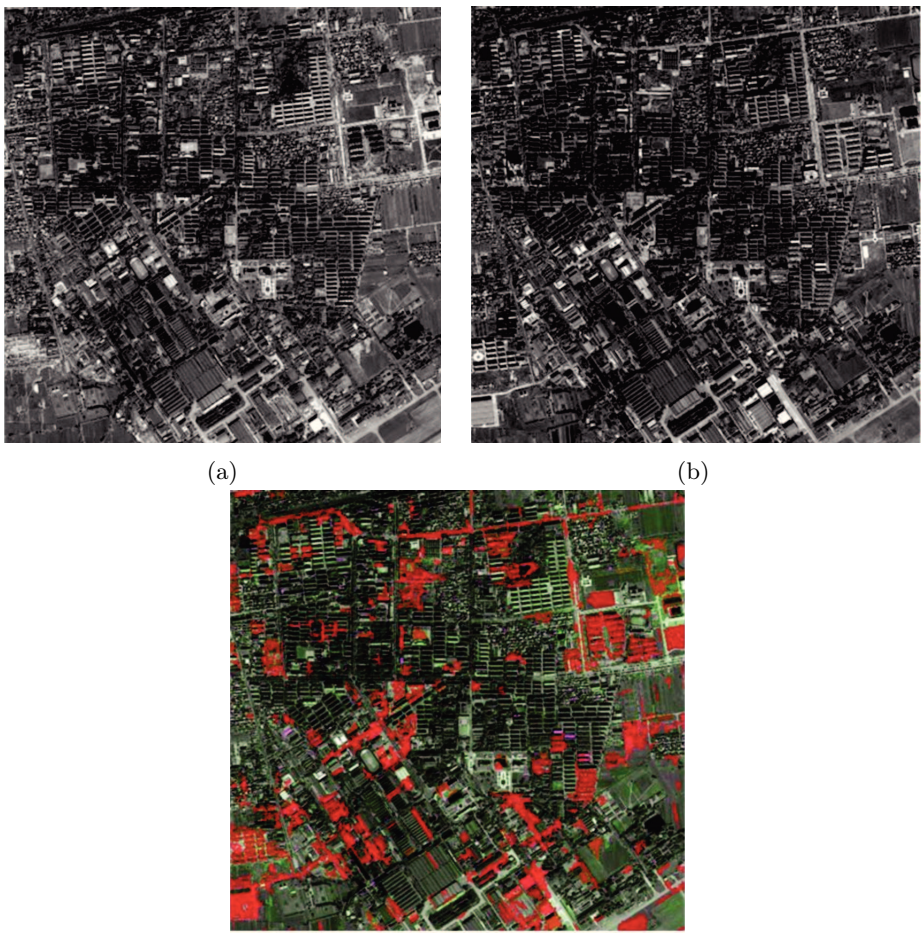
Figure 4.1 – Abnormality detection in mammography. (a) Source image representing a healthy breast; (b) image to be analyzed; (c) detection result through a mixture-based method ([47])

Regarding earth science, agriculture, driving assistance or security-related applications [85], significant means were also employed to develop difference detection techniques on collected data. Earliest applications focused essentially on industrial inspection of manufacturing defects, *e.g* for electronic circuits. Besides video surveillance tasks, remote sensing could definitely be considered as the most investigated case of research in change detection (as illustrated on Figure 4.2).

Given a set of airborne images of the same scene, acquired at different times, the technique lies in the identification of major changes across the various samples. A very rich literature has

been dedicated to elaborate difference detection methodologies and techniques. Hussain *et al.* contributed in a comprehensive overview of existing approaches for remotely-sensed images in [54]. Pixel-based traditional techniques are reported along with newer object-based schemes, while specific limitations and advantages of each approach are listed. Yet, most of the reviewed methods are specifically oriented to remote sensing, making use of particular sensors and multi-spectral channels that are beyond our scope.

Earlier, Radke *et al.* proposed a generic survey of change detection algorithms in [85] where geometric or intensity pre-processing adjustments are also suggested. A few simple and intuitive commonly-utilized techniques are to be reported. Here, we will compare the elementary image differencing scheme versus the so-called rationing approach.



(a) (b)

(c)

Figure 4.2 – Remote sensing images of Beijing taken at two different times (a) and (b); (c) the difference detection result ([80]).

Generally, the most intuitive method for change estimation between two images $\mathbf{I}_1$ and $\mathbf{I}_2$ is to compute an absolute difference:

$$\mathbf{D}(\mathbf{x}) = |\mathbf{I}_1(\mathbf{x}) - \mathbf{I}_2(\mathbf{x})| \tag{4.1}$$

In section 4.5, given the nature of our X-ray images and the specific task to be addressed (irregularity detection), we contribute in introducing a more convenient measure for the detection of differences.

4.3 Binary Segmentation

In some applications, the only visualization of differences is not sufficient. When aiming at detection automation, the output map of difference intensities is segmented in a binary image $\mathbf{B}$. Specific blobs can be identified, indicating notable changes under some defined criteria (connected component analysis). For any $\mathbf{x} \in \Omega$ we have:

$$\mathbf{B}(\mathbf{x}) = \begin{cases} 1 \text{ if } \mathbf{D}(\mathbf{x}) \geqslant \tau \\ 0 \text{ otherwise} \end{cases} \tag{4.2}$$

The threshold τ can be fixed or computed automatically. As motivated in section 4.5, we will use an elementary global thresholding approach. Usually, the core idea is to analyze the histogram of difference intensities and find two or more distinct modes relating to the background and foreground. Probabilistic mixture models are often adopted [85], [47] as well as Markov Random Fields models (MRF) that have also gained popularity [109], [118]. As far as our tasks are concerned (see section 5.8), we resort to the simple Otsu's thresholding method which gives comparable results to a two-class K-means clustering. Furthermore, the global property of Otsu's approach is compensated with the use of hysteresis thresholding.

4.3.1 Otsu's Thresholding

Let us consider the image of differences $\mathbf{D}$. Two modes are to be identified in its histogram of intensities; the left distribution relates to a low change while the right corresponds to a more pronounced difference. The problem is about automatically finding an optimal threshold τ that best separates these two modes, on the one hand the background (where $\mathbf{R} \approx \mathbf{T}_{\mathbf{u}^*}$) and the irregularity detection spots ($\mathbf{R} \neq \mathbf{T}_{\mathbf{u}^*}$) on the other hand. In order to obtain a good separation, the threshold selection should ensure:

1. a small variation of the intensity values within each class
2. a large variation of *interclass* grayscale values.

Actually, the first criteria consists in minimizing the *intraclass* variance $\sigma^2_{intra}(t)$, defined by a weighted sum of the variances of each class:

$$\sigma^2_{intra}(t) = p_e(t)\sigma^2_e(t) + p_d(t)\sigma^2_d(t) \tag{4.3}$$

where the indices 'e' and 'd' respectively refer to the cluster where $\mathbf{R}$ and $\mathbf{T}_{\mathbf{u}^*}$ are considered as equal or different enough. The probability and variance of each class are computed from the histogram and are denoted by p and σ.

By maximizing the *interclass* variance $\sigma^2_{betw}(t)$, the second criteria is fulfilled. With μ, μ_e and μ_d respectively the mean of the whole image, the left and right clusters, the between-class variance is expressed as:

$$\sigma^2_{betw}(t) = p_e(t)(\mu_e(t) - \mu)^2 + p_d(t)(\mu_d(t) - \mu)^2 \tag{4.4}$$

The optimal threshold τ is found by a sequential search over the different t values minimizing $\sigma^2_{intra}(t)$. Note that maximizing $\sigma^2_{betw}(t)$ leads to the same outcome [74].

Concerning the various tasks we had to address, precise specifications were given on the sought objects intensities for the regions of interest analysis. Consequently, a minimal difference threshold τ_{min} can be set and the final threshold τ^* is chosen such that:

$$\tau^* = \max(\tau, \tau_{min}) \tag{4.5}$$

By applying Otsu's method on the map of differences of Figure 4.6, we get the results depicted on the histogram of Figure 4.10 and the binary mask in Figure 4.11 (a).

4.3.2 Hysteresis Thresholding

In order to produce a mask depicting more precisely anomaly regions, a global thresholding technique like Otsu's may not be efficient enough. A famous approach known as hysteresis thresholding is used to expand the identified regions to connected pixels of lower intensity values. This is performed without increasing the false alarm rate as a global method would entail.

Broadly employed for edge detection tasks, it consists in defining two thresholds: T_{low} and T_{high} with $T_{low} < T_{high}$. Given an image $\mathbf{I}$, all pixels below T_{high} are discarded in an initial stage and referred to as *weak* pixels. On the other hand, pixels above T_{high} are designated as *strong* elements. Then, for each *strong* pixel p, located at the edge of a region, all directions are analyzed and any *weak* position, connected to p with an intensity such that $T_{low} \leqslant \mathbf{I}(p) \leqslant T_{high}$ is labelled as *strong* and reassigned to its corresponding connected region.

In our work, an eight-connectivity scheme is adopted. While $T_{high} = \tau^*$, T_{low} is generally chosen such that $T_{low} = \tau^* - \theta$ with $\theta \in]0, 0.1]$. An illustration is provided in Figure 4.11.

4.4 Morphological Operators

After obtaining a mask of differences, the filtering of spurious details is performed by morphological operators. Binary morphology, introduced by G. Matheron and J. Serra [73], is a very complex field of research, and has now been extended to grayscale, color images, videos and graphs. It has very strong theoretical foundations that are beyond the scope of our work. This section only intends to shortly review the main operators we were brought to make use of, to smooth out small details in the binary map and impose specific objects sizes and shapes.

From a general point of view, morphology is referred to as set of image processing operations that process binary images based on shapes. By applying a *structuring element* to the input image, output pixels values are modified by their respective pixels vicinity.

4.4.1 Structuring Element

Let us concentrate on the binary case and consider $E = \mathbb{Z}^2$. Formally, we denote B a subset of E; it actually corresponds to a binary mask that allows the definition of arbitrary neighborhood structures for diverse morphological operations. B is known as the *structuring element*, used to probe an input image with a simple pre-defined shape. Most popular *structuring elements* are the the "cross", square, disk or "diamond" masks.

Let us have $x \in E$, B_x designates the translation of B by x:

$$B_x = \{b + x \mid b \in B\} \tag{4.6}$$

Erosion and dilation represent the most basic morphological operators. Roughly speaking, dilation adds pixels to objects boundaries, making regions more visible. Inversely, erosion is about removing pixels at blobs boundaries; it is especially useful to smooth out small isolated items.

4.4.2 Standard Operators

Erosion

We consider the binary image A defined as a subset of E, $x \in E$, morphological erosion is mathematically expressed via the Minkowski addition:

$$A \ominus B = \{x \mid B_x \subset A\} \tag{4.7}$$

The idea is therefore to keep only pixels $x \in E$ such that the translation of B by x fits inside A. It can also be phrased as follows:

$$A \ominus B = \bigcap_{b \in B} A_{-b} \tag{4.8}$$

Dilation

The dilation operation performs the opposite of erosion. It amounts to taking the union of copies of the translated *structuring element* Bx, centered at every pixel location x in the foreground.

$$A \oplus B = \{x + b \mid b \in B, x \in A\} \tag{4.9}$$

It is often expressed in the following way:

$$A \oplus B = \bigcup_{x \in A} B_x \tag{4.10}$$

Various operators derive from the combination of erosion and dilation such as the widely-used opening and closing.

Opening

While aiming at smoothing out small objects without affecting bigger regions, the opening operator is employed. The opening of A and B is composed by the initial erosion of A by B followed by a dilation through the same *structuring mask* B. Generally, the opening operator is symbolized by $\circ$, and we have:

$$A \circ B = (A \ominus B) \oplus B \tag{4.11}$$

Closing

Correspondingly, the closing operation is generally employed to fill holes in the input binary image or join disconnected neighboring regions. It results from the application of a dilation on A by B and a further erosion of the output image by B. It is written as:

$$A \bullet B = (A \oplus B) \ominus B \tag{4.12}$$

4.5 Our Difference Measure

In this section, we propose two measures to estimate the presence of irregularities following the registration of our X-ray samples. The first is a simple image differencing approach, adapted to the specificity of the addressed detection tasks. The second technique consists in calculating a pixel-wise division of both images under particular constraints. We show that the second technique is more appropriate with regards to change detection capacity, and is more consistent with the very physics of X-ray imaging.

4.5.1 Simple Differences Computation

For notation simplification, the registered source image is denoted $\mathbf{T}(\mathbf{x} + \mathbf{u}^*(\mathbf{x})) := \mathbf{T}_{\mathbf{u}^*}(\mathbf{x})$. Following the registration process, the difference between both deformed and target images, respectively $\mathbf{T}_{\mathbf{u}^*}$ and $\mathbf{R}$, is computed as follows, for any $\mathbf{x} \in \Omega$:

$$\mathbf{D}(\mathbf{x}) = \begin{cases} \mathbf{T}_{\mathbf{u}^*}(\mathbf{x}) - \mathbf{R}(\mathbf{x}) \text{ if } \mathbf{T}_{\mathbf{u}^*}(\mathbf{x}) \geqslant \mathbf{R}(\mathbf{x}) \\ 0 \text{ otherwise} \end{cases} \tag{4.13}$$

As a matter of fact, this choice is motivated by a natural observation. When adding a supplementary object within the freight, the density raise leads to a darker image response, *i.e* a lower intensity value. The quantification of the difference should therefore be performed by subtracting the potentially darker image under inspection $\mathbf{R}$ (lower grayscale value) from the brighter reference $\mathbf{T}_{\mathbf{u}^*}$ (higher grayscale value). Moreover, we only aim at identifying changes stemming from the insertion of a foreign object within the imaged subject $\mathbf{R}$. Accordingly, any denser region in $\mathbf{T}_{\mathbf{u}^*}$ is viewed as irrelevant with respect to the detection application.

Note that $\mathbf{R}$ and $\mathbf{T}$ are normalized throughout the registration process resulting in a difference $\mathbf{D}$ ranging from 0 (for similar regions) to 1 (for extreme changes).

4.5.2 Image Rationing

The measure of differences via the computation of a ratio is a popular alternative for many applications, especially when robustness to illumination is sought (see [85] for further details). $\mathbf{D}$ is then phrased as:

$$\mathbf{D}(\mathbf{x}) = \begin{cases} \frac{\mathbf{R}(\mathbf{x})}{\mathbf{T}_{\mathbf{u}^*}(\mathbf{x})} \text{ if } \mathbf{T}_{\mathbf{u}^*}(\mathbf{x}) \geqslant \mathbf{R}(\mathbf{x}) \text{ and } \mathbf{T}_{\mathbf{u}^*}(\mathbf{x}) \neq 0 \\ 1 \text{ otherwise} \end{cases} \tag{4.14}$$

In order to obtain an error increase as the difference between both images grows, we rather expressed (4.14) in the following way:

$$\mathbf{D}(\mathbf{x}) = \begin{cases} 1 - \frac{\mathbf{R}(\mathbf{x})}{\mathbf{T}_{\mathbf{u}^*}(\mathbf{x})} \text{ if } \mathbf{T}_{\mathbf{u}^*}(\mathbf{x}) \geqslant \mathbf{R}(\mathbf{x}) \text{ and } \mathbf{T}_{\mathbf{u}^*}(\mathbf{x}) \neq 0 \\ 0 \text{ otherwise} \end{cases} \tag{4.15}$$

in compliance with the simple differences method, we only focus on changes that emerge from $\mathbf{R}$. Actually, dividing $\mathbf{T}_{\mathbf{u}^*}(\mathbf{x})$ by $\mathbf{R}(\mathbf{x})$ is physically sound as described in the coming section. It

is indeed inspired from a very efficient merging methodology utilized to build large databases of X-ray images towards CNNs training.

4.5.3 Threat Merging Analogy

As developed in several studies, one can "artificially" insert any object, previously scanned with a blank background (or obtained by simulation), within a given cargo or luggage image I. See [90], [55] or [57] for example, where the *Threat Image Projection* (*TIP*) method is presented. This technique has proven to be valuable for the training of inexperienced custom officers. For research purposes, it is widely adopted in recognition tasks; a CNN can be fine-tuned on large databases made of synthetic images by merging threat patches into various backgrounds and with different orientations (Figure 4.5). Different sorts of mergeable threat images are described on Figure 4.4. The approach consists in simply multiplying I by the threat patch I_{th} at a desired location.

This choice has indeed a strong justification stemming from the physical features of X-ray attenuation. The basic principle of X-ray imaging is the measurement of the X-ray photons which are attenuated when traversing matter. This process is modelled by the Beer-Lambert law, which can be approximated as follows:

$$I_d = I_0 \exp\left(-\sum_{i=1}^{N} \mu_i x_i\right) \tag{4.16}$$

Where I_d is the beam intensity at the detection point after traversing N distinct materials, I_0 stands for the beam intensity at the initial point, and therefore coincides with the white background in our X-ray scans where no absorption occurs. Since most of our data is composed of 16 bit images, the intensity value at these spots is about 65 535. μ_i and x_i stand respectively for the linear attenuation coefficient and thickness of each material. According to (4.16), the beam intensity at the detector point corresponds to the combination of the attenuation due to each material successively penetrated by the X-ray beam (Figure 4.3). Hence, it can be decomposed as:

$$\begin{aligned} I_d &= I_0 \prod_{i=1}^{j} \exp\left(-\mu_i x_i\right) \prod_{i=j+1}^{N} \exp\left(-\mu_i x_i\right) \\ &= I_b \exp\left(-\mu_{th} x_{th}\right) \end{aligned} \tag{4.17}$$

The first product term, denoted I_b corresponds to the beam intensity attenuated by the background (non-illicit cargo without threat constituted of j materials). The second term is written $\exp\left(-\mu_{th} x_{th}\right)$ where the parameters μ_{th} and x_{th} represent the characteristics of the threat element (constituted of $N - j$ materials).

Then, by injecting in (4.17) the additional measure of the incident beam intensity attenuated by the threat I_{th} it comes:

$$I_d = I_b \frac{I_{th}}{I_0} \tag{4.18}$$

with

$$I_{th} = I_0 \exp\left(-\mu_{th} x_{th}\right) \tag{4.19}$$

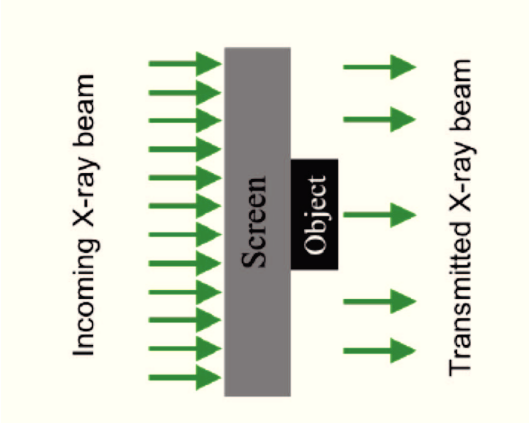


Figure 4.3 – Attenuation of X-Ray beam in materials

Usually, the merged items patches are normalized to be applied on the non-illicit freight by a pixel-wise multiplication ($I_0 = 1$).

In the same way that a term-by-term multiplication is applied to "artificially" insert an illicit object within a scan, a difference could be detected by computing a term-by-term division between the detection and background images I_d and I_b at any pixel:

$$I_{th} = \frac{I_d}{I_b} \tag{4.20}$$

In our case, we consider the image under inspection $\mathbf{R}$ containing a potential threat and the registered empty reference $\mathbf{T}_{\mathbf{u}^*}$ (viewed as pure background against the searched difference), it comes for every $\mathbf{x} \in \Omega$:

$$\mathbf{D}_{th} = \frac{\mathbf{R}(\mathbf{x})}{\mathbf{T}_{\mathbf{u}^*}(\mathbf{x})} \tag{4.21}$$

This expression reflects the attenuation specifically caused by the irregularity (potential threat) $\mathbf{D}_{th}$ as expressed in (4.14).

(a) Gunfire

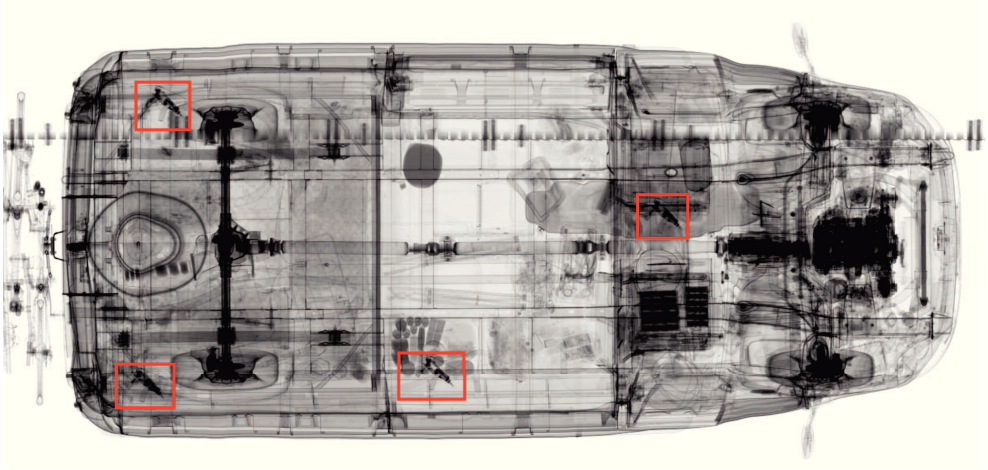
(b) Revolver

(c) Hand-grenade

(d) Mortar

Figure 4.4 – Different sorts of lethal weapons scanned without background by the HCVL system.

(a) Skorpion



(b) Merging

Figure 4.5 – Merging of a Skorpion gun within a loaded vehicle imaged by the HCVL. This procdure enables the insertion of threats in very diverse backgrounds with the objective of data augmentation for CNNs training

4.5.4 Visual and Numerical Results

For illustration purposes, two examples are brought to justify our choice of resorting to rationing instead of simple differencing method. In a first registration case, presented in Figure 4.6 the goal is to find narcotics hidden in a refrigeration unit by analyzing its X-Ray image (HCVS). The second example focuses on HCVL scans of vehicles where several weapons were intentionally smuggled; It is described on Figure 4.7. In Figures 4.8 and 4.9 the graphs of one-dimensional cross-sections of the maps of differences are given for both methods.

In the light of visual and numerical observations, the rationing technique utterly enhances the change intensity and is therefore considered as a prime measure for change evaluation.

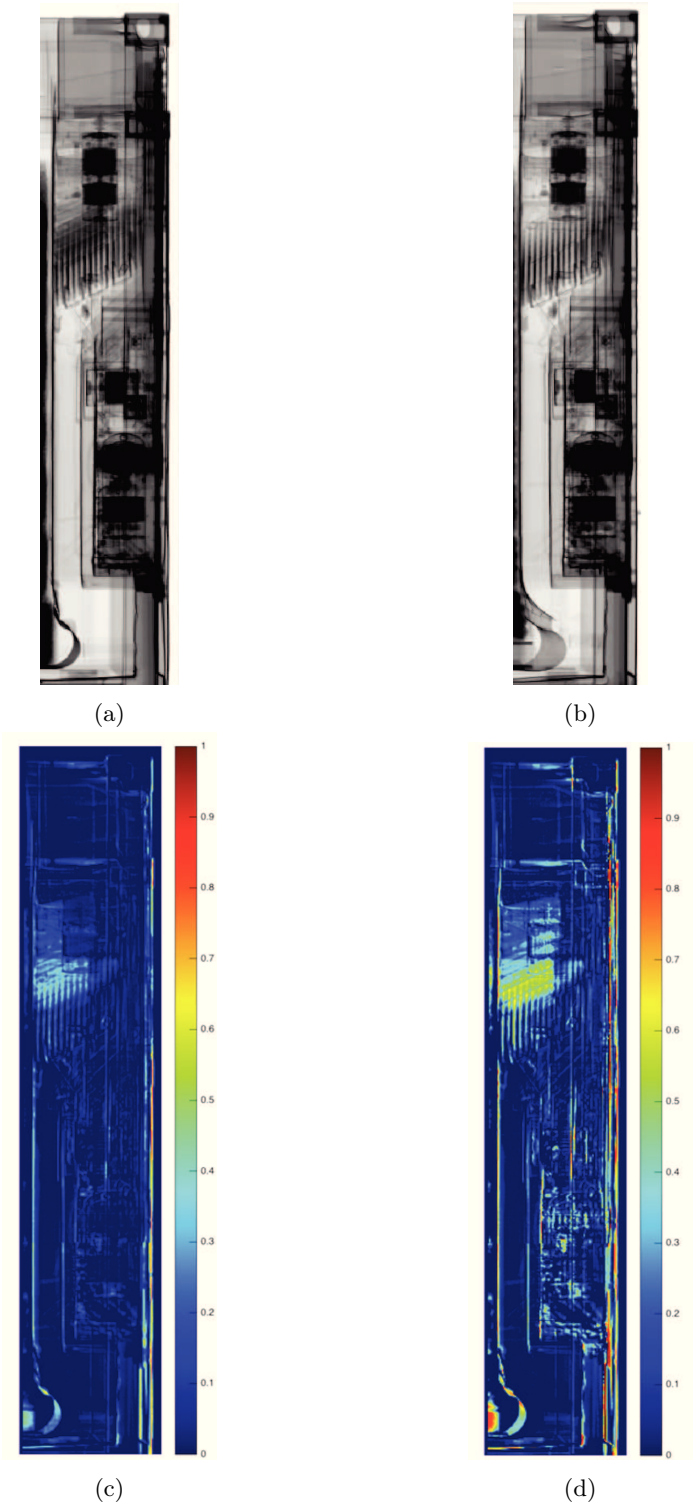
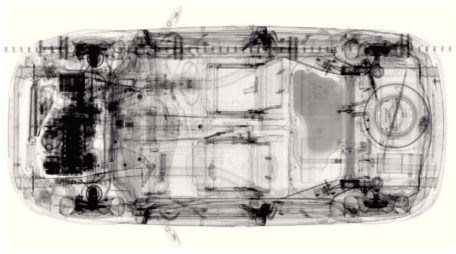
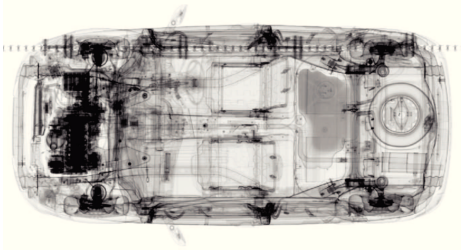


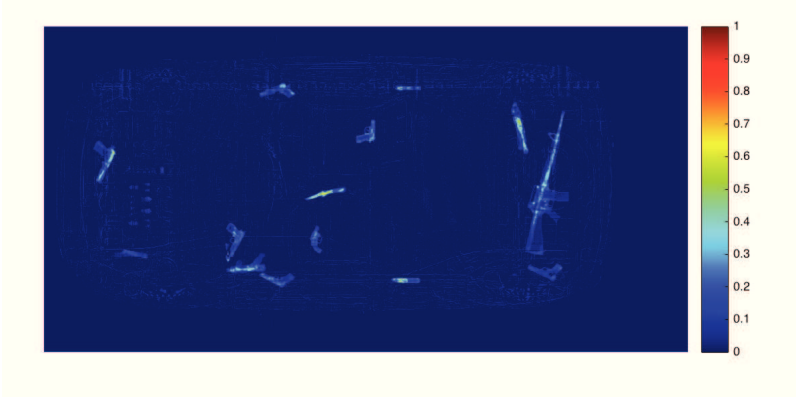
Figure 4.6 – (a) Image under inspection **R** acquired at Rotterdam in 2016. An important amount of narcotics where smuggled in the refrigeration unit by drug dealers; (b) closest image in terms of shape **T**, retrieved in our database via the ASM approach. This reference comes from a different container with a layout similar to **R**; (c) simple differencing following registration; (d) rationing following registration. We conclude that using the ratio of images is more convenient for irregularity detection.



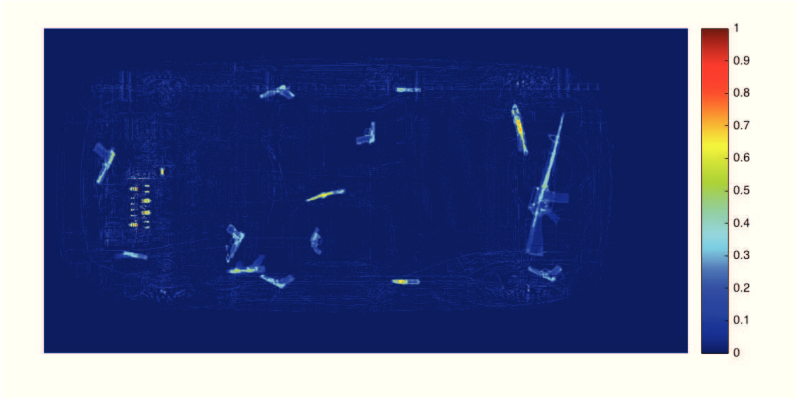
(a)



(b)



(c)



(d)

Figure 4.7 – (a) Inspected car image **R** where several guns and rifles were intentionally arranged in the car, and (b) its empty reference **T** acquired via an HCVL system; (c) simple differencing following registration; (d) rationing following registration. Changes visualization is strongly enhanced by the rationing method. For instance, a gun at the bottom-left of the image is revealed in (d) whereas it is almost invisible in (c).

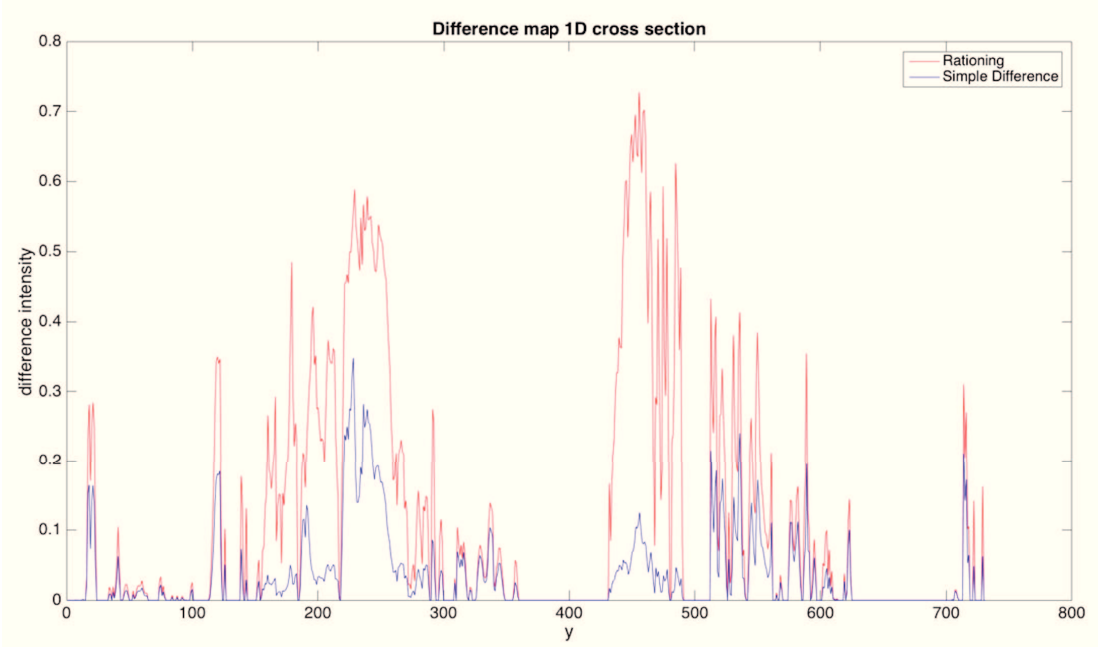


Figure 4.8 – 1D cross-section comparison of simple differencing (blue) and division (red) methods at the threat detection spot ($x = 60$) of the HCVS example presented on Figure 4.6. The division measure shows more sensitiveness to change detection.

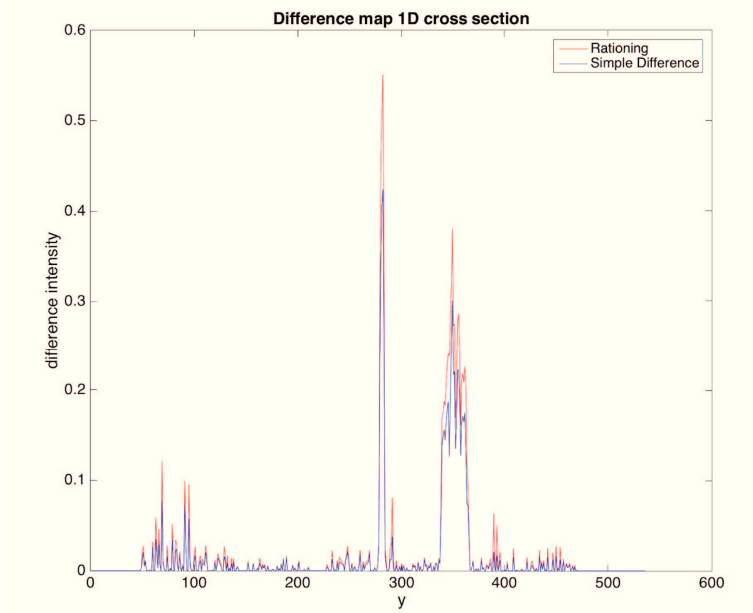


Figure 4.9 – 1D cross-section comparison of simple differencing (blue) and division (red) methods (at $x = 480$) for the HCVL case (Figure 4.7). The division measure seems far more sensitive to changes in the image.

Besides the difference enhancement feature, in Figure 4.6, one can observe in (d) that the irregularity is even visible behind a very dense object thanks to the rationing approach.

Furthermore, it yields a stronger robustness to illumination changes as claimed in [85]. This very feature might be of great interest in our applications: given two systems S_1 and S_2 with similar geometries but different radiation energies, one could choose from a database acquired in S_1 to register an image taken in S_2 for change detection. In fact, it is the case for HCVG systems where too few images are available to build a database while HCVM data is more easily accessible (both systems present comparable geometric settings).

4.5.5 Segmentation

Generally, global thresholding is not viewed as a method of choice due to possible illumination changes across a given scene. Regarding our study, X-Ray images are less prone to this phenomenon and the adopted rationing measure overcomes this precise issue. We therefore employ a simple global automated thresholding approach as detailed in secttion 4.3: Otsu's method paired with an hysteresis strategy (Figure 4.10 and Figure 4.11).

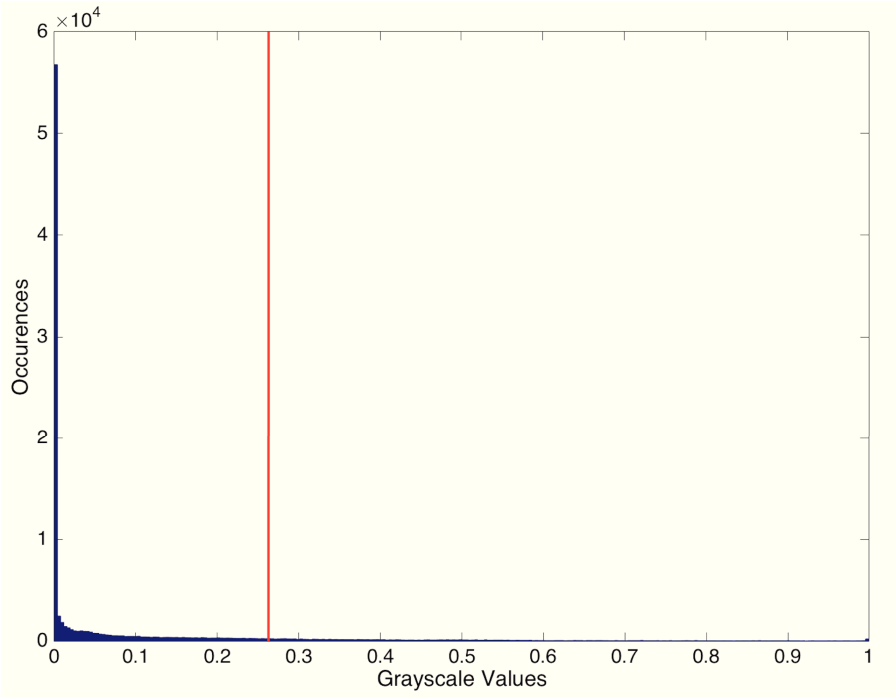
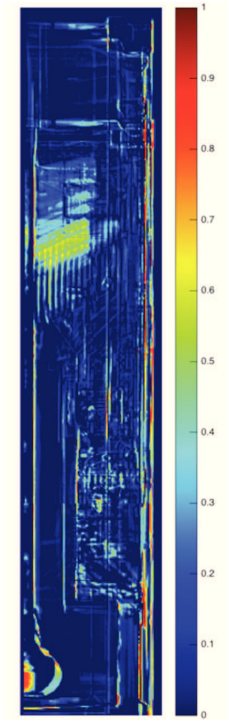


Figure 4.10 – Histogram of differences related to the example of Figure 4.6 with Otsu's automatic thresholding



(a) Map of differences

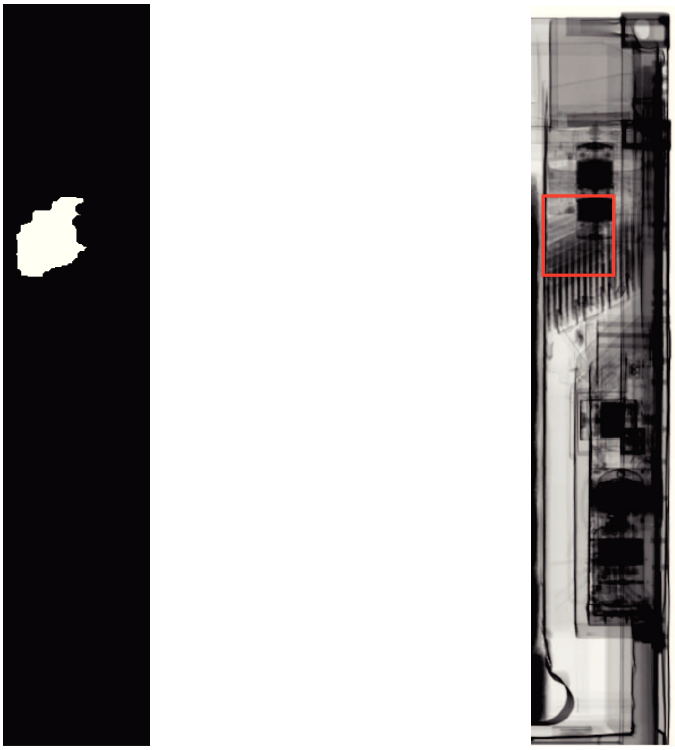
(b) Global Otsu - τ^*

(c) Hysteresis - $\theta = 0.08$

(d) Hysteresis - $\theta = 0.1$

Figure 4.11 – Resulting masks of differences with and without hysteresis thresholding for the refrigeration unit example of Figure 4.6. Detection blobs grow with hysteresis without actually yielding a significant increase of false alarms.

On Figure 4.12, we show the detection results following the filtering of the binary difference map from Figure 4.11 by the application of morphological operations (opening and closing with various structuring elements), as discussed in section 4.4.



(a) (b)

Figure 4.12 – Detection results for Figure 4.6 example, with $\theta = 0.08$. (a) Binary map following opening and closing morphological operations; (b) Bounding box indicating the irregularity position and its size.

4.6 Conclusion

Through this chapter, we presented our image rationing method for the evaluation of changes between $\mathbf{R}$ and $\mathbf{T}_{\mathbf{u}^*}$. **The approach we introduced is particularly relevant for X-ray images and is suitable for our specific irregularity detection task**. It has been compared to a simpler differences calculation scheme (also adjusted to our application), and proved to be more convenient for change detection. This conclusion has been drawn on the basis of a few visual and numerical observations, as well as intrinsic properties of X-ray attenuation (Beer-Lambert's law approximation).

In addition, a brief review on the different tools employed for (but not limited to) change detection was given. So, the main principles of Otsu's thresholding method as well as most common morphological operations have been discussed.

Part II

Applications to X-Ray Images

Chapter 5

Registration Techniques for Irregularity Detection in Refrigeration Units

Outline of the current chapter

5.1 Introduction

In recent reviews issued by the WCO (World Customs Organization), worldwide customs authorities reported the latest trends and patterns in terms of illicit trade [128]. In particular, the so called "rip-off" technique in customs jargon is often singled out, and automatic analysis software applications are especially demanded.

As defined by the United Nations Office on Drugs and Crime (UNODC) [17], a "rip-off" concealment generally designates "a methodology whereby a legitimate shipment, usually containerized, is exploited to smuggle contraband from the country of origin or the transhipment port to the country of destination". In a container, several locations can be used for this purpose, *e.g* the roof, chassis or the door, allowing to a local accomplice, to access easily the smuggled illicit goods.

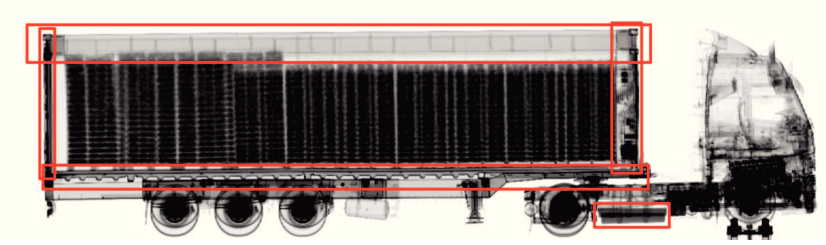


Figure 5.1 – Different rip-off concealments (HCVS X-ray scan).

In the past decade, texture-based detection techniques have emerged to address this issue, with a strong focus on contraband cigarettes detection. The designed algorithms were at the origin of notable recent seizures in Luxembourg and Italy for instance.

Yet, while considering narcotics detection, which is more critical in terms of contraband volume, the issue must be addressed quite differently. Indeed, observation shows that drugs do not display any specific texture signature, as with cigarettes. Due to the different types of narcotics (marijuana, cocaine...), volume and arrangement modes, the use of texture features for detection seems difficult to achieve. On Figure 5.3, the smuggled marijuana displays textures which significantly contrast with cocaine bundles. Furthermore, discrimination techniques show poor results on this type of organic material targetted objects, scanned on a single view with high energies. See Figure 5.2 for further details.

As a consequence, feature-based machine learning techniques must be left out, making room to more traditional irregularity detection approaches such as image comparison.

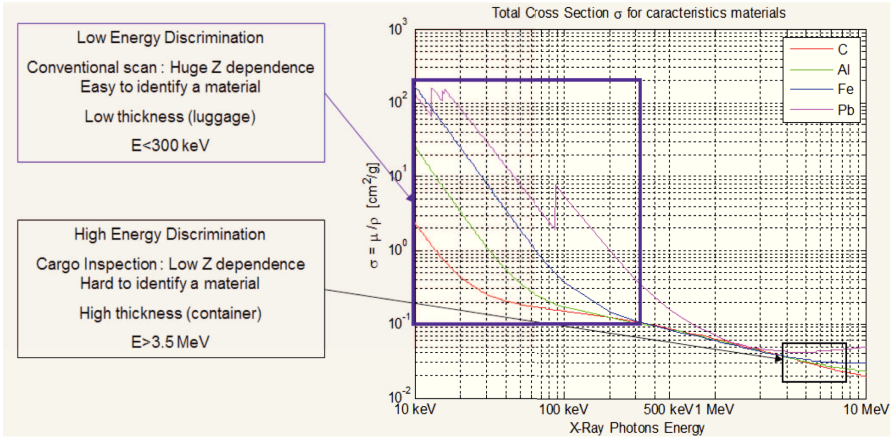
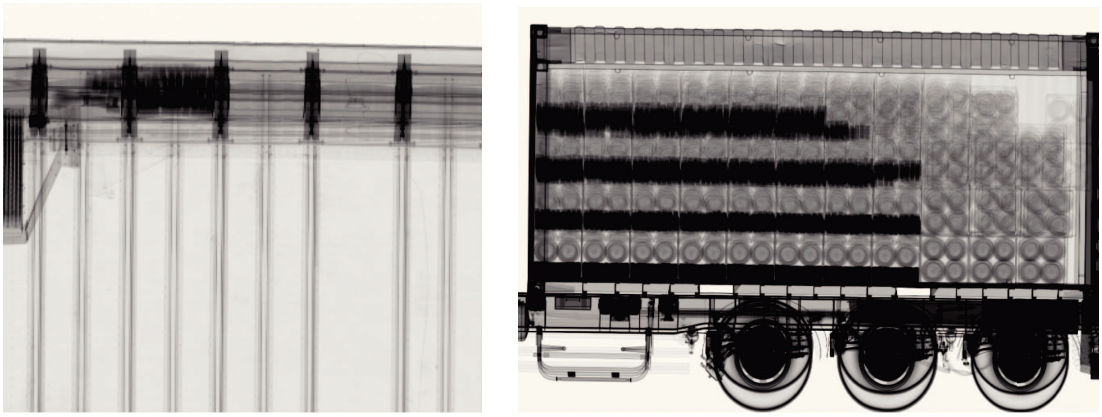


Figure 5.2 – Description of the difficulty to discriminate organic material at high energy ranges. The graph represents σ as a function of the photons energy. σ is the ratio of the attenuation coefficient μ by the material density ρ. μ actually depends on the atomic number Z employed for discrimination. For low photons energies (luggage inspection), the curves of the four listed materials are more distinguishable in comparison with the values at high energy ranges (container inspection).



(a) (b)

Figure 5.3 – (a) Cocaine smuggled on the roof of a refrigerated container; (b) marijuana smuggled between kitchenware (HCVS). This example denotes the complexity of the identification task through texture-based approaches.

In this section, we will concentrate on the analysis of refrigeration units of specific models of containers known as "reefers". Internal refrigeration units (RU) in refrigerated containers or air-coolers (in trailers) are very often subject to drugs smuggling, particularly for cocaine. From a trafficker's point of view, the interest in this region is twofold:

- Refrigeration units offer large free spaces along with dense mechanical structures (compressor, pistons, coils...). It therefore makes visualization far more challenging.

- It represents the most hard-to-reach area of a reefer. RUs are located deep inside the container, in such a way that custom officers have to unload the whole freight to perform manual inspection.

Throughout this chapter, we present a simple and intuitive technique we developed to analyze refrigeration units in containers or similar air-cooler modules while being faced with trailer cases. The core idea is the alignment of the analyzed module with a corresponding reference towards the automatic identification of any suspicious difference. In other words, the refrigeration unit under inspection **R** must be assigned to a specific category, such that its matching empty template **T** can be found. Since **R** is not supposed to contain foreign bodies, any discrepancy detected by alignment and differencing of **R** with respect to **T** may then be considered as potential illicit object.

First, an introduction on the nature of collected data images is given. Then, regions-of-interest (ROI) extraction along with employed classification methods are reported. The necessity for ASM is also demonstrated with regards to the simplification of the ultimate registration task. Different results for both HCVM and HCVS scanning systems are finally brought in the last section.

5.2 Data Collection

Figure 5.4 – Refrigerated ISO container.

A large amount of images is actually required to shed a light on the type of data and variations encountered in refrigeration units for a further analysis using image registration.

The main part of our data has been collected from customs authorities through a European project. All scans originating from a European sea port terminal were acquired through an HCVS stationary system. Thanks to a close partnership with various law-enforcement agencies, another significant part of our database was extracted from HCVM mobile scanning machine. The considered feature could also prove very useful on other X-ray systems such as HCVG or HCVP; yet, too little data is available at the moment.

Both systems have pretty distinct geometries and X-ray energy ranges (accelerators are not placed similarly and the distances to the scanned subject differ strongly), to such an extent that output images are utterly different in size and intensity. Even though a similar overall algorithmic

Figure 5.5 – Refrigerated trailer example. The air-cooler module is placed outside of the trailer.

strategy could be adopted, the registration issue must be addressed separately for mobile and stationary systems. This assertion is particularly true when coming to classification tasks, an imperative prerequisite to the very registration application.

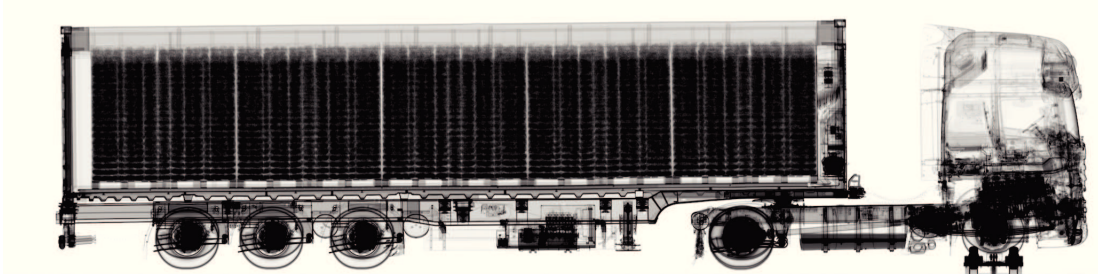


Figure 5.6 – Refrigerated container scanned through HCVS facility. The refrigeration unit is visible at the right of the container.

Among the large volume of images gathered from different sites, around **4 to** 10% only displayed refrigerated containers or semi-trailer trucks. A filtering procedure was therefore necessary to automate the search for X-ray images of refrigeration units.

In cargo screening, two modes of freight transport are to be distinguished:

- ISO or intermodal containers are used for the transport of goods. They are manufactured according to specifications issued by the International Orginization for Standardization (ISO) and are suitable for multiple transportation methods such as truck, rail, or ship. Typical ISO containers are constructed with identical size specification in terms of exterior height, width and length.
  In X-ray lateral view images, ISO containers can be identified by their darker (denser) top left and right corners. These salient components relate to the twistlocks (or corner castings) used to lift the container by shift-to-shore cranes or sidefilters. For ISO refrigerated container, as displayed on Figure 5.4, the refrigeration module is positioned inside the container, right behind the cabin (for electrical connections). Figure 5.6 shows a refrigerated

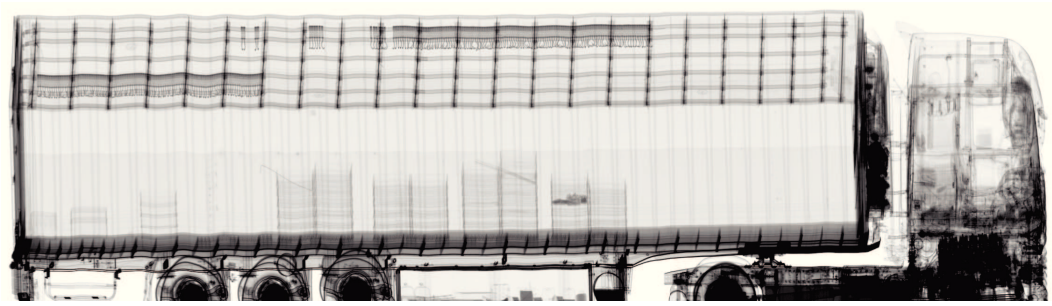


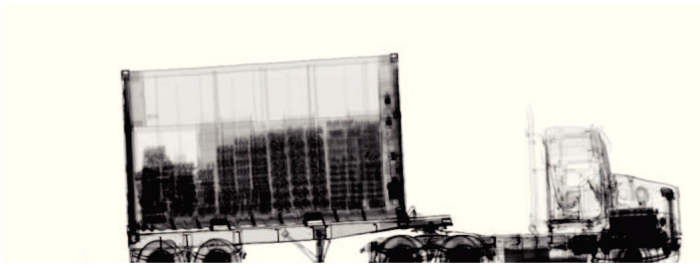
Figure 5.7 – Refrigerated truck scanned by HCVM mobile system. The air-cooler module is placed outside of the container (right side).

ISO container towed by a truck. The HCVS facility from which data was extracted is located at a trading harbor, hence most samples represent ISO containers.

- Alternatively, freight transportation is generally performed by full or semi-trailers attached to a tractor unit (carrier cabin). We will focus on refrigerated trailers for which the cooler unit is placed outside, facing the tractor, as described on Figure 5.5. In Figure 5.7 the X-ray image of a refrigerated semi-trailer is represented.

5.3 Container Analysis and Refrigeration Units Extraction

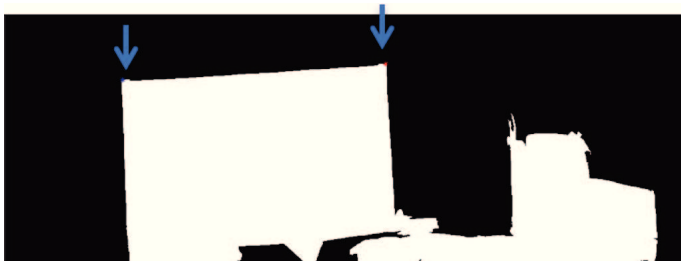
5.3.1 Corners Analysis



(a) Target X-ray image

(b) Binarization

(c) Center of Mass and Geometric Center Comparison



(d) Upper corners detection

Figure 5.8 – Top corners detection stages of an ISO container imaged by HCVM scanning machine. Center of mass and geometric center comparison is performed to check the cabin position.

As shown in the previous examples, ISO containers can be discriminated from trailers by simply analyzing the top right and left corners of the "box" c_r and c_l. Right and left positions are actually defined when the tractor is located at the right of the image.

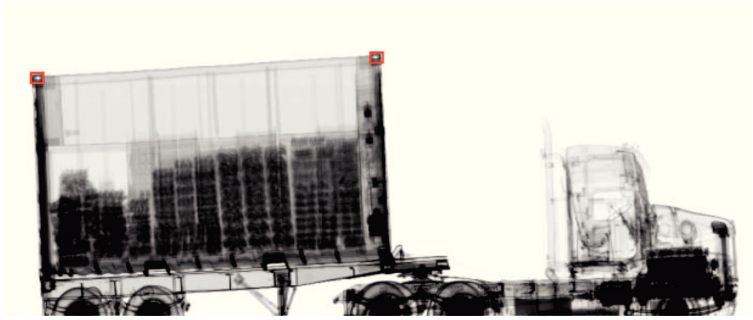


Figure 5.9 – ISO container scanned by the HCVM system. Illustration of corner patches extracted for SVM classification.

Via Otsu's thresholding combined with morphological operations, we transform the input image $\mathbf{I}$ into a binary mask $\mathbf{B}$. The container (resp. semi-trailer) is then isolated from the cabin (only the biggest region is retained) and its top part $\mathbf{B}_t$ is analyzed to compute both corners locations. 40×40 pixels windows centered on c_r and c_l are also sampled (see Figure 5.9). The detection process of upper corners is illustrated on Figure 5.8.

A basic gray-scale value analysis involving mean and standard deviation computations was experimented but led to significant false alarm rates. Rather, **an SVM classifier was directly trained on corner patches**: 260 positive and 206 negative samples yielded a generalization

error below 0.5% on a 800 images test set.

As discussed before, the subject classification as "ISO container" or "trailer" is actually necessary to properly target a potential RU or air-cooler compartment.

5.3.2 Localization of Potential Units

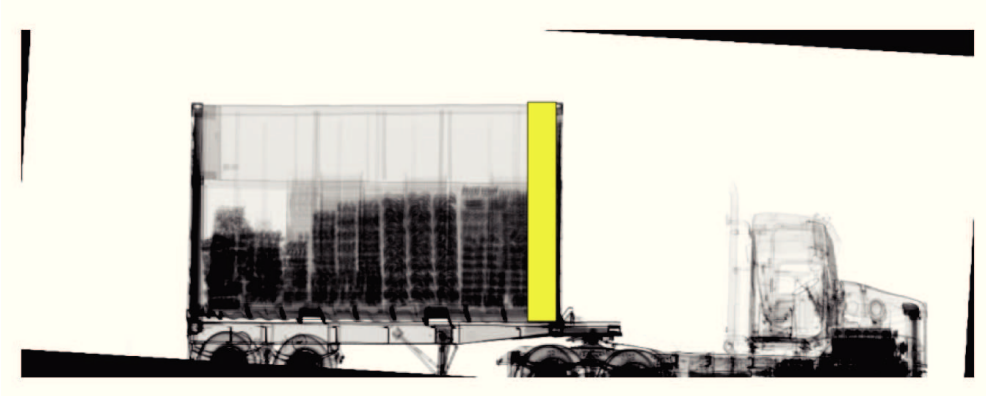


Figure 5.10 – Potential unit searched inside the container for ISO container.

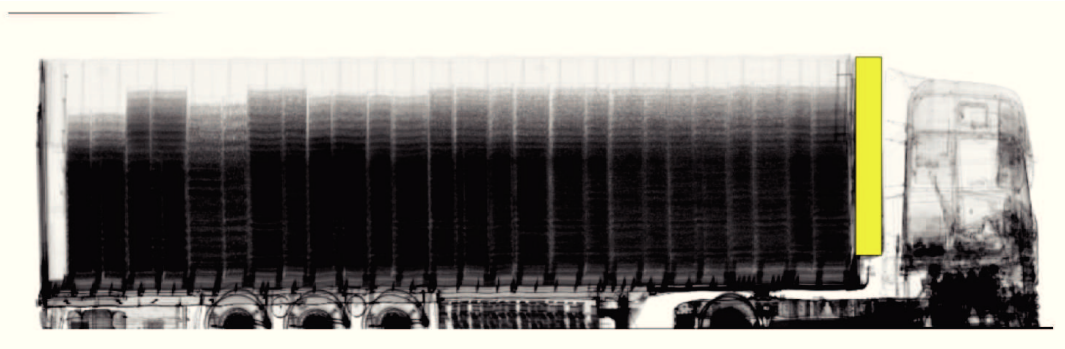


Figure 5.11 – Search window for refrigeration module shifted right-wards in trailer case.

In some situations, the container may be pivoting around the **z** axis. Since all extracted modules are aimed at image registration, they should be preferably set in a common coordinates frame. The rotation angle α of the correction to be applied is computed via c_r and c_l coordinates. Typical bounding box sizes are used for refrigeration units acquired by HCVS facility (145×730) or air-cooler modules imaged through HCVM (270×970). According to the classification output, the frame is placed at the bottom left of c_r for ISO containers, whereas it is translated to its right for the trailer case. On Figures 5.10 and 5.11, the ROI positioning is described.

Note that R-CNN approaches could have been used for the automatic extraction of refrigeration modules. Though, our simple "hand-crafted" technique showed great performance on thousands of examples tested so far.

5.3.3 Non-Refrigeration Modules Filtering

In a preliminary stage, images were extracted at usual locations of refrigeration modules (ROIs). Yet, the obtained samples relatively included a small proportion of true refrigeration units (around 10% for HCVS and 4% on HCVM). **After manually grouping hundreds of images into two categories: RU and non-RU, a binary SVM model was trained on HoG features extracted from samples of both clusters**. By automatically filtering additional non-RU X-ray scans, larger sets of positive and negative samples were generated. Note that each image was represented by a HoG feature vector, using 22×22 cells with blocks constructed from 2×2 cells and with overlapping constraints (as specified in section 2.2.1). A description of HoG features extracted from HCVM ROIs data is provided on Figure example HCVM HoG.

HCVS

For HCVS, non-RU samples relate exclusively to regular freight inside ISO containers. With ~ 1000 instances per class, a final training gave a 100% accuracy on a 1200 images test set. **Thus, a single classification model was employed to discriminate true RU from ordinary freight** (see *e.g* the instances depicted on Figure 5.12).

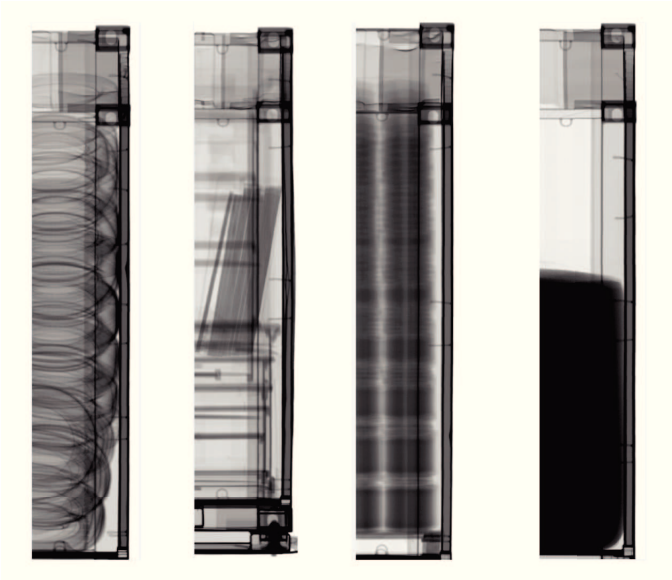


Figure 5.12 – Refrigeration unit ROIs showing regular freight in HCVS images.

HCVM

In contrast, non air-cooler samples designate both regular freight (Figure 5.13) and intermediate area between the trailer and the cabin (Figure 5.14). Therefore, **two distinct SVM models were trained for HCVM samples: one for the outer air-cooler (trailer case) and the other for inner RU (ISO case)**. On much smaller image sets (250 positive and negative samples), a 94% accuracy is obtained.

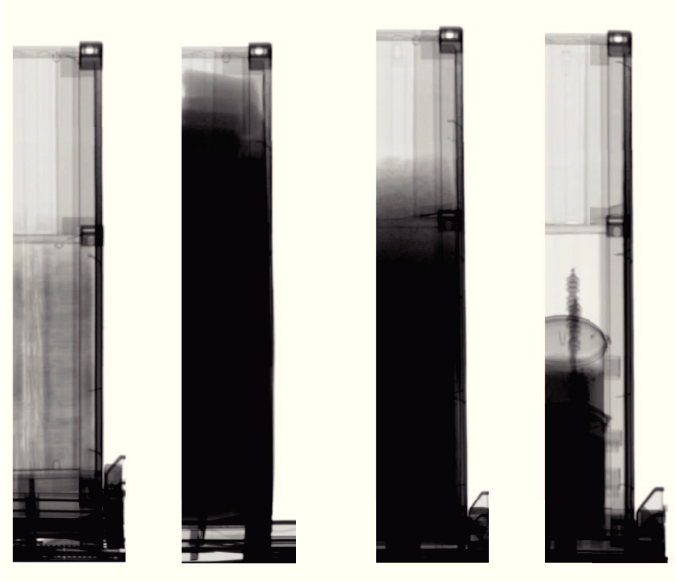


Figure 5.13 – Refrigeration unit ROIs showing regular freight in HCVM images.

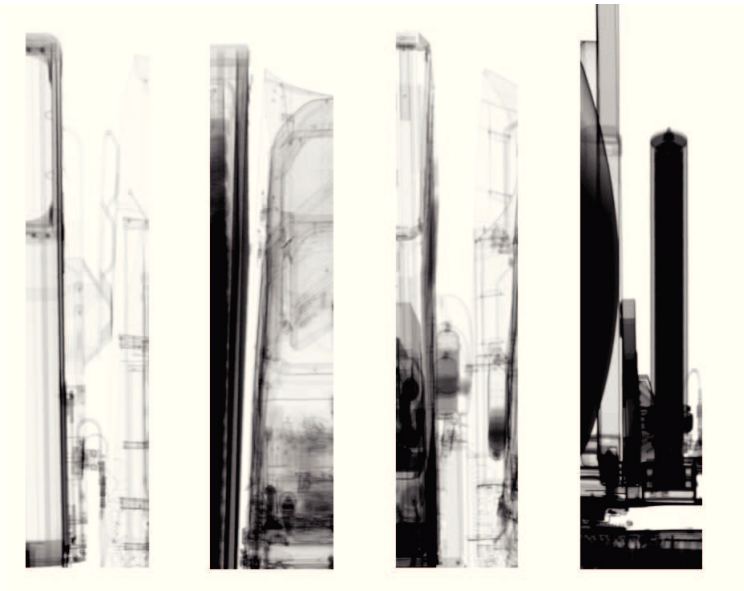
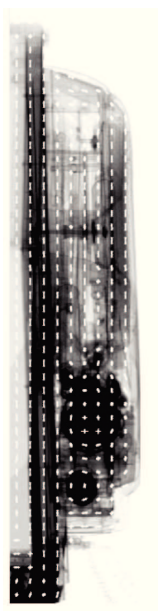
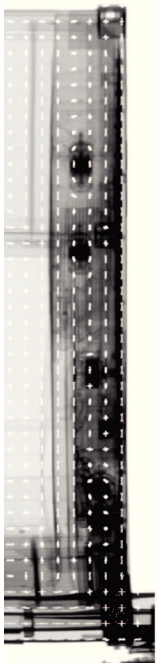


Figure 5.14 – Air coolers ROIs showing intermediate area in HCVM images.

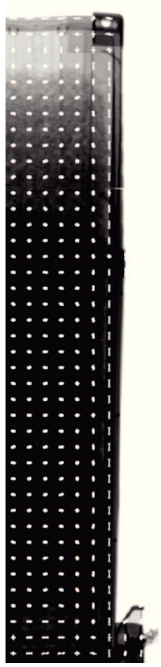


(a) True air cooler

(b) False air cooler



(c) True RU



(d) False RU

Figure 5.15 – HoG features visualization with 22×22 cells on positive and negative HCVM samples.

5.4 Database Construction

As we discussed in section 2.4, **RU and air-cooler databases have been built manually, in combination with K-means unsupervised classification, from several thousands of unlabelled samples**. We obtained about twenty categories of ISO RU and trailer air-coolers. Figure 5.16 and Figure 5.17 respectively show some of the most represented classes found among both HCVS and HCVM datasets.

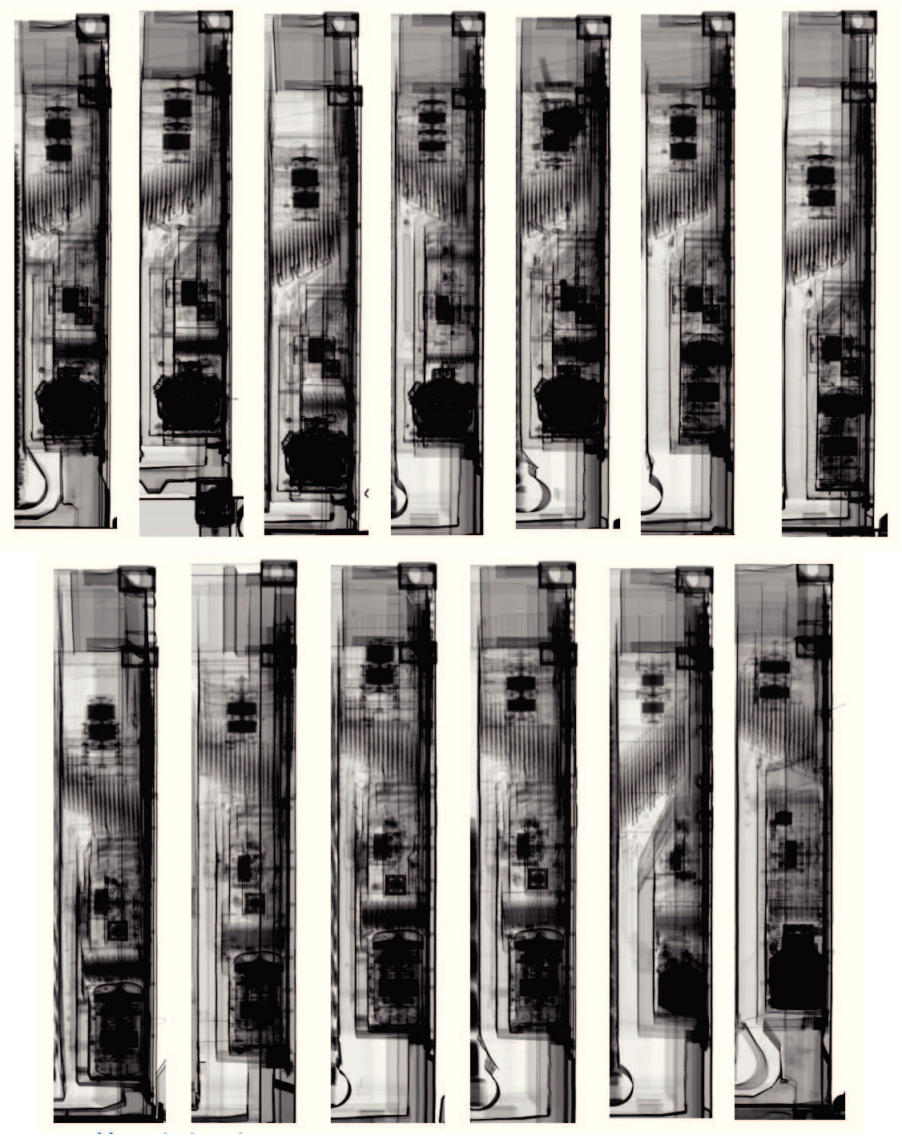


Figure 5.16 – Some HCVS refrigeration units categories.

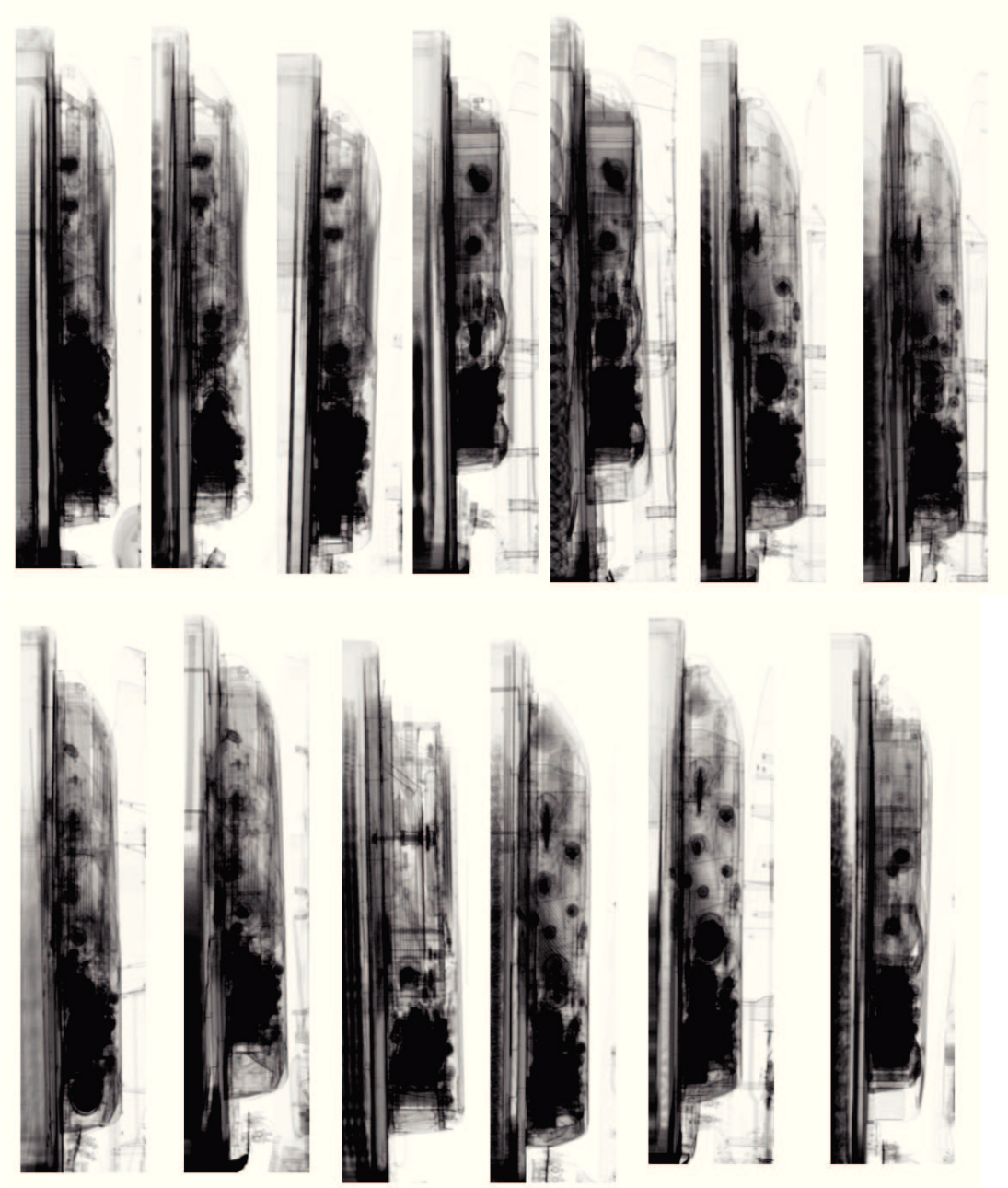


Figure 5.17 – Some HCVM air cooler categories.

HCVM ISO refrigeration units have also been extracted and assigned to different groups. Though, in most instances, the beam angle and possible arm movements affected the resulting images such that strong deformations along with freight superposition are observed. An example is provided on Figure 5.18. As a consequence, using image registration techniques is not relevant, and **we decided to focus on outer air-cooler compartments for HCVM X-ray scans**. Yet, the supply of additional data would help to address this issue in the future.

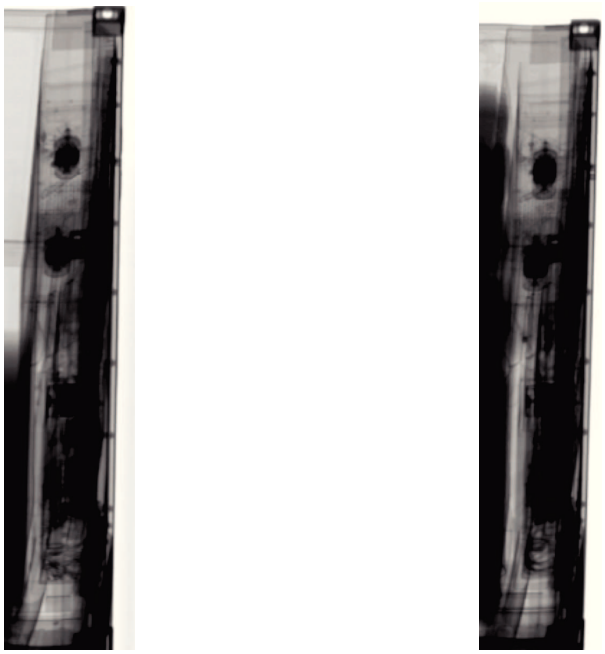


Figure 5.18 – Highly deformed RU samples acquired with HCVM. Parts of the freight can also be superimposed on the RU, avoiding clear visualization and leading to poor registration results.

5.5 Classification

The analysis of input RU scans $\mathbf{R}$ requires to perform an initial classification (or label assignment task) in order to find a corresponding template $\mathbf{T}$ in sight of a future registration process.

Category recognition is performed via the extraction of **pre-trained CNN features** from training images as described in section 2.3.2. We use the *imagenet-vgg-f* network for this task. In order to fit the network, our grayscale images are stacked into three channels (*imagenet* is built from RGB natural images) and resized to certain dimensions (224×224). Each sample indexed by k undergoes a forward pass and the CNN outputs a vector of features denoted $\mathbf{V}_k^t$. The computation of all n_t training feature vectors of same dimensions yield $\mathbf{V}^t = \left(\mathbf{V}_1^t \mid \mathbf{V}_2^t \mid ... \mid \mathbf{V}_{n_t}^t\right)$. In most situations, we were faced with high-dimensional feature vectors, and therefore resorted to PCA for dimension reduction. By computing $\mathbf{F}$ stacking principal eigenvectors of the covariance matrix, any training feature vector $\mathbf{V}_i^t$ is finally represented by:

$$\hat{\mathbf{V}}_i^t = \mathbf{F}^T(\mathbf{V}_i^t - \bar{\mathbf{V}}) \tag{5.1}$$

where $\dim(\hat{\mathbf{V}}_i^t) \ll \dim(\mathbf{V}_i^t)$.

A multi-class **RBF-SVM classifier** is trained on a training set partitioned into the different groups $(C_1, C_2, ..., C_L)$, via the *one-to-one ECOC* scheme. This training set corresponds to a certain proportion of the dataset, and training samples are randomly picked, so are testing instances. Cross-validation (10-fold) is employed during training such that the model does not overfit the data.

Following the classification model learning stage, testing is performed in a similar fashion.

Any testing X-ray sample $\mathbf{R}_s$ is submitted to a single forward pass through our pre-trained CNN. Thus, it can be described by a feature vector $\mathbf{V}_j^s$, $1 \leqslant j \leqslant n_s$, where n_s stands for the cardinality of the testing dataset. Before being classified, $\mathbf{V}_j^s$ goes through PCA (calculated from training instances):

$$\hat{\mathbf{V}}_j^s = \mathbf{F}^T(\mathbf{V}_j^s - \bar{\mathbf{V}}) \tag{5.2}$$

The SVM algorithm outputs a vector of probabilities:

$$p_{svm} = \Big(P(C_1 \mid \hat{\mathbf{V}}_j^s), P(C_2 \mid \hat{\mathbf{V}}_j^s), ..., P(C_L \mid \hat{\mathbf{V}}_j^s)\Big)^T \tag{5.3}$$

defining the posterior probability that $\mathbf{R}_s$ (more exactly its representation $\hat{\mathbf{V}}_j^s$) belongs to each reported label. The result is thus given by the index yielding the maximum posterior probability.

$$l^* = \underset{l \in [\![1,L]\!]}{\arg\max}\, P(C_l \mid \hat{\mathbf{V}}_j^s) \tag{5.4}$$

As we will see in the next section, we estimate that at least 30 images per RU category are required for the ASM algorithm to perform well. Our classifier is therefore only trained on respectively 10 and 13 classes for HCVS and HCVM. with a $70-30\%$ partitioning between training and testing data. **We obtain a recall and precision above** 95% **for both systems**, as described in the reported HCVM and HCVS confusion matrices Figures 5.19 and 5.20.

Actually, the Bag-of-Visual-Words method (BoW) [134] could have been adopted either. Still, larger datasets are expected in the future, so we rather opted for pre-trained CNN-based image descriptors, as motivated in section 2.3.2.

Class	1	2	3	4	5	6	7	8	9	10	11
1	0.996	0.003	0	0	0	0	0	0	0.001	0	0
2	0.12	0.88	0	0	0	0	0	0	0	0	0
3	0	0	1	0	0	0	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0	0
7	0	0	0	0	0	0	1	0	0	0	0
8	0	0	0	0	0	0.03	0	0.97	0	0	0
9	0	0	0	0	0	0	0	0	1	0	0
10	0	0	0	0	0	0	0	0	0	1	0
11	0	0	0	0	0	0	0	0	0	0	1

Figure 5.19 – Confusion matrix obtained on HCVS data.

Class	1	2	3	4	5	6	7	8	9	10	11	12	13
1	1	0	0	0	0	0	0	0	0	0	0	0	0
2	0	1	0	0	0	0	0	0	0	0	0	0	0
3	0	0	0.89	0	0	0	0	0.11	0	0	0	0	0
4	0	0	0	1	0	0	0	0	0	0	0	0	0
5	0	0	0	0	1	0	0	0	0	0	0	0	0
6	0	0	0	0	0	1	0	0	0	0	0	0	0
7	0	0	0	0	0	0.08	0.92	0	0	0	0	0	0
8	0	0	0	0	0	0	0	0.94	0.06	0	0	0	0
9	0	0	0	0	0	0	0	0.10	0.90	0	0	0	0
10	0	0	0	0	0	0	0	0	0	1	0	0	0
11	0	0	0	0	0	0	0	0	0	0	1	0	0
12	0	0	0	0	0	0	0	0	0	0	0	1	0
13	0	0	0	0	0	0	0	0	0	0	0	0	1

Figure 5.20 – Confusion matrix obtained on HCVM data.

5.6 Intra-Class Variability Issue

Once the label of l^* of the input sample has been identified, we need to choose a corresponding template **T** to perform registration. Since it would serve as reference for image comparison, it must be as close as possible to the inspected input **R**. Accordingly, **T** is picked via some similarity criteria.

Maximum number of matched keypoints

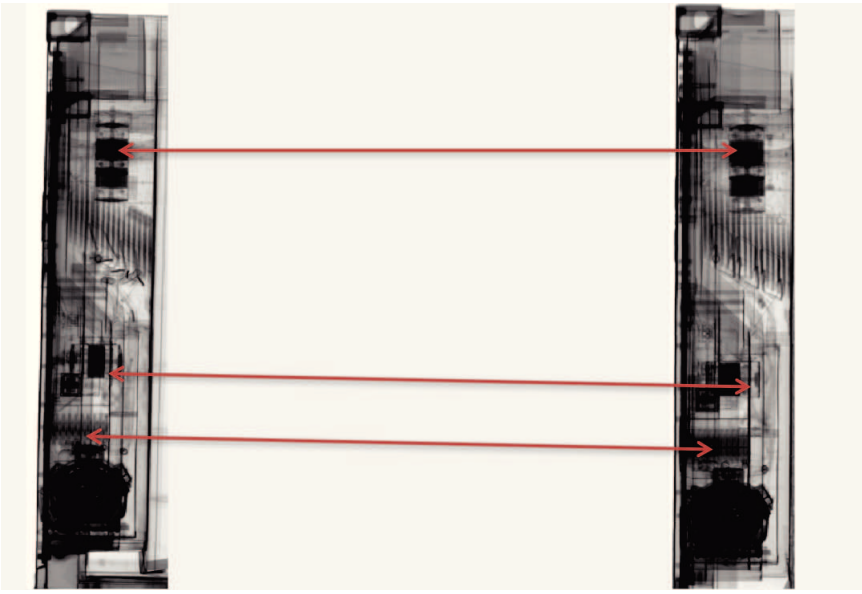


Figure 5.21 – Matching performed by our scientific partners. Significant deformations can be reported on different spots between both samples.

Our scientific partners (under a European project) attempted to extract SIFT features from all training samples from category l^*, as well as from **R**. By matching the keypoints through RANSAC filtering, the training instance resulting in a maximum number of matches was selected

as template **T**. However, this approach lacked of robustness, and did not accounted well for local variations as the reader may observe on Figure 5.21.

Minimal distance between dense descriptors

A simple alternative technique would search for the minimizing distance between feature vectors computed from dataset samples and the input image **R**. Experiments have thus been conducted with HoG and CNN feature vectors, reduced by PCA and using nearest-neighbor matching. 40 input images have been tested against a given matching class of 100 training samples. In terms of Mean-Squared-Error (MSE) minimization, the method yielded satisfactory results as described on Figure 5.22 (in comparison with the ASM approach finally employed).

Still, despite minimizing overall intensity differences, these dense descriptors do not handle efficiently local major differences. As a consequence, several images are mismatched, as they present large structural variations. This is likely to affect the difference detection process after performing image registration (false alarms or anomaly being occluded). Visual examples are provided on Figure 5.23 and 5.24.

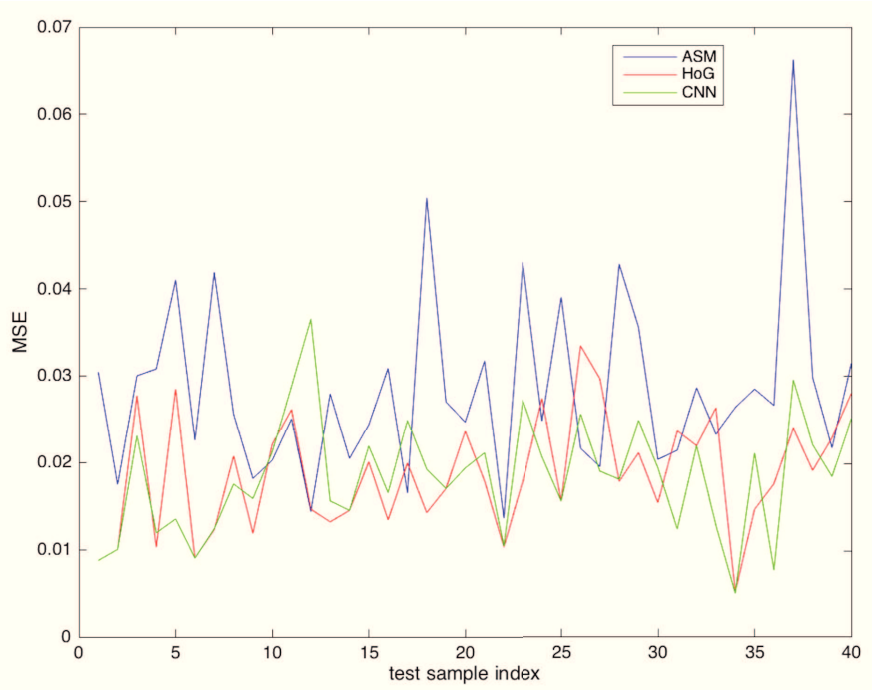


Figure 5.22 – MSE results after matching and rigid registration between the 40 test images and the retrieved templates among the training set.

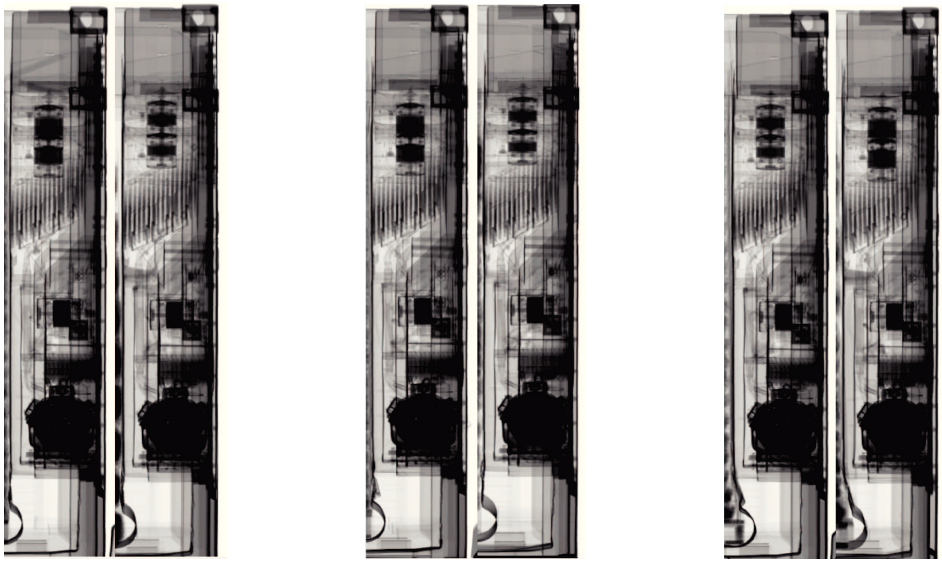


Figure 5.23 – Matching results using HoG descriptors.

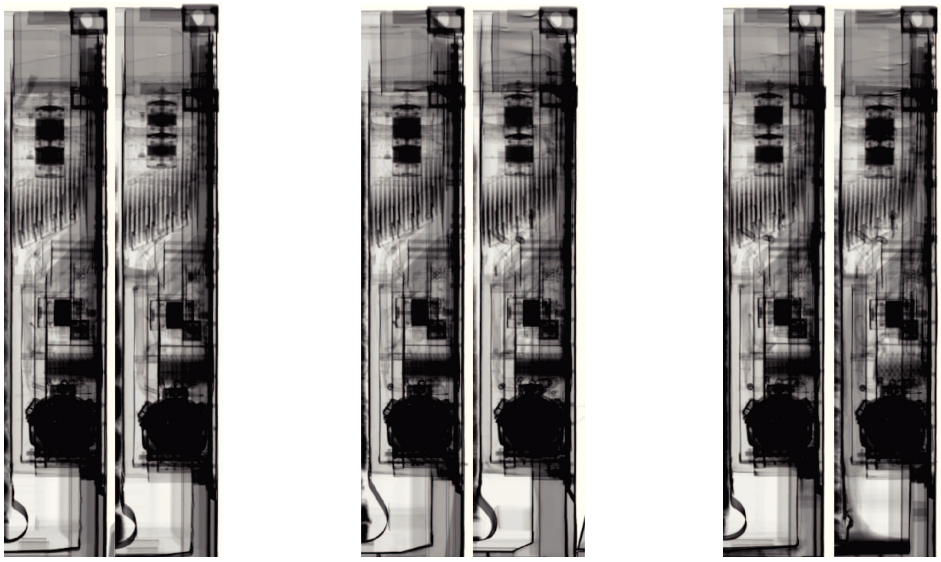


Figure 5.24 – Matching results using pre-trained CNN descriptors.

As mentioned, because of local substantive variability of mechanical structures within a given class, this simple approach does not provide the sought closeness between images. A sample with specific configurations of the compressor and cylinders could be matched badly to a sample showing strong shape differences. In the following depiction, several variations of a refrigeration unit of a similar category are represented. **It designates the "intra-class variability issue"** as illustrated on Figure 5.25.

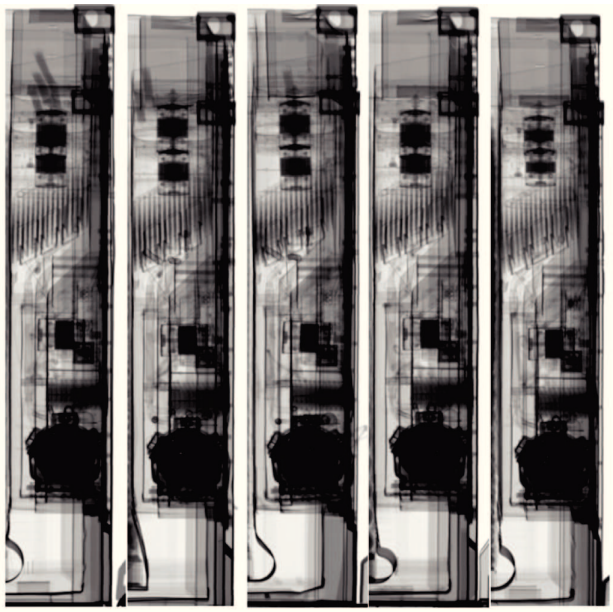


Figure 5.25 – A few variations among refrigeration units of the same class (mainly on the top shape).

ASM adaptation

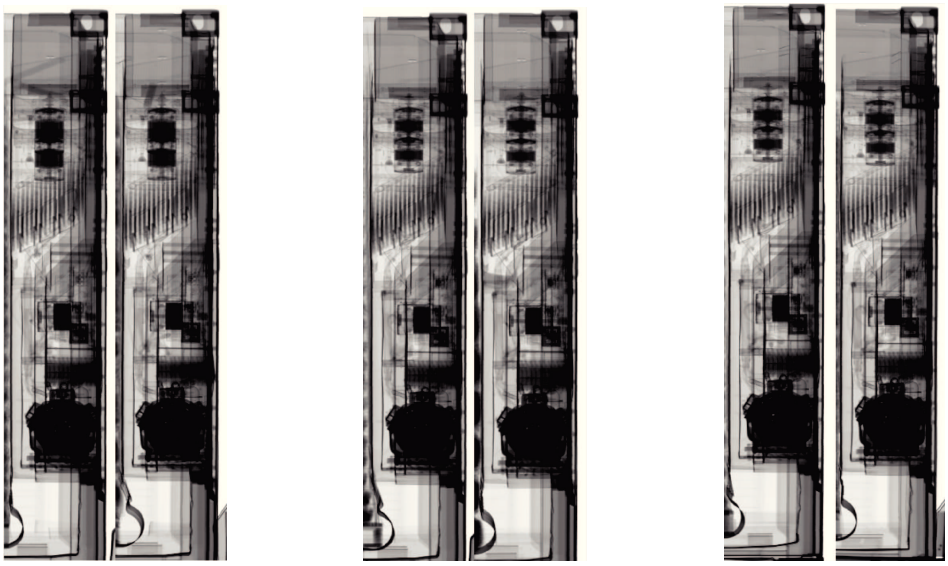


Figure 5.26 – Matching results using ASM closest shape retrieval scheme.

One could also suggest to pick one template for each class and make use of more complex registration methods to handle deformations as those observed in Figure 5.25. Nevertheless, we preferred to adapt the ASM method which, after classification, automatically selects the

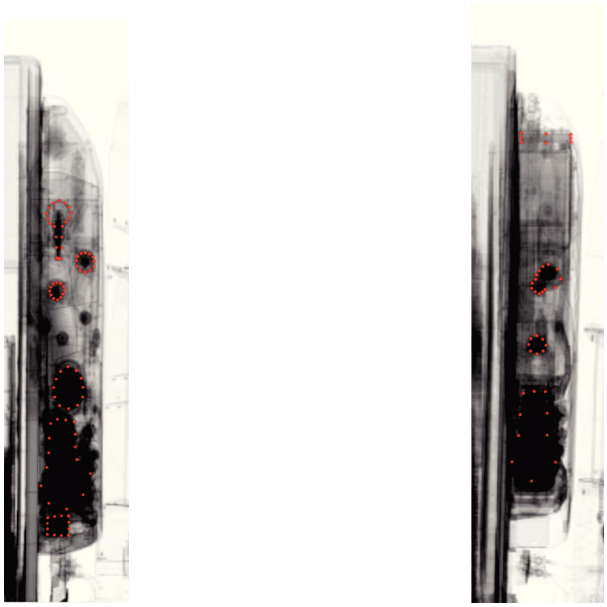
closest image in terms of structure shapes. Accordingly, a straightforward rigid registration scheme might be acceptable for alignment given that both **R** and **T** images are supposed to be very similar. This method assumes that the dataset contains all structural and system-related deformations in the resulting image.

Whenever it seems not to be the case, additional data should be collected to catch more variations. **Our ASM approach is much simpler in the sense that it circumvents the need for examination of the different system properties that might possibly cause deformed outputs (acquisition angle, arm movement...)**. It contributes in focusing on relevant shapes rather than making use of randomly distributed feature points. Moreover, one could claim that our method is actually necessary to address structural deformations of inner mechanical parts in refrigeration units. Indeed, in these situations, finding a plausible transformation to yield an optimal overlay between **R** and **T** is very hard to achieve. As shown on Figure 5.26, the method seems more relevant for template retrieval.

5.7 Active Shape Models

As introduced previously, the closest image retrieval task, in terms of salient shapes, is operated through the Active Shape Models technique. The main concepts have already been discussed in chapter 1, we will therefore go through a simple example of an air cooler module scanned by an HCVM imaging system.

5.7.1 Shapes Annotation



(a) (b)

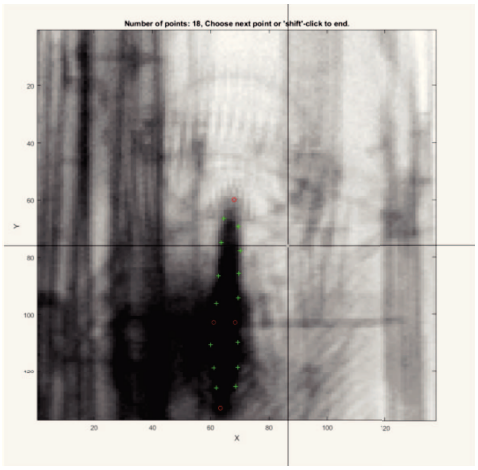
Figure 5.27 – Examples of annotated shapes HCVM air-cooler images from different categories.

An object is described by n points, referred to as landmark points manually determined in a set of s training images. On Figure 5.28, the annotation process is described. See also Figure 5.27 for other examples of annotated shapes. These landmark points are stacked in shape vectors and collected to get a point distribution model:

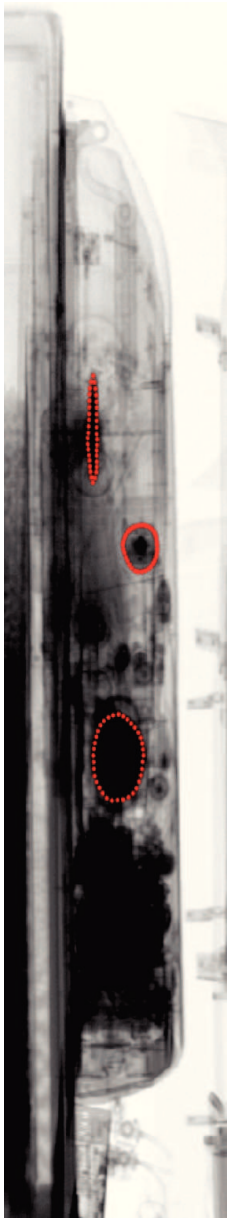
$$\mathbf{x} = (x_1, y_1, ..., x_n, y_n) \tag{5.5}$$

Actually, only a dozen of samples must be manually annotated to create a initial model $\mathbf{M}_i$. Then, larger sets of images can be automatically annotated by fitting $\mathbf{M}_i$ on the new samples, and extracting the position of the converged set of points. This way, the model can be refined and become more robust.

Concerning the RU analysis application, we aim at catching local non-linear deformations or mechanical differences (particular positioning of elements) to match images presenting similar shape features. Hence, the most salient and evenly distributed parts in the image are selected for annotation, *i.e* the compressor, pistons, coils of wires and other dense spots.



(a)



(b)

Figure 5.28 – (a) Landmark manual annotation example on trailer air-cooler. Red circles and green crosses respectively designate landmarks and intermediate points; (b) fully annotated shape on air-cooler with generated points by interpolation.

5.7.2 Statistical Model Building

Prior to model computation, the s shapes undergo Procrustes analysis to be aligned and be uniformly represented across the training set (Figure 5.29). The mean shape $\bar{\mathbf{x}}$ as well as the covariance S are calculated to operate PCA dimensionality reduction on the set of collected

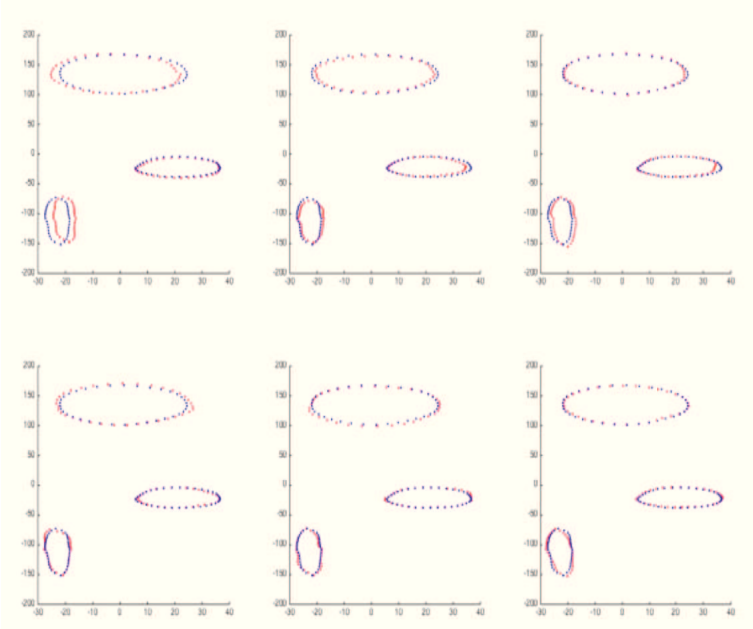


Figure 5.29 – Representation of six different modes of annotated shapes from 5.28 following Procrustes Analysis. The mean shape is described in blue, while the modes are represented in red.

shapes. By taking $r_v = 0.98$ and the t components of $\mathbf{b}$ constrained with $m = 3$, the shape model is given by the approximation:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b} \tag{5.6}$$

The statistical shape model is employed to define any shape $\mathbf{x}$ in lower dimensions $t \ll 2n$. During the training phase, a second statistical model is computed: the appearance model which makes the shape model converge (or diverge if not applied on a corresponding image) during the fitting process.

As widely explained in Chapter 1, the appearance model is built by sampling gray-level profiles for each annotated point $i \in [\![1, n]\!]$. These profiles are centered at the landmark point. For HCVM images, the length of landmark intensity profiles is fixed to 15. Any appearance profile $\mathbf{x}_g$ can be written:

$$\mathbf{x}_g = \bar{\mathbf{x}}_g + \mathbf{P}_g\mathbf{b}_g \tag{5.7}$$

As specified, models computations are operated on two resolution scales for both training and fitting stages.

5.7.3 Model Fitting

Initialization

In early stages, the shape center was initialized at the mean position (x_c, y_c) from all s training images. Yet, model fitting is very sensitive to initialization and despite the use of a coarse-to-fine

multi-resolution approach, the shape often mistakenly diverges.

Rather, d_i, the mean of vertical distances separating the trailer's upper corner c_r and the shape center was computed for each class. Accordingly, for every input image, the upper-corner ordinate c_r^y is computed and the vertical component of the shape center is initially placed at $c_r^y + d_i$. A depiction of the initialization stage is provided on Figure 5.30.

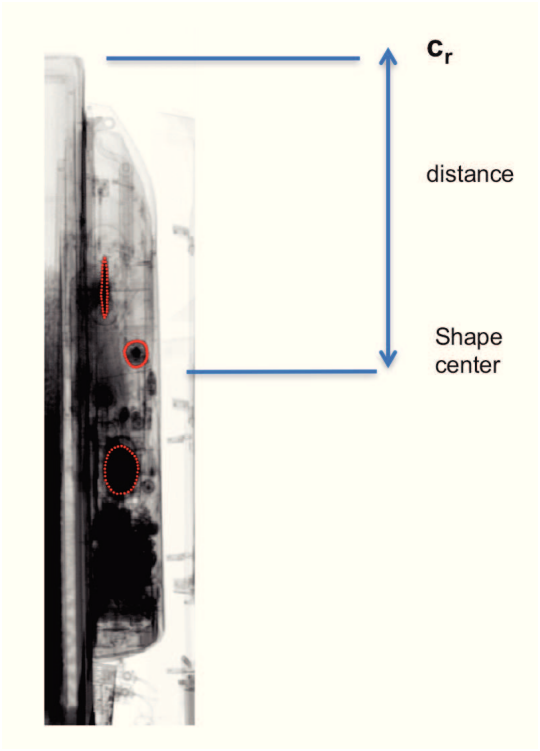
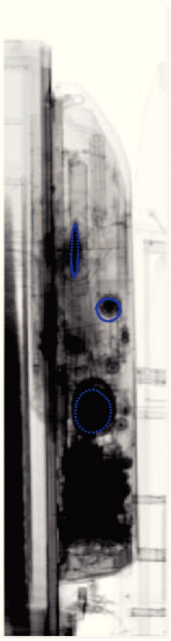


Figure 5.30 – Shape initialization description.

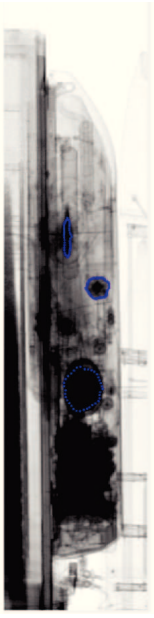
Fitting

During the fitting process, for each iteration, n_s different normal profiles are sampled at every landmark $i \in [\![1, n]\!]$. We set $n_s = 15$ such that disjoint profiles are also computed. Also, the number of search iterations is fixed to 40. The average running time until convergence is ~ 1 seconds.

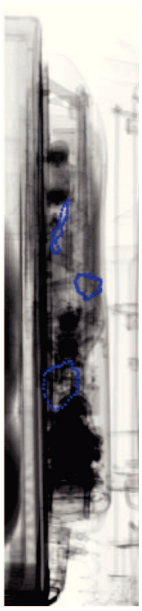
After being applied, a shape may converge or diverge if the input image does not relate to the training set category. It can also be caused by a poor initialization estimate. The deviation from the mean shape can be assessed visually as depicted on Figure 5.31. It can also be numerically quantified by directly looking at the shape parameters of $\mathbf{b} = \mathbf{P}^T(\mathbf{x} - \bar{\mathbf{x}})$. In Figure 5.32 the two first components of $\mathbf{b}$ (with $t = 4$ in this case) are graphically represented and show clearly that converging and diverging shapes can be numerically discriminated (*e.g* using a simple thresholding). The last observation is quite obvious given that $\mathbf{b}$ actually quantifies how much the current shape deviates from the mean $\bar{\mathbf{x}}$.



(a)



(b)



(c)

Figure 5.31 – (a) shape initialization (in blue); (b) converged shape; (c) deviated shape when applied on non-corresponding X-ray sample.

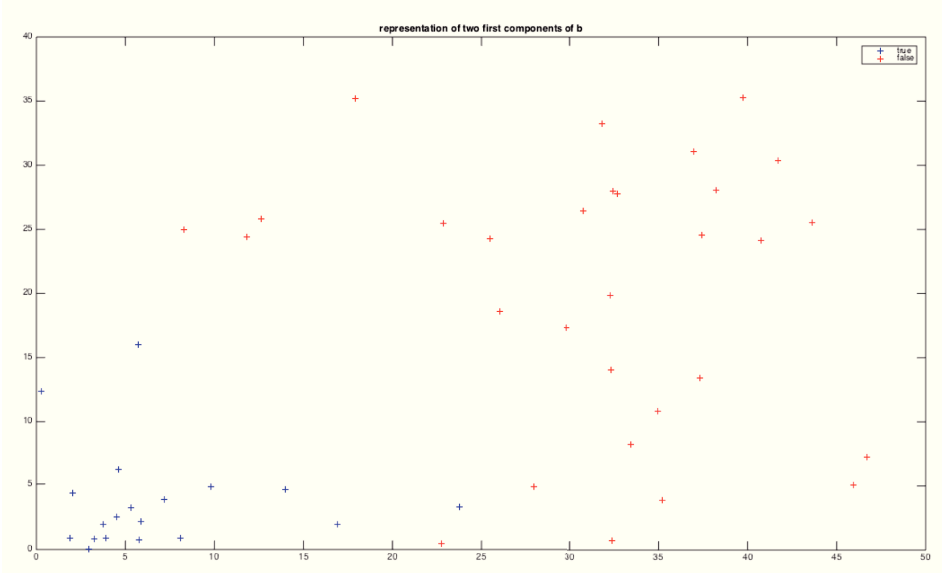


Figure 5.32 – Representation of two first components from resulting **b** vectors. Red points stand for shapes mistakenly applied on non-corresponding images. Blue samples result from converging shapes and are much closer to 0.

5.7.4 ASM as Classifier

In the light of these results, one could suggest to directly employ ASM for air cooler category recognition, as discussed in Chapter 1. A cost combining both appearance and shape parameters would be computed after applying the shapes of every training classes on the input image **R**.

As a consequence, the shape yielding the lowest cost would designate the category to which **R** belongs. Still, the classification output would be obtained after applying all shapes, *i.e* more than 20 seconds whereas the CNN-features extraction and SVM classification are performed in less than a second altogether.

Nevertheless, ASM may indeed play a valuable role as far as RU model recognition is concerned. As mentioned previously, multi-class SVM results in a vector of probabilities $\mathbf{p}_{svm}$ which designates the posterior probabilities that observation **R** belongs to each trained class C_l where $1 \leqslant l \leqslant L$. The input classification is therefore simply performed by selecting the label for which the maximal posterior probability is reached. In some situations, especially where only small training datasets are available, **R** may be misclassified. It may also happen while seeking to discriminate between two very close image categories (as in Figure 5.33 (a) and (b) for example).

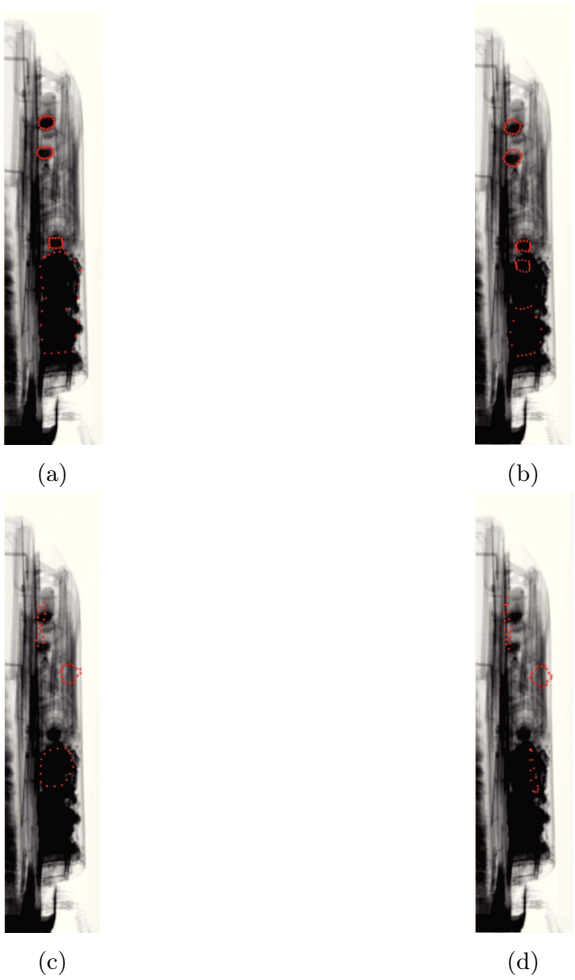
This is where ASM comes in: **both classes with maximal and second highest probabilities are retained and their respective shape models are matched to R**. We calculate a cost taking into account the deviation from the mean shape and the mean appearance, for both categories. Thus, the label resulting in the lowest cost provides the final classification result. If the obtained costs are "too large", **R**'s class is considered as unknown and registration cannot be undertaken.

As outlined in subsection 5.33, the cost is defined by the norm of **f** with:

$$\mathbf{f} = \begin{pmatrix} \bar{\mathbf{v}} \\ \hat{b} \end{pmatrix} \tag{5.8}$$

$\bar{\mathbf{v}}$ stands for the mean of $\mathbf{v} = \left(\left\|\mathbf{b}_g^{1^*}\right\|, \left\|\mathbf{b}_g^{2^*}\right\|, ..., \left\|\mathbf{b}_g^{n^*}\right\|\right)$ using the notations of Chapter 1. $\hat{b}$ is defined as: $\hat{b} = \frac{\|\mathbf{b}\|}{t}$, with t the number of modes.

This vector controls whether the current fitted set of points significantly strives from the appearance and shape model. Whenever $\|\mathbf{f}\| > \upsilon$, the classification is considered as erroneous, where υ stands for a pre-set threshold. We set $\upsilon = 20$.



(a) (b)

(c) (d)

Figure 5.33 – ASM application for classification. (a) Shape correctly fitted to input image; (b) The shape seems to be correctly matched, though appearance data would yield a strong deviation w.r.t (a); (c)-(d) The fitted shapes do not converge. **R** belongs therefore to a different category.

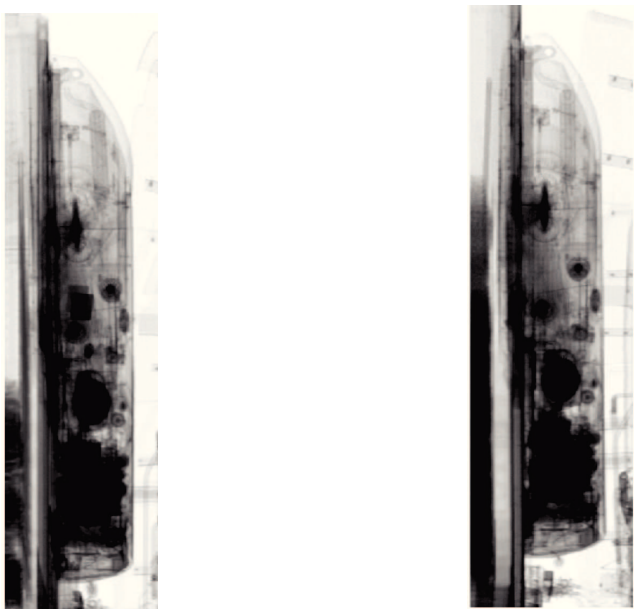
5.7.5 Closest Shape Retrieval

As detailed in the dedicated chapter, following the ASM training stage, the model is fitted to all training images for each separate class of RUs. Thereby, a low-dimensional shape parameter

vector **b** is extracted for each training sample. For each category, these vectors are stacked in a table to be kept in memory just as appearance and shape models.

Accordingly, after applying the shape model on **R**, a vector of shape weights $\mathbf{b}_{new}$ is calculated and compared to the training vectors through a nearest-neighbor scheme. The training X-ray sample whose **b** vector yields the smallest Euclidean distance to $\mathbf{b}_{new}$ is considered as the closest reference image in terms of shape. Here, we do not interfere with appearance information: the objective is to find an image displaying similar structures rather than gray-scale intensity values. Making use of appearance data within the matching process generally results in a weaker emphasis on salient shape similarities.

In the following example, presented on Figure 5.34, a bag of drugs was artificially merged into an air-cooler compartment image. The image **R** was successfully classified and the application of ASM provided the closest template **T** among more than one hundred candidates (from the ASM training set). In the next section, the registration result involving **R** and **T** is represented. Several issues are also discussed.



(a) **R** (b) **T**

Figure 5.34 – (a) Image under inspection **R** with smuggled cocaine (synthetically obtained by merging); (b) **T** retrieved by finding matching ASM **b** vector from the training database.

5.8 Image Registration and Difference Detection

5.8.1 Example

In the algorithm pipeline, we resort to **our fast piece-wise linear registration method**, as introduced in section 3.5. Given the ASM connected sub-shapes, both source and target images are partitioned into sub-images. SURF features are iteratively extracted and matched between sub-images pairs from **R** and **T**. Through MLESAC filtering, rigid transform matrices - to be applied on each sub-part of **T** - are computed. The deriving dense deformation field is then

regularized by applying an isotropic diffusion scheme at the boundaries separating sub-images (as explained in section 3.5). Using the images of Figure 5.34, a description of the piece-wise feature matching process is given in Figure 5.35. The smoothing effect is shown on Figure 5.36.

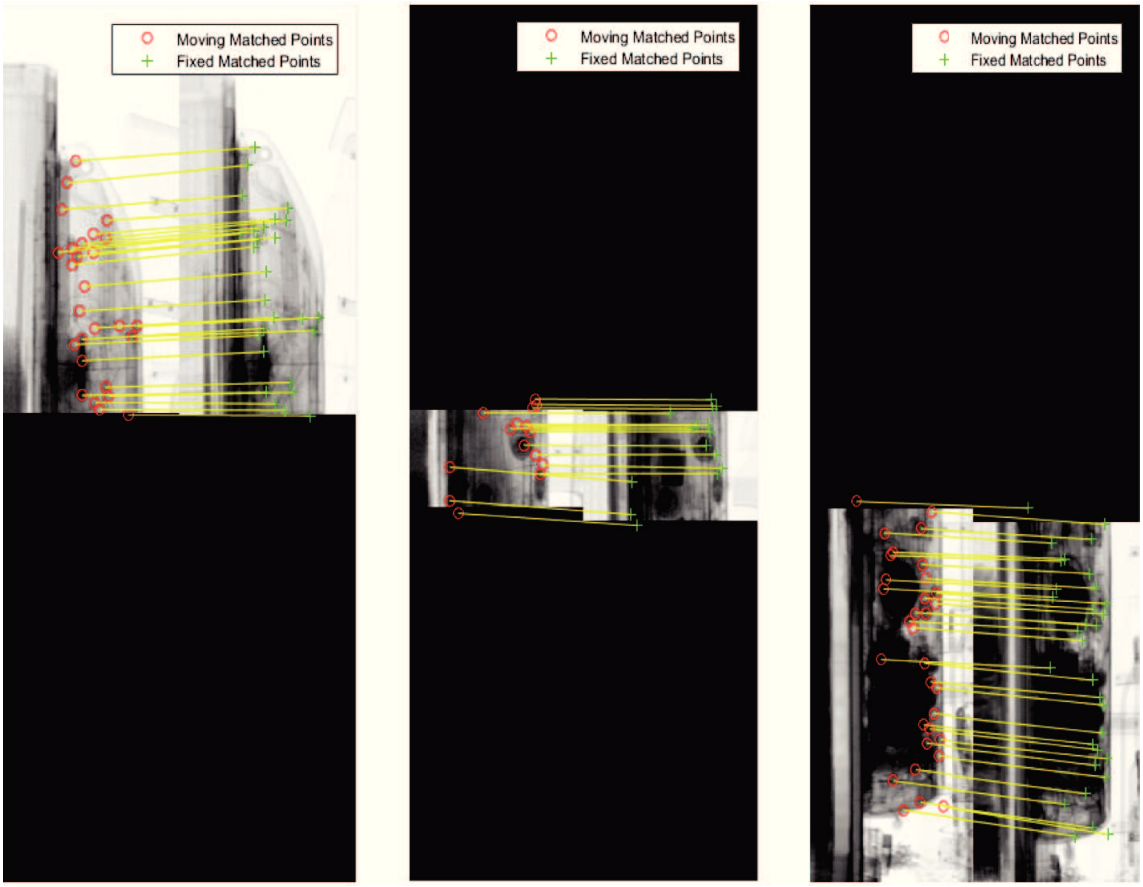
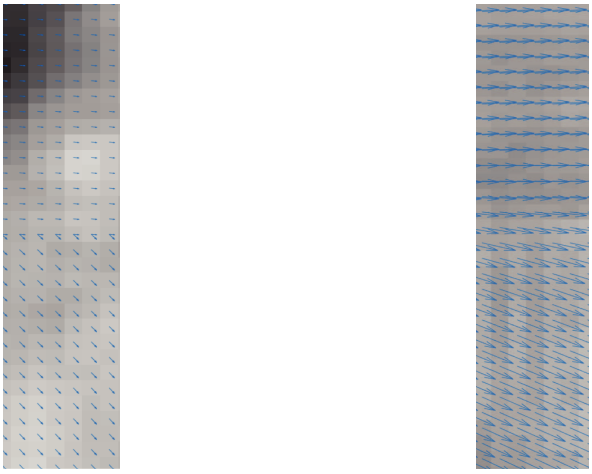


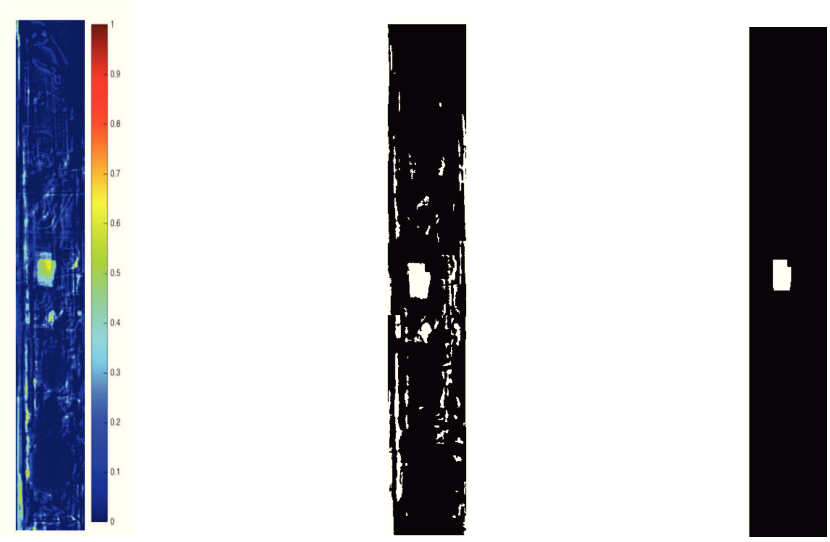
Figure 5.35 – SURF features matching (after MLESAC filtering) for each pair of **R** and **T** sub-images.



(a) before regularization (b) after regularization

Figure 5.36 – Dense field of displacement comparison before and after diffusion regularization. At boundaries, (b) shows a smoother direction change.

The map of differences is computed through our image rationing approach, as motivated in subsection 4.5.2. Using Otsu's method, in combination with hysteresis thresholding, a binary map $\mathbf{B}$ is obtained and filtered by morphological operations, to yield the final detection mask $\mathbf{B}_d$. Actually, the type of illicit goods being searched were assigned a minimal size **specification of** 20×20 **pixels, such that structuring elements can be chosen accordingly**. In the following figures, the difference detection process is depicted as well as the suspicious item localization output.



(a) map of differences (b) resulting mask (c) filtered mask

Figure 5.37 – Difference detection process

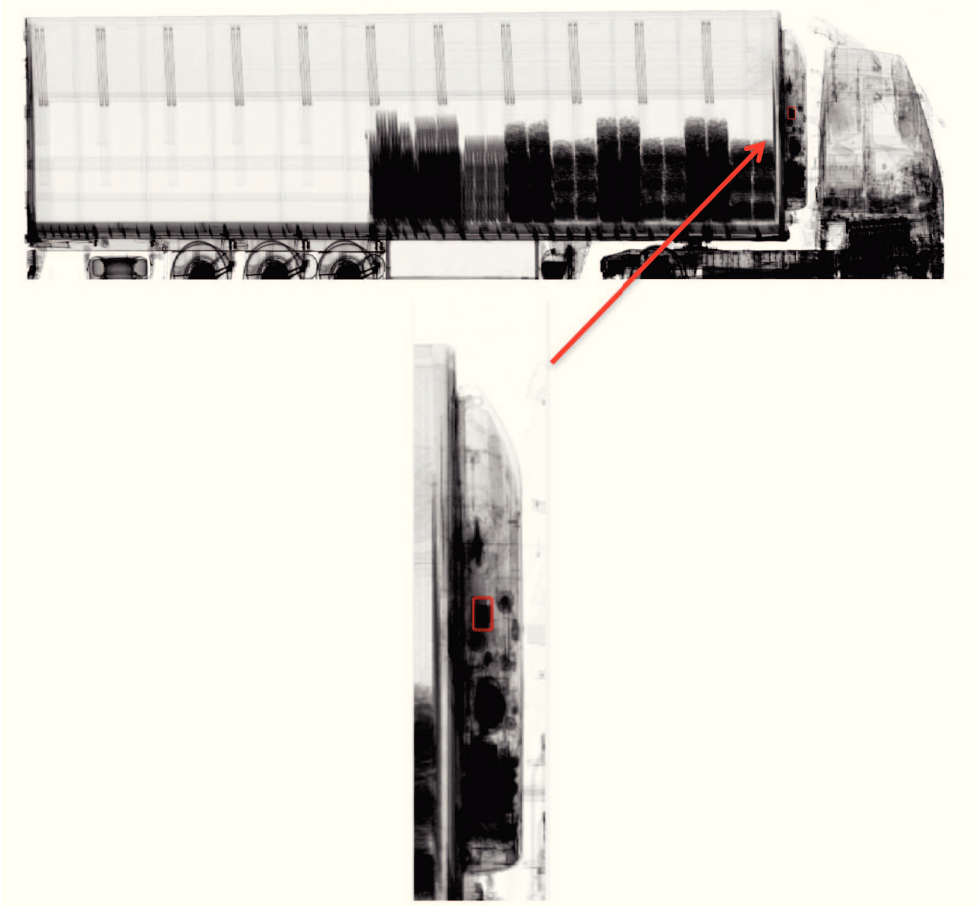


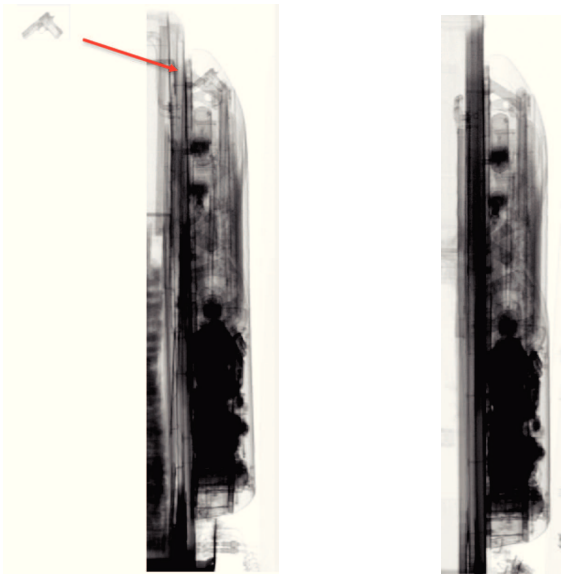
Figure 5.38 – Detection result indicated by a red bounding box

5.8.2 Additional Experiments

Rigid vs piecewise-rigid schemes

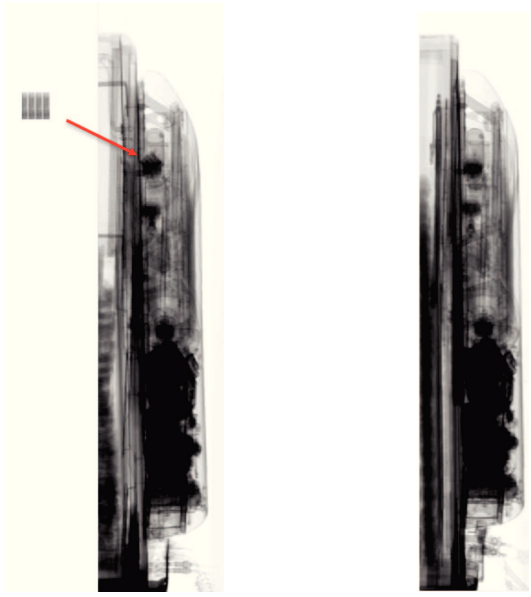
Because of a limited amount of data, an important proportion of images was allocated to training tasks rather than testing. Nevertheless, various tests have still been conducted with HCVM samples. The access to X-ray data with real targetted objects smuggled in air-cooler compartments being an issue; we produced synthetic images, as realistic as possible, by merging targetted objects (threats, narcotics), initially acquired without background. In this paragraph, two registration illustrations are provided, and results obtained by rigid and piece-wise rigid methods are reported. In certain cases, rigid registration results in similar performances. Yet, our piecewise-rigid paradigm often leads to a more accurate detection.

In the first example (Figure 5.39), a gun was inserted at the top of the air-cooler compartment. In the second example of Figure 5.40, a set of small cocaine bricks were merged in a very dark area of the ROI, making it almost invisible.



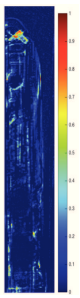
(a) **R** generated by merging (b) **T** retrieved via ASM

Figure 5.39 – (a) **R** generated by merging the gun X-ray patch into a regular air-cooler compartment; (b) **T** is the closest image in terms of shape as defined by our ASM application.

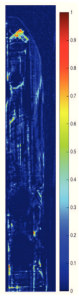


(a) **R** generated by merging (b) **T** retrieved via ASM

Figure 5.40 – (a) **R** generated via inserting a cocaine patch into a regular air-cooler compartment; (b) **T** is the closest image in terms of shape as defined by our ASM application.



(a) map of differences: rigid



(b) map of differences: piece-wise rigid

(c) difference mask: rigid

(d) difference mask: piece-wise rigid

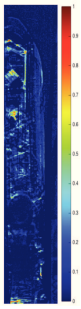
(e) filtered mask: rigid

(f) filtered mask: piece-wise rigid

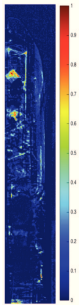
(g) result: rigid

(h) result: piece-wise rigid

Figure 5.41 – Difference detection steps for rigid and piece-wise rigid paradigms, relative to the example of Figure 5.39.



(a) map of differences: rigid



(b) map of differences: piece-wise rigid

(c) difference mask: rigid

(d) difference mask: piece-wise rigid

(e) filtered mask: rigid

(f) filtered mask: piece-wise rigid

(g) result: rigid

(h) result: piece-wise rigid

Figure 5.42 – Difference detection steps for rigid and piece-wise rigid paradigms, relative to the example of Figure 5.40.

In the gun-merging illustration, both methods yield similar detection results (Figure 5.41). However, regarding the cocaine bricks example, **the subdivision-based approach provides a better visualization of the threat in the map of differences**. As a consequence, after applying morphological operations on the binary mask, narcotics are still detectable as shown in Figure 5.42 (f) and (h). Besides, the global rigid registration scheme results in an extra false-alarm at the bottom of the module. **In this case, employing a piece-wise registration scheme allows to address local deformations between R and T in a more effective way**.

Switching to a global-linear approach

Still, one may be faced with the computation of non-consistent transformations accross sub-images with our the piece-wise linear procedure. The phenomenon is described on Figure 5.43. By this non-consistency issue, the transform matrices of two consecutive sub-images show significant differences. Thus, while running the procedure, we subtract consecutive transform matrices (from top to bottom for instance) and analyze the resulting differences. If the offset between translation and rotation parameters is too large, the piece-wise approach is aborted and a simpler global linear scheme is used instead.

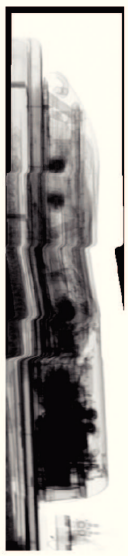


Figure 5.43 – Application of non-consistent transform matrices on sub-images of **T**.

5.9 Tests on Real Data

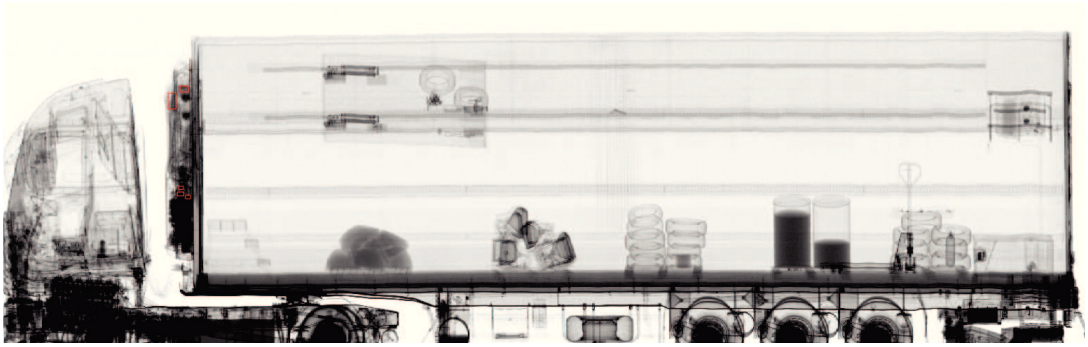
5.9.1 Tests on HCVM Real Data

Nonetheless, a very few real tests were made possible thanks to the European program we have been part of. In June 2017, a refrigerated semi-trailer was rented and scanned through an available HCVM facility in Vitry-sur-Seine. Also, weapons and explosive simulants were utilized during this testing campaign. Despite having only a few corresponding templates in our database, the algorithm performed well and led to true detection results. As illustrated on Figure 5.44, a few acquisitions were conducted with an AK-47 Kalachnikov rifle manually smuggled from the

outer door of the air-cooler module. In addition, we opened a trap from the inside of the trailer (located at the top left) and inserted explosive simulants. A few false alarms were also observed as in Figure 5.44 (b) at the bottom of the cooler unit.

(a) kalachnikov

(b) explosives



(c) kalachnikov and explosives

Figure 5.44 – Results from a demonstration in Vitry (June 2017). An AK-47 Kalachnikov rifle was manually smuggled in the cooler compartment, as well as explosive simulants.

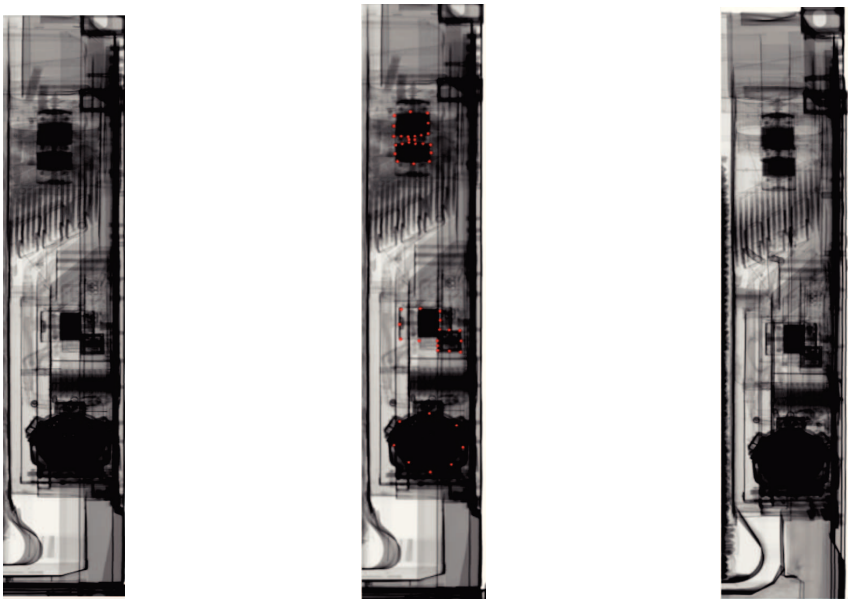
5.9.2 Tests on HCVS Real Data

Our algorithm has also been tested on HCVS real data at a meeting with customs authorities in 2016. A dozen of X-ray images of cocaine bundles smuggled within various refrigeration unit

models were provided to be processed by our algorithm. **Our proposed strategy demonstrated high detection performances without any false negative output**.

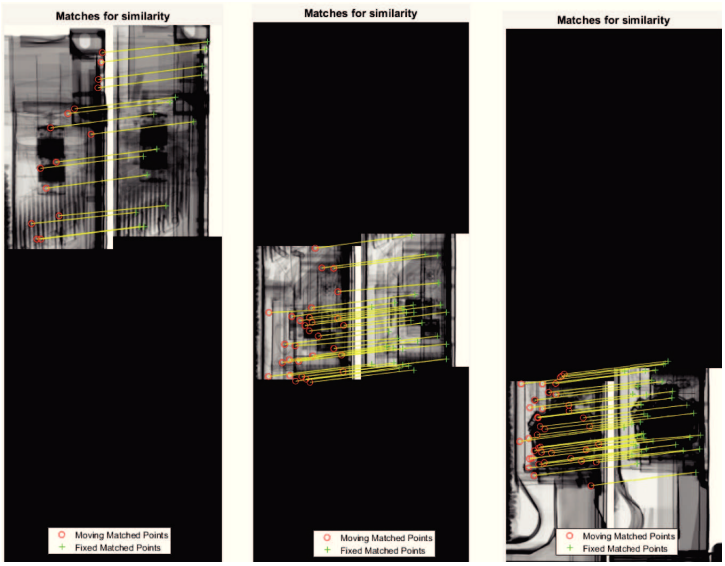
In this subsection, we describe the full analysis pipeline, applied on a X-ray sample of the dataset under consideration. Each processing stage is clearly illustrated: on Figure 5.45 the ASM fitting as well as the closest template retrieval are reported. Figure 5.46, the feature-based piecewise linear registration is described and a displacement field discontinuity can be observed. The smoothing process is depicted on Figure 5.47. Figure 5.48 displays the computation of the image of differences, the resulting filtered binary along with the final detection output. At the end, after applying our method on the dataset remainder, irregularity detection results are supplied on Figure 5.49.

Note that different ASM parameters are used for both training and fitting steps woth HCVS data. Indeed, due to the types of observed variations, together with the number of available training samples per class, appearance profile length is set to 20 pixels and the search length for optimal point positioning, in both normal directions is fixed to 10 pixels. More training samples are supplied, so the robustness of appearance models for each contour point can be increased by using longer profiles. Initialization is easier on high-energy images (less deformations occur) such that a smaller search length can be chosen. Computations are performed on two resolution scales and the number of iterations until convergence is set to $N_{iter} = 30$.

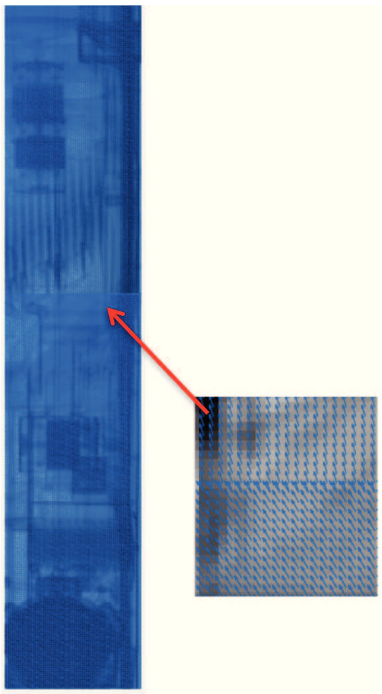


(a) **R** with smuggled cocaine. (b) ASM (c) Retrieved **T**

Figure 5.45 – (a) True X-ray image of cocaine bundle hidden in refrigeration unit concealment (acquired by HCVS); (b) ASM fitting to **R**; (c) closest template found in our database **T**.

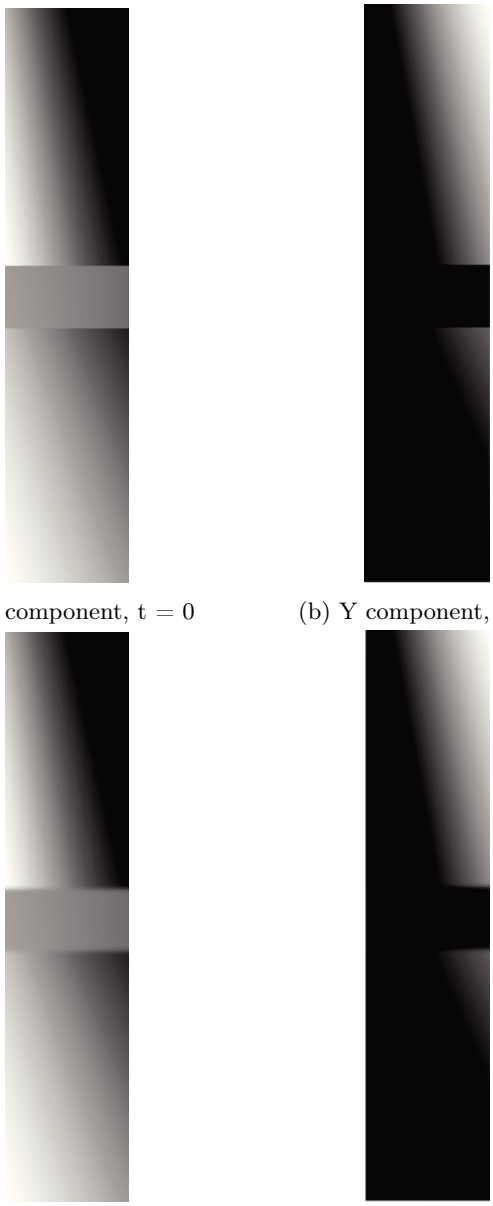


(a) SURF features extraction and matching



(b) Discontinuity issue

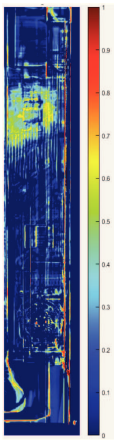
Figure 5.46 – (a) Extraction and matching of SURF feature points from pairs of sub-images; (b) Field of displacement discontinuity at sub-image borders



(a) X component, t = 0 (b) Y component, t = 0

(c) X component, t = 50 (d) Y component, t = 50

Figure 5.47 – (a),(b) Vertical and horizontal components of displacement field before direction diffusion at sub-images boundaries; (c),(d) following regularization, the displacement field is smoother at borders.

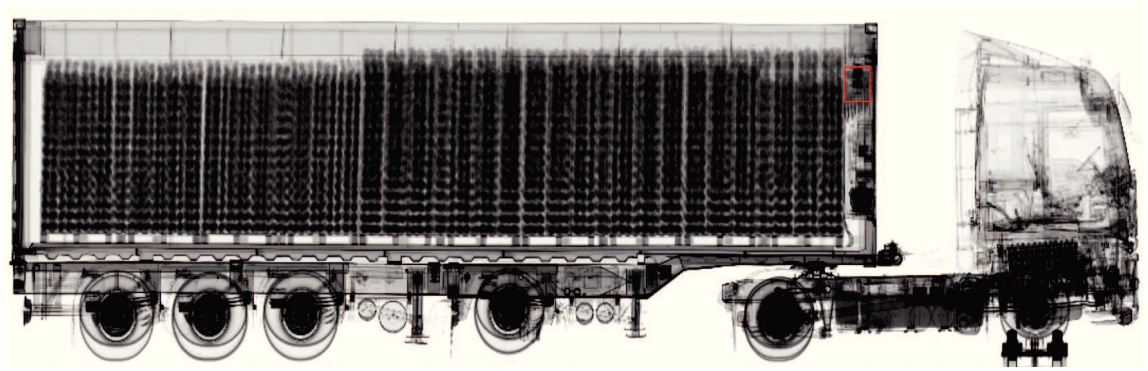


(a) Map of differences

(b) Binary mask

(c) Filtered mask

(d) Detection Result



(e) Full detection Result

Figure 5.48 – Different steps leading to the detection result.

(a) (b) (c) (d) (e) (f)

Figure 5.49 – Other tests performed on HCVS scans containing true cocaine bundles.

5.10 False Alarms Handling

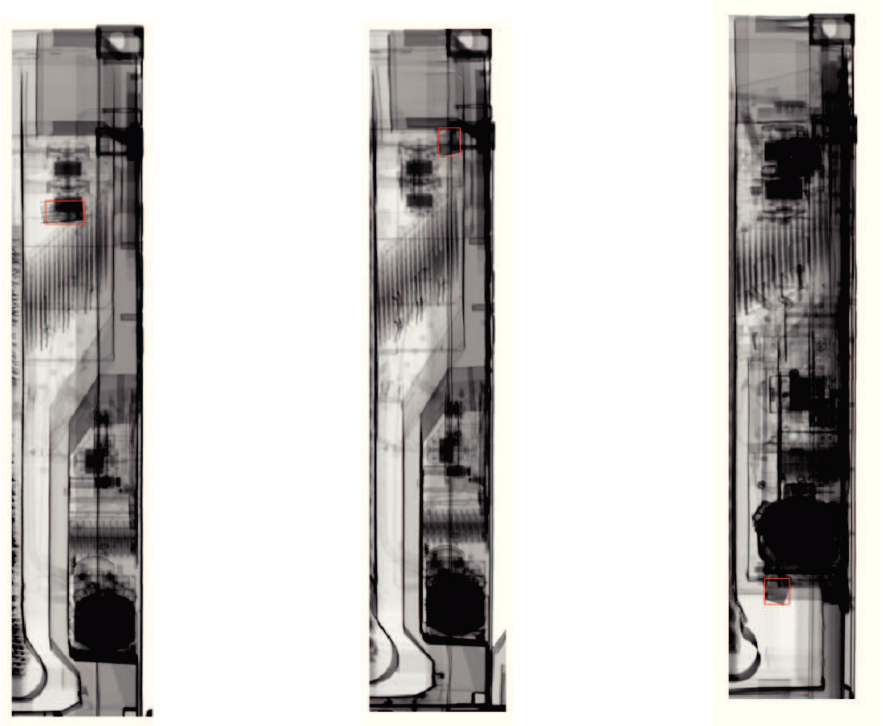
The meeting with customs authorities was indeed very informative, officers shared their field-based experience, providing more precise specifications and raising new problems. At the same time, a small dataset of cargo images displaying apparent irregularities in the refrigeration unit were supplied. **In particular, different settings of electric coils or ice block formation are very common, and may cause confusion**. Similarly, at the bottom-left of refrigeration modules, a thin metal sheet is placed to isolate the RU from the rest of the reefer container. The freight is sometimes prone to crush and utterly deform the metal plate as shown on Figure 5.50 (b). Figure 5.50 describes a few illustrations of false alarms caused by the above-mentioned phenomenons.

(a) ice formation (b) metal sheet being crushed (c) electric coils arrangement

Figure 5.50 – Different false alarms cases.

Further informations were also provided on areas within RUs which are unlikely to hide a threat. Despite the lack of substantive amounts of true threat data, we generated typical narcotics patches (by using rationing differences between **R** and **T**) on the basis of real scans of drugs. These different patches were randomly merged into regular RUs (different positions and orientations).

Through this method, thousands of detection samples have been created. They were used to build a binary **SVM classifier**, based on simple features such as **mean**, **standard deviation** and **position** of the irregularity being detected through registration. The training step was performed on 1460 positive and 1520 negative samples. **On a distinct 500 images test set we get a precision of** 93.26% **and recall of** 93.1%.



(a) (b) (c)

Figure 5.51 – Detection of merged patches of narcotics.

In the previous example, note that irregularities that are unlikely to be identified by human vision are detected on Figure 5.51 (a). Obviously, the whole method should be refined by processing new real data. Yet, his extra step could not be carried out on HCVM in the meantime, for which most images are used for training.

5.11 Algorithm Pipeline

The full pipeline of our algorithm is summarized on Figure 5.52.

5.12 Similar Applications

The strategy presented in this chapter is not confined to the refrigeration analysis task specifically. It can be widened to other regions of interest such as the chassis, as well as the roof, the wheels or the cabin (see Figures 5.53, 5.54 and 5.55).

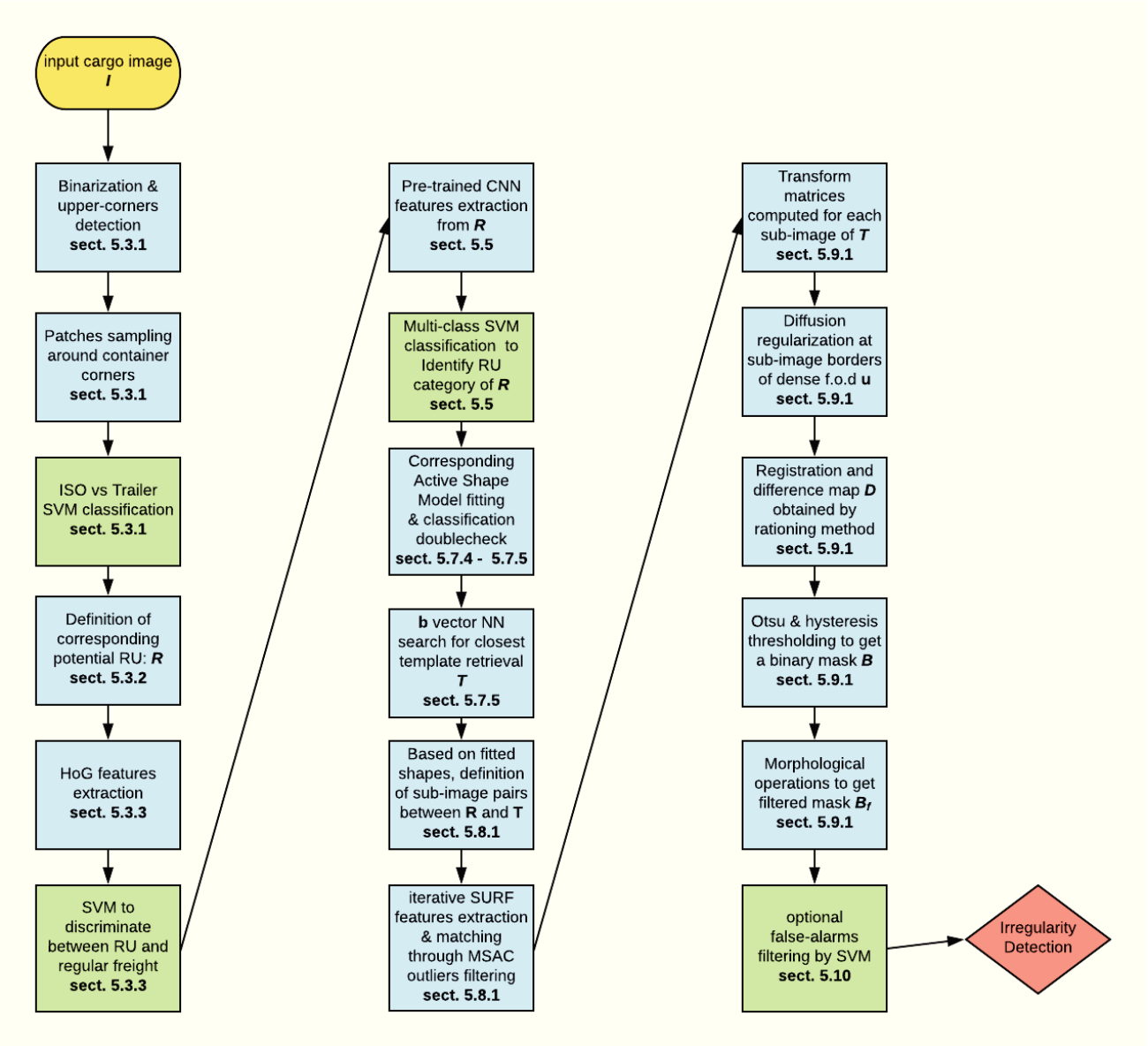


Figure 5.52 – Full algorithm pipeline. Classification steps are indicated in green.

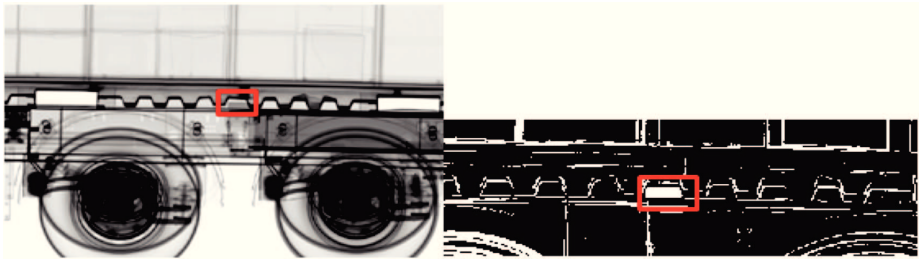
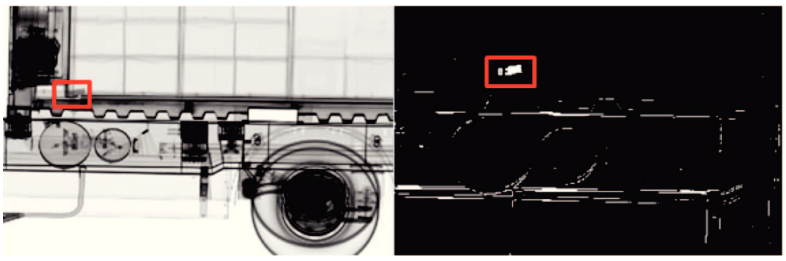


Figure 5.53 – Chassis inspection using registration

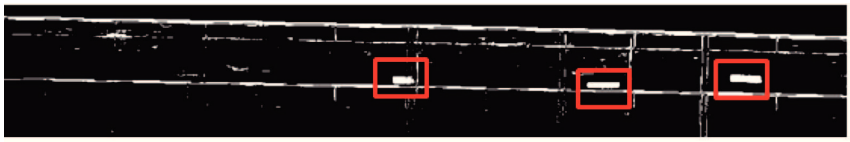
(a) drug introduced in sidewall



(b) detection by registration

Figure 5.54 – Detection example on sidewalls.

(a) drug introduced on the roof



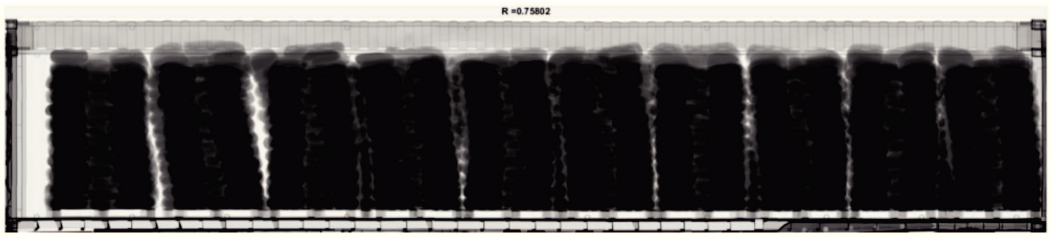
(b) detection by registration

Figure 5.55 – Detection of narcotics smuggled on the roof of a container.

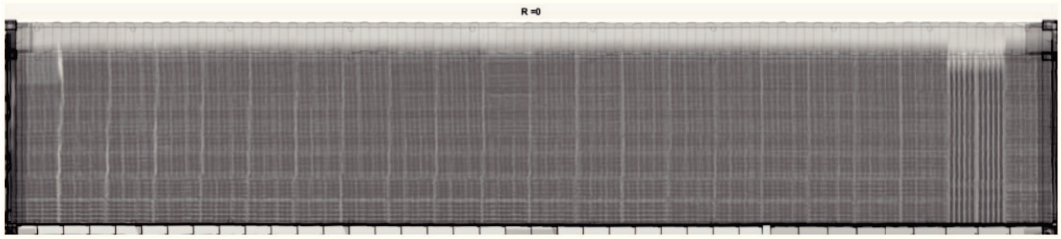
A more thorough study has been carried out for a door inspection assistant tool. The motivation is to design a simple method that could allow the analysis of the doors of a container. Generally, we are faced with two situations:

- the door seems clear: any detected discrepancy is considered as a potential threat (Figure 5.56 (a)).

- The door is cluttered: the fret encroaches on the door region. By using comparison methods, we check if the detected irregularity represents the continuity of the fret at the vicinity of the door (Figure 5.56 (b)).



(a) clear door



(b) cluttered door

Figure 5.56

A simple method to determine the door status would be to binarize the freight image by thresholding and analyze the vicinity of the door. A clearance ratio is defined by $r = \frac{l}{L}$ as described on Figure 5.57. Whenever r falls below 0.7, the door is considered as cluttered.

We observed several categories of doors with different mechanical structures. The analysis of more than 13000 HCVS images yielded a total of about 3000 clear doors which were used to build a classifier. See Figure 5.58 where a few instances are reported.

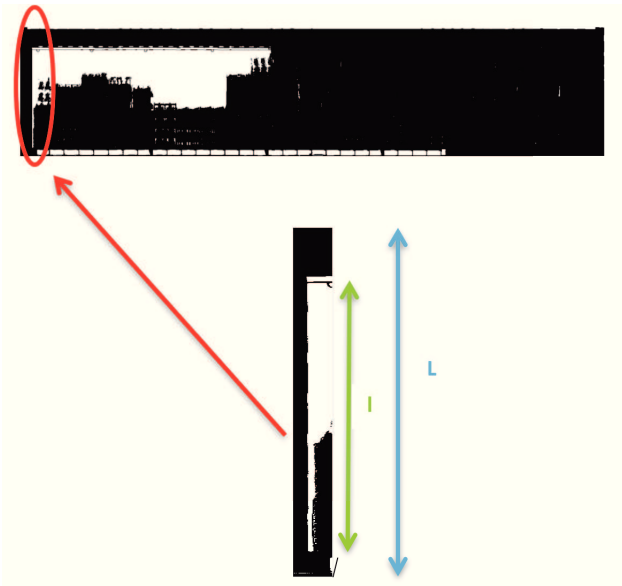


Figure 5.57 – Description of the clearance ratio.

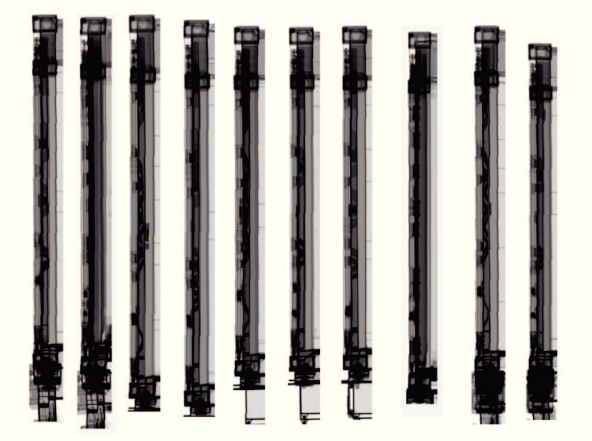
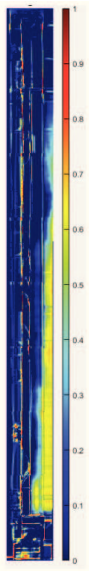


Figure 5.58 – Different door categories found across our HCVS dataset.

A bag of words approach (BoW) was employed to build a multi-class SVM classifier (due to the dimensions of door regions, BoW has been preferred to the CNN-features based SVM classifier). The overall accuracy on a 500 images test set reaches 90%. Similarly, classification was combined with a "closest template retrieval method" which consists in finding the most similar image within the dataset of a given class. HoG feature vectors are extracted from the current image and the dataset images of the corresponding class. Thus, the nearest vector determines the closest reference image for further registration tasks.

(a) **R** with corresponding **T**



(b) map of differences

(c) binary mask

(d) irregularity detection

Figure 5.59 – Door analysis by registration example

The analysis scheme is very similar to the refrigeration unit inspection process, as depicted on Figure 5.59. SURF features are extracted from both current and reference images. Matching points are filtered by MLESAC and an affine transform matrix is computed. To get a map of differences, we divide the current image by its registered reference. An adaptive thresholding along with morphological operations give a precise localization of the irregularity.

Figure 5.60 – Possible irregularity checking technique in case of cluttered region.

In a cluttered-door situation, an idea would be to extract Gabor features within the detected region and compare it with one or two windows selected rightwards on the longitudinal axis. Whenever the texture signatures are "different enough", the detected region would be marked as a potential threat (Figure 5.60). This example was brought to discuss the possibilities of addressing similar applications by drawing on our refrigeration unit analysis scheme. Door inspection algorithm combining Gabor features comparison is beyond the scope of the present work.

5.13 Conclusion

In this chapter, we developed a strategy to address automatic inspection of cargo refrigeration units and air cooler modules. The core idea of the algorithm stands on the combination of three central techniques:

- SVM classification with images being represented through pre-trained CNN features.
- On the basis of a prior training phase, Active Shape Models are used to find the closest sample from a specific class set (in terms of shape). It often enables to refine classification outputs.
- A piece-wise registration scheme employed between sub-image pairs of $\mathbf{R}$ under inspection and $\mathbf{T}$, the closest template found. It handles well local changes and allows for further irregularity detection.

The method has been validated on real data, and yielded very promising results on both HCVM and HCVS X-ray imaging systems. Still, for more robustness and accuracy, the different models require larger amounts of data.

Finally, the presented method is general and can be applied to a wider range of applications. In the specific cargo screening context, doors, chassis, roof and other rip-off locations could be analyzed through similar algorithmic pipelines.

Chapter 6

Registration of X-ray Images of Vehicles

Outline of the current chapter

6.1 Introduction

In the previous chapter, we dealt with scanning machines designed for containers and trucks screening. Other systems are dedicated to lighter vehicles, producing a top-view scan of the inspected subject. This type of stationary machine, known as HCVL, is widely used at borders, harbors, channel tunnels, sensitive sites such as embassies, other diplomatic representations and so on. Its efficiency lies also in its high scanning rate that reaches more than one vehicle per minute.

When a vehicle enters the machine, the driver goes out and waits in a remote place (protected from radiations). The car is trailed by a conveyor and is smoothly displaced through the machine.

Figure 6.1 – HCVL system representation. Each vehicle is trailed slowly to be imaged.

We aim at making use of registration techniques to push further the end-user visualization capabilities. More particularly, we wish to find hidden objects at dense locations where more advanced detection approaches would probably fail. We will therefore assume that for an inspected image $\mathbf{R}$, a corresponding $\mathbf{T}$ representing an empty vehicle of the same model can be retrieved.

First, we will show that rigid registration is not adapted to our problem, despite the fact that vehicles are rigid by nature. Then, different experiments will be conducted through state-of-the-art parametric and non-parametric approaches, yet without achieving satisfying performances.

Thus, the origins of non-linear deformations observed in the X-ray images are analyzed, and lead to simplifying assumptions expressed on the field of displacement $\mathbf{u}$. We formulate a very simple one-dimensional horizontal registration scheme that seems to address the non-linear deformations issue very nicely.

Then, a second misregistration phenomenon is reported, principally occurring when the scanned car in $\mathbf{R}$ is laterally shifted with respect to the reference vehicle's position in $\mathbf{T}$. In

fact, the radiated beam is pyramidal, which entails a stereoscopic effect. We therefore add a 1-D unconstrained optimization to the previous scheme for a y-direction correction. We also demonstrate that applying the vertical correction following our 1-D optimization in the x-axis yields better results than performing a simultaneous minimization on both components.

Finally, throughout our different experiments, incompressibility constraints are studied and their necessity for our detection task is strongly underlined.

6.2 Database Construction

Considering the top-view scans registration issue, our HCVL dataset was collected in different manners:

- Most of our images were provided by customs partners in different regions of the world (Europe, South America...). This way, we had access to thousands of unlabelled X-ray images. Though, for our registration task, organizing the dataset into categories is crucial. In order to label the inputs by car model, license plates information (automatically extracted by the scanning machine) was used. A total of more than three thousands models of vehicles have been identified so far, on the basis of the usable gathered data.

- At Smiths Detection's site, in Vitry-sur-Seine, an HCVL system, usually employed as a demonstrator for customers, was also available. This is how a couple of acquisition campaigns were conducted, in which employees were invited to place their vehicles at our disposal. In certain cases, various types of cars and trucks have directly been hired, allowing to image a very wide range of light vehicles. Besides, demilitarized weapons as well as drugs and explosives simulants were made available to us. Therefore, each vehicle was scanned twice to get a suspicious sample $\mathbf{R}$ as well as its matching "clean" template $\mathbf{T}$.

Before registering $\mathbf{R}$ and $\mathbf{T}$, the trailer outside the car is filtered out by a simple closing operation on the binary mask of the image. This pre-processing step is necessary, considering that the trailer may bias the search for an optimal alignment between both X-ray scans.

Model recognition of inspected vehicles is beyond the scope of the present study. Though, some ideas are briefly discussed in section 6.8 of the current chapter.

6.3 Rigid Registration

In the meantime, let us assume that, given an image $\mathbf{R}$, we are able to find its corresponding empty reference $\mathbf{T}$.

By definition, a vehicle is viewed as a stiff body made of mechanical parts that can barely be deformed. Thus, when coming to image registration of top-view scans of vehicles, employing a rigid scheme would seem to be a natural choice.

However, a quick observation of the database recurrently shows substantive non-linear deformations in X-ray output samples. For instance, such distortions are found in different regions of the following images:

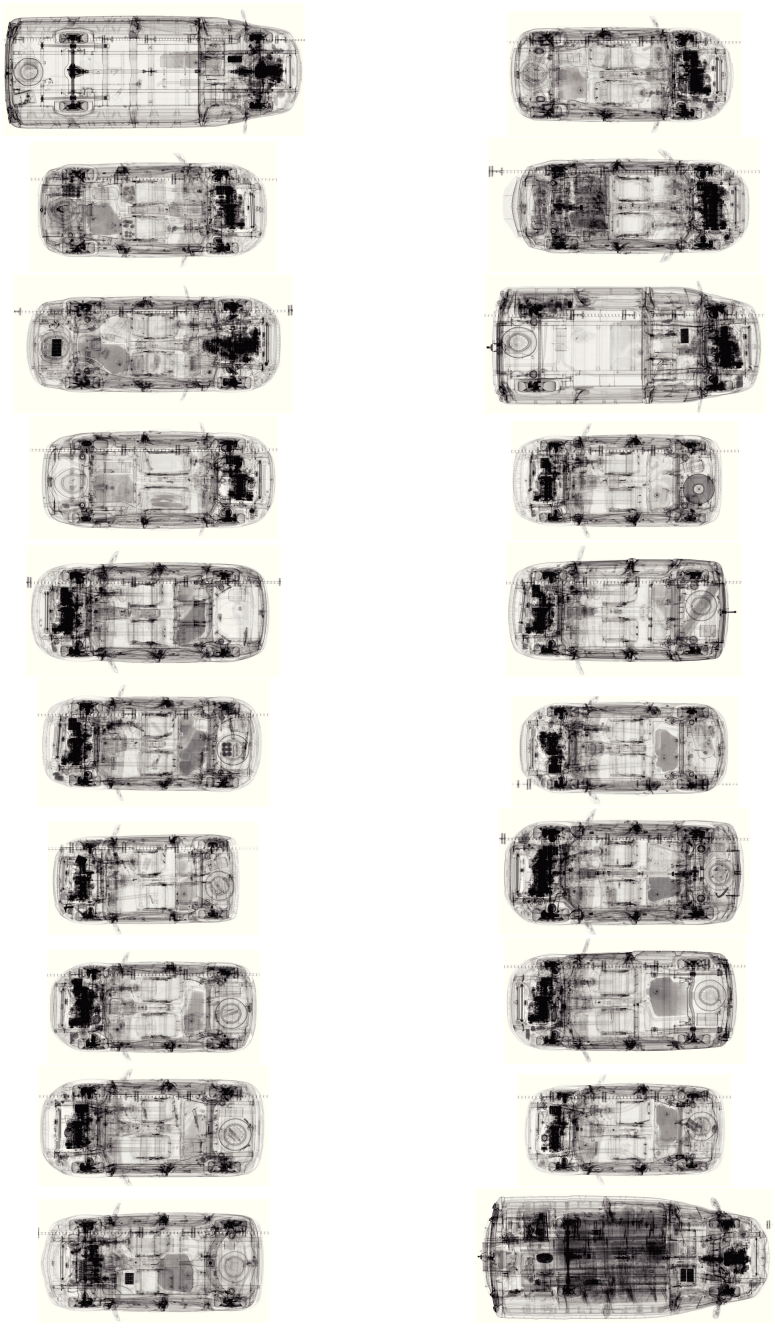
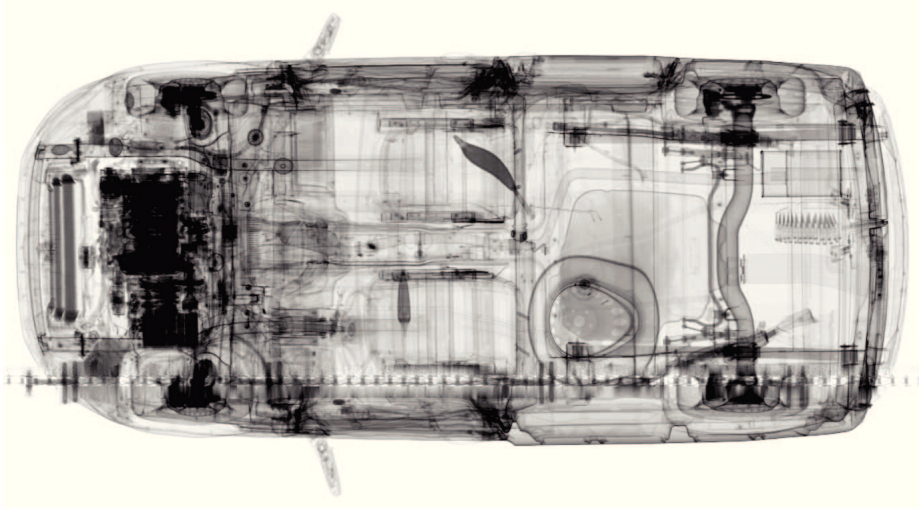
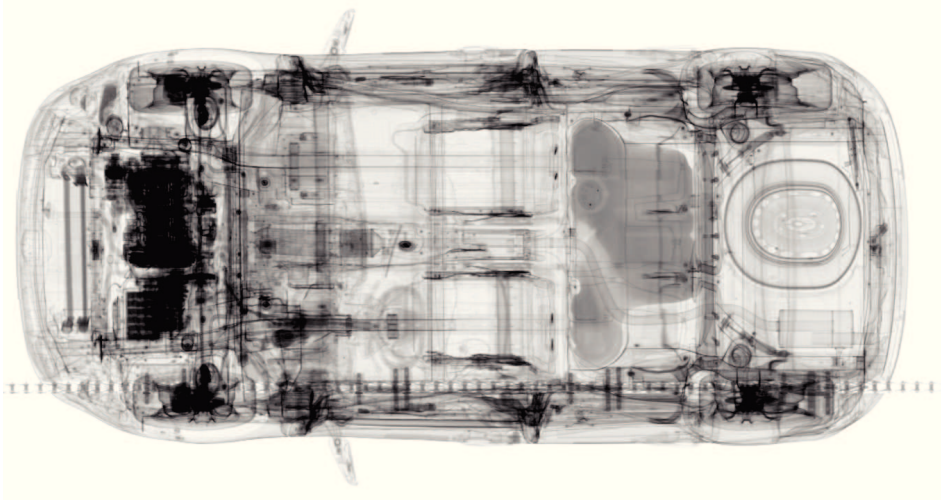


Figure 6.2 – Thousands of collected HCVL X-ray samples.



(a)



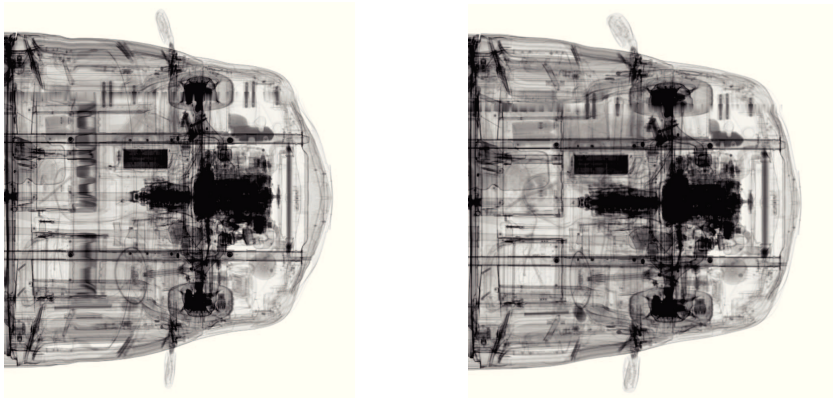
(b)

Figure 6.3 – HCVL top-view X-ray images of cars. Some regions *e.g* the spare wheel or weapons in (a) show non-rigid deformations resulting from the scanning process

Hence, a simple, rigid approach for registration may turn to be non-efficient as shown in the forthcoming example.

We focus on the front part of a car to conduct our experiments. Two HCVL scans of the same model but acquired on different cars (different license plates, at different times), were chosen to be registered.

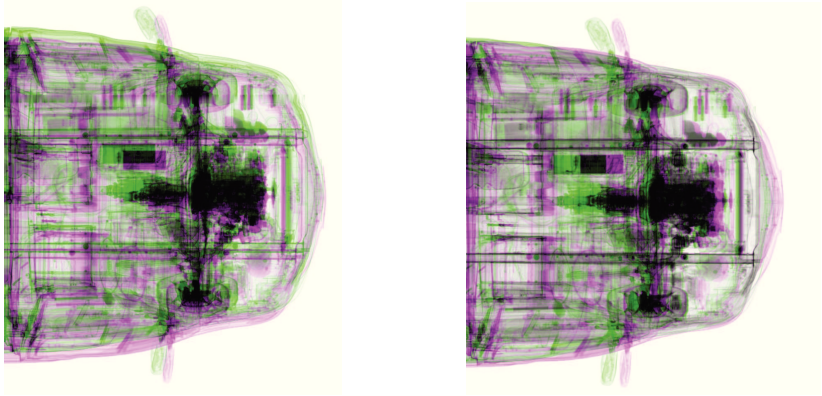
In this case, the inspection of the front part of a given vehicle is considered (Figure 6.4 (a)). A corresponding empty reference scan of the same model but acquired on a different car is used for registration (Figure 6.4 (b)).



(a) Target image being inspected **R** (b) Empty source image **T**

Figure 6.4 – Front-part images of the same model acquired on different vehicles.

We extract and match SURF features from **T** and **R** and extract a similarity transform via RANSAC filtering. After the application of the linear transform on **T** to align as good as possible with **R**, the overlay of both scans is provided on the following figure:



(a) overlay before rigid registration (b) overlay following rigid registration

Figure 6.5 – Overlay obtained before and after rigid registration (similarity transform). **Green and pink colors depict respectively the differences originating from the moving template T and target image R**.

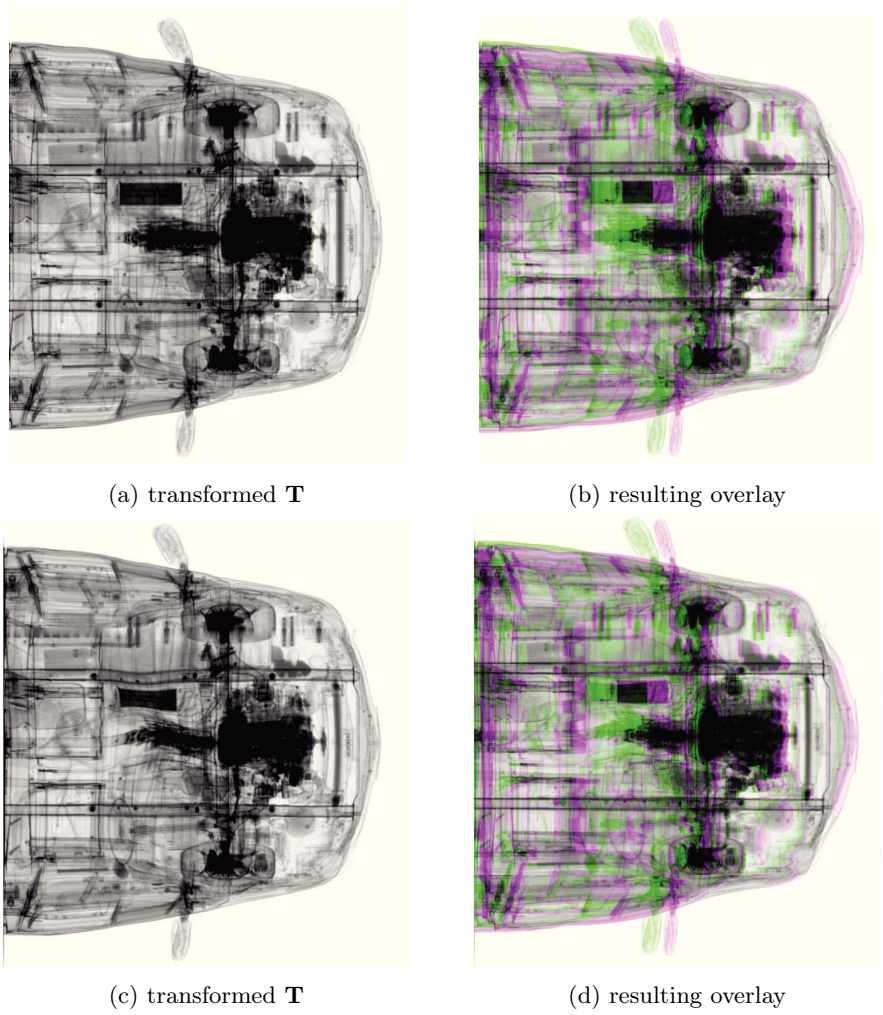
Thus, major visual differences are easily identified in the overlay map describing rigid registration (Figure 6.5 (b)). In particular, the rear-view mirrors, front wheels, the gearbox, the steering wheel, a few liquid tanks as well as the battery spots are misregistered.

In an initial stage, we did not strive to analyze the roots of non-linear distortions in resulting images. We rather tested parametric and non-parametric approaches to register our car fronts instances **R** and **T**, and observed the behavior of both algorithmic methods.

6.4 Testing Parametric and Non-Parametric Methods

6.4.1 B-Spline

The parametric registration method employed is the B-spline method. **R** and **T** are first pre-registered through a rigid transform. Concerning the numerical resolution of the optimization problem, l-BFGS is used along with a multi-resolution scheme (three levels). The algorithm was run for two different smoothing parameters controlling the "thin sheet of metal" term presented in 3.3: 0.1 and 0.01. The results are provided in the following figures:



(a) transformed **T** (b) resulting overlay

(c) transformed **T** (d) resulting overlay

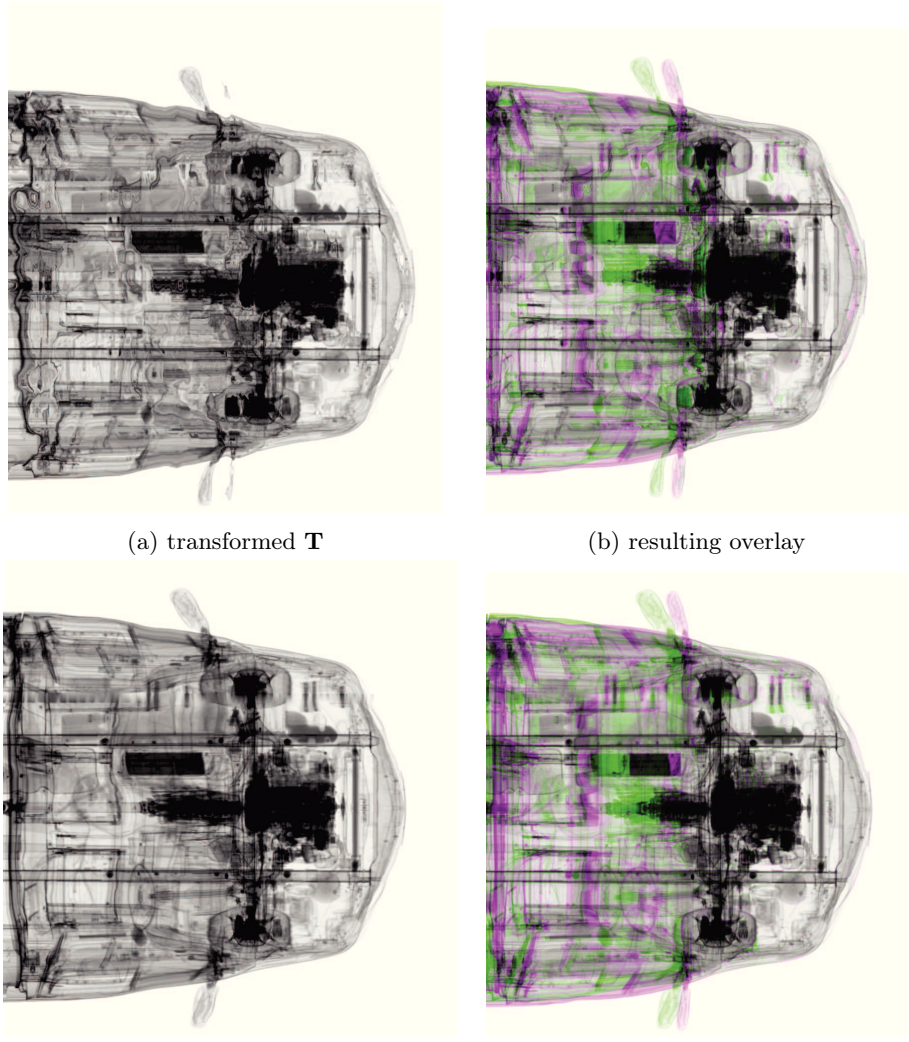
Figure 6.6 – Results obtained by **B-spline** run with regularization parameters 0.1 for (a) and (b) and 0.01 for (c) and (d)

Besides computational cost considerations (B-spline registration can be significantly time-consuming), two observations are to be made. First, the B-spline technique does not yield a perfect alignment as most misregistered regions remain. In addition, the intrinsic stiffness feature of the car is not respected since non-linear deformations are permitted. In Figure 6.6 (c)

for instance, the low regularization parameter entails a strong nonrigid distortion, which is far from being plausible with respect to the typical structure of a vehicle.

6.4.2 Demons Registration

Among the non-parametric approaches, we tested Thirion's demons method with a multilevel strategy (three scales) and two different smoothing parameters: 0.5 and 2.



(a) transformed **T** (b) resulting overlay

(c) transformed **T** (d) resulting overlay

Figure 6.7 – **Demons** registration outputs computed with smoothing parameters (a),(b) 0.5 and (b), (c) 2

In Figures 6.7 (b) and (d) a few regions display a better alignment than these obtained by B-spline (Figure 6.6) or rigid registration (Figure 6.5). It applies to the liquid reservoirs, the engine and can be observed at the bumper spot as well. Yet, many pink and green regions are

still visible after running demons registration, which is notable on the rear-view windows and the battery.

Besides, while adopting a low regularization parameter, the registration scheme is not diffeomorphic. The deformed version of $\mathbf{T}$ is not "reasonable" in terms of rigidity and topology of the objects, as depicted on Figure 6.7 (a). In the light of these experiments, the standard demons approach does not achieve accurate registration for the particular case we must address hereby. Numerical results (mean-square errors) are reported in section 6.6.3.

Instead, one should focus on examining the scanning system properties and specific constrains behind the non-linear warping observed in Figure 6.3. The idea would be to formulate an alternative optimization scheme that integrates the inner vehicle stiffness along with the mechanical process causing the non-rigid deformation.

6.5 Cause of Non-Linear Deformations

In HCVL X-ray imaging systems, the inspected vehicle is trailed by a conveyor facility via rollers pushing on its back wheels (Figure 6.8). The conveyor speed is meant to remain constant and fixed to 12 m/min.

Though, the car often rolls off the trailer equipment as a result of shocks between the conveyor rollers and the wheels. Thus, the vehicle speed undergoes disturbances affecting the scanning process: a slowdown (resp. acceleration) implies a local “over-sampling” (resp. “sub-sampling”) in specific regions of the car (see Figure 6.3).

Let us consider two X-ray images of the same car model (not necessarily the same vehicle): we assume that pose estimation has already been performed such that both images were linearly registered through a rigid transform. We formulate two strong hypothesis:

- **Hypothesis 1**: A *columnwise-constant deformation*. We make the reasonable assumption that the field of displacement is uniform along each column. In fact, the car is scanned with a constant sampling frequency while entering the X-ray beam line, hence each column of the resulting image corresponds to a lateral cut of the vehicle at a given time (Figure 6.10). Any speed disturbance would therefore affect each separate column in a uniform fashion. The idea is illustrated on Figure 6.9.

- **Hypothesis 2**: The deformation direction is perpendicular to the X-ray beam line (the vertical component of $\mathbf{u}$ is null). This assumption is an accurate approximation since the car can hardly strive from the conveyor during scanning (Figure 6.10).

Figure 6.8 – The conveyor rollers (in yellow) applying a pushing force on the rear wheels before scanning.

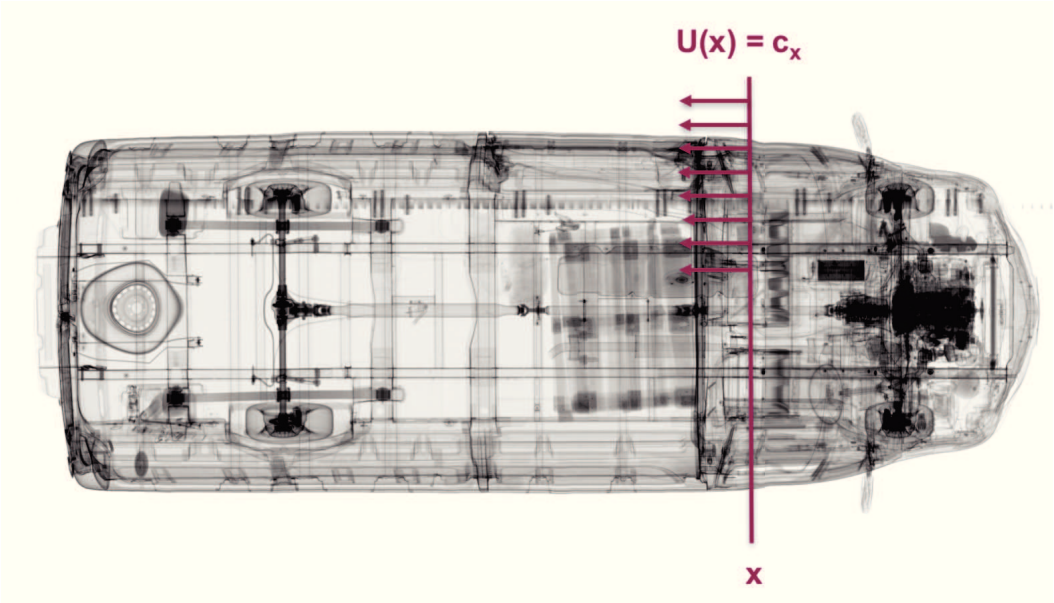


Figure 6.9 – Description of the column-wise constancy constraint. For each column, the deformation is meant to be equal for every ordinate y.

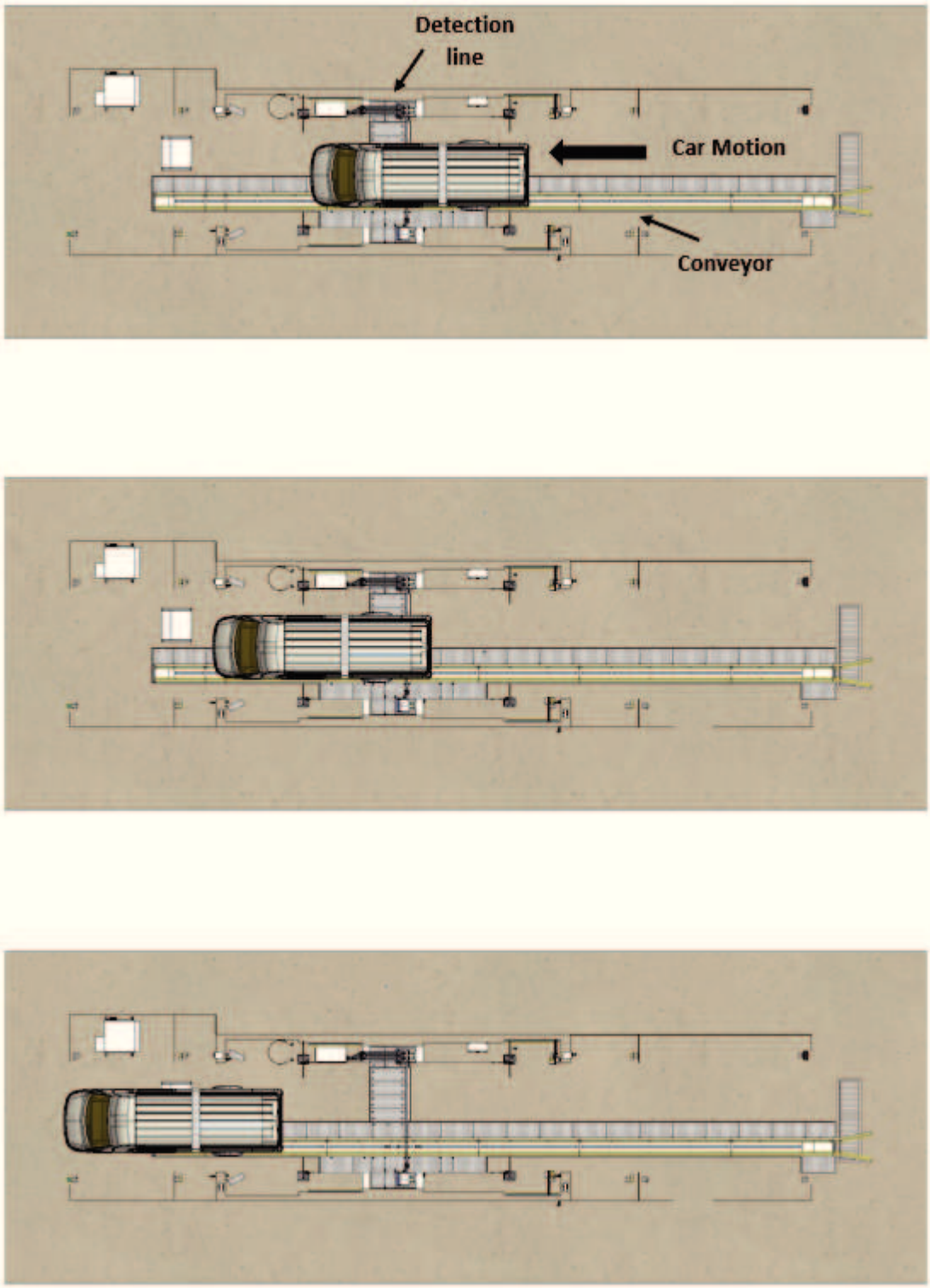


Figure 6.10 – Top-view description of the car motion during scanning in the HCVL system.

As we will show in the next section, by employing the variational framework, these simplification hypotheses turn the 2-D optimization scheme into a simple unidimensional minimization problem.

6.6 One-Dimensional Optimization Scheme

6.6.1 Problem Setup

Let us formulate the 1-D optimization problem. Given two images $\mathbf{R}$ and $\mathbf{T}$, pose estimation is performed by applying a rigid transform on $\mathbf{T}$. For notation simplification, we keep using $\mathbf{T}$ to designate the moving image following this pre-registration warping. We define x and y as the horizontal and vertical coordinates of the position $\mathbf{x} \in \Omega$:

$$\mathbf{x} = (x, y) \tag{6.1}$$

Similarly, the displacement field is defined by:

$$\mathbf{u} = (u_x(x, y), u_y(x, y)) \tag{6.2}$$

Hypothesis 1 implies that:

$$u_x(x, y) = u_x(x) \tag{6.3}$$

And **Hypothesis 2** yields:

$$u_y(x, y) = 0 \tag{6.4}$$

Eventually, combining (6.3) and (6.4) gives:

$$\mathbf{u} = (u_x(x), 0) \tag{6.5}$$

Thus, by employing an *SSD* distance and a diffusion regularization, the 1-D registration scheme is formulated as follows:

$$\begin{aligned} &\text{find } u_x(x) \text{that minimizes the objective function:} \\ &\mathcal{J}[u_x] = \frac{1}{2}\|\mathbf{T}(u_x) - \mathbf{R}\|_2^2 + \frac{\alpha}{2}\|\nabla_x u_x\|_2^2 \end{aligned} \tag{6.6}$$

6.6.2 Numerical Resolution

A first-degree descent scheme would be relevant for this 1-D optimization problem (low complexity). We resort to the gradient descent method, in combination Armijo's backtracking line search method [78]. We evolve the following equation with Dirichlet boundary conditions:

$$\begin{aligned} &\frac{\partial u_x(x,t))}{\partial t} = -\nabla_{u_x}\mathcal{J}[u_x] \text{ for } x \in \Omega \setminus \partial\Omega \\ &\text{with } u_x(x) = 0 \text{ for } x \in \partial\Omega \end{aligned} \tag{6.7}$$

As developed in section 3.4, the gradient of $\mathcal{J}$ is obtained by computing its Gâteaux derivative. Since the directional derivative of the *SSD* data fitting term $\mathcal{D}$ has already been computed in

Chapter 3, we will focus on the diffusion smoothing term $\mathcal{S}[u_x] = \|\nabla_x u_x\|_2^2$. Let us address the two-dimensional case whose result will easily be customized to the one-dimensional scheme.

$$\begin{aligned} d_{\mathbf{u},\mathbf{q}}\mathcal{S}[\mathbf{u}] :&= \lim_{h\to 0}\frac{1}{h}(\mathcal{S}[\mathbf{u}+h\mathbf{q}] - \mathcal{S}[\mathbf{u}]) \\ &= \lim_{h\to 0}\frac{1}{h}(\|\nabla\mathbf{u}+h\nabla\mathbf{q}\|_2^2 - \|\nabla\mathbf{u}\|_2^2) \\ &= \lim_{h\to 0}\frac{1}{h}\langle h\nabla\mathbf{q}, h\nabla\mathbf{q}+2\nabla\mathbf{u}\rangle \\ &= \lim_{h\to 0}\frac{1}{h}\langle h\nabla\mathbf{q}, 2\nabla\mathbf{u}\rangle + \mathcal{O}(h) \\ &= \langle\nabla\mathbf{q}, 2\nabla\mathbf{u}\rangle \\ &= \langle\mathbf{q}, -2\Delta\mathbf{u}\rangle \end{aligned} \tag{6.8}$$

The last result is obtained via a simple integration by parts (see [8] for further details). On the basis of this computation, the gradient of $\mathcal{J}$ is expressed as follows in two dimensions:

$$\nabla_{\mathbf{u}}\mathcal{J}[\mathbf{u}] = \nabla_{\mathbf{u}}\mathbf{T}(\mathbf{x}+\mathbf{u})[\mathbf{T}(\mathbf{x}+\mathbf{u}) - \mathbf{R}(\mathbf{x})] - \alpha\Delta\mathbf{u} \tag{6.9}$$

In one-dimension, following the column-wise constancy constraint, we obtain, for each column x:

$$\nabla_{u_x}\mathcal{J}[u_x(x)] = \nabla_{u_x}\mathbf{T}(x+u_x(x))[\mathbf{T}(x+u_x(x)) - \mathbf{R}(x)] - \alpha\frac{d^2u_x(x)}{dx^2} \tag{6.10}$$

At the $k+1^{th}$ iteration, the displacement field update $u_x(x)^{(k+1)}$ for each column $x \in [0, N]$ is therefore given by (under Dirichlet boundary conditions):

$$\begin{aligned} &u_x(x)^{(k+1)} = u_x(x)^{(k)} - \rho^{(k)}\nabla_{u_x}\mathcal{J}\left[u_x(x)^{(k)}\right] \\ &u_x(0)^{(k+1)} = 0 \text{ and } u_x(N)^{(k+1)} = 0 \end{aligned} \tag{6.11}$$

With the step size $\rho^{(k)}$ obtained via Armijo's backtracking line-search. By abuse of notation: $\mathbf{T}(x+u_x(x))$ corresponds to the column of $\mathbf{T}$ at $x+u_x(x)$. Likewise, calculating $\nabla_{u_x}\mathbf{T}(x+u_x(x))$ amounts to computing the horizontal gradient of the source image $\mathbf{T}$ at $x+u_x(x)$. In practice, the transformed image $\mathbf{T_u}$ is first computed and the horizontal gradient is extracted.

A multi-scale approach is used to speed up the registration process and in order to avoid convergence at local minima (see Figure 6.11). Thereby, the displacement $u_x^{(p)}$ obtained at level p gives a strong initialization for the displacement estimate $u_x^{(p-1)}$ at the finer level $p-1$. The number of levels is fixed to $l = 4$.

A smoothing parameter α set to 10^{-4} for all scales yields accurate results in most simple cases (when a few sparse differences appear). The choice of the smoothing parameter is discussed in the coming sub-part. The stopping criterion is defined by a pre-set threshold $\epsilon = 10^{-4}$: whenever $\|\nabla_t\mathbf{u}\|_2^2$ falls below ϵ, the search for a minimizer is aborted. In parallel, the descent process is stopped if the line-search condition is not fulfilled. At the end, an estimate of $\mathbf{T}$ transformed via the optimal displacement field $\mathbf{u}^*$ is calculated using cubic interpolation (refer to Chapter 3 for more thorough information).

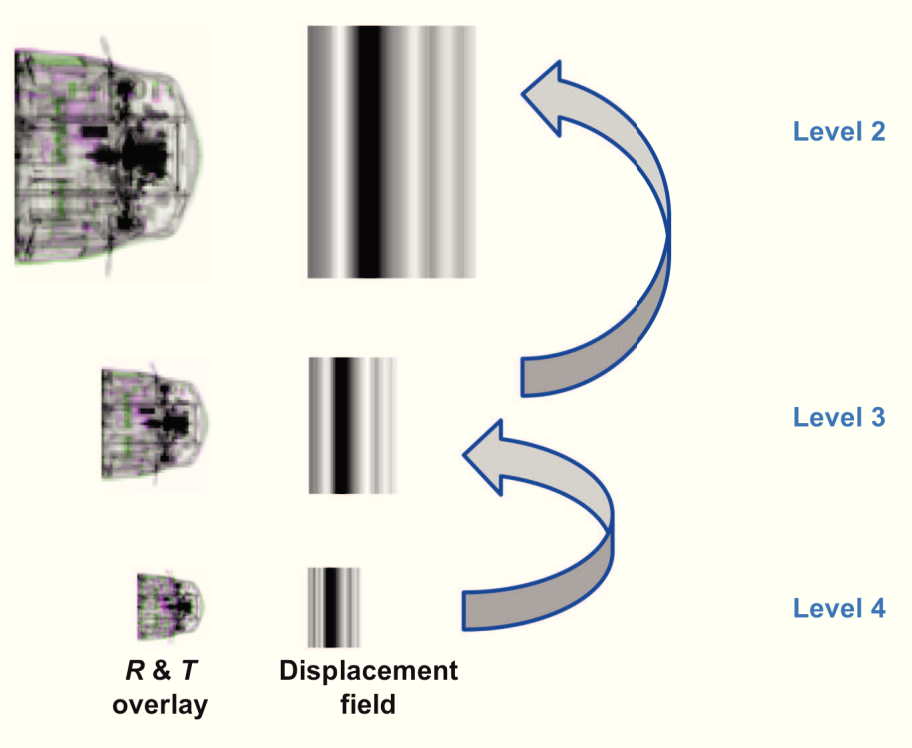


Figure 6.11 – Pyramidal coarse-to-fine approach description

6.6.3 Preliminary Results

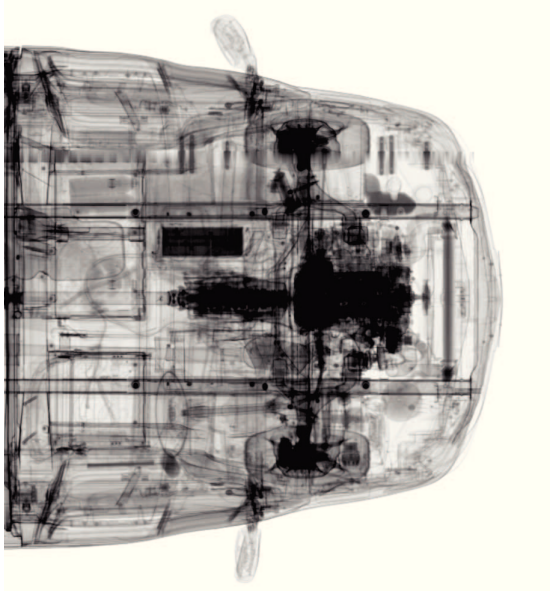
The next figures describe the results obtained by our unidimensional registration approach, applied on the front-hood pair of images.

In Figure 6.12, a significant shrinkage of the battery, front wheels and the engine can be spotted. Correspondingly, a conspicuous thinning of both rear-view windows regions are to be noted. In contrast, a strong dilation of the top-right liquid tank is observed.

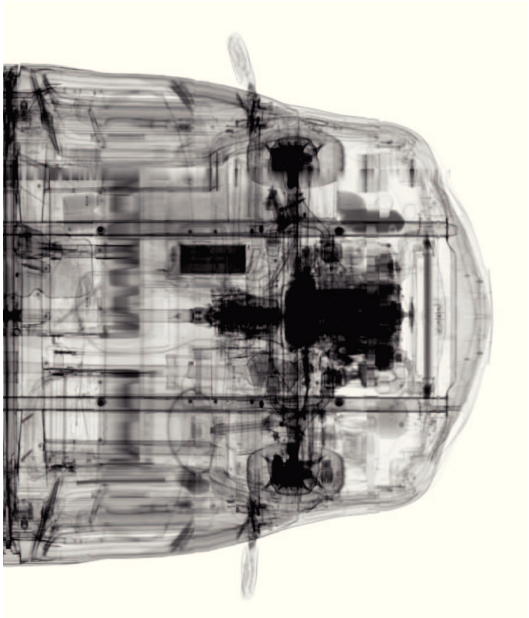
The overlay maps depicted on **Figure 6.13** show that most of **major differences have completely vanished** with our unidimensional constrained diffusion registration scheme.

The resulting field of displacement $\mathbf{u}^*$ obtained by gradient descent (6.11) is displayed on Figure 6.14. Because of the column-wise constancy constraint, it actually boils down to a 1-D vector. Its smoothness, ensured by a careful selection of α can be noticed as well.

Finally, a graph reporting the mean-square errors (MSE) achieved for the different tested methods is provided on Figure 6.15. The lowest cost is reached by our approach, which numerically demonstrates its advantage regarding the registration problem we must address.

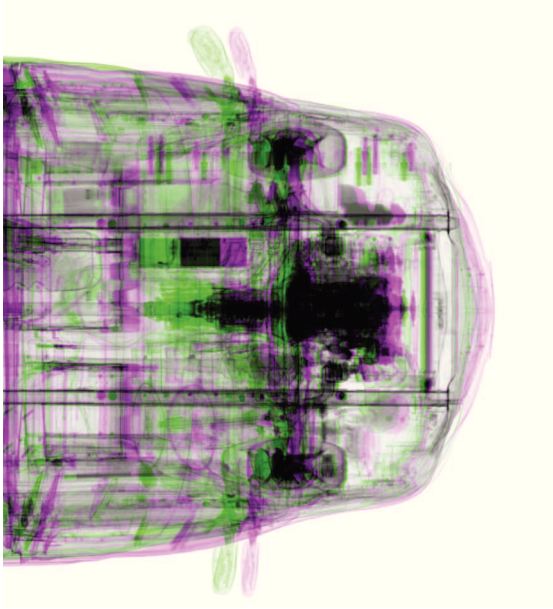


(a) initial $\mathbf{T}$

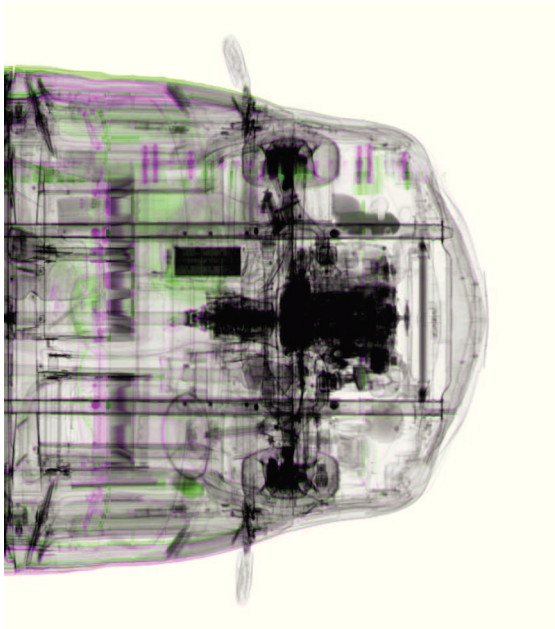


(b) $\mathbf{T}_{\mathbf{u}^*}$

Figure 6.12 – Comparison between source image $\mathbf{T}$ and $\mathbf{T}_{\mathbf{u}^*}$ after 1-D registration. Multiple non-rigid deformations can be observed *e.g* the shrinking of the battery and the rear-view mirrors



(a) Rigid Registration



(b) 1-D Non-Linear Registration

Figure 6.13 – Comparison of **T** and **R** overlays following rigid registration and unidimensional non-linear scheme. The main differences described by pink and green colors have disappeared in (b).

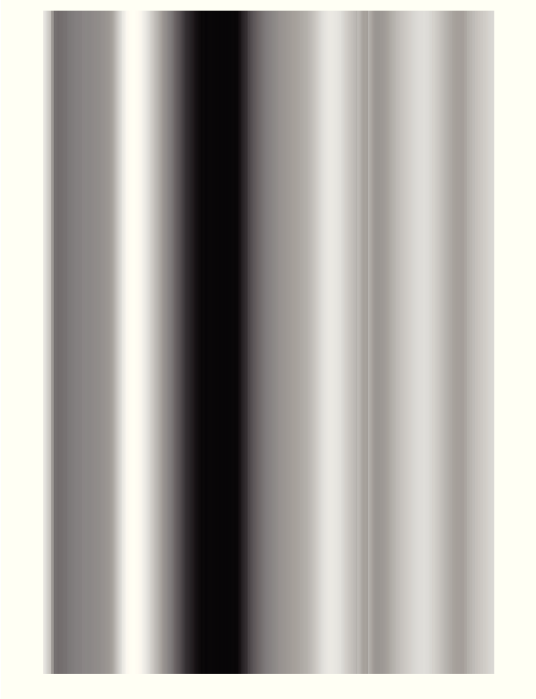


Figure 6.14 – Representation of the resulting column-wise constant displacement field $\mathbf{u}^*$.

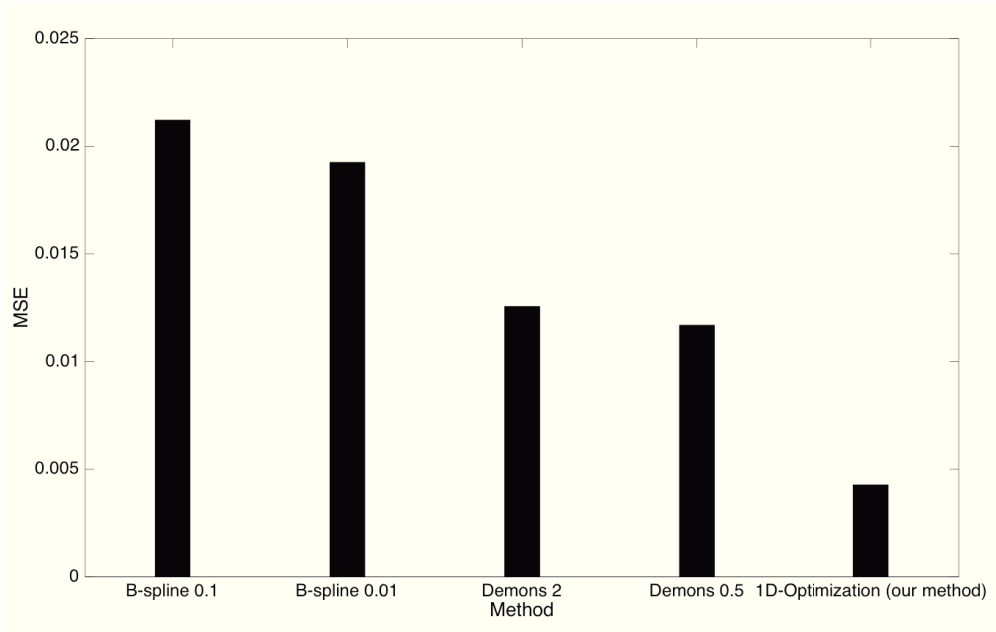


Figure 6.15 – Mean-square errors reached for each tested method (with its corresponding smoothing parameter)

6.6.4 Diffeomorphism Issues

Choice of smoothing parameter α

Rather than simply fixing a smoothing parameter α, a more stable and intuitive approach is adopted. Instead, a desired "smoothing ratio" r_α is set for all levels such that:

$$r_\alpha = \frac{\mathcal{D}[\mathbf{u}]}{\alpha \mathcal{S}[\mathbf{u}]}$$

At the first iteration of each level, α^p is thus estimated (for the given level $p \in [1, l]$) by simply computing:

$$\alpha^p = \frac{\mathcal{D}\left[\mathbf{u}^{k=0}\right]}{r_\alpha \mathcal{S}\left[\mathbf{u}^{k=0}\right]}$$

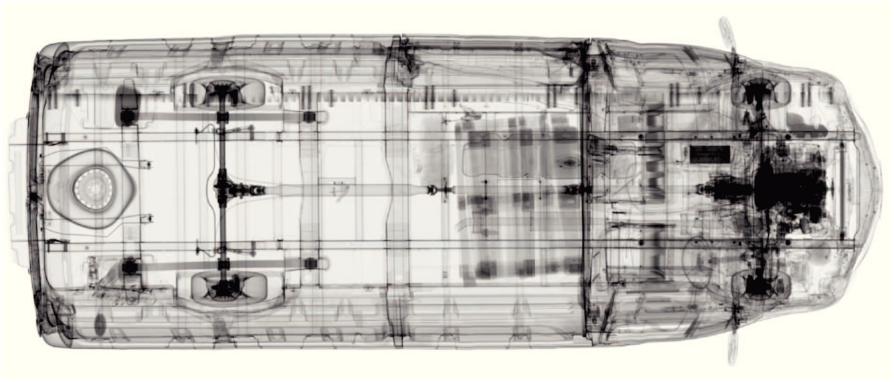
By doing so, the regularization can be applied with respect to a fixed proportion between the data fitting term and the smoothing component. Small ratios indicate substantial regularization while a larger r_α would entail a lower smoothing. For the coarsest level l, the initial displacement $\mathbf{u}_l^{k=0}$ is null, so is $\mathcal{S}$. We therefore empirically choose $\alpha = 10^{-4}$ for a rough estimate of the displacement at this level, thus allowing to relax the regularization constraints.

A few preliminary tests of the 1-D method carried out with different smoothing ratios show that r_α must be properly chosen. In fact, a too weak regularization entails non-diffeomorphic transforms resulting in very unlikely deformed images $\mathbf{T}_{\mathbf{u}^*}$. As mentioned previously, this occurs especially in presence of differences in $\mathbf{R}$. The source image $\mathbf{T}$ would inevitably tend to minimize the data fitting term and "imitate" the differences in an unreasonable way, causing serious artifacts in the resulting image.

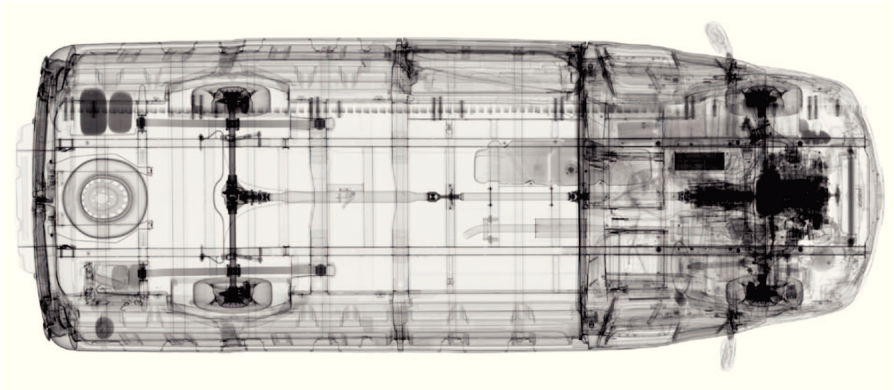
The next images describe the registration of a pair of whole X-ray images (full version of previous images, on Figure 6.16), using various ratio parameters in order to highlight the non-diffeomorphic transformation issue in specific regions. Thus, Figure 6.17 describes the evolution of the superposition of $\mathbb{R}$ and $\mathbf{T}_{\mathbf{u}^*}$ according to r_α. Figure 6.18 depicts more precisely the effect of a low regularization constraint. In $\mathbf{R}$ (Figure 6.16 (a)), several wooden pallets are found behind the passengers seats while $\mathbf{T}$ is empty. $\mathbf{T}$ will therefore be prone to deform such that $\mathbf{T}_{\mathbf{u}}$ matches as much as possible with $\mathbf{R}$. When a low smoothing parameter (larger r_α ratio) is picked, a non-reasonable resulting image cannot be avoided as observed on 6.18 (d) and (e). Indeed, as noted on the representation of the transform function ϕ (Figure 6.19), it is piece-wise constant, meaning that a single column in the source image $\mathbf{T}$ is repeated several times in the deformed result $\mathbf{T}_{\mathbf{u}^*}$. Hence, besides being continuous, ϕ is not strictly monotically increasing and cannot be viewed as a smooth bijective function.

Yet, with a more stringent diffusion regularization constraint, ϕ is maintained smooth and refrained from being piece-wise constant, as described on Figure 6.20. Accordingly, any column in $\mathbf{T}$ has a single correspondence in $\mathbf{T}_{\mathbf{u}^*}$ and no major artifacts are to be reported in the output (see for instance Figure 6.17 (b), (c) and Figure 6.18 (a), (b), (c)).

Figure 6.21 reports a few MSE results for different r_α values. A larger regularization ratio actually produces a lower error (the data fitting term is better minimized), to the detriment of a plausible transformed image $\mathbf{T}_{\mathbf{u}^*}$.

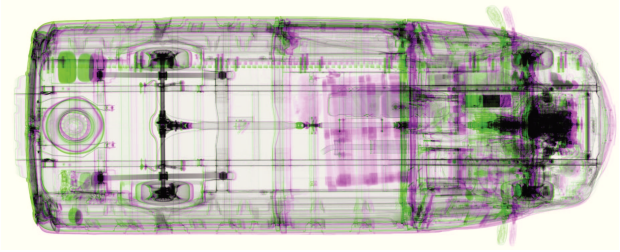


(a) Target image being inspected **R**

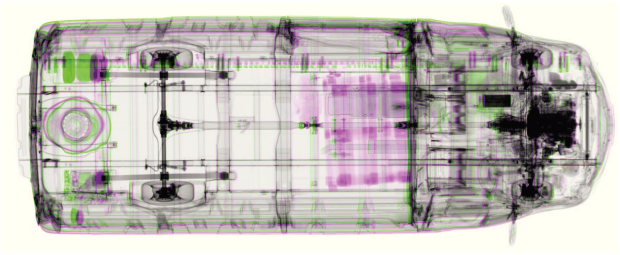


(b) Empty source image **T**

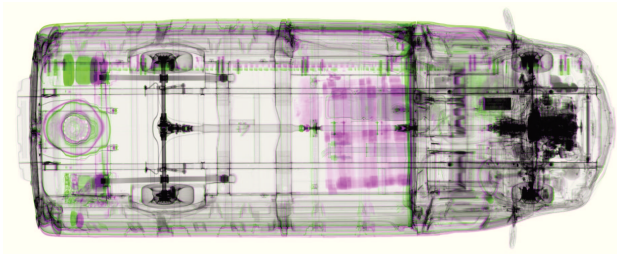
Figure 6.16 – Image pair to be registered. The scans were acquired on different vehicles of the same model.



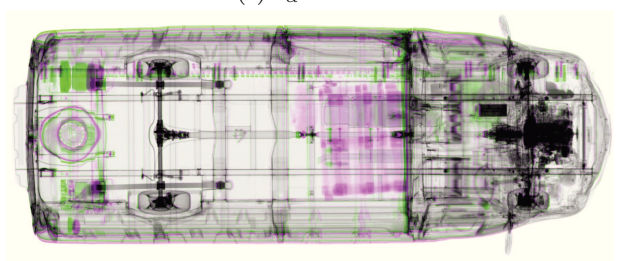
(a) rigid registration



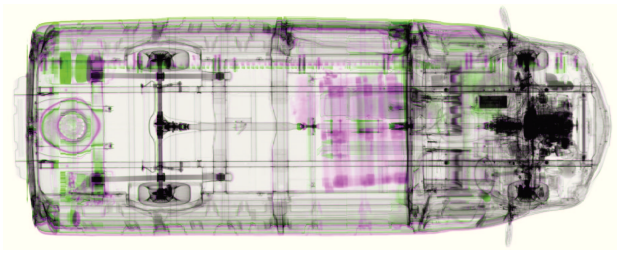
(b) $r_\alpha = 5.10^{-5}$



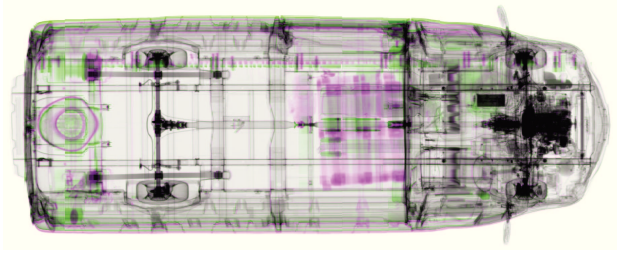
(c) $r_\alpha = 10^{-4}$



(d) $r_\alpha = 10^{-3}$

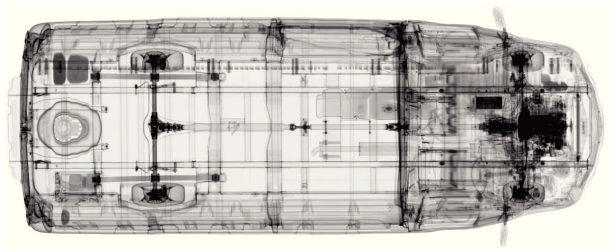


(e) $r_\alpha = 10^{-2}$

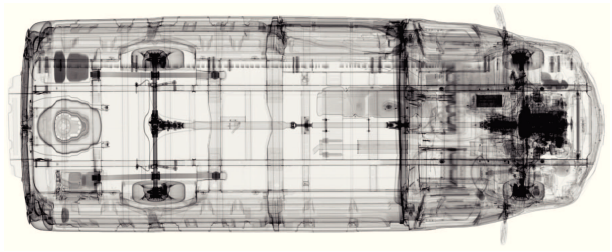


(f) $r_\alpha = 10^{-1}$

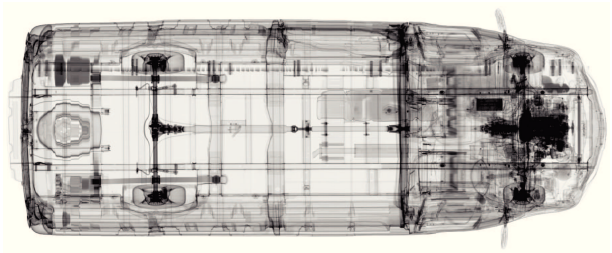
Figure 6.17 – Description of $\mathbf{R}$ and $\mathbf{T}_{\mathbf{u}^*}$ overlay for different r_α values.



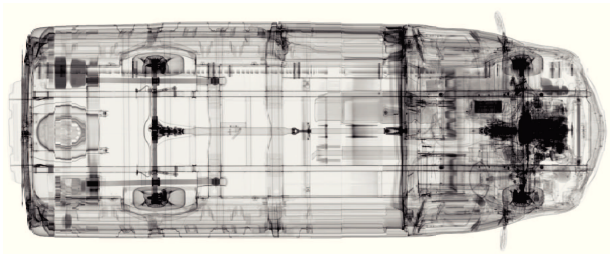
(a) $r_\alpha = 5.10^{-5}$



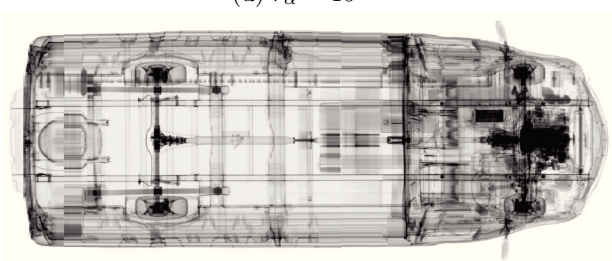
(b) $r_\alpha = 10^{-4}$



(c) $r_\alpha = 10^{-3}$

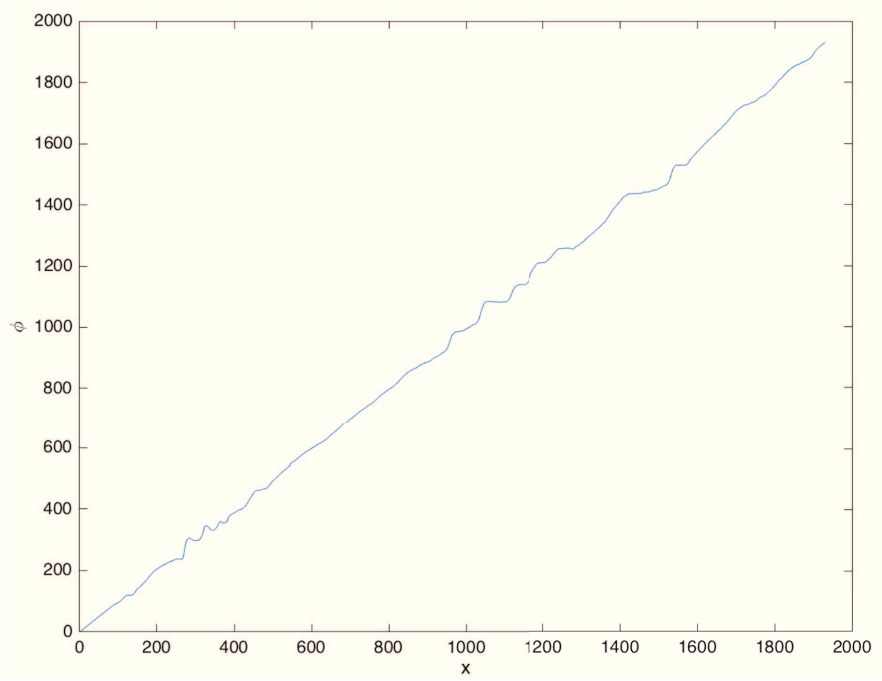


(d) $r_\alpha = 10^{-2}$

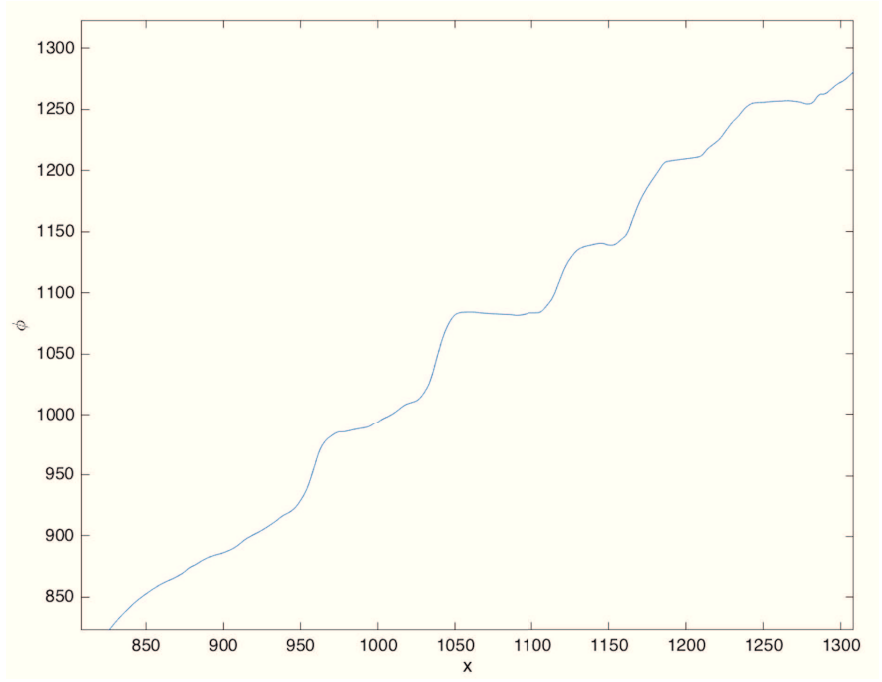


(e) $r_\alpha = 10^{-1}$

Figure 6.18 – Resulting transformed image $\mathbf{T}_{\mathbf{u}^*}$ for different regularization ratios r_α.

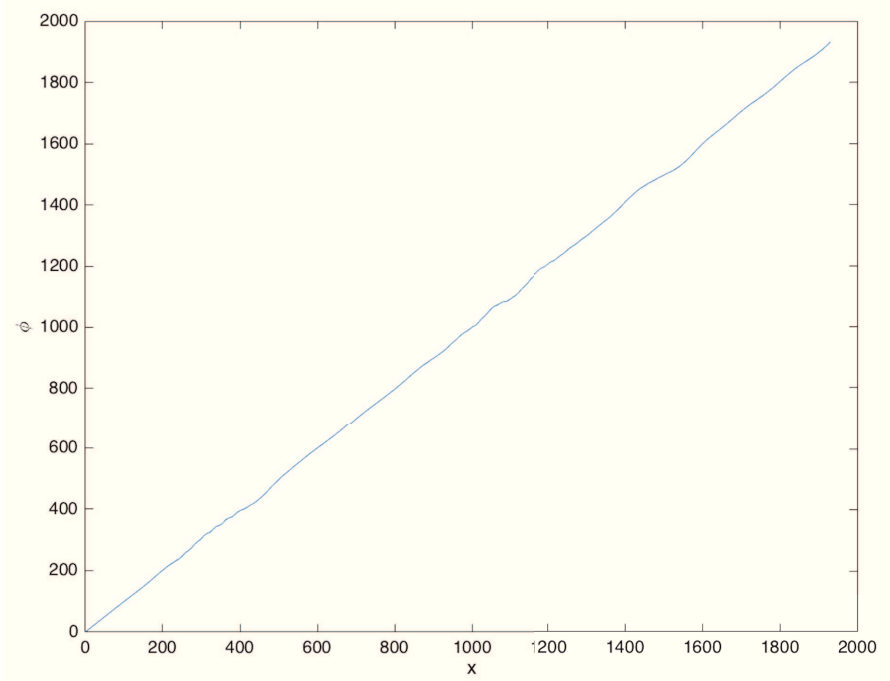


(a) ϕ representation for $r_\alpha = 10^{-1}$

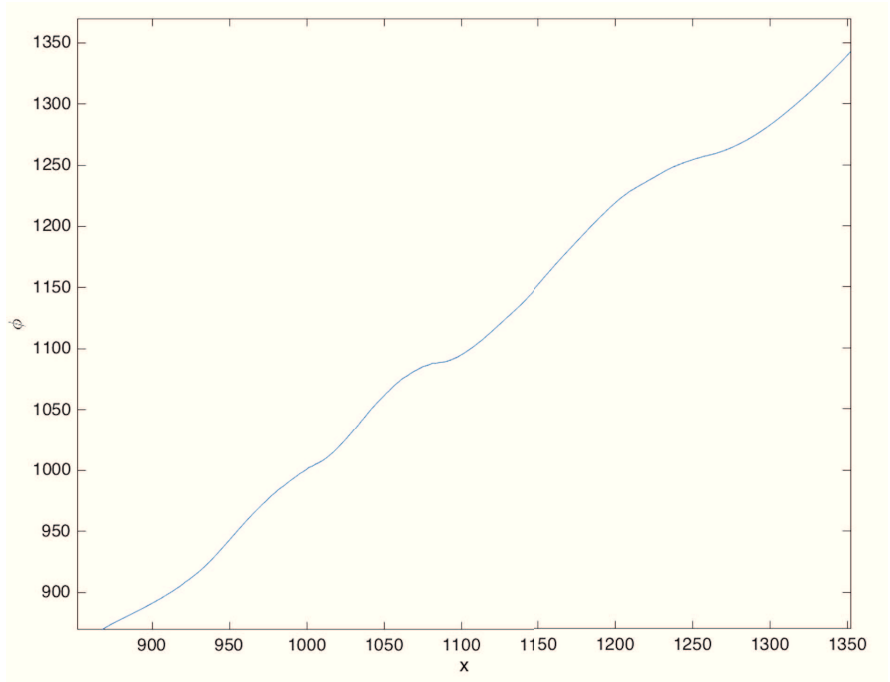


(b) zoom

Figure 6.19 – Representation of the transformation function $\phi(x) = u_x(x) + x$ obtained with $r_\alpha = 10^{-1}$. On (b), a zoomed view shows clearly that ϕ is not diffeomorphic.



(a) ϕ representation for $r_\alpha = 10^{-3}$



(b) zoom

Figure 6.20 – Representation of the transformation function $\phi(x) = u_x(x) + x$ obtained with $r_\alpha = 10^{-3}$. The curve is smoother and strictly monotically increasing. It is thus assimilated to a bijection between the columns of the source image $\mathbf{T}$ and its deformed version $\mathbf{T}_{\mathbf{u}^*}$.

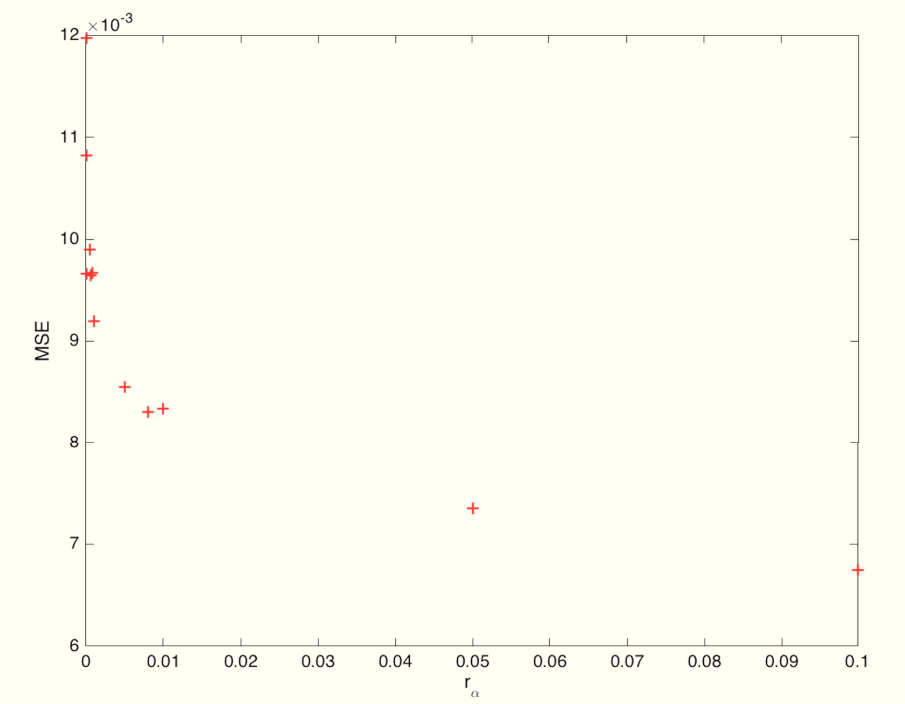


Figure 6.21 – Mean square errors obtained for different r_α values.

Flow of diffeomorphisms

The diffeomorphic registration proposed in [131] and presented in section 3.4.4 has also been experimented. For our 1-D case, by using the notations of 3.4.4 it comes:

$$d_{\phi,w}\mathcal{J}\left[\phi\right] = \int_{\Omega} \left\langle \nabla_y \mathbf{T}(y) \left[\mathbf{T}(y) - \mathbf{R}(\phi^{-1}(y))\right], w(\phi^{-1}(y))\right\rangle_{\mathbb{R}^d} \left[\phi^{'}\left(\phi^{-1}(y)\right)\right]^{-1} dy \tag{6.12}$$

Hence, $\vartheta_\phi \circ \phi^{-1}$ should be solution of:

$$Lg(y) = \nabla_y \mathbf{T}(y) \left[\mathbf{T}(y) - \mathbf{R}(\phi^{-1}(y))\right] \left[\phi^{'}\left(\phi^{-1}(y)\right)\right]^{-1} \tag{6.13}$$

As a result, the first step consists in computing:

$$\theta(y) = \nabla_y \mathbf{T}(y) \left[\mathbf{T}(y) - \mathbf{R}(\phi^{-1}(y))\right] \left[\phi^{'}\left(\phi^{-1}(y)\right)\right]^{-1} \tag{6.14}$$

L being the second derivative with respect to x, we easily solve $Lg(y) = \theta(y)$ (implicit smoothing). On the basis of the boundary conditions $g_k(0) = 0$ and $g_k(N) = 0$, we get:

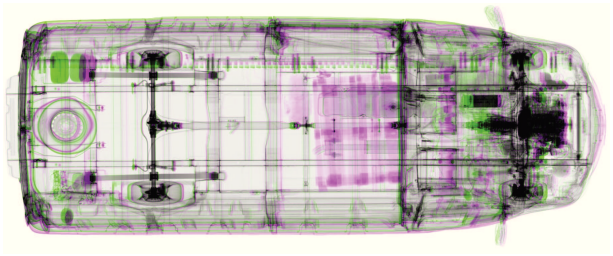
$$g_k(y) = \int_0^y \int_0^q \theta_k(t)dtdq - \frac{1}{N}\left[\int_0^N \int_0^q \theta_k(t)dtdq\right] y \tag{6.15}$$

The displacement can then be updated as follows:

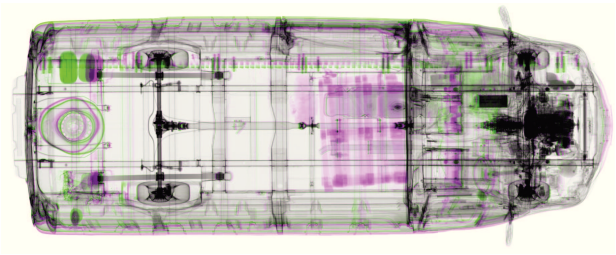
$$\phi_{k+1} = \phi_k + \rho^{(k)} g_k \circ \phi_k \tag{6.16}$$

$\rho^{(k)}$ is computed by Armijo's backtracking line-search. The initial step size ρ_i must be small enough to ensure that ϕ is a flow of diffeomorphisms as asserted in section 3.4.4. A multi-level approach is also employed to find an optimizer ϕ^* (or $\mathbf{u}^* = \phi^* - id$) of the objective function $\mathcal{J}$.

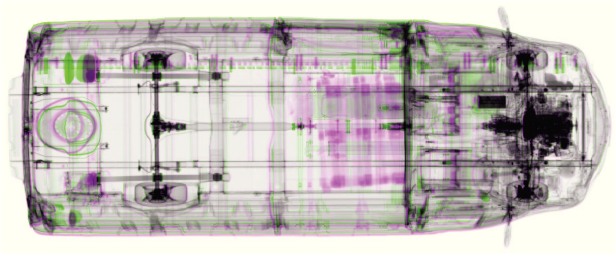
Several experiments were conducted with different ρ_i values. As shown in the following images, setting a too large ρ_i entails a non-diffeomorphic registration. In Figure 6.23 (d) and (c), the appearance of artificial streaks can be noted for $\rho_i \geqslant 10^{-1}$. Concerning the flow of diffeomorphisms technique, a different kind of artifact is highlighted: several columns of $\mathbf{T}$ tend to gather into one column of $\mathbf{T}_{\mathbf{u}^*}$ (and disappear). This folding effect is well illustrated on Figure 6.23 where the vertical tubular structure located at the immediate left of the middle of the car is progressively vanishing as ρ_i is growing. It can also be observed on the sharp discontinuities of ϕ representation in Figure 6.24 (b). Note that the strict monotonic increase of the function is not respected since it also shows many sudden decreases.



(a) $\rho_i = 10^{-5}$

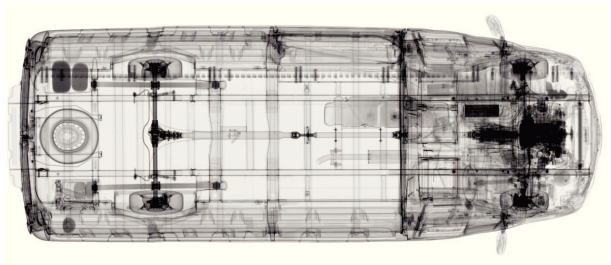


(b) $\rho_i = 10^{-3}$

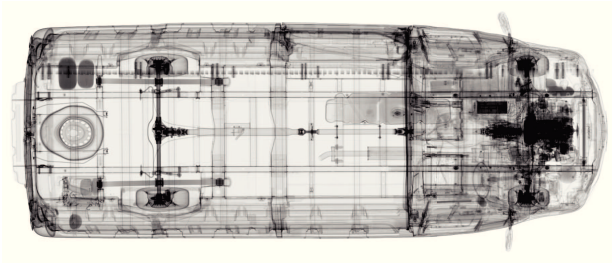


(c) $\rho_i = 1$

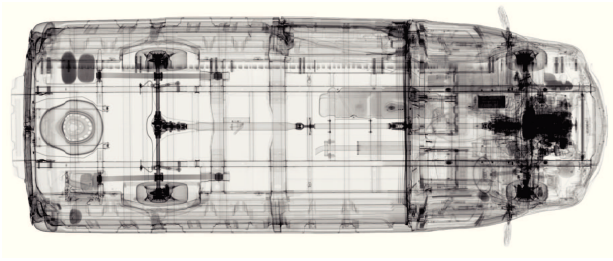
Figure 6.22 – Description of $\mathbf{R}$ and $\mathbf{T}_{\mathbf{u}^*}$ overlay for different ρ_i values.



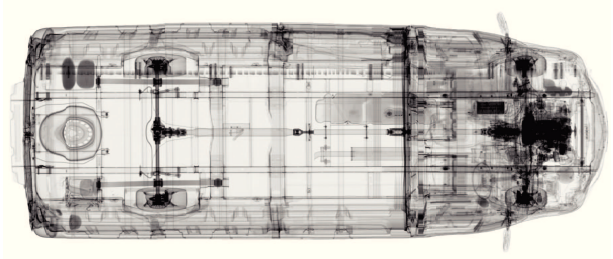
(a) $\rho_i = 10^{-5}$



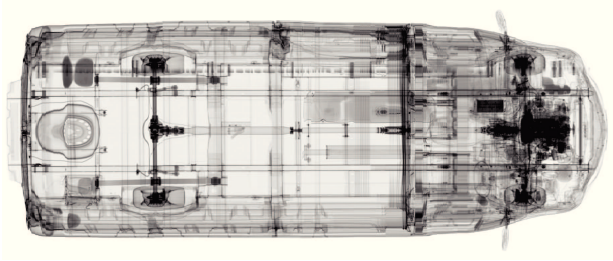
(b) $\rho_i = 10^{-3}$



(c) $\rho_i = 10^{-2}$

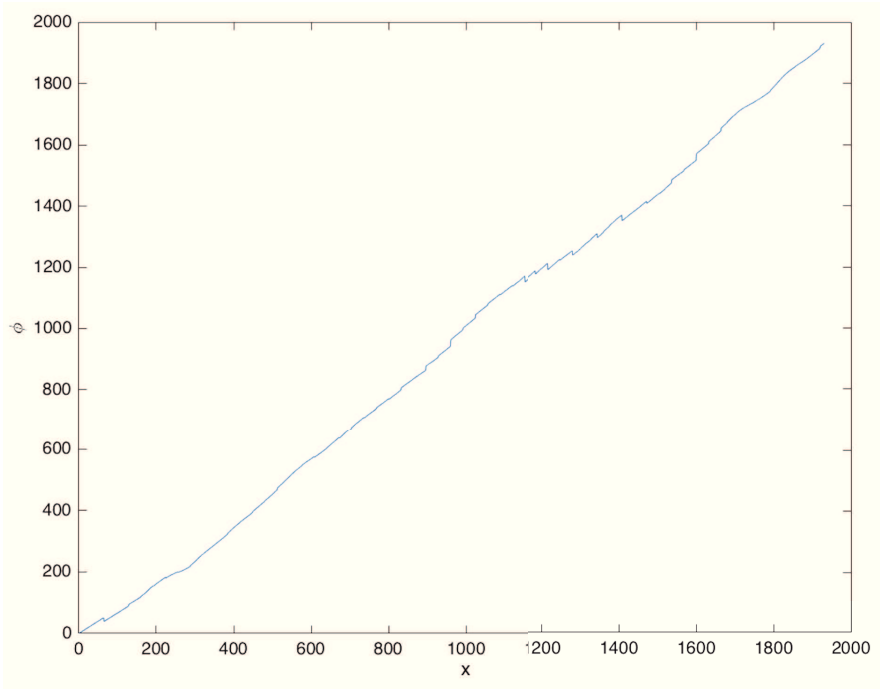


(d) $\rho_i = 10^{-1}$

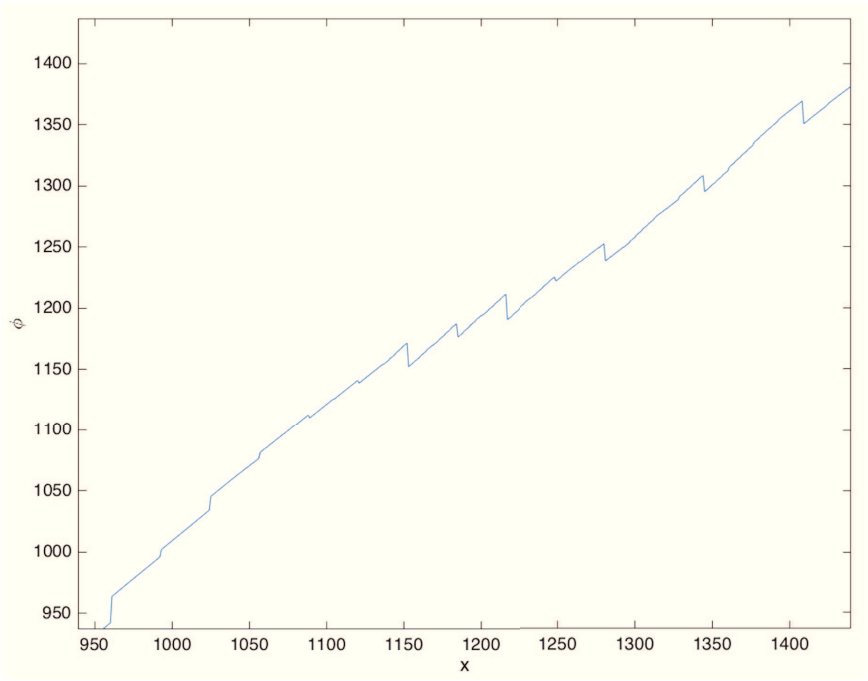


(e) $\rho_i = 1$

Figure 6.23 – Resulting transformed image $\mathbf{T}_{\mathbf{u}^*}$ for different initial step sizes ρ_i.

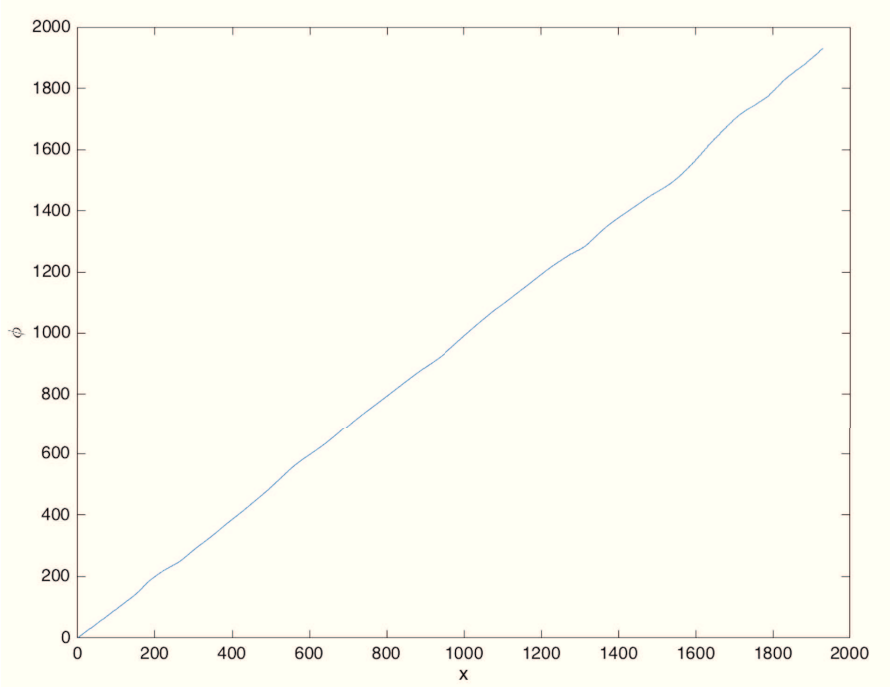


(a) ϕ representation for $\rho_i = 1$

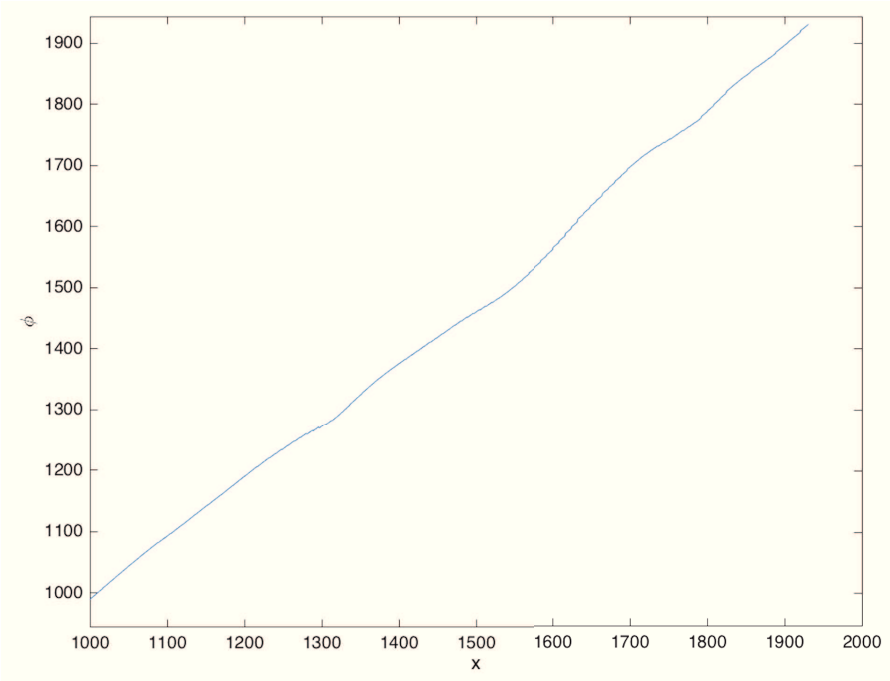


(b) zoom

Figure 6.24 – Representation of the transformation function $\phi(x) = u_x(x) + x$ obtained with $\rho_i = 1$. On (b), a zoomed view shows clearly that ϕ is not diffeomorphic.



(a) ϕ representation for $\rho_i = 10^{-3}$



(b) zoom

Figure 6.25 – Representation of the transformation function $\phi(x) = u_x(x) + x$ obtained with $\rho_i = 10^{-3}$. The curve is smoother and strictly monotically increasing. It is thus assimilated to a bijection between the columns of the source image $\mathbf{T}$ and its deformed version $\mathbf{T}_{\mathbf{u}^*}$.

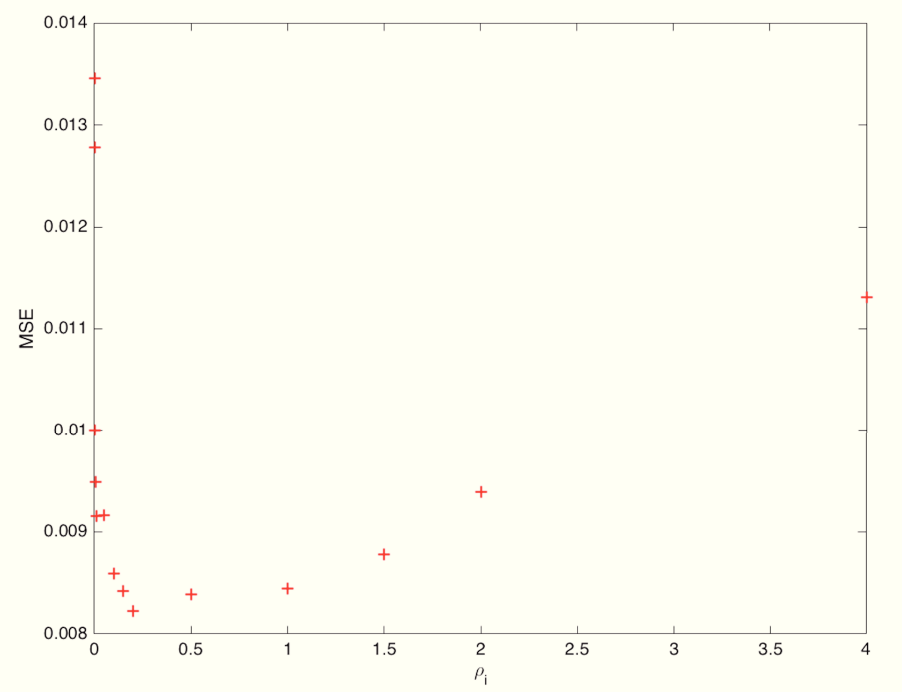


Figure 6.26 – Mean square errors obtained for different ρ_i values.

As a matter of fact, this method is less flexible than the regular 1-D variational approach where one can directly control the balance between data fitting and regularization in a intuitive manner (see Figure 6.26 and Figure 6.22). In addition, its computational complexity is higher, leading to calculations sometimes taking up to twice as long (on MATLAB, without using MEX files). We therefore adopted our one-dimensional diffusion approach for the continuation of the study.

6.6.5 Additional Examples

Our unidimensional constrained registration scheme has been tested on hundreds of image pairs with various smoothing ratios. Most often, $r_\alpha = 10^{-1}$ is a good compromise between data fitting cost and a plausible resulting $\mathbf{T}$ (for simple situations). Nevertheless, the proportion defined by r_α could optionally be refined by the custom officer, according to the load cluttering or the specific region wished to be registered (trunk, front-hood...). Indeed, this criterion is quite intuitive and can be clearly understood by the end-user.

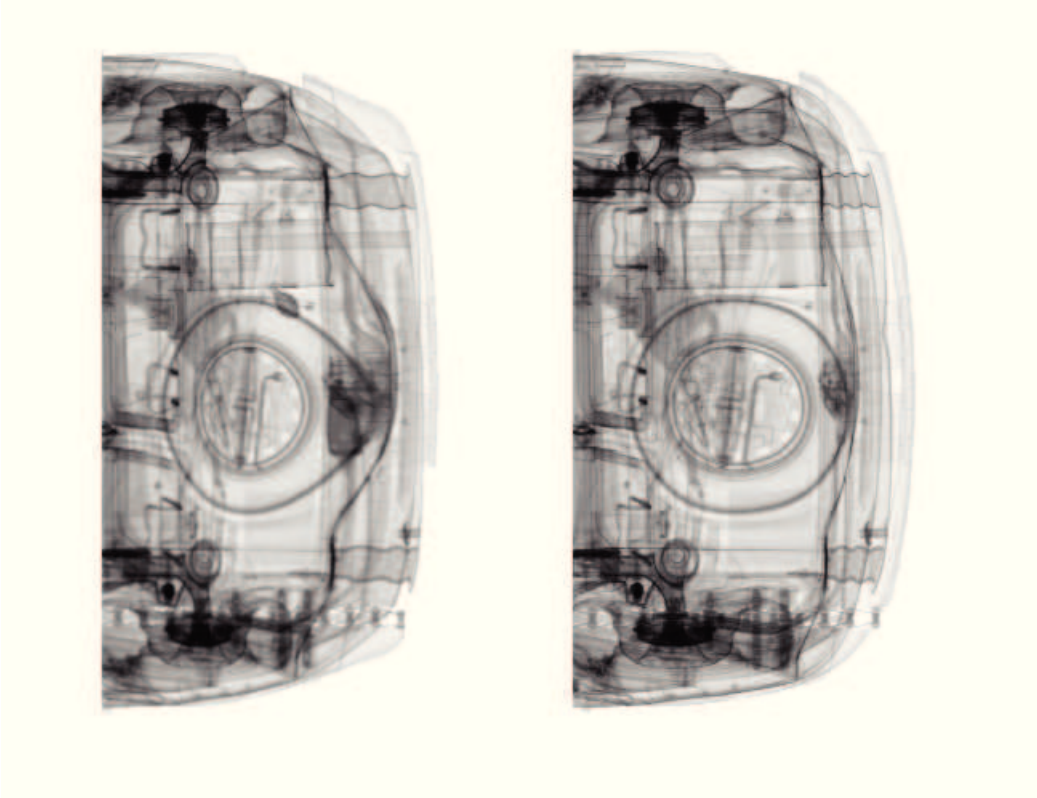
Trunk registration

Two supplementary examples are described here with the setting $r_\alpha = 10^{-3}$. In the first illustration, two scans of a same car were taken at different times, while a few threats were intentionally added inside the trunk (a **grenade** and a **small mortar**). We therefore focus our analysis on the trunk region and test *parametric* (B-spline) and *non-parametric* (demons) approaches under different smoothing constraints, as performed earlier for the front-hood registration case.

The overlay outputs are respectively displayed on Figure 6.28 and Figure 6.29. The example

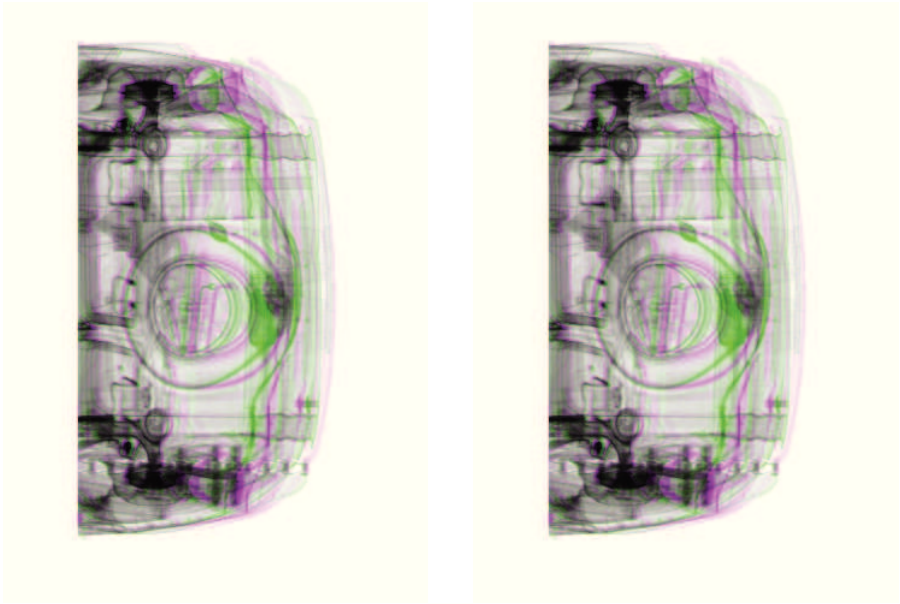
demonstrates clearly that our approach outperforms the non-constrained techniques, on both numerical and visual aspects, as described on Figures 6.30 and 6.32.

The stiffness property, inherent to the vehicle, is maintained while an optimal alignment is achieved. It follows that the different foreign objects can easily be identified on the overlay map (Figure 6.30), whereas this would be much tougher when employing standard parametric and non-parametric approaches.



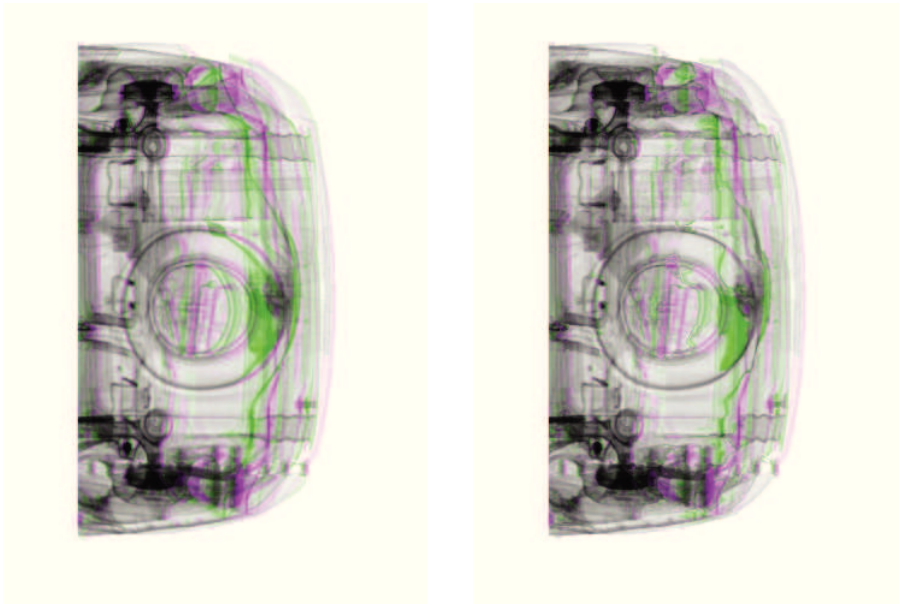
(a) Target image being inspected **R** (b) Empty source image **T**

Figure 6.27 – Two acquisitions of trunk images of the same vehicle. A notable non-linear deformation appears in **R** where illicit items were added.



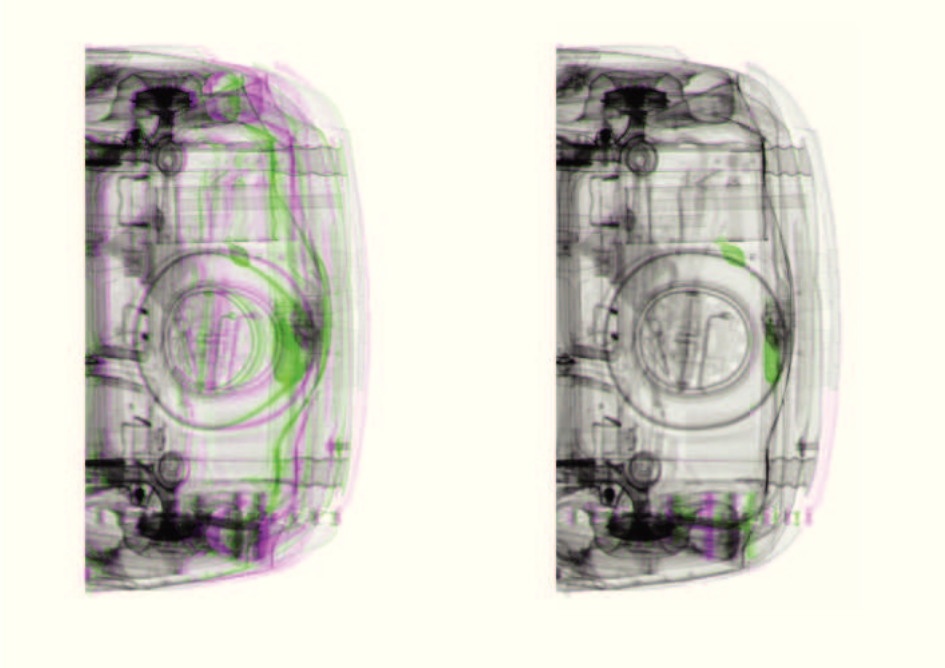
(a) Overlay with B-spline - 0.1 (b) Overlay with B-spline - 0.01

Figure 6.28 – Overlay results obtained via **B-spline** registration with different smoothing parameters.



(a) Overlay with demons - 2 (b) Overlay with demons - 0.5

Figure 6.29 – Overlay results obtained by **Thirion's demons** approach with different smoothing parameters.



(a) Rigid Registration (b) 1-D Non-Linear Registration

Figure 6.30 – Comparison of **T** and **R** overlays following rigid registration and 1-D non-linear scheme. The main differences described by pink and green colors have disappeared in (b). **Yet, the hand grenade as well as the mortar are made more easily identifiable**.

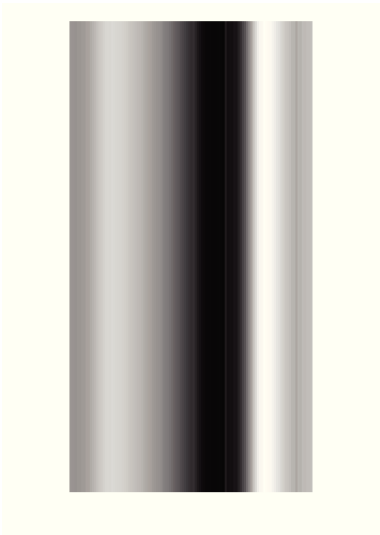


Figure 6.31 – Representation of the resulting column-wise constant displacement field $\mathbf{u}^*$.

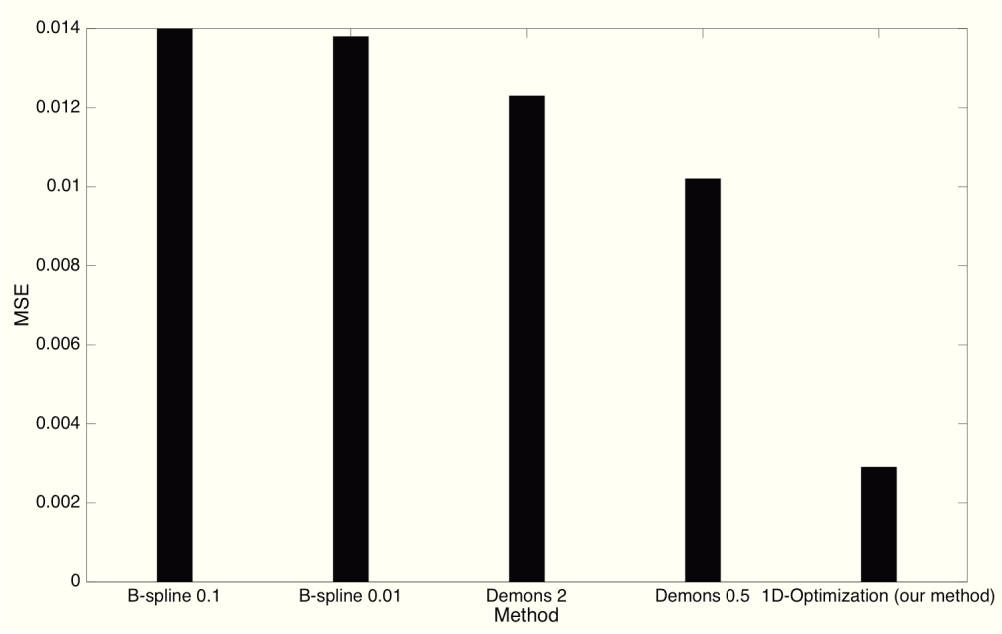
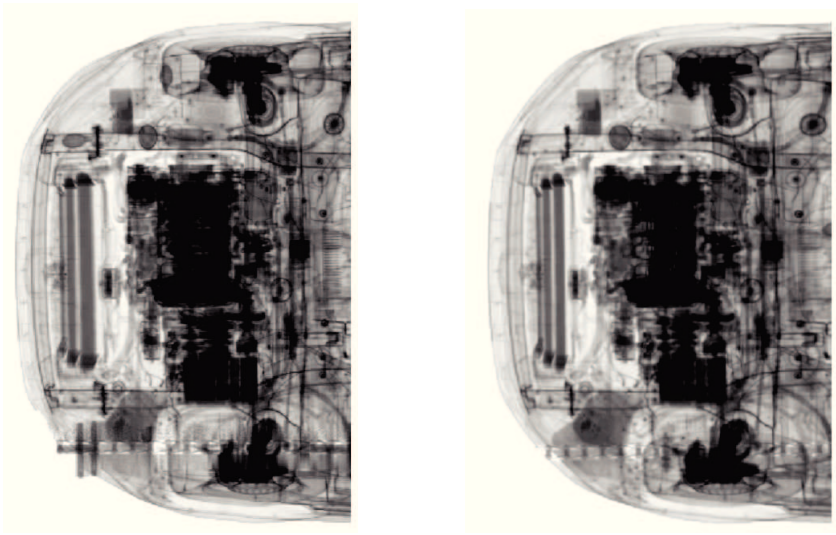


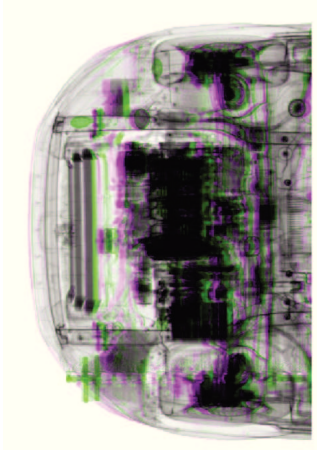
Figure 6.32 – Mean-square errors reached for each tested method (with its corresponding smoothing parameter)

Hand-grenades detection

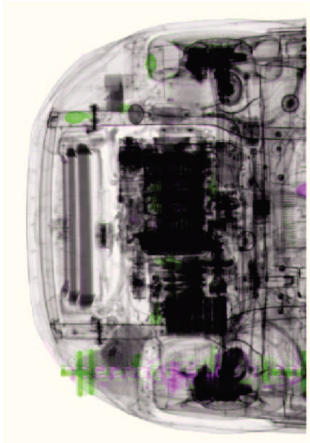


(a) Target image being inspected **R** (b) Empty source image **T**

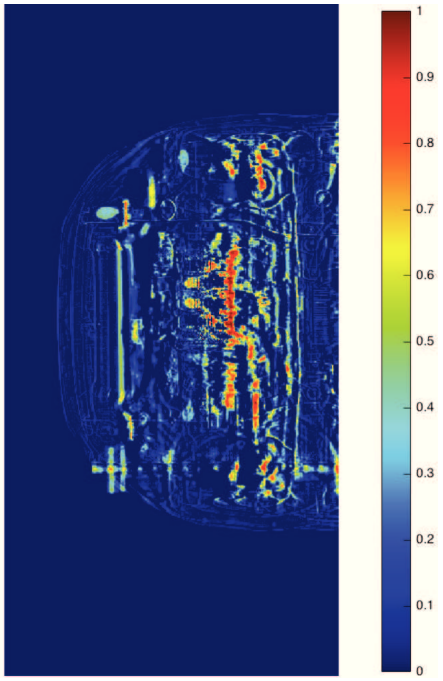
Figure 6.33 – Two acquisitions of front-hood images of the same vehicle. **Eight hand-grenades** (a mix of offensive and defensive sorts) were introduced for the acquisition represented in **R**. Only two or three threats can be visually pointed out on the X-ray image, when comparing to **T**.



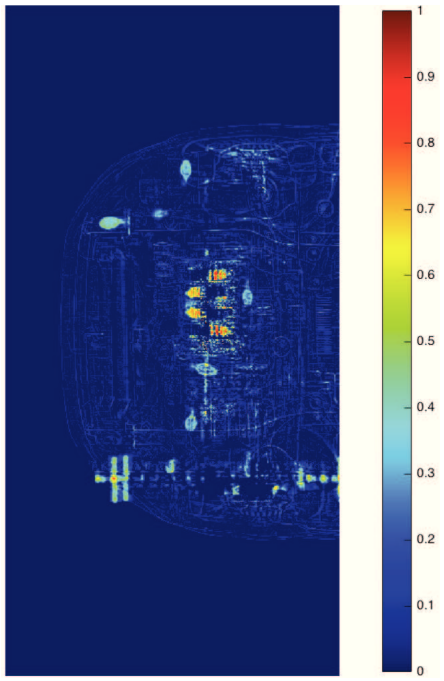
(a) Overlay following rigid registration



(b) Overlay following our 1-D approach



(c) Rationing difference following rigid registration



(d) Rationing difference following our 1-D approach

Figure 6.34 – Visual appraisal of the method compared to rigid registration for threat detection in the front-hood of a vehicle. At least **seven grenades** are detectable in a region where advanced machine learning techniques would probably fail.

In the second example, the complementarity of deep-learning recognition techniques and our registration application is firmly stressed. It shows that a well-tuned model can achieve detection of objects beyond standard human vision, particularly in extremely dense regions such as the engine. The registration problem of two front-hood images (same vehicle) is addressed through our approach. In **R** (Figure 6.33 (a)), standing as the sample to be analyzed, eight defensive

and offensive hand grenades were hidden. Rigid registration yields very poor results in terms of detection capacity as depicted on Figures 6.34 (a) and (c). On the other hand, threat detection is strongly enhanced via the application of our 1-D optimization scheme. The overlay on Figure 6.34 (b) displays an almost-perfect alignment and the map of differences enables the identification of more than seven hand grenades (Figure 6.34 (d)) in a region where visibility is strongly reduced.

Conclusion

The technique we proposed addresses the registration problem of top-view X-ray scans from different vehicles of the same model. It especially aims at dealing with **non-linear deformations** induced by possible **speed variations of the vehicle** during scanning.

Our simple and intuitive solution is described: in a first stage, a pose estimation between both scans is performed. Then, assuming that the field of displacement is column-wise constant and parallel to the car motion, we formulate a one-dimensional optimization scheme. Both visual and numerical results presented in this paper demonstrate the necessity for our method as well as its high performances with regard to registration accuracy.

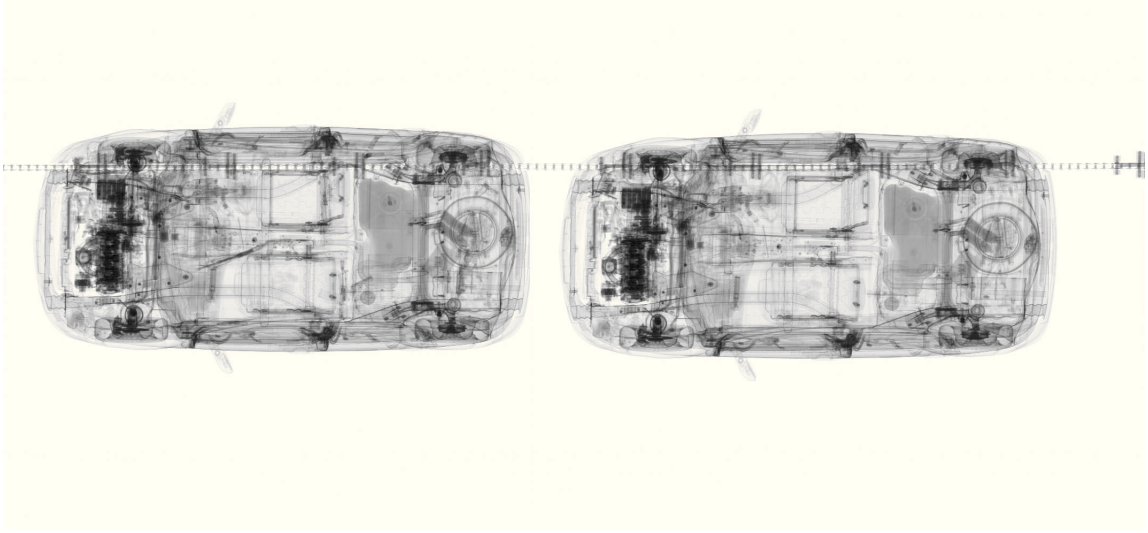
Yet, as we will describe in the next section, in some situations, the method might achieve lower performances in terms of alignment.

6.7 Vertical Correction

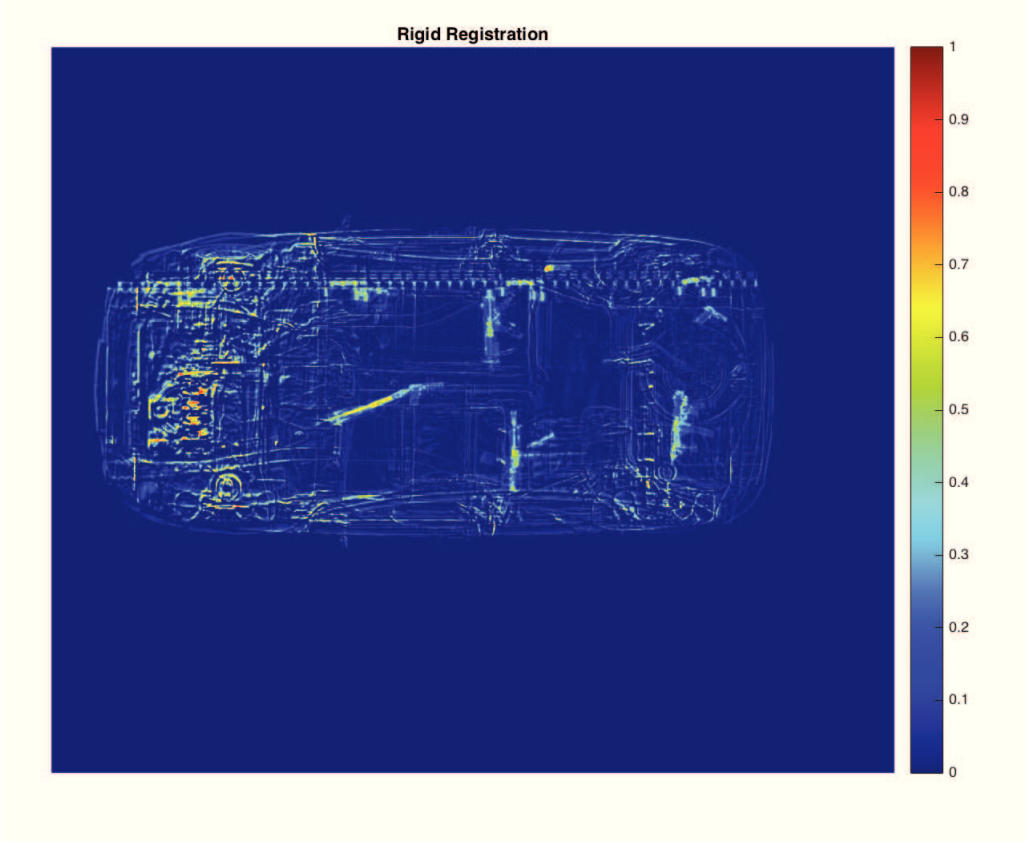
In this section, we show that the method we suggested in section 6.6 is not sufficient in case of significant **vertical shifts** between the cars of both moving and static images. In fact, the radiated beam is triangular, thus any translated object is projected differently according to its distance to the X-ray source. We therefore add a 1-D unconstrained optimization to the previous scheme for a "y-direction" correction. We also demonstrate that applying the vertical correction following our 1-D optimization in the x-axis yields better results than performing a simultaneous minimization on both components. Finally, the possible apparition of artifacts in the deformed image throughout the optimization process is analyzed. Diffusion and volume-preserving schemes are considered and compared in this regard.

6.7.1 Side-wards Shift: Other Misregistration Cause

Let us first consider the following registration problem where $\mathbf{T}$ and $\mathbf{R}$ are acquired on the same car by the same system, but at different times (Figures 6.35 (a), (b)). In $\mathbf{R}$, several firearms were scattered all over the vehicle. By applying the unidimensional diffusion presented in the previous section, quite a few vertical differences tend to disappear as can be observed by comparing Figure 6.35 (c) and Figure 6.36. However, substantive alignment artifacts can still be seen on the difference map presented on Figure 6.36..



(a) **R** (b) **T**



(c) Rigid registration

Figure 6.35 – (a) **R** containing threats (several gunfires); (b) source image **T**; (c) the change map resulting from pose estimation (rigid registration).

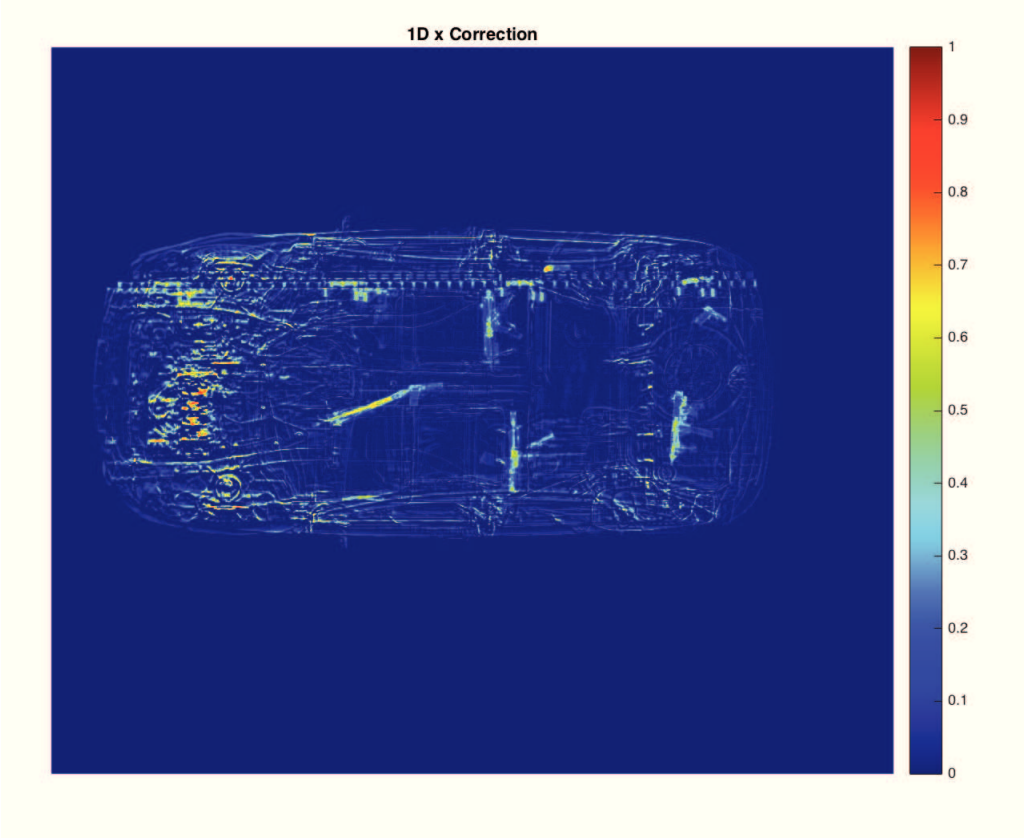


Figure 6.36 – Difference output of unidimensional registration on x-axis. The substantial vertical shift between both acquisitions entails alignment artifacts, particularly on the front part, making inspection impracticable.

In fact, the scanning system allows for an additional degree of freedom: the car translation along the trailer width (Figure 6.38). Besides, the X-ray source is located at the detection line center (above the incoming vehicles) and generates a pyramidal beam as indicated on Figure 6.37. Any significant translation with respect to the other image therefore leads to important deformations in the y-axis.

More particularly, in an analogous fashion to stereo-vision problems, the deformation magnitude depends on the distance/depth separating the radioactive source from the scanned object as shown in Figure 6.39. The more the shifted object is close to the source, the more it will be dilated in the projection image. **Thus, a row-wise constancy constraint on the vertical component of u should not apply to this situation** (unlike **Hypothesis 1** formulated in the previous section).

Figure 6.37 – Illustration of X-ray beam during scanning.

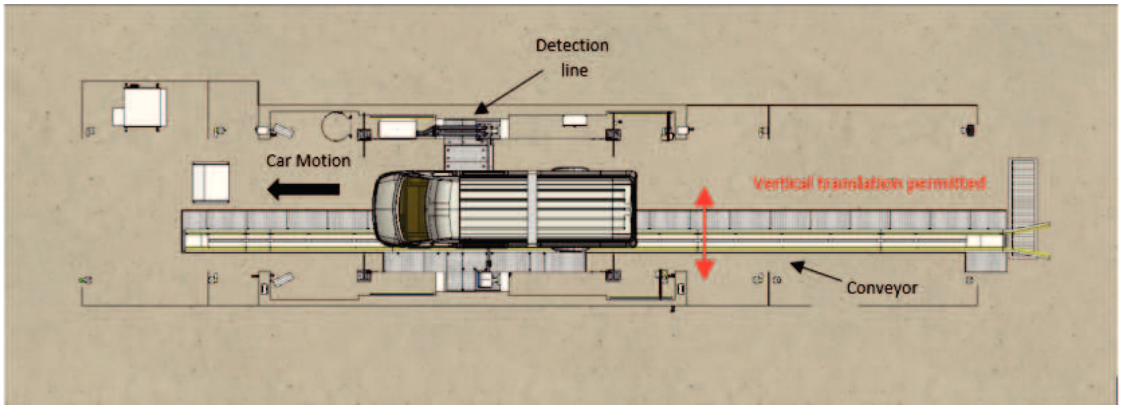


Figure 6.38 – Possible side-wards shift of the vehicle during the scanning process.

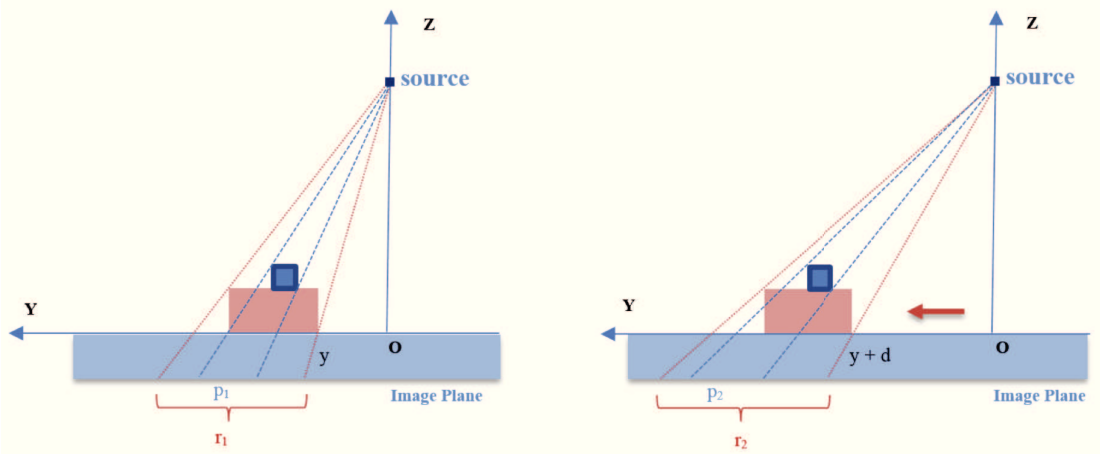


Figure 6.39 – Vertical shift effect description. The projected distances r_2 and p_2 after displacement are larger than r_1 and p_1. Similarly to the stereo-vision problem, this difference grows with the object proximity to the X-ray source: $p_2/p_1 > r_2/r_1$

6.7.2 Solution Description

We remain in the variational framework, and simply intend to apply a supplementary 1-D registration over the vertical component of the displacement field. Unlike our previous scheme exposed in section 6.6, we do not assume any specific priors or constraints on the vertical displacement component since the required correction varies accordingly to the objects depth (Figure 6.39). As mentioned earlier, a row-wise constancy restriction on u_y is therefore irrelevant. Hence, for this specific vertical correction, we have:

$$\mathbf{u} = (0, u_y(x, y)) \tag{6.17}$$

Diffusion registration

Using a diffusion regularization, the objective function is written as follows:

$$\begin{aligned} &\text{find } u_y(x, y) \text{ that minimizes} \\ &\mathcal{J}[u_y] = \frac{1}{2} \|\mathbf{T}(u_y) - \mathbf{R}\|_2^2 + \frac{\alpha}{2} \|\nabla u_y\|_2^2 \end{aligned} \tag{6.18}$$

We also wish to compare both simultaneous and successive optimization paradigms for horizontal and vertical corrections. Let us therefore formulate the above problems in two dimensions, maintaining the column-wise constancy constraint for the horizontal component of the displacement field. Thus, for $x, y \in \Omega$ we consider:

$$\mathbf{u} = (u_x(x), u_y(x, y)) \tag{6.19}$$

The two-dimensional version of the diffusion registration 6.18 is given by:

$$\begin{aligned} &\text{find } \mathbf{u} \text{ minimizing} \\ &\mathcal{J}[\mathbf{u}] = \frac{1}{2}\|\mathbf{T}(\mathbf{u}) - \mathbf{R}\|_2^2 + \frac{\alpha}{2}\left(\|\nabla u_x\|_2^2 + \|\nabla u_y\|_2^2\right) \end{aligned} \tag{6.20}$$

Volume-preserving soft constraint

A volume-preserving approach is also tested by adding a soft constraint to the above formula. The joint functional is then given by:

$$\begin{aligned} &\text{find } u_y(x,y) \text{ that minimizes} \\ \mathcal{J}[u_y] &= \frac{1}{2}\|\mathbf{T}(u_y) - \mathbf{R}\|_2^2 + \frac{\alpha}{2}\|\nabla u_y\|_2^2 + \frac{\beta}{2}\|\det(\nabla\phi) - 1\|_2^2 \\ &= \frac{1}{2}\|\mathbf{T}(u_y) - \mathbf{R}\|_2^2 + \frac{\alpha}{2}\|\nabla u_y\|_2^2 + \frac{\beta}{2}\|\det(\nabla u_y + Id) - 1\|_2^2 \end{aligned} \tag{6.21}$$

The soft constrained approach ([91], [19]) is preferred over hard constraints or diffeomorphisms because of its flexibility and implementation simplicity. The β parameter allows control over the balance between diffusion registration and more stringent volume preservation constraints. In alternative methods, parameter tuning turns into a tougher task.

The two-dimensional volume-preserving optimization scheme is also written as follows:

$$\begin{aligned} &\text{find } \mathbf{u} \text{ minimizing} \\ &\mathcal{J}[\mathbf{u}] = \frac{1}{2}\|\mathbf{T}(\mathbf{u}) - \mathbf{R}\|_2^2 + \frac{\alpha}{2}\left(\|\nabla u_x\|_2^2 + \|\nabla u_y\|_2^2\right) + \frac{\beta}{2}\|\det(\nabla\mathbf{u} + Id) - 1\|_2^2 \end{aligned} \tag{6.22}$$

6.7.3 Numerical Resolution

We will describe the numerical resolution of the 2-D problems (6.20) and (6.22). The resolution of (6.18) and (6.21) is then straightforward since it amounts to exclusively optimizing the vertical displacement field component after obtaining an optimal u_x^* by solving (6.6).

In the meantime, we resort to the l-BFGS quasi-Newton optimization method.

For the diffusion-regularization problem (6.20), the directional derivatives of the joint functional yield (as proven in equation (6.8):

$$\nabla_{\mathbf{u}}\mathcal{J}[\mathbf{u}] = \nabla_{\mathbf{u}}\mathbf{T}(\mathbf{x}+\mathbf{u})\left[\mathbf{T}(\mathbf{x}+\mathbf{u}) - \mathbf{R}(\mathbf{x})\right] - \alpha\Delta\mathbf{u} \tag{6.23}$$

with $\mathbf{x} = (x, y) \in \Omega$ and $\mathbf{u}$ as defined in (6.19).

Let us now give a look at the volume-preserving soft constraint introduced in (6.22). With a linearization approximation, valid for small displacements, we get:

$$\det(\nabla\mathbf{u} + Id) - 1 \approx \partial_x u_x + \partial_y u_y = \operatorname{div}(\mathbf{u}) \tag{6.24}$$

Consequently, under Dirichlet boundary conditions, the Gâteaux derivative of the soft incompressibility constraint $\mathcal{S}^{VP}[\mathbf{u}] = \|\det(\nabla\mathbf{u} + Id) - 1\|_2^2$ becomes:

$$
\begin{aligned}
d_{\mathbf{u},\mathbf{q}}\mathcal{S}^{VP}[\mathbf{u}] :&= \lim_{h\to 0}\frac{1}{h}(\mathcal{S}^{VP}[\mathbf{u}+h\mathbf{q}] - \mathcal{S}^{VP}[\mathbf{u}]) \\
&= \lim_{h\to 0}\frac{1}{h}(\|\mathrm{div}(\mathbf{u}+h\mathbf{q})\|_2^2 - \|\mathrm{div}(\mathbf{u})\|_2^2) \\
&= \lim_{h\to 0}\frac{1}{h}\langle h\mathrm{div}(\mathbf{q}), 2\mathrm{div}(\mathbf{u}) + h\mathrm{div}(\mathbf{q})\rangle \\
&= \langle \mathrm{div}(\mathbf{q}), 2\mathrm{div}(\mathbf{u})\rangle \\
&= \langle \mathbf{q}, -2\nabla.\mathrm{div}(\mathbf{u})\rangle
\end{aligned}
\tag{6.25}
$$

Hence, the directional derivative of (6.22) leads to the following expression:

$$
\nabla_{\mathbf{u}}\mathcal{J}[\mathbf{u}] = \nabla_{\mathbf{u}}\mathbf{T}(\mathbf{x}+\mathbf{u})[\mathbf{T}(\mathbf{x}+\mathbf{u}) - \mathbf{R}(\mathbf{x})] - \alpha\Delta\mathbf{u} - \beta\nabla.\mathrm{div}(\mathbf{u}) \tag{6.26}
$$

This is an interesting result since it actually corresponds to the derivative of the elastic registration functional (see [37]). In fact, by noting $\beta = \alpha + \gamma$ we retrieve the well-known Lamé coefficients α and γ (for more thorough details, *e.g* [1]'s appendix).

We usually run the descent algorithm with a 4-levels Gaussian pyramid. Armijo's backtracking line-search technique is also integrated into the process to yield an optimal step size for each update of $\mathbf{u}$. Both α and β are defined for each scale by fixing desired ratios r_α and r_β between the data fitting term and the regularizers. As explained in section 6.6.4, the parameters are estimated at the first iteration and hold for the whole level.

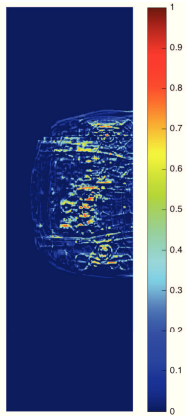
Stopping criteria are identical to those mentioned for the 1-D horizontal registration. Computations are performed on a 8 Go RAM - 3,1 GHz Intel Core i7 MacBook. The average calculation time on 992×1186 images with MATLAB 2016b is about $50s$ for the successive minimization scheme and can reach more than two minutes for the simultaneous one.

In the next section, we present the results of both approaches for a simultaneous and consecutive optimization fashion. We will use the example brought in Figure 6.36 and focus on the front part of the vehicle for a visual and numerical assessment of the different techniques.

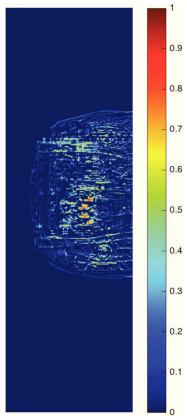
6.7.4 Horizontal-then-vertical optimization

This approach consists simply in rectifying first both $\mathbf{R}$ and $\mathbf{T}$ by employing our 1-D horizontal registration. In a second stage, an unconstrained registration involving the vertical component $u_y(x, y)$ of $\mathbf{u}$ is performed. In other words, once columns are matched between the source and target images (image *rectification*), a 1-D vertical correction within each column may me undertook.

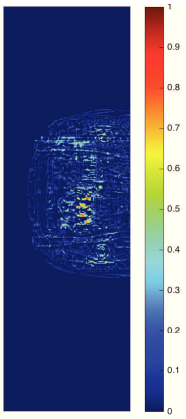
In Figure 6.40 and Figure 6.41, we observe that the smoothing parameters/ratios must be chosen carefully. A too large regularization entails a poor registration accuracy (Figure 6.40 (a) and Figure 6.41 (a), (b)) while a too soft smoothing makes objects disappear in the difference map. The last phenomenon can be observed on Figure 6.40 (e), (f) where the threat cannot be longer detected. more explanations are given in subsection 6.7.6.



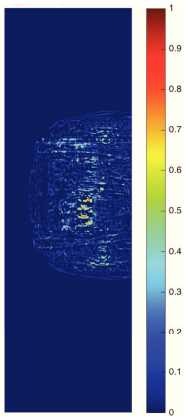
(a) 1-D horizontal registration



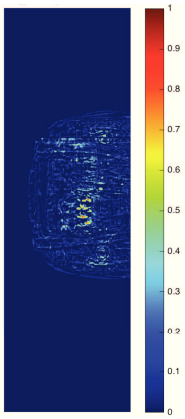
(b) $r_\alpha = 0.1$



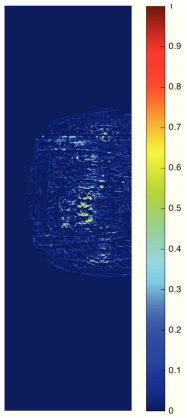
(c) $r_\alpha = 1$



(d) $r_\alpha = 1.5$

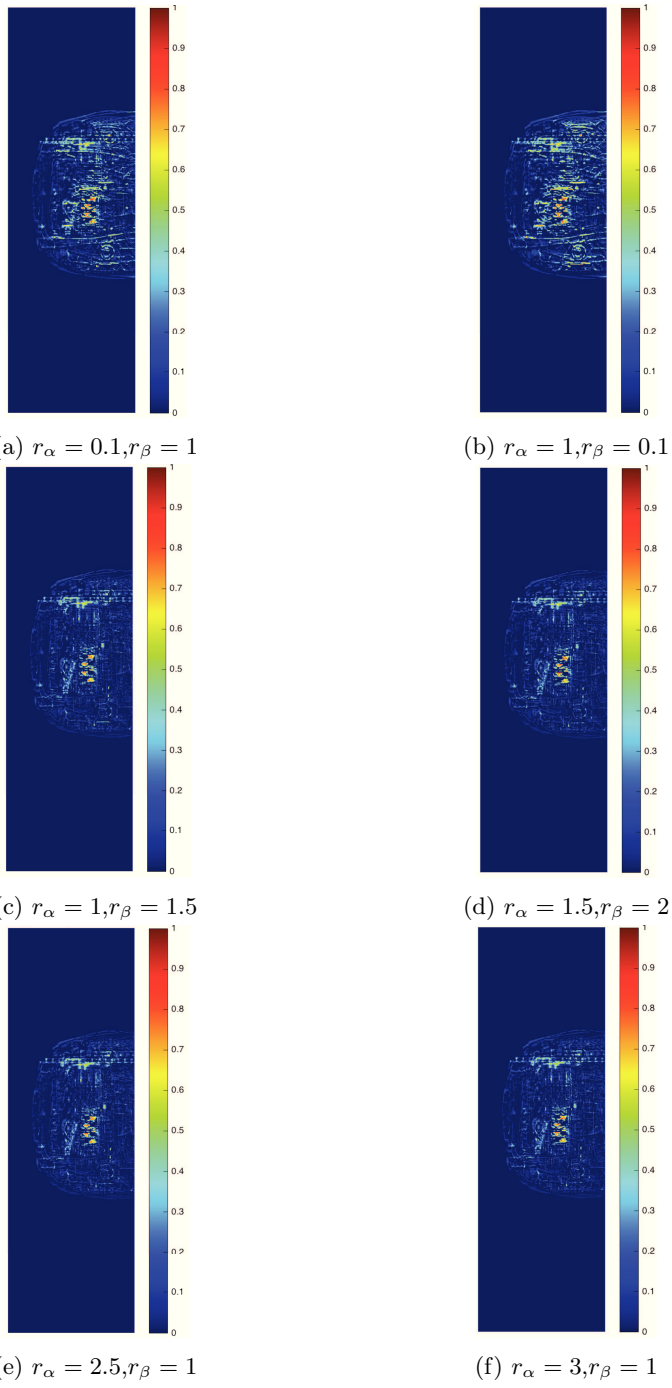


(e) $r_\alpha = 2$



(f) $r_\alpha = 2.5$

Figure 6.40 – Front part difference maps after 1D horizontal correction (a); Diffusion registration was then applied on the vertical component of $\mathbf{u}$ with $r_\beta = \infty$ and various r_α parameters.



(a) $r_\alpha = 0.1, r_\beta = 1$ (b) $r_\alpha = 1, r_\beta = 0.1$

(c) $r_\alpha = 1, r_\beta = 1.5$ (d) $r_\alpha = 1.5, r_\beta = 2$

(e) $r_\alpha = 2.5, r_\beta = 1$ (f) $r_\alpha = 3, r_\beta = 1$

Figure 6.41 – Difference maps outputs of elastic registration performed on the vertical component of the displacement after horizontal 1-D correction. A few examples with different combinations of r_β and r_α are shown. Different sets of parameters may yield similar results. In some of the images, *i.e* (c), (d), (e), (f), a gun shape is easily recognizable while its identification was much tougher in Figure 6.40 (a).

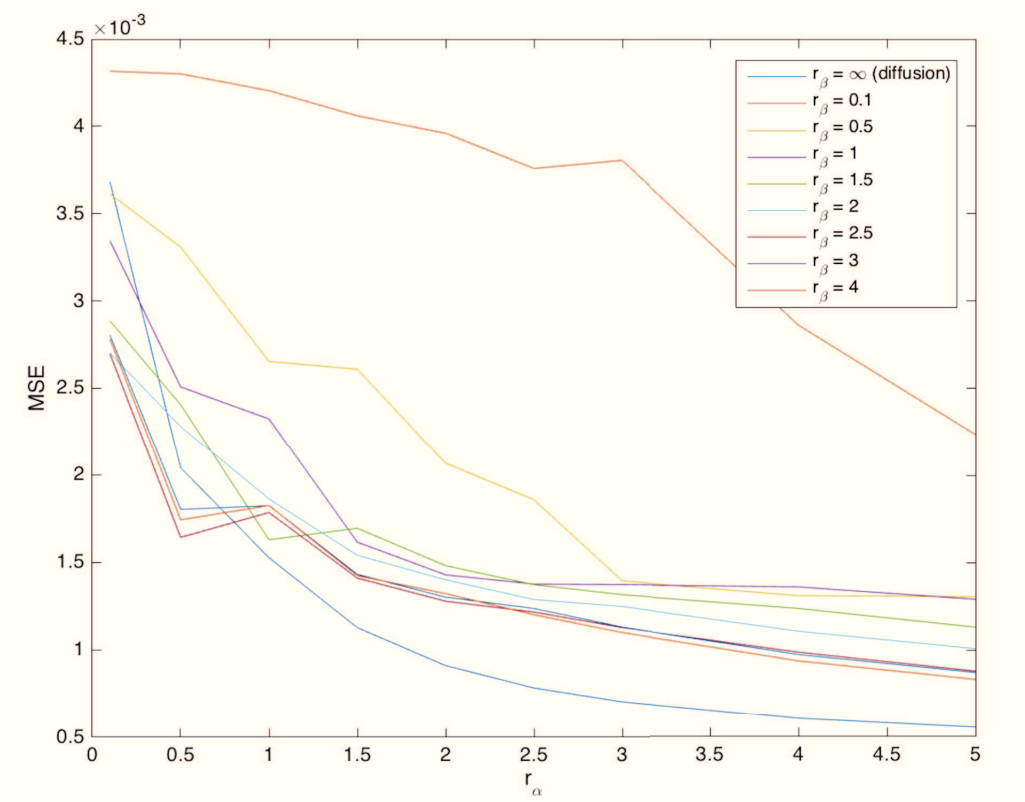


Figure 6.42 – Evolution of mean square errors according to the set of regularization ratios r_α and r_β

In Figure 6.41 (c), the addition of a vertical volume-preserving constraint gives an accurate and reasonable difference map in contrast with the output obtained with the same r_α ratio in Figure 6.40 (c). Yet, a too small r_β would lead to a stronger regularization of $\mathbf{u}$ as depicted on Figure 6.40 (b).

Overall, elastic registration yields good results by better combining threat visualization and MSE minimization at the same time, as detailed in the graph presented on Figure 6.42. In fact, diffusion registration gives lower MSE scores but does not enable to monitor efficiently the regularity of ϕ, causing deformation artifacts in $\mathbf{T}$.

6.7.5 Simultaneous optimization

In contrast with the previous approach, we also intended to perform registration in a simultaneous optimization fashion. Rather than applying two separate horizontal and vertical corrections, we run the algorithm with a direct constraint formulated on both components of $\mathbf{u}$ at the same time, as expressed in (6.19).

However, the upcoming results show that "cleaner" difference maps are obtained with the horizontal-then-vertical registration approach where columns are first rectified to enable a further vertical correction (Figure 6.44).

In addition, by using analogous parameters, the simultaneous method yields much larger MSE

distances after convergence as shown in the graph displayed on Figure 6.43. These results are quite straightforward since the method does not allow for a separate setting of the parameters w.r.t vertical and horizontal components of the displacement **u**.

Indeed, both deformations are completely different by nature: in the x-direction, oversampling issues are behind the non-linear deformations of objects, whereas regarding the y-component, a stereoscopic effect occurring when the vehicle is shifted side-wards in **R** was singled-out. Both problems should therefore be addressed in a separate fashion, with their respective specificity.

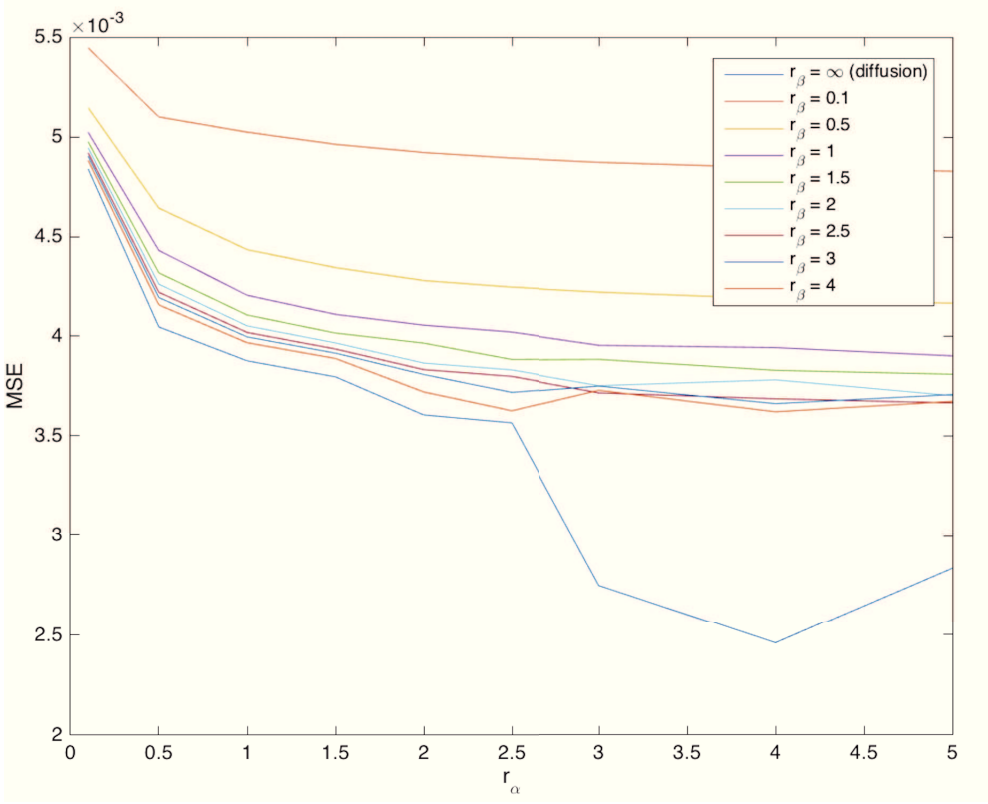
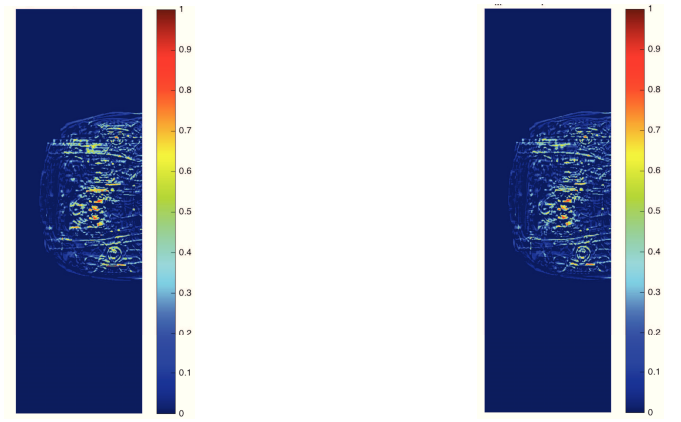
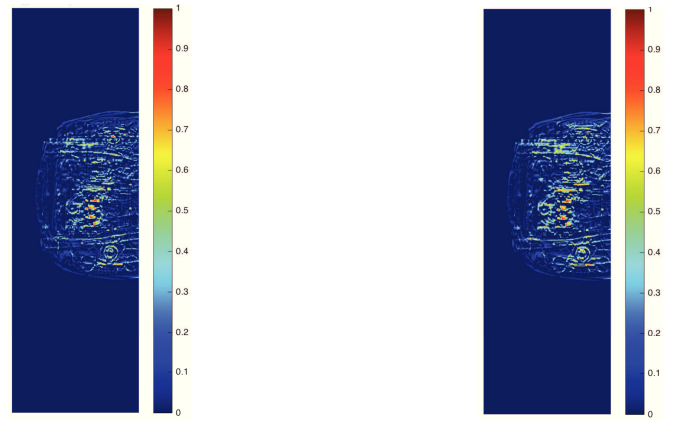


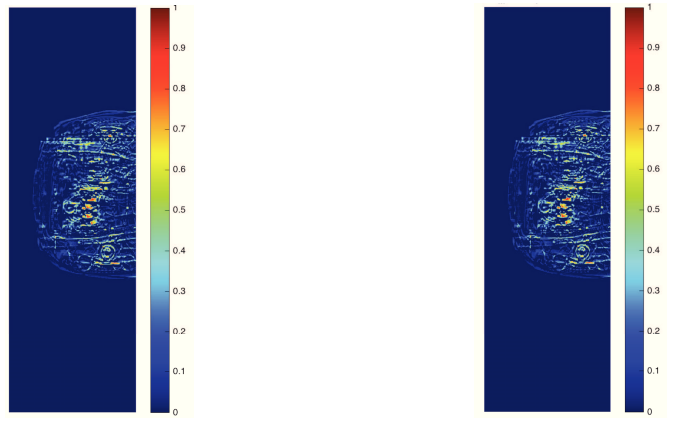
Figure 6.43 – Evolution of mean square errors according to the set of regularization ratios r_α and r_β



(a) $r_\alpha = 0.1, r_\beta = \infty$ (b) $r_\alpha = 1.5, r_\beta = \infty$



(c) $r_\alpha = 2.5, r_\beta = \infty$ (d) $r_\alpha = 0.1, r_\beta = 1$



(e) $r_\alpha = 1, r_\beta = 1.5$ (f) $r_\alpha = 3, r_\beta = 1$

Figure 6.44 – Difference images obtained with simultaneous optimization scheme for various sets of regularization ratios. (a) - (c): diffusion registration; (d) - (f): elastic registration

Regular 2-D registration

An unconstrained registration scheme was also tested where no prior constraint was imposed on the displacement field $\mathbf{u} = (u_x(x, y), u_y(x, y))$. Yet, we observed that the topology of the objects to be identified was not respected in the resulting difference map, as was indeed observed by the use of Thirion's demons in 6.4.2. We conclude that the constraints on $\mathbf{u}$ are utterly needed to produce a map where objects can be recognized by their geometrical setting.

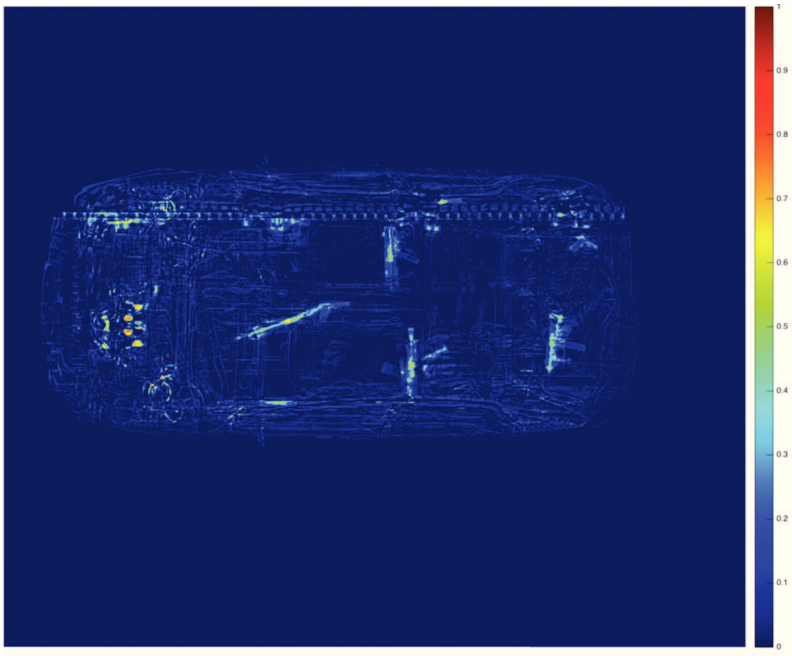


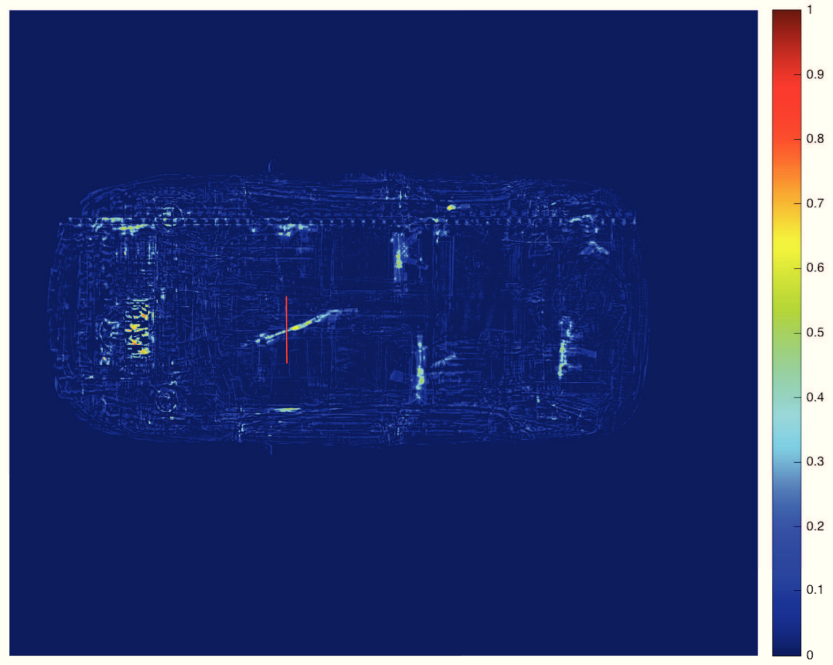
Figure 6.45 – Difference map obtained with regular two-dimensional registration for $r_\alpha = 1$,$r_\beta = 1$.

6.7.6 Volume Preservation

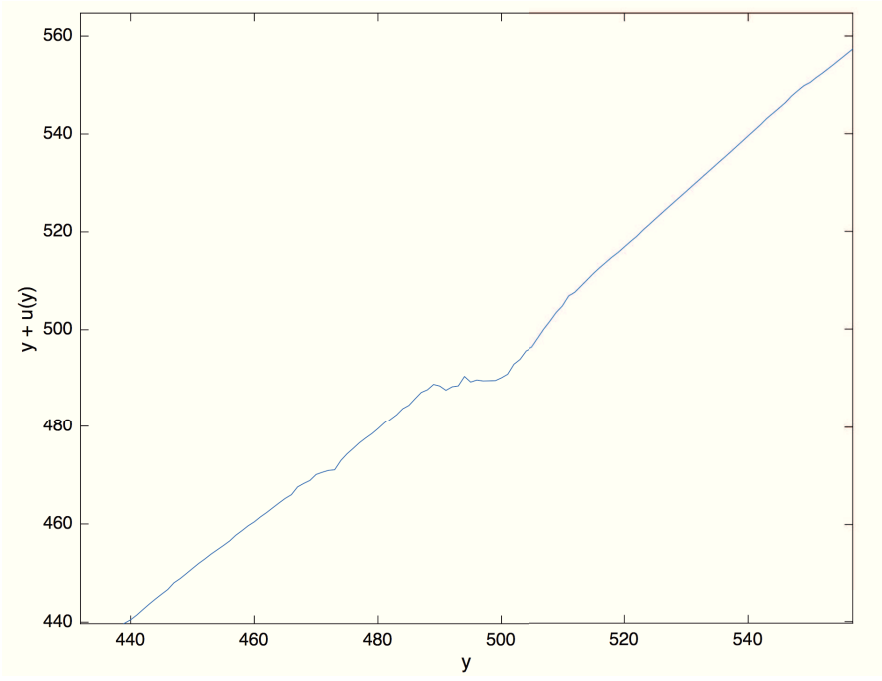
As mentioned, regularization should be tuned carefully, a too weak smoothing may give a better registration accuracy at the expense of unreasonable deformations. This is especially true when the algorithm tends to deform $\mathbf{T}$ in a non-diffeomorphic fashion to "imitate" the added objects in $\mathbf{R}$.

Figure 6.46 describes the resulting difference map with a weakly-regularized registration. The transformation is piece-wise constant (Figure 6.46 (b)), meaning that, in a given column, a pixel in the original image $\mathbf{T}$ is repeated several times in the resulting $\mathbf{T}_{\mathbf{u}^*}$. The transform ϕ is therefore non-diffeomorphic.

By using proper parameters, we prevent the transformation from being non-diffeomorphic. See the localized 1-D-cross section of ϕ in Figure 6.47 (b): the curve is smoother and strictly increasing.

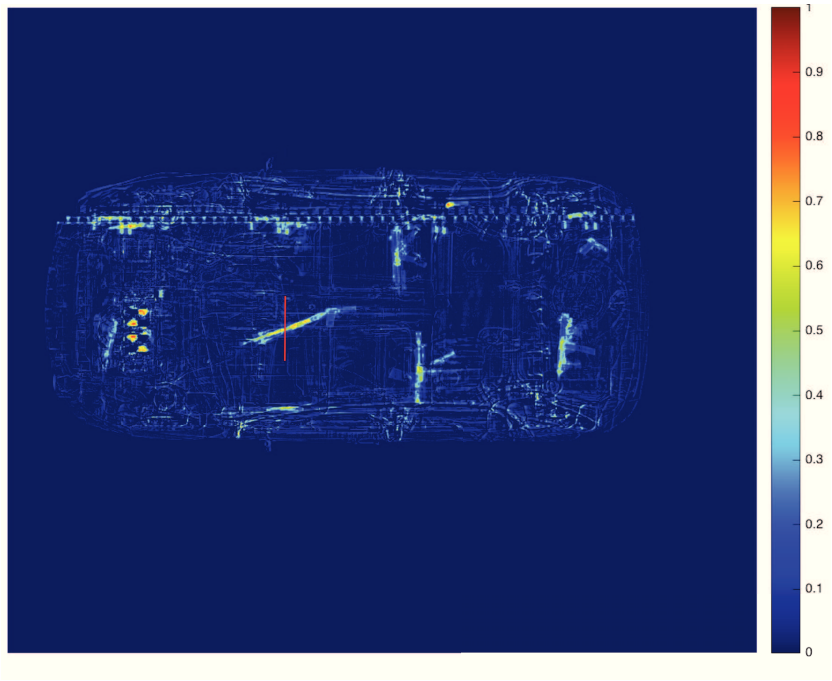


(a) $r_\alpha = 5$,$r_\beta = 5$

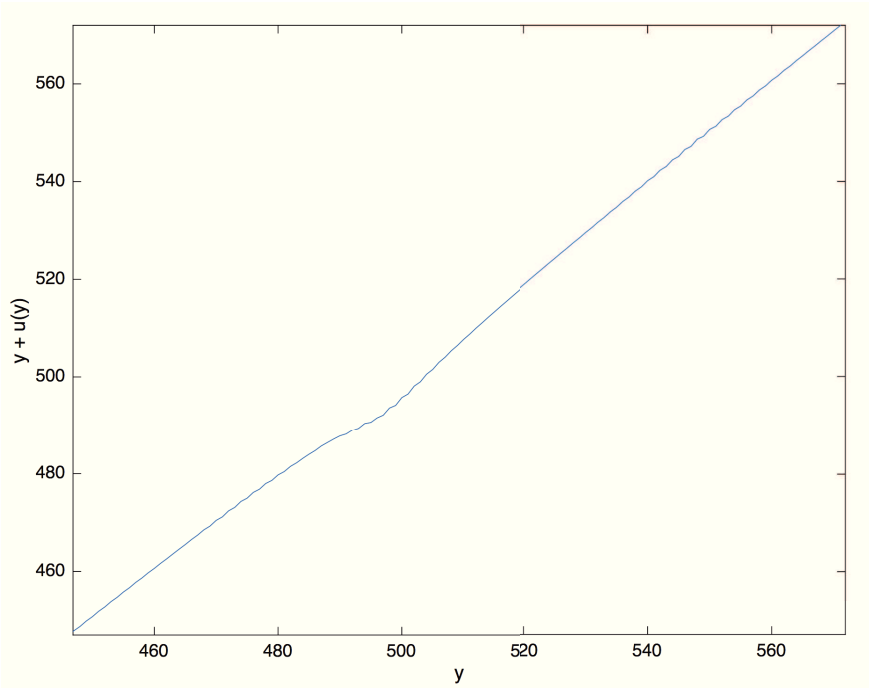


(b) ϕ representation

Figure 6.46 – (a) Difference map obtained by elastic registration with a low smoothing of $\mathbf{u}$; (b) ϕ representation on a local 1-D cross-section.

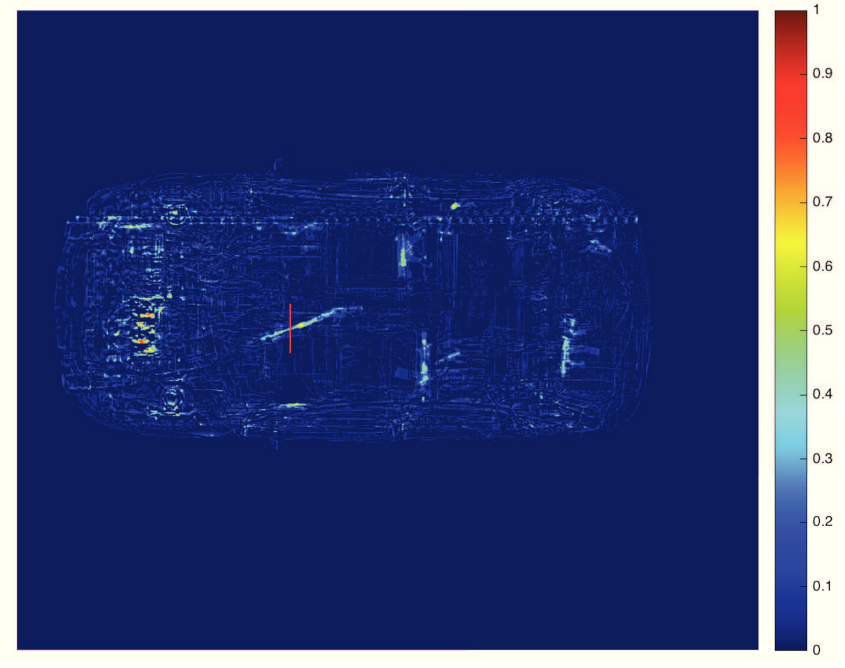


(a) $r_\alpha = 1.5$,$r_\beta = 2$

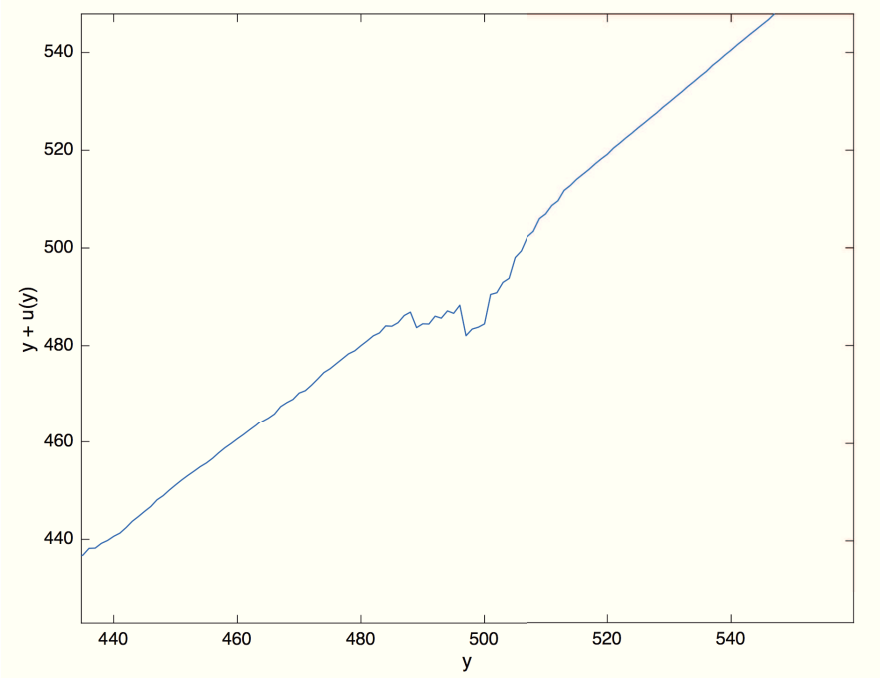


(b) ϕ representation

Figure 6.47 – (a) Difference map obtained by elastic registration with a larger smoothing of $\mathbf{u}$; (b) ϕ representation on a local 1-D cross-section.



(a) $r_\alpha = 1.5, r_\beta = \infty$

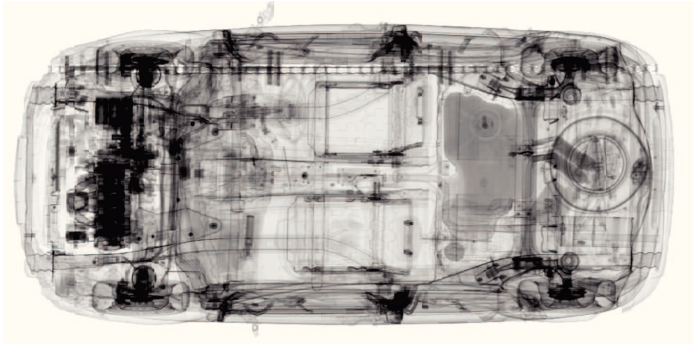


(b) ϕ representation

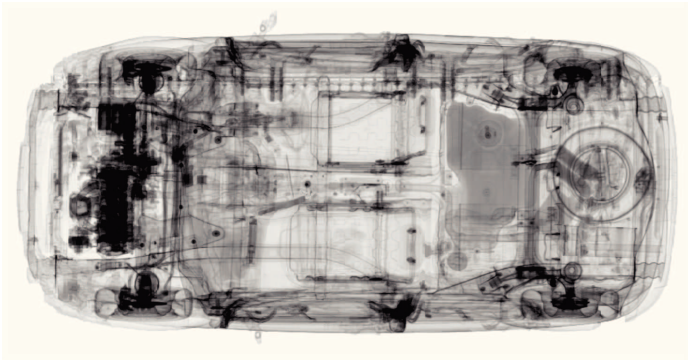
Figure 6.48 – (a) Difference map obtained by diffusion registration with a larger smoothing of $\mathbf{u}$; (b) ϕ representation on a local 1-D cross-section.

In Figure 6.48, we show the relevance of the additional volume-preserving soft constraint with respect to the diffusion registration result. On a visual aspect, the preservation of volumes is not maintained, leading to a weaker detectability of objects in the difference map. The transformation lacks regularity and is clearly non-diffeomorphic.

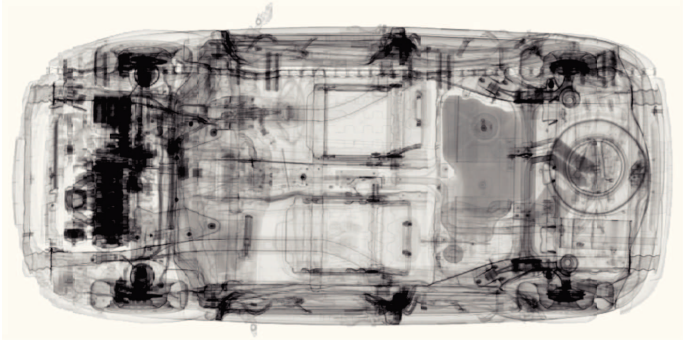
In the following images, we show **the "artificial" appearance of weapons in $\mathbf{T}_{\mathbf{u}^*}$ when adopting a too weak regularization**, whereas it is avoided with a more careful setting of r_α and r_β.



(a) $\mathbf{T}$ after horizontal 1-D registration

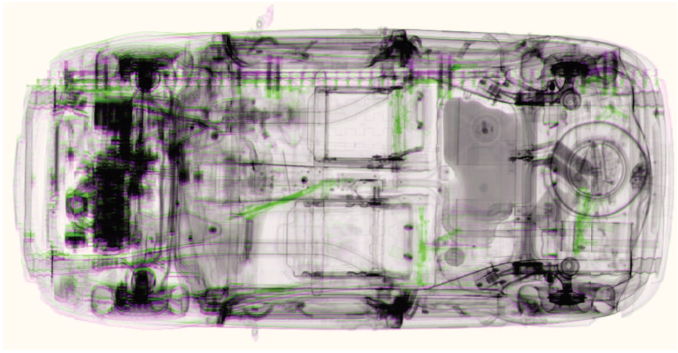


(b) $\mathbf{T}_{\mathbf{u}^*}$ with $r_\alpha = 5$,$r_\beta = 5$

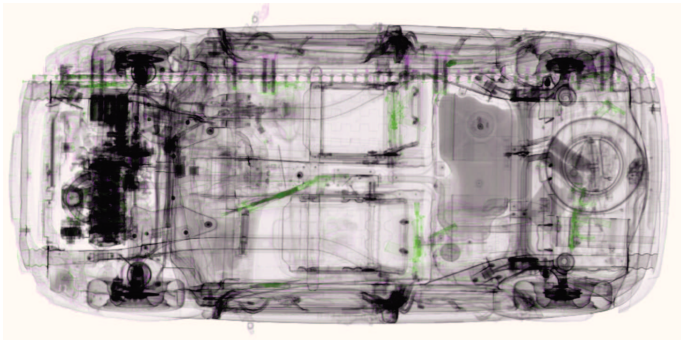


(c) $\mathbf{T}_{\mathbf{u}^*}$ with $r_\alpha = 1.5$,$r_\beta = 2$

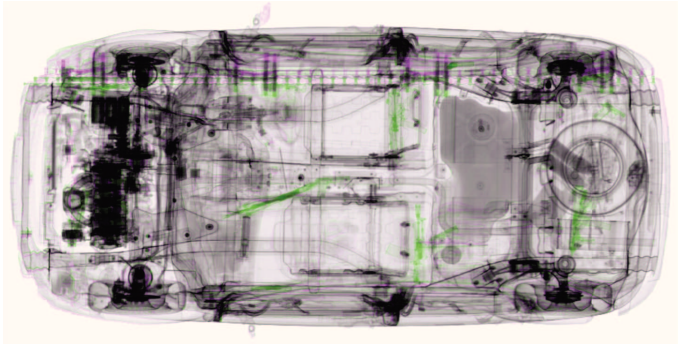
Figure 6.49 – Visualization of $\mathbf{T}_{\mathbf{u}^*}$ following vertical registration without volume preservation (b); (c) with incompressibility soft constraint. With regards to the initial $\mathbf{T}$. In (b), several objects appear, causing the difference reduction visible on Figure 6.46.



(a) $\mathbf{R}$ and $\mathbf{T}$ overlay after horizontal 1-D registration



(b) $\mathbf{R}$ and $\mathbf{T}_{\mathbf{u}^*}$ overlay with $r_\alpha = 5$,$r_\beta = 5$



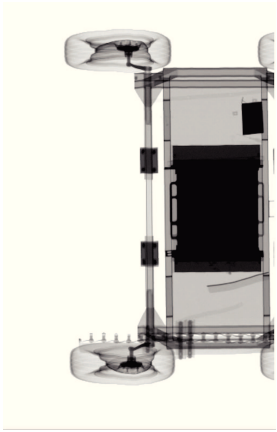
(c) $\mathbf{R}$ and $\mathbf{T}_{\mathbf{u}^*}$ overlay with $r_\alpha = 1.5$,$r_\beta = 2$

Figure 6.50 – Visualization of $\mathbf{R}$ and $\mathbf{T}_{\mathbf{u}^*}$ overlay following vertical registration without volume preservation (b); (c) with incompressibility soft constraint. In (b), the differences in green are less visible.

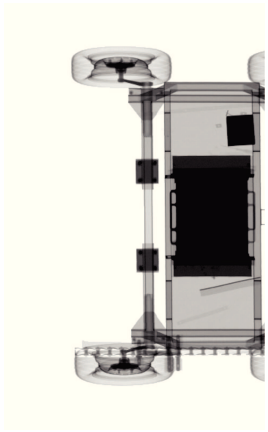
6.7.7 More Results

After validating our method on hundreds of scan pairs, we empirically obtain stable and accurate results with the ratio parameters $r_\alpha = 1.5$ and $r_\beta = 2$ for an optimal detectability of the objects of interest being targeted. A few instances are presented in this section.

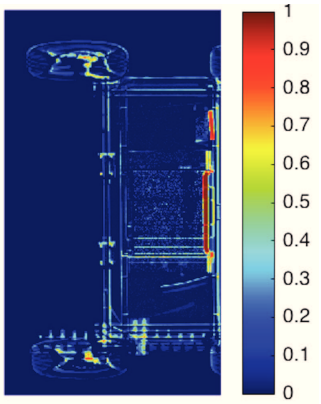
First, the registration example of images of a carrier is given.



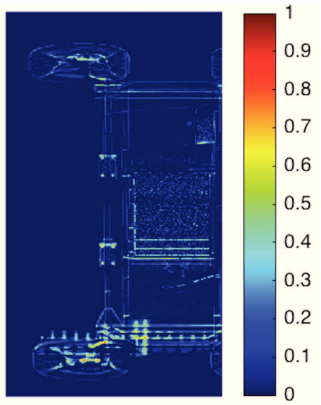
(a) **R**



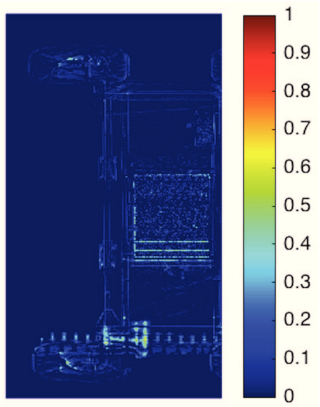
(b) **T**



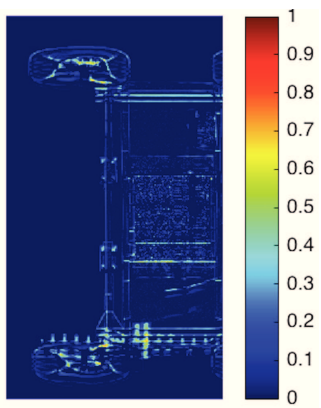
(c) rigid registration



(d) horizontal 1-D registration



(e) 2-stages elastic registration



(f) simultaneous elastic registration

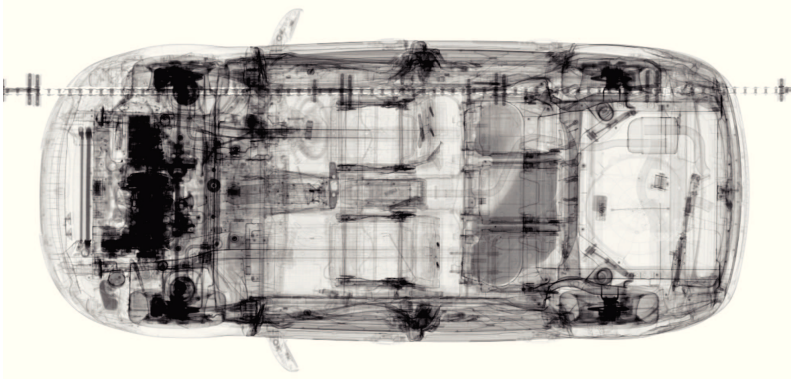
Figure 6.51 – Alignment of two carrier images with the different methods. Elastic registration was performed with $r_\alpha = 1.5$ and $r_\beta = 2$. In (a) and (b), a significant vertical shift is observed (note the wheel position w.r.t the trailer).

On this simpler image set too, the consecutive minimization scheme shows a better correction than the simultaneous approach. Visually, the major differences at the wheels or the frame borders are removed (Figure 6.51). It also achieves the lowest MSE as reported on Table 6.1.

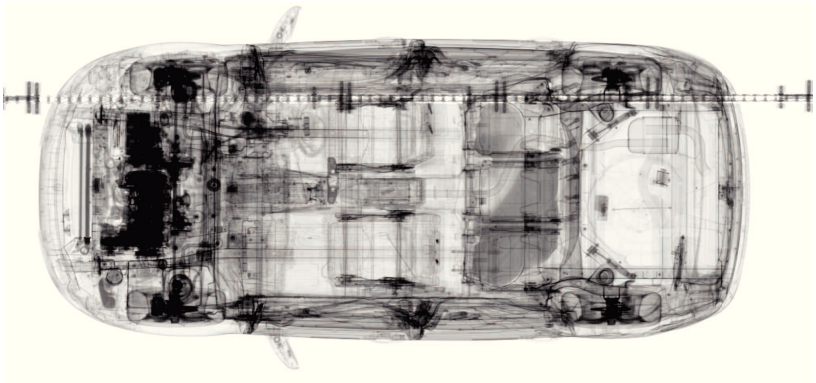
Table 6.1 – Mean Square Error measures obtained with the different registration methods in the example of Figure 6.51

	Rigid Registration	**X-axis Correction**	**X-then-Y Correction**	**Simultaneous**
MSE	1.55E-02	4.14E-03	1.11E-03	4.90E-03

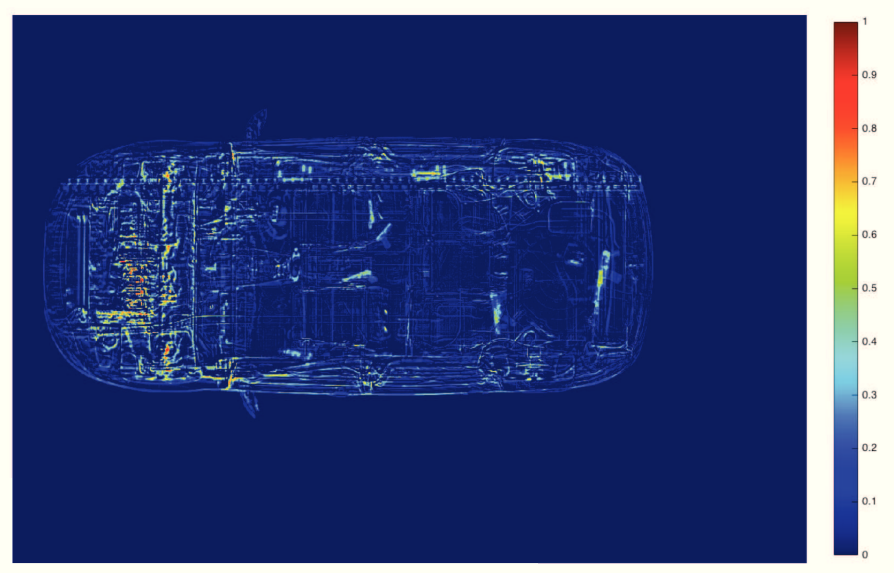
Further examples show the necessity for an additional vertical correction in case of significant shift of the car to the sides while being scanned.



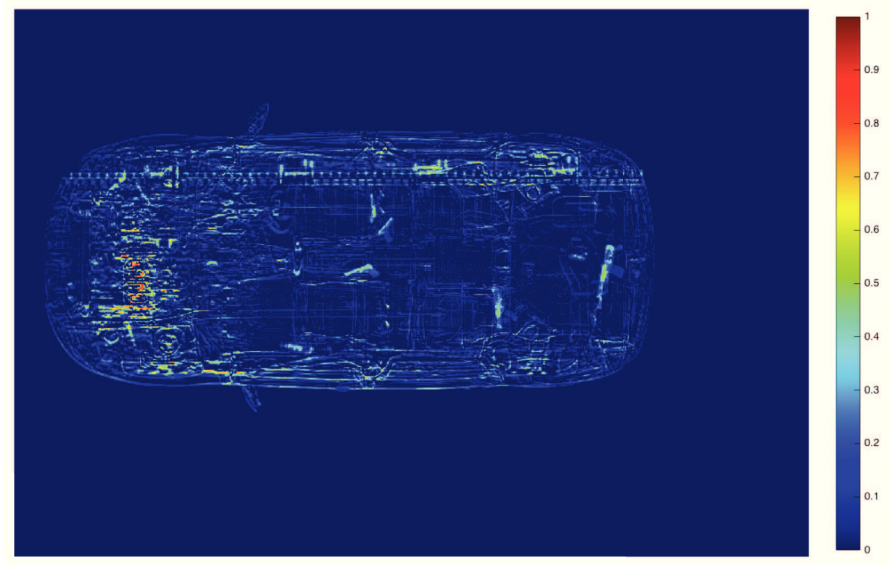
(a) **R**



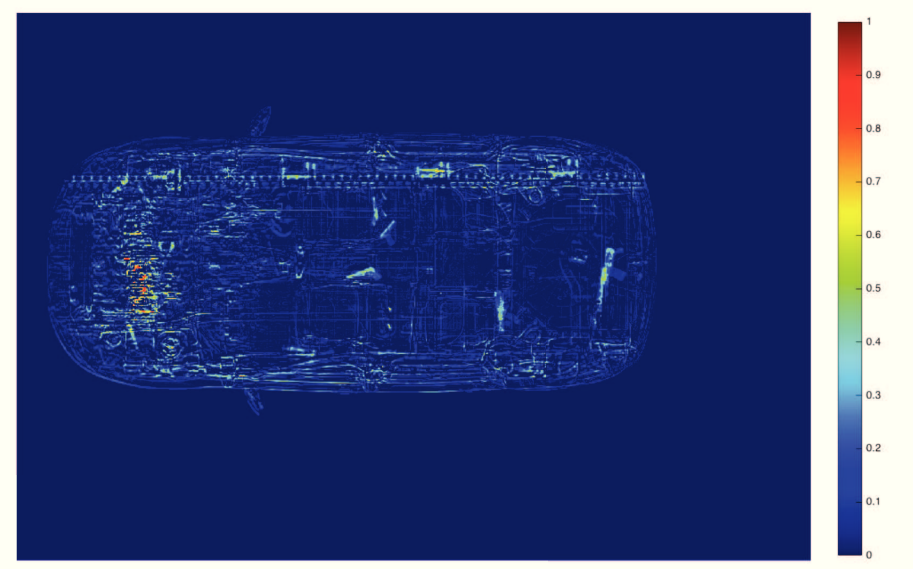
(b) **T**



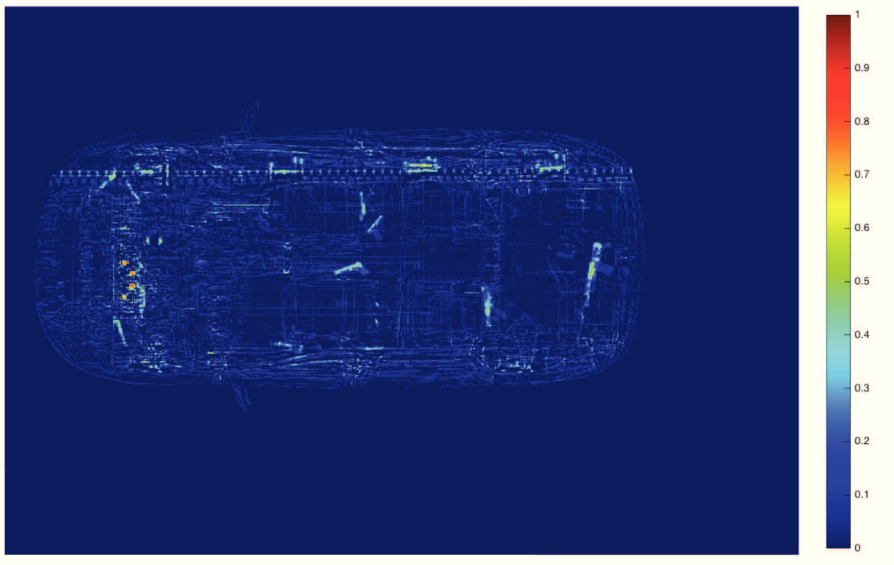
(c) rigid registration



(d) horizontal 1-D registration



(e) simultaneous registration

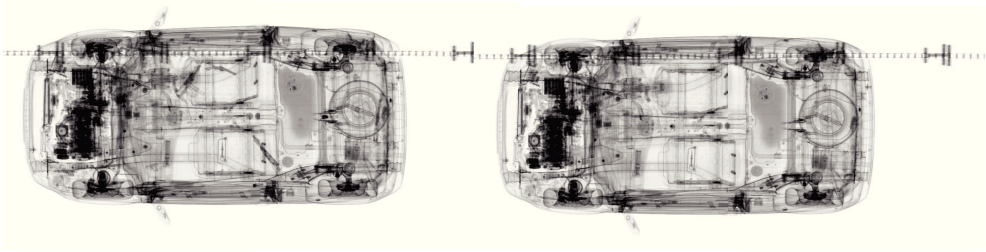


(f) two-stages elastic registration

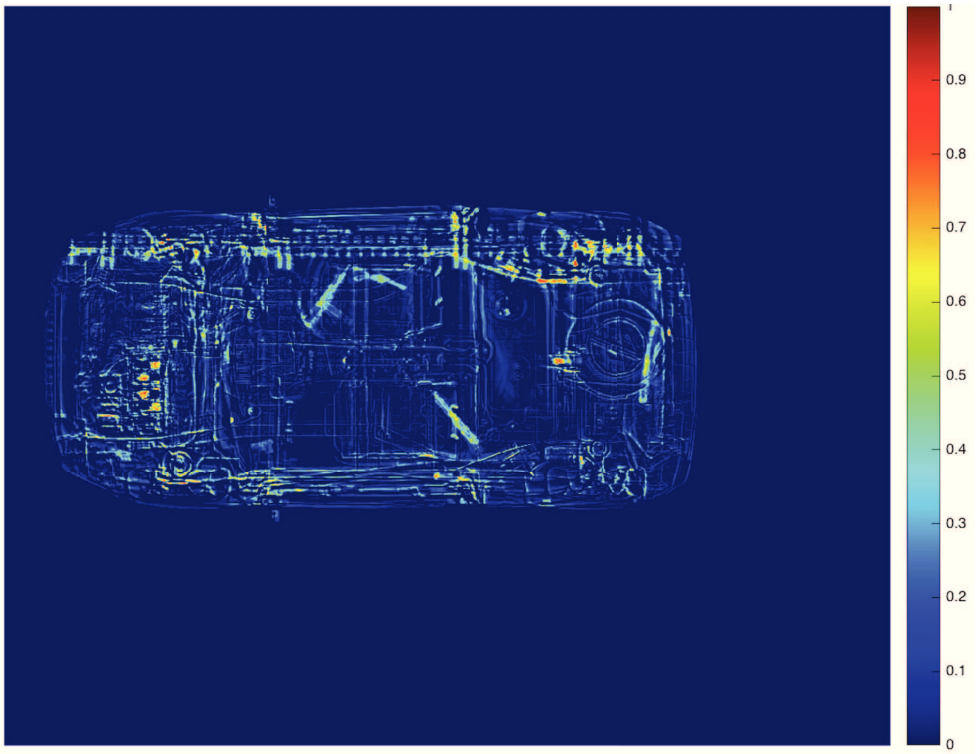
Figure 6.52 – Visualization enhancement obtained with our horizontal-then-vertical elastic correction. Simultaneous and consecutive elastic registration were performed with $r_\alpha = 1.5$ and $r_\beta = 2$.

Table 6.2 – Mean Square Error measures obtained with the different registration methods in the example of Figure 6.52

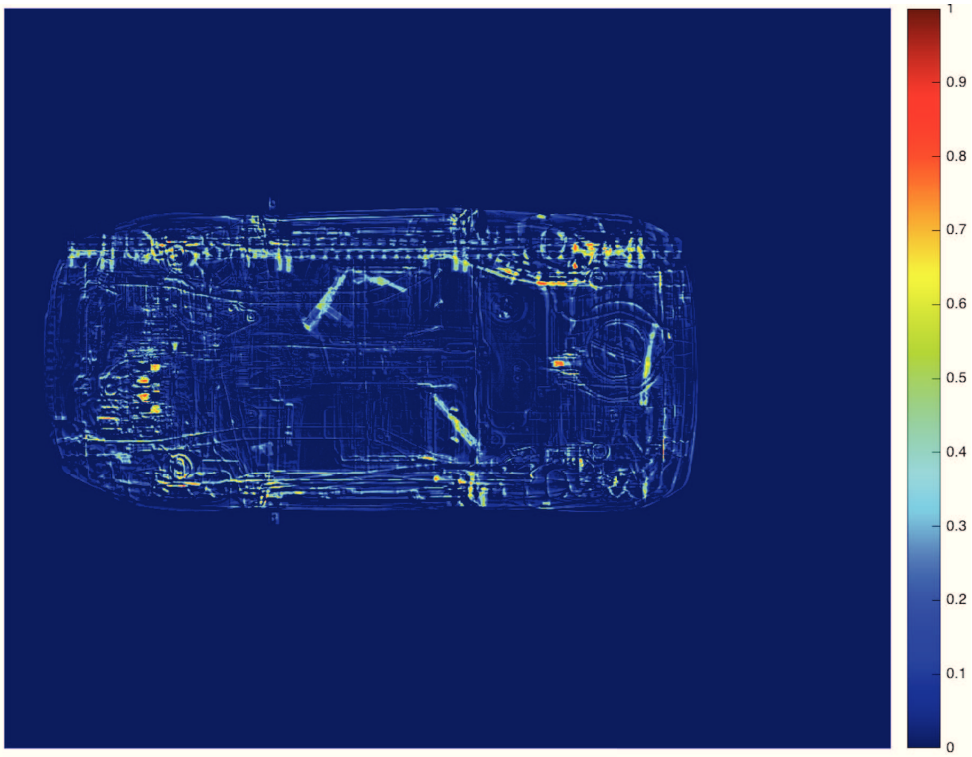
	Rigid Registration	**X-axis Correction**	**X-then-Y Correction**	**Simultaneous**
MSE	3.40E-02	2.80E-03	1.10E-03	2.10E-03



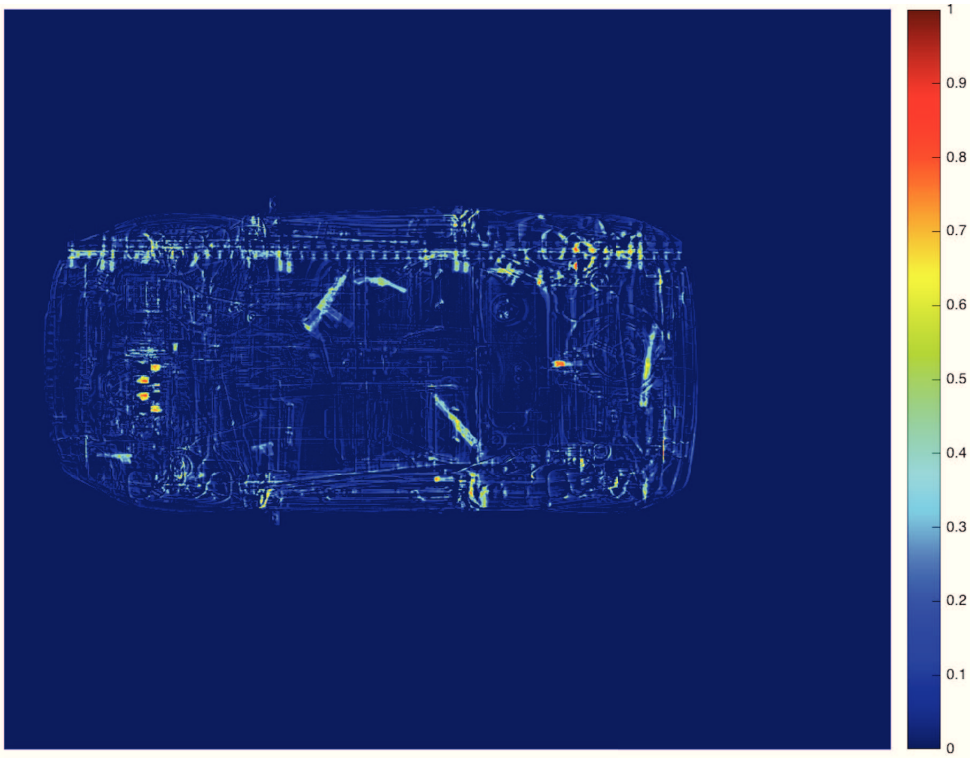
(a) **R** (b) **T**



(c) rigid registration



(d) horizontal 1-D registration



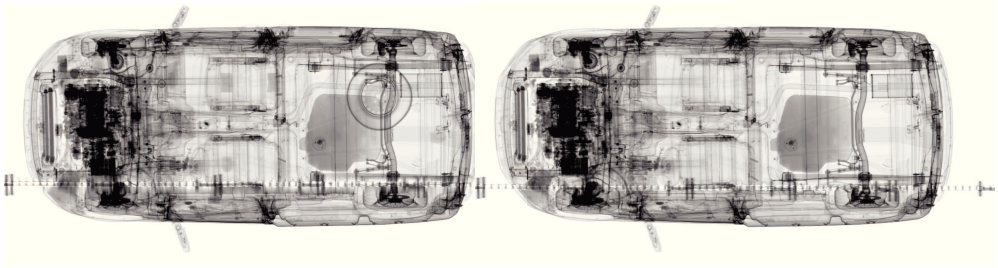
(e) 2-stages registration

Figure 6.53 – Visualization enhancement obtained with our horizontal-then-vertical elastic correction.

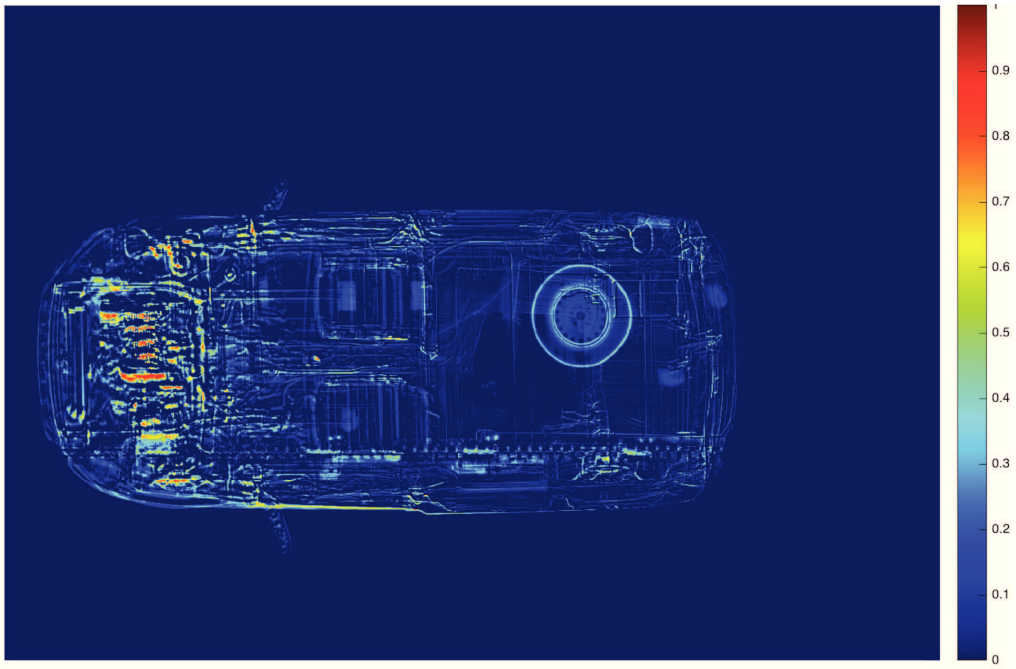
Table 6.3 – Mean square error measures obtained with the different registration techniques.

	Rigid Registration	**X-axis Correction**	**X-then-Y Correction**
MSE	4.2E-03	3.4E-03	2.4E-03

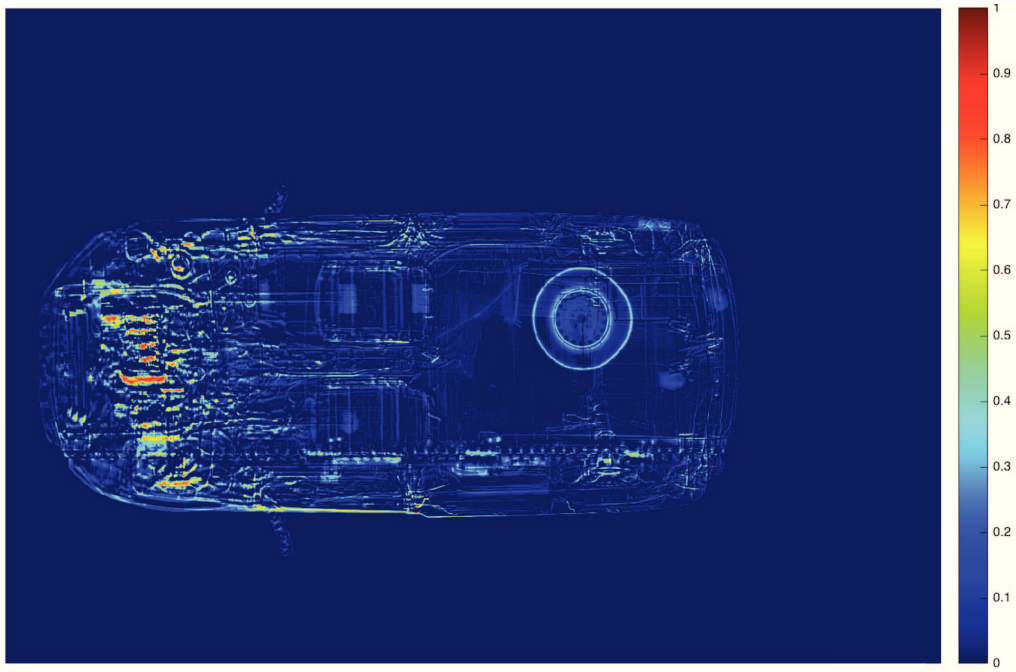
In the following case, drugs and explosives simulants were placed all over a civilian vehicle. Those introduced within the front of the car are almost undetectable on Figure 6.54 (a) under analysis. The vertical shift occurring between **T** and **R** scans causes severe registration errors as shown in Figure 6.54 (c). Misregistration effects are particularly conspicuous at the front-hood of the car, where dense mechanical parts make our rationing difference measure highly sensitive. Threats placed at this spot are thus very unlikely to be detected by custom officers using either rigid or our constrained 1-D horizontal registration approach (Figure 6.54 (d)). The relevance of our additional vertical elastic correction is demonstrated on Figure 6.54 (e) where discrepancies should be visually noticed at the front-hood top and bottom.



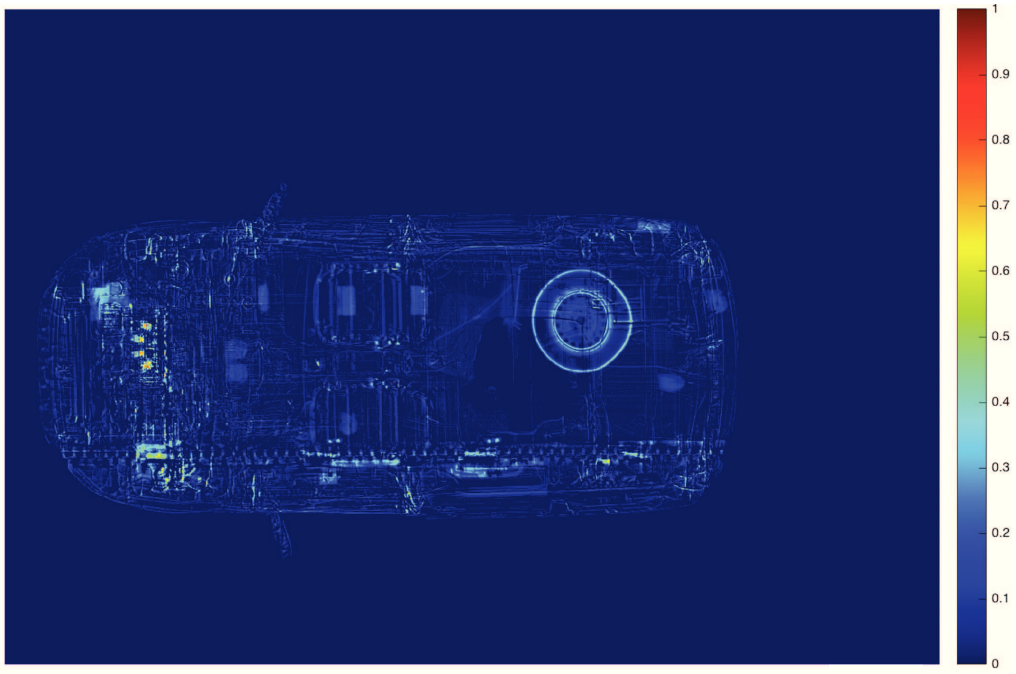
(a) **R** (b) **T**



(c) rigid registration



(d) horizontal 1-D registration



(e) 2-stages registration

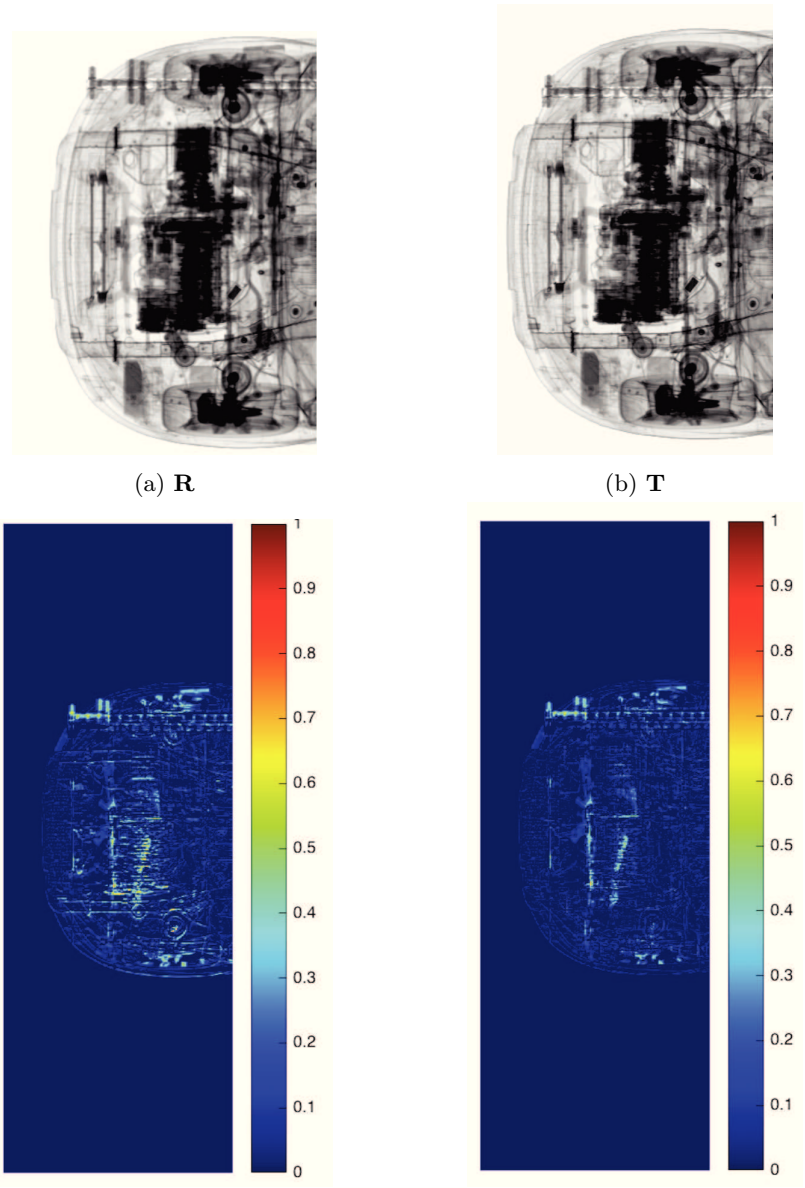
Figure 6.54 – Strong visualization enhancement obtained with our horizontal-then-vertical elastic correction using $r_\alpha = 1.5$ and $r_\beta = 2$.

Table 6.4 – Mean square error measures obtained with the different registration techniques.

	Rigid Registration	**X-axis Correction**	**X-then-Y Correction**
MSE	4E-03	3.6E-03	1.2E-03

Sometimes, vertical correction may be employed for visualization enhancement purposes, although foreign objects are already visible through our horizontal correction step. Removing noisy differences could turn to be particularly useful, given that a further automation stage should be considered in the future (binarization and morphological operations).

The images displayed on Figure 6.55 (a) and (b) represent the front-hood hood of a given vehicle scanned twice. In **R**, three guns as well as an AK-47 rifle were smuggled whereas **T** stands for the empty reference. A substantive visibility improvement is obtained by the application of an further vertical correction in Figure 6.55 (d) where two guns and a revolver are visible at the right of the assault rifle.

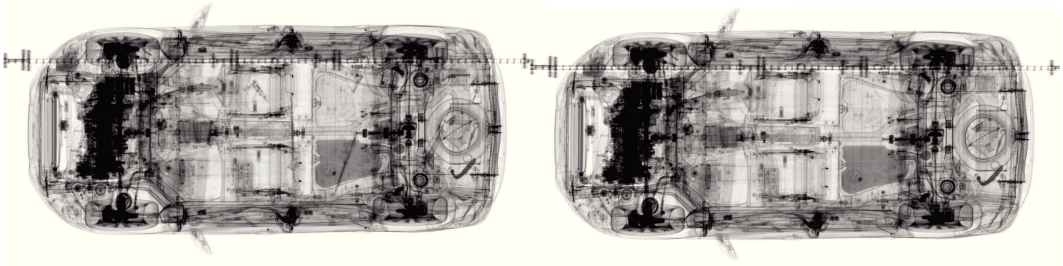


(a) **R** (b) **T**

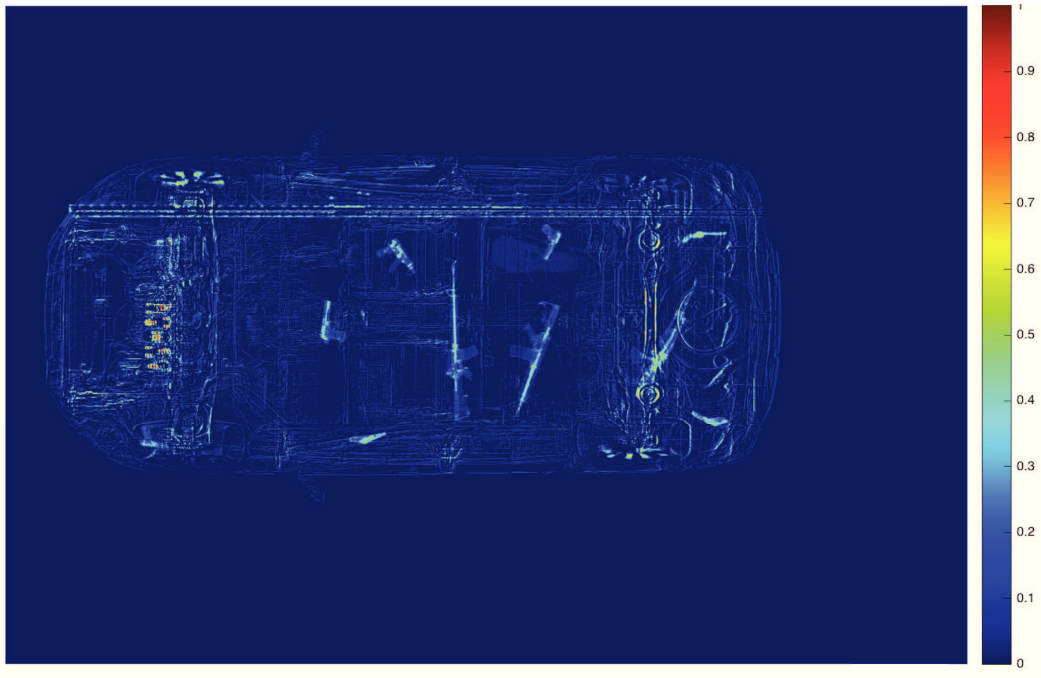
(c) horizontal 1-D registration (d) additional vertical correction

Figure 6.55 – Strong visualization enhancement obtained with our horizontal-then-vertical elastic correction using $r_\alpha = 1.5$ and $r_\beta = 2$.

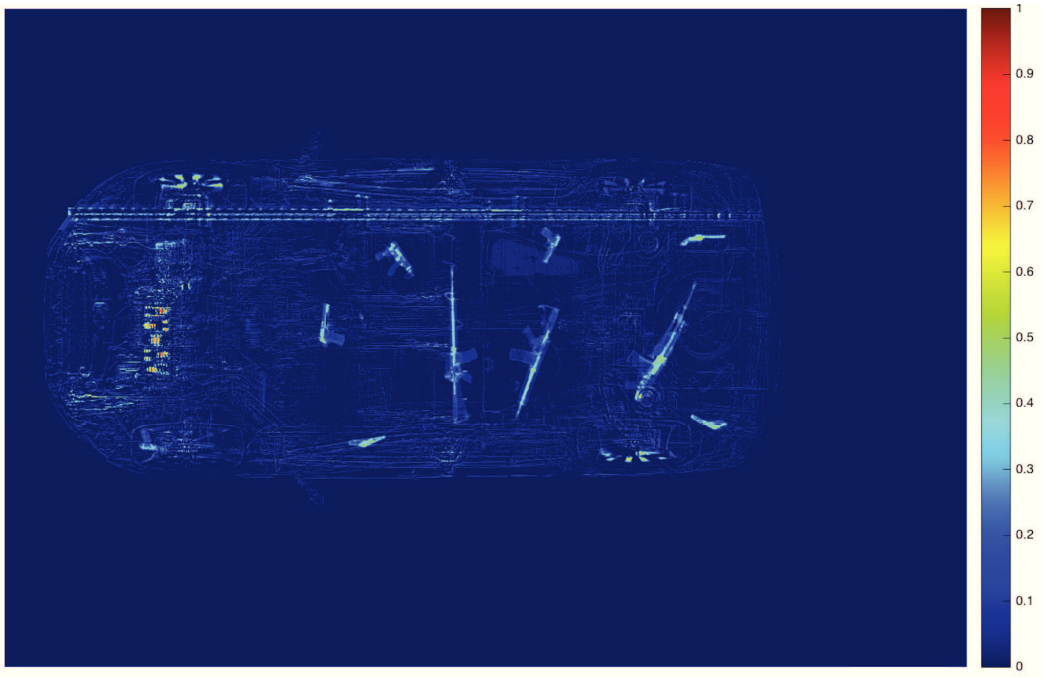
In the next example, gunfires are visible in throughout the car on the difference map. The additional correction in the vertical direction smooths out noisy spurious details as described by Figure 6.56 (e).



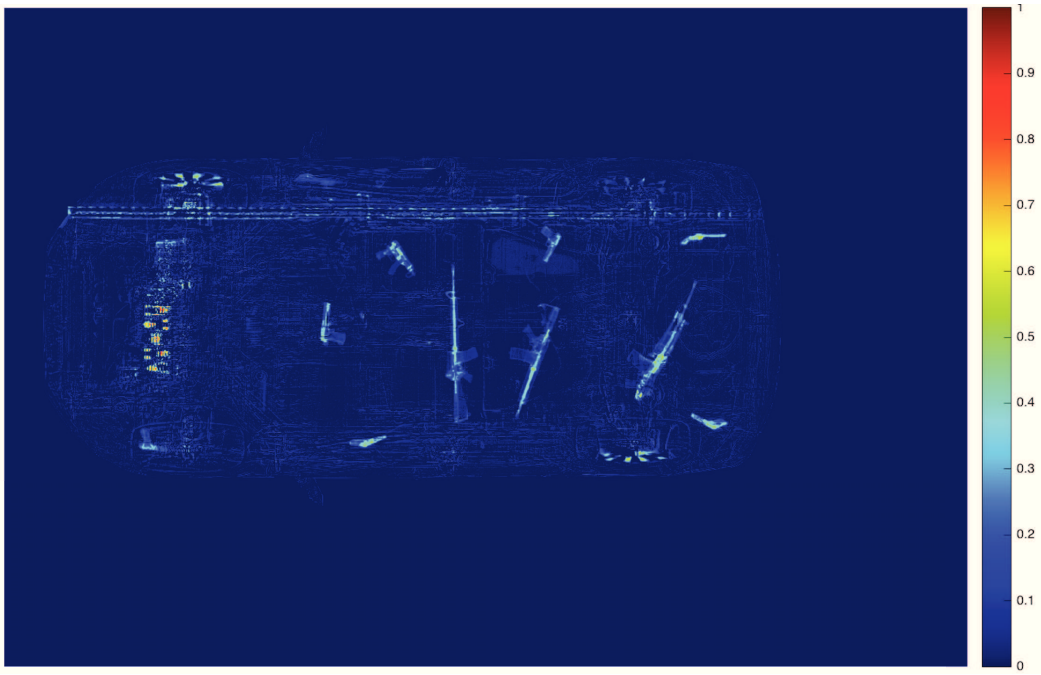
(a) **R** (b) **T**



(c) rigid registration



(d) horizontal 1-D registration



(e) additional vertical correction

Figure 6.56 – Strong visualization enhancement obtained with our horizontal-then-vertical elastic correction using $r_\alpha = 1.5$ and $r_\beta = 2$.

Table 6.5 – Mean square error measures obtained with the different registration techniques.

	Rigid Registration	**X-axis Correction**	**X-then-Y Correction**
MSE	9.1E-03	6.4E-03	4.6E-03

6.8 Classification Issues

As mentioned earlier, the application may be limited to different scopes.

Thanks to homeland security programs, custom officers are able to automatically recognize the exact car model on the basis of its license plate number (OCR). Thus, a database composed of empty reference scans could be directly built by themselves, and queried whenever a vehicle is being scanned. If a vehicle model is matched to a pre-scanned model of the database, image registration can be carried out towards difference detection. As a result, the whole car could be analyzed.

Figure 6.57 – Snapshot taken on each scanned vehicle, for license plate OCR recognition. These images represent valuable information (front of the car), which can be jointly utilized with X-ray samples for classification.

Another approach consists in focusing only on specific parts of the vehicle *e.g* the trunk or front-hood. Thanks to field of displacements extracted from a bank of image pairs, one could create hundreds of variations for each real sample of data and then obtain a very large database. A CNN could be trained from scratch or a pre-trained model employed for fine-tuning on the basis of the created samples. Regular images (non X-ray) acquired for OCR plate recognition could also be jointly employed with scan samples in order to achieve a more accurate recognition of car models (see Figure 6.57). Actually, getting an empty reference scan for each car model is not obvious, that is why we would prefer to make recognition and registration tasks focus on regions that are generally empty (*e.g* the front-hood). The automatic extraction of regions is well-handled by R-CNN approaches for instance.

6.9 Conclusion

In this chapter, we addressed the registration problem of top-view scans $\mathbf{R}$ and $\mathbf{T}$ acquired on vehicles of a same model. After obtaining unsuccessful results with rigid registration, standard nonlinear methods were tested. They also showed poor performances by generally violating the intrinsic rigidity feature of a car and achieving low registration accuracy.

Through a simple analysis of the scanning machine properties, we stated two hypotheses: the field of displacement $\mathbf{u}$ must be column-wise constant and parallel to the car motion. We formulated these assumptions, turning the two-dimensional variational paradigm into a simple unidimensional scheme to perform a horizontal correction through the search of an optimal $\mathbf{u}^*$.

However, some experiments showed that the proposed method seems to be uneffective in case of significant vertical shifts between the cars of both moving and static images. In fact, the radiated beam is triangular, projecting differently any translated object according to its distance to the X-ray source.

We proposed an add-on to our previous contribution through a vertical correction under a simple elastic registration paradigm (diffusion scheme together with a linearized approximation of the soft volume-preservation constraint). Two approaches are thus put forward. The first consists in applying simultaneously the registration on both components of the displacement

field (with the column-wise constancy constraint on the horizontal term). In the alternative, more efficient method, we run the horizontal registration first, followed by the vertical correction in a separate fashion. The results are less noisy, yielding a better registration accuracy on both numerical and visual aspects, especially when elastic regularization is employed.

We also highlight the importance of an appropriate choice for the regularization parameters with the risk of smoothing out suspicious items in the difference map.

Conclusion and Future Improvements

Throughout this thesis, we proposed simple and intuitive schemes to address different registration problems of images obtained by cargo screening systems. We globally contributed in designing new methods that would offer either a full-automation or enhanced visualization for irregularity detection in various kinds of vehicles. We also show that registration approaches may prove very efficient in extreme situations, especially at dense spots yielding a very strong attenuation of the X-ray intensity.

Our first contribution consisted in using ASM to find, among a set of images, the closest template $\mathbf{T}$ to the inspected pre-labelled input image $\mathbf{R}$. On the basis of the ASM representation of shapes, we were also able to formulate a cost combining both shape and appearance offset from the trained model, and employ it to perform a classification refinement when needed (double-check multi-class SVM output). In order to address local variations, we partition $\mathbf{R}$ and $\mathbf{T}$ into paired sub-images on which a fast feature-based linear registration is achieved. An additional regularization is applied on the resulting field of displacement $\mathbf{u}$, usually subject to strong discontinuities at sub-images boundaries. To this end, a diffusion process, confined to the discontinuity regions is performed.

The differences between the transformed empty image $\mathbf{T_u}$ and $\mathbf{R}$ are measured by our adapted rationing approach, in compliance with X-ray physical laws (Beer-Lambert), enhancing notably the detection capacity with respect to simple differencing methods. Given the map of differences, straightforward thresholding and morphological operations allow for a precise localization of irregularities. Eventually, a post-processing filtering can be applied to decrease the number of false alarms.

The strategy, despite being presented through the specific issue of refrigeration units inspection in containers, its concepts are quite generic and can be applied to a wider range of problems in X-ray cargo imaging. For instance, the chassis, along with the roof, side walls, doors and wheels could be extracted via R-CNN methods and undergo a similar algorithmic pipeline for discrepancy detection.

A few improvements could be made for a better accuracy and robustness. First, concerning the ASM process, we could use strictly connected shapes $\mathbf{x}_j$ instead of a shape $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, .., \mathbf{x}_c)$ constructed from c connected sub-shapes. The search of the nearest template would then amount to directly identify closest sub-images $\mathbf{T}_j$ in terms of shape. Doing so would allow to compensate, in a more robust fashion, for the various local deformations in X-ray scans.

One could also think about using more sophisticated thresholding or segmentation methods to identify the substantive changes through the map of differences.

Our strategy uses a simple linear registration scheme. As a consequence, non-linear deformations are meant to be considered by finding a very similar empty reference that displays relatively same distortions. The approach therefore relies on the "variability completeness" within our dataset. In order to be robust to all modes of deformation, we need to gather the main observed distortions across a very large number of samples. Accordingly, the current database must be significantly consolidated by the supply of additional data. It also applies for images containing true illicit goods or threats to perform a more stringent validation of the algorithm.

Our second application deals with a very specific registration task on top-view scans of same-model light vehicles. While linear and standard non-linear registration schemes fail in providing an accurate alignment of $\mathbf{R}$ and $\mathbf{T}$, we contributed in formulating a simple customized optimization paradigm. Indeed through the examination of the HCVL system, we assumed that the horizontal component of the field of displacement $\mathbf{u}$ should be column-wise constant ($u_x(x, y) = u_x(x)$) and its vertical component neglected ($u_y(x, y) = 0$). By resorting to the unified variational registration framework, we turn a 2-D problem into a straightforward 1-D diffusion registration scheme. A fast numerical solution is supplied by the use of gradient descent associated with Armijo's backtracking line-search and a coarse-to-fine approach. In most cases, the model handles well observed non-linear distortions, as described visually and numerically.

Yet, in some situations, a lateral translation of a vehicle during scanning with respect to a previous acquisition entails a sterescopic effect, reducing notably the detection capacity in some regions (*e.g* front-hood). We show that the best solution is provided by the implementation of two separate stages:

- a 1-D horizontal registration with the column-wise constancy constraint $u_x(x, y) = u_x(x)$. It is similar to the well-known rectification pre-processing step in stereo vision. Columns from $\mathbf{T}$ and $\mathbf{R}$ are therefore aligned.
- For each column, a 1-D vertical correction by searching for an optimal $u_y(x, y)$. Since volume preservation is a critical issue in threat detection, an additional soft incompressibility constraint is introduced in the diffusion scheme, yielding an elastic registration model.

Actually, without using an adequate regularization, the deformable template $\mathbf{T}$ will tend to "mimic" added objects in $\mathbf{R}$ in a non-diffeomorphic fashion. This would decrease the differences amplitudes and put the threat detection algorithm at risk. The numerical solution used in the proposed approach is the l-BFGS also combined with Armijo's criteria and a multi-resolution scheme.

Still, our method is time consuming and we will need to speed-up the different processes in the near future. Moreover, the choice of smoothing parameters plays an essential role throughout the registration process, and depends utterly on the loading of the vehicles represented in $\mathbf{R}$ and $\mathbf{T}$. Indeed, two vehicles displaying a few differences would require much less regularization than when $\mathbf{R}$ is full (and a low smoothing would then lead to a non-realistic deformation). A future improvement of our algorithm should inevitably go through the design of a method that would automatically define the smoothing parameters to use. It could be performed by an histogram-comparison approach for instance.

Yet, this issue arises only if we aim at registering entire vehicles. If not, one could consider extracting specific regions of interest by R-CNN and compare them by registration techniques. For the front-hood case *e.g*, the congestion level is similar for all acquisitions of a same model. Similarly, the classification problem of HCVL scans of vehicles has only been discussed briefly and must be addressed by precisely defining the scope of the application.

An interesting alternative method is to directly use the dimensions of the HCVL system to compute a 3-D representation of the car scanned twice (or different instances of a same model). It amounts to considering both acquisitions as the same object being imaged through different points-of-view. As for stereo images, triangulation would give an estimate of the car in 3-D. One could constrain the cloud of 3-D points to be in compliance with the system properties and the car model. By back-projecting the obtained 3-D data into 2-D images, we could iteratively perform image registration under these specific regularization constraints.

Finally, besides the visual and numerical examples reported along our study, we should create ground-truth difference maps by using different scans of a same vehicle and synthetically merge threats all over it. A mask gathering all the inserted threats would stand for a ground-truth detection image. By doing so, we would be able to validate our approach in a more rigorous way.

Our works contribute in advancing the field of cargo-vision and provide interesting tools for law-enforcement officers to go further in their strive against illicit trade, crime and terror. In the future, one could also think about considering image comparison and deep-learning recognition methods as two complementary tools for threat detection.

List of Publications

The work done during this PhD led to the following publications and patent.

Peer-Reviewed Conference Papers

- Abraham Marciano, Laurent D Cohen, and Najib Gadi. "Vehicle X-ray scans registration: A one-dimensional optimization problem". In: *International Conference on Scale Space and Variational Methods in Computer Vision*, pages 578–589. Springer, 2017
- Abraham Marciano, Laurent D. Cohen, and Najib Gadi. "Vehicle X-ray images registration". In: *Energy Minimization Methods in Computer Vision and Pattern Recognition: 11th International Conference, EMMCVPR2017, Venice, Italy, October 30–November 1, 2017, Revised Selected Papers*, volume 10746, page 278. Springer, 2018
- Abraham Marciano, Laurent D. Cohen, and Najib Gadi. "Recalage d'images radiographiques de véhicules : un problème d'optimisation unidimensionelle". In *ORASIS 2017, Colleville-sur-Mer, France, Revised Selected Papers*, 2017
- Wicher Visser, Adrian Schwaninger, Diana Hardmeier, Alexander Flisch, Marius Costin, Caroline Vienne, Frank Sukowski, Ulf Hassler, Irène Dorion, Abraham Marciano, et al. "Automated comparison of X-ray images for cargo scanning". In: *Security Technology (ICCST), 2016 IEEE International Carnahan Conference on Security Technology, Orlando, Florida, USA, 24-27 October, 2016*, pages 1–8. IEEE, 2016

Journal Articles

- Selina Kolokytha, Alexander Flisch, Thomas Lüthi, Mathieu Plamondon, Adrian Schwaninger, Wicher Vasser, Diana Hardmeier, Marius Costin, Caroline Vienne, Frank Sukowski, Abraham Marciano et al. "Improving customs border control by creating a reference database of cargo inspection X-ray images". In: *Advances in Science, Technology and Engineering Systems* Journal Vol. 2, No. 3, 60-66, 2017
- Selina Kolokytha, Alexander Flisch, Thomas Lüthi, Mathieu Plamondon, Wicher Visser, Adrian Schwaninger, Diana Hardmeier, Marius Costin, Caroline Vienne, Frank Sukowski, Abraham Marciano et al. "Creating a reference database of cargo inspection X-ray images using high energy radiographs of cargo mock-ups". In: *Multimedia Tools and Applications*, 77(8):9379–9391, 2018

Patent Pending

- Abraham Marciano, Laurent D. Cohen, and Najib Gadi. "Detection of Irregularities Using Registration". In: *UK Application* 1620098.2, November 2016.

Bibliography

[1] T. Albrecht, A. Dedner, M. Lüthi, and T. Vetter. "Finite element surface registration incorporating curvature, volume preservation, and statistical model information". In: *Computational and mathematical methods in medicine* 2013 (2013).

[2] A. Aldroubi, M. Eden, and M. Unser. "Discrete Spline Filters for Multiresolutions and Wavelets of l_2". In: *SIAM Journal on Mathematical Analysis* 25.5 (1994), pp. 1412–1432.

[3] V. Arsigny, O. Commowick, X. Pennec, and N. Ayache. "A log-euclidean framework for statistics on diffeomorphisms". In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer. 2006, pp. 924–931.

[4] D. Arthur and S. Vassilvitskii. "k-means++: The advantages of careful seeding". In: *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms*. Society for Industrial and Applied Mathematics. 2007, pp. 1027–1035.

[5] J. Ashburner. "A fast diffeomorphic image registration algorithm". In: *Neuroimage* 38.1 (2007), pp. 95–113.

[6] A. Asthana, S. Zafeiriou, S. Cheng, and M. Pantic. "Robust discriminative response map fitting with constrained local models". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2013, pp. 3444–3451.

[7] G. Aubert and P. Kornprobst. *Mathematical problems in image processing: partial differential equations and the calculus of variations*. Vol. 147. Springer Science & Business Media, 2006.

[8] J.-F. Aujol. "Calculus of variations in image processing". In: *Notes de cours de Master* (2008).

[9] A. Babenko, A. Slesarev, A. Chigorin, and V. Lempitsky. "Neural codes for image retrieval". In: *European conference on computer vision*. Springer. 2014, pp. 584–599.

[10] E. Bardinet, L. D. Cohen, and N. Ayache. "A parametric deformable model to fit unstructured 3D data". In: *Computer vision and image understanding* 71.1 (1998), pp. 39–54.

[11] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. "Speeded-up robust features (SURF)". In: *Computer vision and image understanding* 110.3 (2008), pp. 346–359.

[12] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes. "Computing large deformation metric mappings via geodesic flows of diffeomorphisms". In: *International journal of computer vision* 61.2 (2005), pp. 139–157.

[13] S. Benayoun. "Calcul local du mouvement: applications a l'imagerie medicale multidimensionnelle". PhD thesis. Paris 9, 1994.

[14] C Bishop. "Pattern Recognition and Machine Learning (Information Science and Statistics), 1st edn. 2006. corr. 2nd printing edn". In: *Springer, New York* (2007).

[15] C. Broit. "Optimal registration of deformed images". PhD thesis. University of Pennsylvania, 1981.

[16] C. J. Burges. "A tutorial on support vector machines for pattern recognition". In: *Data mining and knowledge discovery* 2.2 (1998), pp. 121–167.

[17] *CCP glossary of terms (UNODC)*. `https://www.unodc.org/ropan/en/BorderControl/container-control/ccp-glossary-of-terms.html`. Accessed: 2017-12-25.

[18] K. Chatfield, K. Simonyan, A. Vedaldi, and A. Zisserman. "Return of the devil in the details: Delving deep into convolutional nets". In: *arXiv preprint arXiv:1405.3531* (2014).

[19] G. E. Christensen and H. J. Johnson. "Consistent image registration". In: *IEEE transactions on medical imaging* 20.7 (2001), pp. 568–582.

[20] G. E. Christensen, R. D. Rabbitt, and M. I. Miller. "Deformable templates using large deformation kinematics". In: *IEEE transactions on image processing* 5.10 (1996), pp. 1435–1447.

[21] G. E. Christensen. "Deformable shape models for anatomy". PhD thesis. Sener Institute, Washington University, 1994.

[22] D. C. Ciresan, U. Meier, J. Masci, L. Maria Gambardella, and J. Schmidhuber. "Flexible, high performance convolutional neural networks for image classification". In: *IJCAI Proceedings-International Joint Conference on Artificial Intelligence*. Vol. 22. 1. Barcelona, Spain. 2011, p. 1237.

[23] T. Cootes, M. Roberts, K. Babalola, and C. Taylor. "Active shape and appearance models". In: *Handbook of Biomedical Imaging*. Springer, 2015, pp. 105–122.

[24] T. Cootes, E. Baldock, and J Graham. "An introduction to active shape models". In: *Image processing and analysis* (2000), pp. 223–248.

[25] T. F. Cootes, G. J. Edwards, and C. J. Taylor. "Active appearance models". In: *IEEE Transactions on pattern analysis and machine intelligence* 23.6 (2001), pp. 681–685.

[26] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. "Active shape models-their training and application". In: *Computer vision and image understanding* 61.1 (1995), pp. 38–59.

[27] T. F. Cootes, C. J. Taylor, et al. *Statistical models of appearance for computer vision*. 2004.

[28] K. Crammer and Y. Singer. "On the algorithmic implementation of multiclass kernel-based vector machines". In: *Journal of machine learning research* 2.Dec (2001), pp. 265–292.

[29] D. Cristinacce and T. F. Cootes. "Feature Detection and Tracking with Constrained Local Models." In: *BMVC*. Vol. 1. 2. 2006, p. 3.

[30] G. Csurka, C. Dance, L. Fan, J. Willamowski, and C. Bray. "Visual categorization with bags of keypoints". In: *Workshop on statistical learning in computer vision, ECCV*. Vol. 1. 1-22. Prague. 2004, pp. 1–2.

[31] N. Dalal and B. Triggs. "Histograms of oriented gradients for human detection". In: *Computer Vision and Pattern Recognition, 2005. CVPR 2005. IEEE Computer Society Conference on*. Vol. 1. IEEE. 2005, pp. 886–893.

[32] N. Dalal, B. Triggs, and C. Schmid. "Human detection using oriented histograms of flow and appearance". In: *European conference on computer vision*. Springer. 2006, pp. 428–441.

[33] *Did British police just find a surface-to-air missile in a local arms cache? - The Washington Post*. https://www.washingtonpost.com/news/checkpoint/wp/2017/02/17/did-british-police-just-find-a-surface-to-air-missile-in-a-local-arms-cache/?noredirect=on&utm_term=.6e0922d21736. Accessed: 2018-04-10.

[34] J. Donahue et al. "Decaf: A deep convolutional activation feature for generic visual recognition". In: *International conference on machine learning*. 2014, pp. 647–655.

[35] B. Fischer and J. Modersitzki. "A unified approach to fast image registration and a new curvature based registration technique". In: *Linear Algebra and its applications* 380 (2004), pp. 107–124.

[36] B. Fischer and J. Modersitzki. "Fast diffusion registration". In: *Contemporary Mathematics* 313 (2002), pp. 117–128.

[37] B. Fischer and J. Modersitzki. "Fast image registration: a variational approach". In: *Proceedings of the International Conference on Numerical Analysis & Computational Mathematics, G. Psihoyios (ed.), Wiley*. 2003, pp. 69–74.

[38] B. Fischer and J. Modersitzki. "Ill-posed medicine—an introduction to image registration". In: *Inverse Problems* 24.3 (2008), p. 034008.

[39] K. Fukushima. "Neocognitron: A hierarchical neural network capable of visual pattern recognition". In: *Neural networks* 1.2 (1988), pp. 119–130.

[40] R. Girshick, J. Donahue, T. Darrell, and J. Malik. "Rich feature hierarchies for accurate object detection and semantic segmentation". In: *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2014, pp. 580–587.

[41] G. H. Golub and C. Reinsch. "Singular value decomposition and least squares solutions". In: *Numerische mathematik* 14.5 (1970), pp. 403–420.

[42] C. Goodall. "Procrustes methods in the statistical analysis of shape". In: *Journal of the Royal Statistical Society. Series B (Methodological)* (1991), pp. 285–339.

[43] I. Goodfellow et al. "Generative adversarial nets". In: *Advances in neural information processing systems*. 2014, pp. 2672–2680.

[44] E. Haber and J. Modersitzki. "A multilevel method for image registration". In: *SIAM Journal on Scientific Computing* 27.5 (2006), pp. 1594–1607.

[45] E. Haber and J. Modersitzki. "Image registration with guaranteed displacement regularity". In: *International Journal of Computer Vision* 71.3 (2007), pp. 361–372.

[46] E. Haber and J. Modersitzki. "Numerical methods for volume preserving image registration". In: *Inverse problems* 20.5 (2004), p. 1621.

[47] M. Hachama, A. Desolneux, and F. Richard. "Combining registration and abnormality detection in mammography". In: *International Workshop on Biomedical Image Registration*. Springer. 2006, pp. 178–185.

[48] R. Hameeteman, J. F. Veenland, and W. J. Niessen. "Volume preserving image registration via a post-processing stage". In: *Medical Imaging 2008: Image Processing*. Vol. 6914. International Society for Optics and Photonics. 2008, 69140Y.

[49] P. C. Hansen. *Rank-deficient and discrete ill-posed problems: numerical aspects of linear inversion*. Vol. 4. Siam, 2005.

[50] B. Hariharan, P. Arbeláez, R. Girshick, and J. Malik. "Hypercolumns for object segmentation and fine-grained localization". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2015, pp. 447–456.

[51] C.-W. Hsu, C.-C. Chang, C.-J. Lin, et al. "A practical guide to support vector classification". In: *Technical report* (2003).

[52] C.-W. Hsu and C.-J. Lin. "A comparison of methods for multiclass support vector machines". In: *IEEE transactions on Neural Networks* 13.2 (2002), pp. 415–425.

[53] D. H. Hubel and T. N. Wiesel. "Receptive fields of single neurones in the cat's striate cortex". In: *The Journal of physiology* 148.3 (1959), pp. 574–591.

[54] M. Hussain, D. Chen, A. Cheng, H. Wei, and D. Stanley. "Change detection from remotely sensed images: From pixel-based to object-based approaches". In: *ISPRS Journal of Photogrammetry and Remote Sensing* 80 (2013), pp. 91–106.

[55] N. Jaccard, T. W. Rogers, E. J. Morton, and L. D. Griffin. "Tackling the X-ray cargo inspection challenge using machine learning". In: *Anomaly Detection and Imaging with X-Rays (ADIX)*. Vol. 9847. International Society for Optics and Photonics. 2016, 98470N.

[56] A. K. Jain. "Data clustering: 50 years beyond K-means". In: *Pattern recognition letters* 31.8 (2010), pp. 651–666.

[57] S. Kolokytha et al. "Improving customs' border control by creating a reference database of cargo inspection X-ray images". In: *Advances in Science, Technology and Engineering Systems* 2 (2017), pp. 60–66.

[58] A. Krizhevsky, I. Sutskever, and G. E. Hinton. "Imagenet classification with deep convolutional neural networks". In: *Advances in neural information processing systems*. 2012, pp. 1097–1105.

[59] M Krückhans, L. Plümer, and D. I. J. Schmittwilken. "Ein Detektor für Ornamente auf Gebäudefassaden auf Basis des" histogram-of-orientedgradients"-Operators". PhD thesis. Master's thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, 2010.

[60] Y. LeCun, Y. Bengio, and G. Hinton. "Deep learning". In: *Nature* 521.7553 (2015), pp. 436–444.

[61] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner. "Gradient-based learning applied to document recognition". In: *Proceedings of the IEEE* 86.11 (1998), pp. 2278–2324.

[62] Y. LeCun et al. "Handwritten digit recognition with a back-propagation network". In: *Advances in neural information processing systems*. 1990, pp. 396–404.

[63] M. Lefebure and L. D. Cohen. "A multiresolution algorithm for signal and image registration". In: *Image Processing, 1997. Proceedings., International Conference on*. Vol. 3. IEEE. 1997, pp. 252–255.

[64] T. Lindeberg. "Scale-space theory: A basic tool for analyzing structures at different scales". In: *Journal of applied statistics* 21.1-2 (1994), pp. 225–270.

[65] C. Lindner, P. A. Bromiley, M. C. Ionita, and T. F. Cootes. "Robust and accurate shape model matching using random forest regression-voting". In: *IEEE transactions on pattern analysis and machine intelligence* 37.9 (2015), pp. 1862–1874.

[66] C. Lindner, S Thiagarajah, J Wilkinson, T. Consortium, G Wallis, and T Cootes. "Fully automatic segmentation of the proximal femur using random forest regression voting". In: *IEEE transactions on medical imaging* 32.8 (2013), pp. 1462–1472.

[67] D. G. Lowe. "Object recognition from local scale-invariant features". In: *Computer vision, 1999. The proceedings of the seventh IEEE international conference on*. Vol. 2. Ieee. 1999, pp. 1150–1157.

[68] J. MacQueen et al. "Some methods for classification and analysis of multivariate observations". In: *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. 14. Oakland, CA, USA. 1967, pp. 281–297.

[69] F. Maes, A. Collignon, D. Vandermeulen, G. Marchal, and P. Suetens. "Multimodality image registration by maximization of mutual information". In: *IEEE transactions on medical imaging* 16.2 (1997), pp. 187–198.

[70] A. Mang and G. Biros. "An Inexact Newton–Krylov Algorithm for Constrained Diffeomorphic Image Registration". In: *SIAM journal on imaging sciences* 8.2 (2015), pp. 1030–1069.

[71] A. Mang and G. Biros. "Constrained Hˆ1-Regularization Schemes for Diffeomorphic Image Registration". In: *SIAM Journal on Imaging Sciences* 9.3 (2016), pp. 1154–1194.

[72] P. Martins, J. F. Henriques, R. Caseiro, and J. Batista. "Bayesian constrained local models revisited". In: *IEEE transactions on pattern analysis and machine intelligence* 38.4 (2016), pp. 704–716.

[73] G. Matheron and J. Serra. "The birth of mathematical morphology". In: *Proc. 6th Intl. Symp. Mathematical Morphology*. Sydney, Australia. 2002, pp. 1–16.

[74] D. Mery. "Computer Vision for X-Ray Testing". In: *Switzerland: Springer International Publishing* (2015).

[75] J. Modersitzki. *FAIR: Flexible Algorithms for Image Registration*. Vol. 6. SIAM, 2009.

[76] J. Modersitzki. *Numerical methods for image registration*. Oxford University Press on Demand, 2004.

[77] *Multimodal image fusion software for MRI CT and PET*. `http://www.medicalexpo.com/prod/hermes-medical-solutions-inc/product-100595-677505.html`. Accessed: 2017-09-26.

[78] J. Nocedal and S. Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

[79] T. Ojala, M. Pietikainen, and T. Maenpaa. "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns". In: *IEEE Transactions on pattern analysis and machine intelligence* 24.7 (2002), pp. 971–987.

[80] F. Pacifici. *Very High Spatial Resolution Imagery for Urban Application*. `https://earthzine.org/2011/06/27/very-high-spatial-resolution-imagery-for-urban-applications/`. Accessed: 2018-04-26.

[81] *Pictured: Huge weapons haul seized by Spanish police - BBC News*. `http://www.bbc.com/news/world-europe-39264664`. Accessed: 2018-03-21.

[82] J. C. Platt, N. Cristianini, and J. Shawe-Taylor. "Large margin DAGs for multiclass classification". In: *Advances in neural information processing systems*. 2000, pp. 547–553.

[83] R. Prevost. "Méthodes variationnelles pour la segmentation d'images à partir de modèles: applications en imagerie médicale". PhD thesis. Paris 9, 2013.

[84] S. J. Prince. *Computer vision: models, learning, and inference*. Cambridge University Press, 2012.

[85] R. J. Radke, S. Andra, O. Al-Kofahi, and B. Roysam. "Image change detection algorithms: a systematic survey". In: *IEEE transactions on image processing* 14.3 (2005), pp. 294–307.

[86] W. Rawat and Z. Wang. "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review". In: *Neural Computation* (2017).

[87] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi. "You only look once: Unified, real-time object detection". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016, pp. 779–788.

[88] J. Redmon and A. Farhadi. "YOLO9000: Better, Faster, Stronger". In: *arXiv preprint arXiv:1612.08242* (2016).

[89] S. Ren, K. He, R. Girshick, and J. Sun. "Faster R-CNN: Towards real-time object detection with region proposal networks". In: *Advances in neural information processing systems*. 2015, pp. 91–99.

[90] T. W. Rogers, N. Jaccard, E. J. Morton, and L. D. Griffin. "Automated X-ray image analysis for cargo security: Critical review and future promise". In: *Journal of X-ray science and technology* 25.1 (2017), pp. 33–56.

[91] T. Rohlfing, C. R. Maurer, D. A. Bluemke, and M. A. Jacobs. "Volume-preserving nonrigid registration of MR breast images using free-form deformation with an incompressibility constraint". In: *IEEE transactions on medical imaging* 22.6 (2003), pp. 730–741.

[92] E. Rosten, R. Porter, and T. Drummond. "Faster and better: A machine learning approach to corner detection". In: *IEEE transactions on pattern analysis and machine intelligence* 32.1 (2010), pp. 105–119.

[93] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. "ORB: An efficient alternative to SIFT or SURF". In: *Computer Vision (ICCV), 2011 IEEE international conference on*. IEEE. 2011, pp. 2564–2571.

[94] D. Rueckert and P Aljabar. "Non-rigid registration using free-form deformations". In: *Handbook of Biomedical Imaging*. Springer, 2015, pp. 277–294.

[95] D. Rueckert, L. I. Sonoda, C. Hayes, D. L. Hill, M. O. Leach, and D. J. Hawkes. "Nonrigid registration using free-form deformations: application to breast MR images". In: *IEEE transactions on medical imaging* 18.8 (1999), pp. 712–721.

[96] J. Rühaak, L. König, F. Tramnitzke, H. Köstler, and J. Modersitzki. "A matrix-free approach to efficient affine-linear image registration on CPU and GPU". In: *Journal of Real-Time Image Processing* 13.1 (2017), pp. 205–225.

[97] T. Salimans, I. Goodfellow, W. Zaremba, V. Cheung, A. Radford, and X. Chen. "Improved techniques for training gans". In: *Advances in Neural Information Processing Systems*. 2016, pp. 2234–2242.

[98] M. Y. Sallam and K. W. Bowyer. "Registration and difference analysis of corresponding mammogram images". In: *Medical image analysis* 3.2 (1999), pp. 103–118.

[99] B. Schölkopf and A. J. Smola. *Learning with kernels: support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[100] N. Shapovalova, C. Fernández, F. X. Roca, and J. Gonzàlez. "Semantics of human behavior in image sequences". In: *Computer Analysis of Human Behavior*. Springer, 2011, pp. 151–182.

[101] A. Sharif Razavian, H. Azizpour, J. Sullivan, and S. Carlsson. "CNN features off-the-shelf: an astounding baseline for recognition". In: *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*. 2014, pp. 806–813.

[102] A. Sotiras, C. Davatzikos, and N. Paragios. "Deformable medical image registration: A survey". In: *IEEE transactions on medical imaging* 32.7 (2013), pp. 1153–1190.

[103] S. Srinivas, R. K. Sarvadevabhatla, K. R. Mopuri, N. Prabhu, S. S. Kruthiventi, and R. V. Babu. "A taxonomy of deep convolutional neural nets for computer vision". In: *arXiv preprint arXiv:1601.06615* (2016).

[104] L. H. Staib and J. S. Duncan. "Boundary finding with parametrically deformable models". In: *IEEE transactions on pattern analysis and machine intelligence* 14.11 (1992), pp. 1061–1075.

[105] M. B. Stegmann. "Active appearance models: Theory, extensions and cases". In: *Informatics and Mathematical Modelling* (2000), p. 262.

[106] I. Steinwart and A. Christmann. *Support vector machines*. Springer Science & Business Media, 2008.

[107] C. Studholme, D. L. Hill, and D. J. Hawkes. "An overlap invariant entropy measure of 3D medical image alignment". In: *Pattern recognition* 32.1 (1999), pp. 71–86.

[108] F. Suard, A. Rakotomamonjy, A. Bensrhair, and A. Broggi. "Pedestrian detection using infrared images and histograms of oriented gradients". In: *Intelligent Vehicles Symposium, 2006 IEEE*. IEEE. 2006, pp. 206–212.

[109] A. Sundaresan, P. K. Varshney, and M. K. Arora. "Robustness of change detection algorithms in the presence of registration errors". In: *Photogrammetric Engineering & Remote Sensing* 73.4 (2007), pp. 375–383.

[110] R. Szeliski. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

[111] B. Tang, G. Sapiro, and V. Caselles. "Direction diffusion". In: *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*. Vol. 2. IEEE. 1999, pp. 1245–1252.

[112] *Terrorisme: les douaniers alertent sur la «destruction» de leurs effectifs*. `http://www.lefigaro.fr/emploi/2015/08/25/09005-20150825ARTFIG00287-terrorisme-les-douaniers-alertent-sur-la-destruction-de-leurs-effectifs.php`. Accessed: 2018-03-20.

[113] D. Terzopoulos, A. Witkin, and M. Kass. "Constraints on deformable models: Recovering 3D shape and nonrigid motion". In: *Artificial intelligence* 36.1 (1988), pp. 91–123.

[114] P. Thévenaz, T. Blu, and M. Unser. "Image interpolation and resampling". In: *Handbook of medical imaging, processing and analysis* 1.1 (2000), pp. 393–420.

[115] J.-P. Thirion. "Image matching as a diffusion process: an analogy with Maxwell's demons". In: *Medical image analysis* 2.3 (1998), pp. 243–260.

[116] P. H. Torr and A. Zisserman. "MLESAC: A new robust estimator with application to estimating image geometry". In: *Computer Vision and Image Understanding* 78.1 (2000), pp. 138–156.

[117] D. Tschumperle and R. Deriche. "Vector-valued image regularization with PDEs: A common framework for different applications". In: *IEEE transactions on pattern analysis and machine intelligence* 27.4 (2005), pp. 506–517.

[118] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. Paragios. "Simultaneous registration and change detection in multitemporal, very high resolution remote sensing data". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*. 2015, pp. 61–69.

[119] B. Van Ginneken, A. F. Frangi, J. J. Staal, B. M. ter Haar Romeny, and M. A. Viergever. "Active shape model segmentation with optimal features". In: *IEEE transactions on medical imaging* 21.8 (2002), pp. 924–933.

[120] V. N. Vapnik and S. Kotz. *Estimation of dependences based on empirical data*. Vol. 40. Springer-Verlag New York, 1982.

[121] V. N. Vapnik and V. Vapnik. *Statistical learning theory*. Vol. 1. Wiley New York, 1998.

[122] T. Vercauteren, X. Pennec, A. Perchant, and N. Ayache. "Diffeomorphic demons: Efficient non-parametric image registration". In: *NeuroImage* 45.1 (2009), S61–S72.

[123] M. A. Viergever, J. A. Maintz, S. Klein, K. Murphy, M. Staring, and J. P. Pluim. "A survey of medical image registration–under review". In: *Medical image analysis* 33 (2016), pp. 140–144.

[124] P. Viola and W. M. Wells III. "Alignment by maximization of mutual information". In: *International journal of computer vision* 24.2 (1997), pp. 137–154.

[125] C. R. Vogel. *Computational methods for inverse problems*. SIAM, 2002.

[126] C. Vondrick, A. Khosla, T. Malisiewicz, and A. Torralba. "Hoggles: Visualizing object detection features". In: *Computer Vision (ICCV), 2013 IEEE International Conference on*. IEEE. 2013, pp. 1–8.

[127] Z. Wang and X. Xue. "Multi-class support vector machine". In: *Support Vector Machines Applications*. Springer, 2014, pp. 23–48.

[128] *WCO Illicit Trade Report 2016*. `http://www.wcoomd.org/-/media/wco/public/global/pdf/topics/enforcement-and-compliance/activities-and-programmes/illicit-trade-report/itr-2016-en.pdf?la=en`. Accessed: 2018-03-25.

[129] S. Wold, K. Esbensen, and P. Geladi. "Principal component analysis". In: *Chemometrics and intelligent laboratory systems* 2.1-3 (1987), pp. 37–52.

[130] J. Yosinski, J. Clune, Y. Bengio, and H. Lipson. "How transferable are features in deep neural networks?" In: *Advances in neural information processing systems*. 2014, pp. 3320–3328.

[131] L. Younes. *Invariance, déformations et reconnaissance de formes*. Vol. 44. Springer Science & Business Media, 2003.

[132] A. L. Yuille, P. W. Hallinan, and D. S. Cohen. "Feature extraction from faces using deformable templates". In: *International journal of computer vision* 8.2 (1992), pp. 99–111.

[133] M. D. Zeiler and R. Fergus. "Visualizing and understanding convolutional networks". In: *European conference on computer vision*. Springer. 2014, pp. 818–833.

[134] J. Zhang, M. Marszałek, S. Lazebnik, and C. Schmid. "Local features and kernels for classification of texture and object categories: A comprehensive study". In: *International journal of computer vision* 73.2 (2007), pp. 213–238.

[135] Q. Zhu, M.-C. Yeh, K.-T. Cheng, and S. Avidan. "Fast human detection using a cascade of histograms of oriented gradients". In: *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on*. Vol. 2. IEEE. 2006, pp. 1491–1498.

Résumé

La société contemporaine fait face à un niveau de menace sans précédent depuis la seconde guerre mondiale. La lutte contre le trafic illicite mobilise aussi l'ensemble des organes de police, visant à endiguer le financement du crime organisé. Dans cet effort, les autorités s'engagent à employer des moyens de plus en plus modernes, afin notamment d'automatiser les processus d'inspection.

L'objectif de cette étude est de développer des outils de vision par ordinateur afin d'assister les officiers de douanes dans la détection d'armes et de narcotiques. Le travail présenté examine l'emploi de techniques avancées de classification et de recalage d'images pour l'identification d'irrégularités dans des acquisitions radiographiques de fret.

Plutôt que de recourir à la reconnaissance par apprentissage, nos méthodes revêtent un intérêt particulier lorsque les objets ciblés présentent des caractéristiques visuelles variées. De plus, elles augmentent notablement la détectabilité d'éléments cachés dans des zones denses, là où même les algorithmes de reconnaissance n'identifieraient pas d'anomalie. Nos travaux détaillent l'état de l'art des méthodes de classification et de recalage, explorant aussi diverses pistes de résolution. Les algorithmes sont testés sur d'importantes bases de données pour apprécier visuellement et numériquement leurs performances.

Mots Clés

Analyse d'Images, Images Radiographiques, Détection de Menaces, Recalage d'Images, Classification, Méthodes de Minimisation d'Énergie, Méthodes Variationnelles

Abstract

Our societies, faced with an unprecedented level of security threat since WWII, must provide fast and adaptable solutions to cope with a new kind of menace. Illicit trade also, often correlated with criminal actions, is viewed as a defining stake by governments and agencies. Enforcement authorities are thus very demanding in terms of technological features, as they explicitly aim at automating inspection processes.

The main objective of our research is to develop assisting tools to detect weapons and narcotics for law-enforcement officers. In the present work, we intend to employ and customize both advanced classification and image registration techniques for irregularity detection in X-ray cargo screening scans.

Rather than employing machine-learning recognition techniques, our methods prove to be very efficient while targeting a very diverse type of threats from which no specific features can be extracted. Moreover, the proposed techniques significantly enhance the detection capabilities for law-enforcement officers, particularly in dense regions where both humans or trained learning models would probably fail. Our work reviews state-of-the art methods in terms of classification and image registration. Various numerical solutions are also explored. The proposed algorithms are tested on a very large number of images, showing their necessity and performances both visually and numerically.

Keywords

Image Analysis, X-ray Images, Threat Detection, Image Registration, Classification, Energy-Minimization Methods, Variational Methods