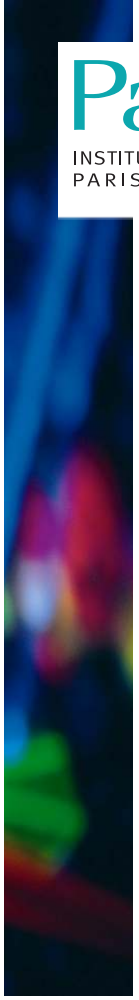


ParisTech
INSTITUT DES SCIENCES ET TECHNOLOGIES
PARIS INSTITUTE OF TECHNOLOGY

TELECOM ParisTech

2018-ENST-0020

EDITE - ED 130

Doctorat ParisTech

T H È S E

pour obtenir le grade de docteur délivré par

TELECOM ParisTech

Spécialité Signal et Images

présentée et soutenue publiquement par

Cristian Felipe OCAMPO-BLANDÓN

le 12 Avril 2018

Fusion d'Images par Patchs pour la Photographie Computationnelle

Patch-Based Image Fusion for Computational Photography

Jury

M. Charles KERVRANN, Directeur de recherche, INRIA	Rapporteur
M. Nicolas PAPADAKIS, Chargé de recherche, CNRS - Université Bordeaux 1	Rapporteur
Mme. Agnès DESOLNEUX, Directrice de recherche, CNRS - ENS Paris Saclay	Examinatrice
M. Frédéric DUFAUX, Directeur de recherche, CNRS - CentraleSupelec - UPSud	Examinateur
M. Wolf HAUSER, Ingénieur de recherche, DXO Labs	Invité
M. Yann GOUSSEAU, Professeur, Télécom ParisTech	Directeur
M. Saïd LADJAL, Maître de Conférences, Télécom ParisTech	Directeur

T H È S E

TELECOM ParisTech
école de l'Institut Mines-Télécom - membre de ParisTech
46 rue Barrault 75013 Paris - (+33) 1 45 81 77 77 - www.telecom-paristech.fr

A mi querida familia.

Jose Edgar y *Flor Lucero*.
Juanda, Loise, Valentina y *Vanessa*.
Majo, Sofi y *Juanfe*.

Table of contents

Résumé (Français)

Au cours des dernières années, de multiples avancées technologiques ainsi que la miniaturisation des circuits, ont été les principaux facteurs de la croissance soutenue de la photographie numérique. De nos jours, il est courant d'avoir accès à des appareils photo numériques et la possibilité d'enregistrer les événements de la vie quotidienne est un fait acquis. Tout le monde peut être photographe et aucune notion importante n'est nécessaire pour générer des images de bonne qualité.
Cette tendance est générée soit par l'ambition des fabricants, les besoins des consommateurs ou la curiosité des scientifiques, elle promeut constamment les standards de qualité de l'image au moyen d'augmentations de résolution, d'amélioration de couleur, de contraste ou de détail.

Une alternative au renforcement de la caméra est d'étendre les limites physiques de l'appareil photo par le développement d'algorithmes plus intelligents. La collection de ces algorithmes intelligents définit ce que l'on appelle généralement comme la *Photographie Computationnelle*. Parmi ces techniques, l'amélioration de l'image concerne les outils de calcul qui améliorent le contenu des images, lorsque l'amélioration porte sur une ou plusieurs caractéristiques de l'image. Ils se spécialisent dans des tâches telles que l'amélioration des couleurs, la normalisation de l'éclairage, la réduction du bruit, l'amélioration de mise au point et des détails, etc. Par conséquent, ces techniques étendent les capacités des appareils photo numériques conventionnels.

Le but de cette thèse est de développer des techniques basées sur des patchs pour la photographie numérique. Ici, nous nous concentrons particulièrement sur deux applications: *Fusion d'images Multi-exposition* et *Fusion d'images Multifocus* et leur extension aux paramètres dynamiques. A partir d'une pile d'images acquises avec différents réglages, la première méthode vise à capturer toute la gamme dynamique d'une scène, tandis que la seconde vise à produire une image nette partout.

Motivation

Le premier problème que nous nous intéressons à résoudre se pose avec de fortes variations spatiales dans l'éclairage. L'éclairage est un facteur critique lors de l'acquisition d'images. Sur des scénarios non contrôlés, il peut présenter des variations qui modifient la visibilité et le contraste des objets et, par conséquent, provoquent un rendu imparfait des images par le capteur. Pour les scènes à très haute dynamique, l'éclairage peut atteindre 20 stops, ce qui surpasse les capacités des capteurs commerciaux modernes limités à environ 15 stops [1]. Même si les caméras spécialisées sont plus robustes à une dynamique plus grande, en intégrant des circuits plus puissants, elles sont chères et la plupart du temps inaccessibles. En ce qui concerne les caméras ordinaires, la façon typique de surmonter cette limitation est d'utiliser plusieurs captures d'une scène.

Pour l'acquisition de scènes à haute dynamique, deux courants principaux sont apparus: Imagerie HDR (High Dynamic Range) et fusion d'exposition.

L'imagerie HDR est une technique qui a été popularisée depuis le milieu des années 1990 [32], et est capable de rendre des images vives avec des détails abondants sur des environnements qui présentent des forts changements d'illumination, figure 1. Cette approche consiste en deux étapes, la création d'une carte HDR et le Tone Mapping. La carte HDR est une reproduction numérique de l'irradiance des objets de la scène. Il est estimé en utilisant comme entrée plusieurs images qui sont capturées avec des temps d'exposition variables.
L'intention d'utiliser plusieurs images est de compenser les intensités qui saturent le capteur (intensités très lumineuses) ou les objets qui ne sont pas bien éclairés avec les réglages normaux du capteur. Étant donné que les images naturelles présentent souvent de fortes variations de contraste dans le spectre des tons foncés à clairs, la carte HDR possède une nature dynamique élevée non linéaire qui dépasse la gamme dynamique standard des écrans ou des dispositifs d'impression courants. En effet, les nouveaux dispositifs d'affichage HDR sont encore limités, avec une dynamique supportée typique de moins de 10 stops. Pour cette raison, un traitement supplémentaire, appelé Tone Mapping [37], est nécessaire pour équilibrer correctement les intensités de couleurs et obtenir des images standard de 8 bits qui sont facilement visualisables sur des écrans couramment commercialisés.

Au contraire, la Fusion d'Exposition [91] ou Fusion d'images Multi-exposition (MEIF) est une technique beaucoup plus simple qui passe directement à l'étape de fusion sans qu'il soit nécessaire d'estimer la carte HDR. Comme pour l'imagerie HDR, l'algorithme utilise plusieurs images avec différents niveaux d'exposition et suppose que tous les objets apparaissent correctement éclairés dans au moins une des images d'entrée.

(a) 1/1000 (b) 1/500 (c) 1/250

(d) 1/125 (e) 1/60 (f) 1/30

(g) 1/15 (h) 1/8 (i) 1/4

(j) HDR + tone mapping. (k) Fusion d'Exposition.

Fig. 1 **(a-i)**: Pile originale des images d'exposition bracketées et les temps d'exposition correspondants. (j) Tone mappé image avec [42]. (k). Fusion d'Exposition [91]. Images issues de Fattal et al. [42].

Par exemple, vérifiez l'ensemble des images d'exposition bracketées de la figure 1. Pour les expositions de longue durée, les objets très lumineux semblent saturés, tandis que pour les expositions de courte durée, ils présentent de bonnes intensités. Un phénomène similaire se produit pour les objets sombres sur des expositions plus longues. Étonnamment, la Fusion de l'Exposition parvient à utiliser une simple combinaison pondérée pour générer une image acceptable, figure 1(k), qui n'est pas très différente du résultat du HDR, figure 1(j).

Dans cette méthode, la combinaison de la pile d'images est basée sur les composantes laplaciennes de chaque image en inspectant la qualité de leur contenu dans le domaine RGB, comme suit:

$$\mathcal{L}_l(R) = \sum_{i=1}^{N} \mathcal{G}_l(W_i)\mathcal{L}_l(I_i) \ , \tag{1}$$

où $\mathcal{L}_l(I_i)$ est la pyramide Laplacienne [20] de l'image I_i au niveau l et $\mathcal{G}_l(W_i)$ est la pyramide Gaussienne de la carte des poids de qualité W_i au niveau l. L'image fusionnée R est ensuite recomposée à partir de la pyramide Laplacienne qui en résulte.

Malgré des progrès considérables dans ces deux méthodes, l'imagerie HDR et la Fusion d'Exposition, elles présentent encore certaines limites. C'est-à-dire que, malgré son effet réaliste, l'imagerie HDR est connue pour présenter des artefacts qui proviennent principalement de l'étape de Tone Mapping [112, 41]. Parmi eux, les artefacts halo, les inversions de gradient ou les conditions de couleur, comme l'excès irréaliste de détails, sont les plus courants. Sans oublier que la plupart des opérateurs de Tone Mapping dépendent de la scène et que l'étalonnage des paramètres doit être effectué indépendamment pour chaque image afin d'obtenir de meilleurs résultats.

Dans le cas de la fusion d'exposition, la qualité de l'image fusionnée dépend de la façon dont les images d'entrée capturent toute la dynamique de la scène. Cela s'explique en particulier par le fait que la méthode n'utilise que des intensités bien exposées, sans traitement supplémentaire, pour synthétiser les résultats.

Le deuxième problème que nous abordons dans cette thèse vise à rendre une image nette, tout en s'appuyant sur des images de profondeur focale différente. Pour cela, nous utilisons la fusion d'images pour combiner les régions nettes de chaque image, un processus communément connu sous le nom de *Fusion d'images Multifocus* (MFIF). La fusion d'images multifocus ou *Focus Stacking* est une technique classique qui cherche à récupérer et à ajouter des détails aux images qui souffrent de la présence de flou. Un tel flou peut être le résultat de multiples facteurs tels que des techniques d'acquisition inappropriées, des

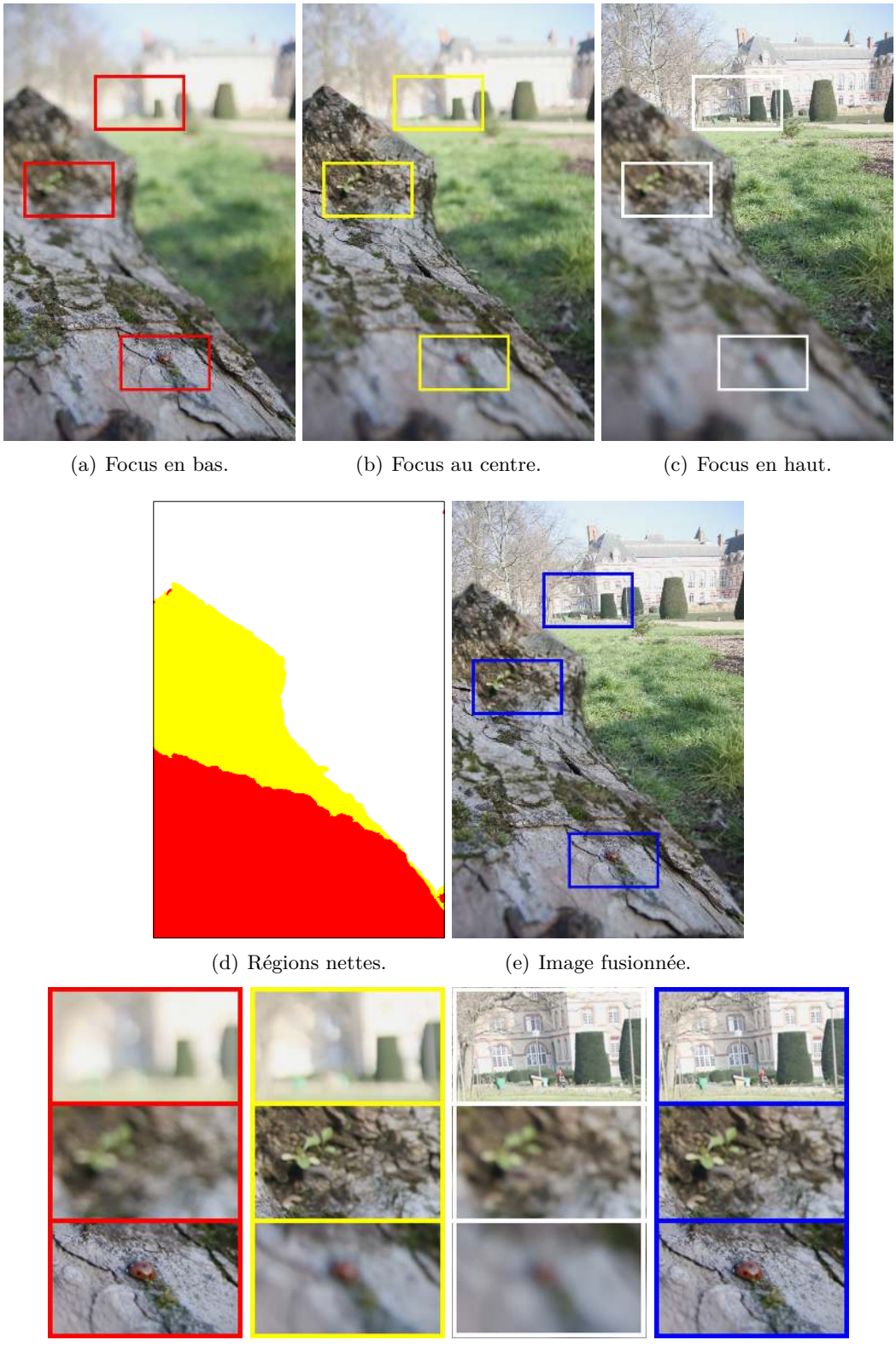
conditions environnementales ou plus probablement à cause d'objets tombant en dehors de la profondeur de champ typique du capteur. On suppose également qu'il apparaît avec un niveau de distorsion variant dans l'espace de l'image.

Suivant un principe similaire à celui de la Fusion d'images Multi-exposition, Focus Stacking prend le meilleur de chaque image pour la composition d'une sortie plus nette. Il synthétise une image en mélangeant de façon transparente toutes les régions qui contiennent plus de détails parmi les images, voir figure 2(a)-2(c). En conséquence, l'image de sortie (Figure 2(e)) affiche plus de détails éparpillés sur son domaine spatial. Étant donné que le processus est soumis au contenu des images d'entrée, une bonne méthode MFIF devrait non seulement être capable d'identifier correctement les régions nettes, mais aussi de les transférer sans réduction de détail.

Ceci est illustré dans la figure 2. La pile originale d'images, figures 2(a)-2(c), contient une quantité excessive de flou, à l'exception des zones localisées qui sont exclusives à une image de l'ensemble. De telles zones nettes peuvent être mieux vues sur la carte dans la figure 2(d), où le rouge, le jaune et le blanc sont les codes de couleur assignés aux régions nettes sur les images 2(a), 2(b) et 2(c), respectivement. Ici, les régions in-focus de la carte sont extraites via la Variation Totale Locale, comme mesure du niveau d'activité, et fusionnées de manière multirésolution avec la pyramide Laplacienne.

D'un point de vue général, la Fusion d'images Multi-exposition (MEIF) et la Fusion d'images Multifocus (MFIF) peuvent être décomposée en deux étapes fondamentales similaires: la sélection de la région candidate et la fusion.

La première tâche concerne l'identification locale des pixels dont la qualité de contraste (pour les applications MEIF) ou la qualité de mise au point (pour les applications MFIF) est meilleure que celle du pixel correspondant dans les images restantes. Dans le premier cas, MEIF, cela est généralement réalisé en utilisant des mesures qui répondent proportionnellement à l'éclairage, comme par exemple une combinaison de Contraste/Saturation et Bien-Exposé (voir [91]). Dans le cas du Focus Stacking, l'evaluation des objets focalisés est généralement réalisé avec des mesures qui analysent l'énergie de variation locale, ce qui est un indicateur raisonnable de netteté. Certaines mesures pour estimer la netteté comprennent la variance, l'énergie du gradient d'image, l'énergie du Laplacien, la fréquence spatiale, etc. Néanmoins, aucun consensus standard n'a été atteint pour une mesure de flou.
La deuxième tâche, la fusion, consiste à joindre de façon transparente toutes les régions marquées avec une bonne qualité, de sorte qu'aucune discontinuité ou artefact n'est produite au cours du processus. Pour la fusion d'images, il est courant de trouver des schémas



(a) Focus en bas. (b) Focus au centre. (c) Focus en haut.

(d) Régions nettes. (e) Image fusionnée.

Fig. 2 (a-c). Pile originale d'images avec différents réglages de mise au point. (d) Carte des régions nettes de l'image. (e) Image fusionnée. **Images en bas**: zones zoomées extraites des images d'entrée (rouge, jaune et blanc) et de l'image fusionnée (bleu).

où des combinaisons pondérées sont utilisées. Les approches multirésolution comme les pyramides Laplaciennes et Gaussiennes, comme dans l'équation (1), et les ondelettes, avec différentes familles d'ondelettes, sont également une alternative populaire.

Les deux applications, MEIF et MFIF, peuvent produire des artefacts associés à des mouvements provenant de déplacements de caméra ou d'objets lors de l'acquisition des images d'entrée, figure 3. Ces effets d'artefacts ou *ghosts*, apparaissent comme des objets semi transparents et émergent en raison de divergences locales pendant la fusion, où différents objets sont combinés.

Certaines solutions pour faire face au mouvement comprennent: l'enregistrement d'images ou simplement l'utilisation de trépieds pendant la prise de vue, ce qui pourrait être ennuyeux à transporter et à éviter les scènes rapides. Alors que de telles méthodes assurent une correspondance globale entre les objets immobiles, les objets en mouvement peuvent encore présenter de grands déplacements et provoquer des artefacts.

Dans le cadre de l'imagerie MEIF et HDR, de multiples contributions ont été proposées

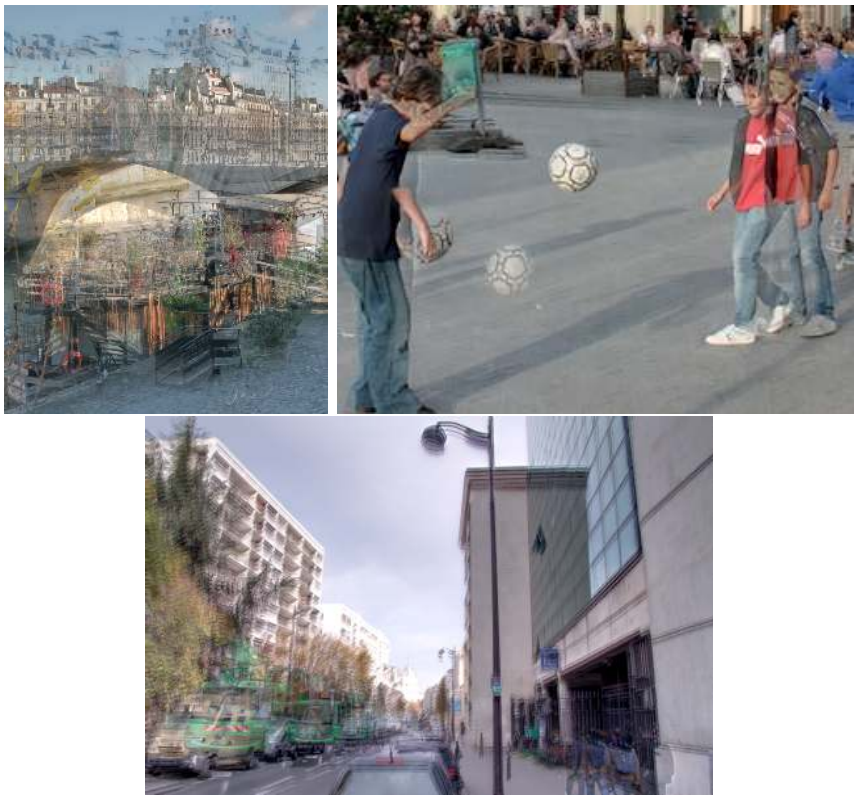


Fig. 3 Exemples d'artefacts "ghosts" générés avec MEIF sur des scénarios dynamiques. Images issues de [3, 102].

pour assurer la robustesse au mouvement. Certaines stratégies comprennent des techniques de déghosting ou d'appariement de similarité optimal.
Quant au MFIF, même s'il a fait l'objet d'études approfondies au cours des dernières décennies, aucune contrainte de mouvement n'a été prise en compte de manière satisfaisante. Ce problème est particulièrement difficile, car une fois les détails sont perdus à cause du flou, il n'y a pas d'informations géométriques à faire correspondre sur d'autres images.

En conclusion, les tâches du MEIF et MFIF sont affectés par deux types de mouvements: les mouvements de caméra et les mouvements d'objets. La technique la plus fiable pour résoudre une variété de transformations est l'enregistrement d'images avec des homographies. Cependant, en raison des distorsions de l'objectif de la caméra ou des conditions d'acquisition (c'est-à-dire si la scène n'est pas faite d'un seul plan et que le centre optique s'est déplacé), l'estimation d'une telle transformation peut donner lieu à des erreurs d'enregistrement. De plus, les mouvements d'objets ne sont pas résolus.

Cette thèse examine comment traiter le mouvement dans des conditions physiques variables comme les éclairages, la profondeur de champ et les transformations géométriques.

Problème de la thèse

Dans les réglages statiques, MEIF et MFIF peuvent s'appuyer sur l'algorithme standard en deux étapes (sélection de la région candidate et fusion) afin d'obtenir des résultats satisfaisants. En effet, la caméra et les objets conservent des positions spatiales identiques parmi les images, ce qui permet une extraction fiable des mesures de qualité locales.

Sur les réglages dynamiques, cependant, le problème est plus complexe car les objets peuvent changer de position ou de géométrie, et on s'attend à ce qu'ils changent le contraste et la mise au point. Ces facteurs font qu'il n'est non viable de s'appuyer sur la procédure standard, qui génère des artefacts sur les conditions susmentionnées. En fait, les artefacts dus au mouvement ne seraient pas un problème si nous savions à l'avance où les objets correspondants sont localisés après les déplacements, ou différemment, où extraire les mesures de qualité qui viennent des objets correspondants parmi les images.

Au lieu de cela, nous nous appuyons sur un ***alignement géométrique*** pour forcer les objets à partager la même position parmi les images, produisant des images qui sont parfaitement enregistrées. Pour cela, nous avons besoin de localiser les objets avec une précision de pixel sur les images, un problème que l'on appelle la dense correspondence (pour des images avec les mêmes conditions d'illumination et de mise au point).

Pour la localisation d'objets, nous partons de l'hypothèse que, même si le contenu des images a changé dans l'espace, le contenu local interne des objets reste identique ou partiellement inchangé sur les images. Un exemple peut être vu dans les figures 4, où le ballon de football présente des changements spatiaux entre les images sans déformation interne substantielle. Ceci s'applique également aux structures de plus petite échelle, qui contiennent suffisamment d'informations uniques pour être reconnues avec précision sur les images, comme par exemple, les chaussures ou les mains des enfants. C'est pour cette raison que nous travaillons avec de petites portions d'images ou des voisinages locaux, désormais appelés *patches*, à rechercher sur les images. En particulier, nous utilisons des mesures qui répondent proportionnellement au contenu de similarité des patchs, dans un contexte qui devrait être significativement robuste à l'éclairage ou au flou.

(a) Sous-exposée. (b) Normal exposé.

(c) Sur-exposée.

Fig. 4 (a-c). Images d'exposition bracketées avec des objets en mouvement. Images issues de [102].

Pour MEIF et MFIF, la localisation d'objets en mouvement peut être formulée comme un problème d'optimisation où nous essayons de minimiser l'erreur entre tous les patchs possibles d'une image et leurs candidats potentiels dans d'autres images, comme suit :

$$f(x) = \underset{y \in \Omega_S}{\arg\min} \, \|P_R(x) - \mathcal{T}_R\{P_S(y)\}\|^2, \tag{2}$$

ici $P_R(x)$ et $P_S(y)$ sont les patches centrés à la position x et y, à partir d'une image perturbée R et d'une image S, respectivement. Ω_S est le domaine spatial de l'image S, et l'opérateur $\mathcal{T}_R\{S\}$ est une transformation de qualité qui s'applique à l'image S qui calibre le type de dégradation, soit l'exposition ou le flou, à celui de l'image R. En d'autres termes, cette transformation renforce la ressemblance de patchs entre les images, ce qui permet une invariance à la perturbation. Il convient de noter que la nature de la transformation est différente pour chaque cas, soit du MEIF ou du MFIF. Dans la pratique, l'estimation de la transformation est un problème complexe en raison de variables inconnues pendant l'acquisition, comme les changements d'illumination non linéaires ou le flou. Il est donc estimé différemment et peut être appliqué à un patch ou à l'image entière selon le cas.

En bref, la résolution de l'équation (2) nous aide à trouver des correspondances denses de voisinages locaux dans les images R et S. Par conséquent nous récupérons une carte de coordonnées $f(x)$ où les objets sont identiques ou similaires.
La solution à ce problème est présentée dans les chapitres de cette thèse correspondant à chaque application, où nous décrivons en détail comment traiter des images multiples.

Les avantages de l'utilisation de méthodes basées sur les patchs sont nombreux. Premièrement, les patchs sont hautement redondants, ce qui signifie que des patchs similaires peuvent être trouvés partout dans les images, profitant de l'autosimilarité des images naturelles. Cette propriété s'est montrée très utile pour la réduction du bruit et a été incorporée dans le travail séminal de Buades et al. [16] pour le débruitage d'image. Ici, nous étendons cette approche pour le raffinement des cartes de correspondance d'une manière agrégée. Deuxièmement, grâce à des contenus discriminants locaux, des patchs ayant une géométrie similaire peuvent être trouvés de manière cohérente sur des images qui souffrent de différentes dégradations. Cette propriété est particulièrement utile pour les problèmes d'estimation de mouvement, où les objets communs entre les images peuvent être localisés, ce qui rend les patchs appropriés pour notre objectif d'invariance au mouvement.
Dans l'esprit de la dernière propriété, on pourrait agrandir la taille du patch pour ajouter suffisamment d'informations supplémentaires pour le patch matching sous distorsions de flou, comme nous l'avons fait pour notre deuxième application.

Reconstruction d'images à base de patchs

Dans ce travail, nous utilisons l'algorithme de Patchmatch [10] pour trouver efficacement les correspondances de patch et pour synthétiser une image de référence avec le contenu d'une image supplémentaire, un processus que nous appelons reconstruction d'image. Ici, nous montrons que l'utilisation de patches résout les désalignements géométriques locaux pour des images avec des conditions d'acquisition similaires, et expliquons pourquoi l'extraction standard du voisin le plus proche s'avère inadéquate sous des changements d'éclairages ou des perturbations de flou.

Étant donné les images $R : \Omega_R \to \mathbb{R}^3$ et $S : \Omega_S \to \mathbb{R}^3$, pour référence et source, respectivement. Supposons qu'un champ du voisin le plus proche ou une carte de déplacement $M : \Omega_R \to \Omega_S$ est extrait (ici avec Patchmatch [10]), où pour chaque patch dans R la carte renvoie la coordonnée du voisin le plus proche dans S, donc la reconstruction est comme suit:

$$\widetilde{R}(p) = S(M(p)) \ , \tag{3}$$

où $p \in \Omega_R \subset \mathbb{Z}^2$ et $M(p) \in \Omega_S \subset \mathbb{Z}^2$. Fondamentalement, chaque pixel sur l'image $\widetilde{R}$ est synthétisé avec seulement le pixel lié à la position $M(p)$ dans la source.

L'algorithme de Patchmatch

L'algorithme de patchmatch [10] est une méthode très efficace pour l'extraction des voisins les plus proches (ou ici désigné sous le nom de cartes de déplacement). Initialement proposé pour l'estimation des correspondances denses, il est utilisé pour l'édition d'images, *completion* ou *retargeting*. L'algorithme commence par une initialisation aléatoire, il fait une correspondance automatique entre tous les quartiers possibles de l'image de référence et le patch le plus similaire de l'image source. Ceci est fait en effectuant simplement deux opérations consécutives qui réduisent une énergie de similarité (définie à l'origine par la norme L^2-norm) après chaque cycle.

Plus précisément, l'algorithme Patchmatch vise à résoudre le problème suivant:

$$\underset{M(x)\in\Omega_S}{\text{minimize}} \ E(M) = \sum_{x\in\Omega_R} D(P_R(x), P_S(M(x))) \ , \tag{4}$$

avec quelques contraintes sur $M(x)$, comme $M(x)$ n'étant pas trop loin de x. Ici, Ω_R et Ω_S sont les domaines spatiaux des images R et S, respectivement, et $M(x)$ est la carte des correspondances (NNF) au pixel x. La distance D est habituellement calculée comme la distance L^2 entre les patches $P_R(x)$ et $P_S(y)$ centrés à x et y.

En bref, les étapes fondamentales de Patchmach sont: l'initialisation, la propagation et la recherche aléatoire des voisins les plus proches.

Patchmatch sous diverses perturbations d'image

Les capacités de reconstruction de Patchmatch ont été étudiées sur des cas où les images subissent diverses perturbations réelles. En particulier, l'algorithme est évalué sur des images prises à partir d'une même scène avec une caméra tenue à la main et sous différents changements, tels que: translations, rotations, illumination et variations de mise au point.

Pour les expériences, nous avons utilisé plusieurs configurations de l'algorithme Patchmatch: standard, basé sur les descripteurs et la version généralisée. Ci-dessous, nous ne présentons que les résultats avec la version standard. Pour la description complète des expériences, voir la Section 3.2.

La première expérience concerne les images en présence de perturbations de **translation** avec des changements d'éclairage inaperçus. Des objets en mouvement, dont certains varient en échelle, sont également présents. Remarquez que l'image source, figure 5(b), est déplacée vers la droite.

(a) Image de référence. (b) Image source.

Fig. 5 Ensemble d'images d'entrée avec translation visible vers la droite.

Notez que cette version standard, figure 6, construit une carte de déplacement qui n'est pas très cohérente, mais qui est capable de rendre avec précision les objets au niveau de détail, ainsi que les objets qui n'apparaissent pas sur la source. Ceci est dû au fait que l'algorithme associe des patches qui n'appartiennent pas nécessairement au même objet dans les deux images, mais plutôt à d'autres régions similaires ou répétées qui sont

dispersées dans l'image.

(a) Reconstruction. PSNR = 26.94. (b) NNF carte.

Fig. 6 Reconstruction d'image et carte des correspondances pour l'image perturbée par une translation.

Même si la précision des détails est très bonne pour l'algorithme de Patchmatch, l'image reconstruite (figure 6(a)) souffre d'artefacts visibles *jitter* qui sont réduits avec plusieurs techniques de raffinement (décrites dans la Section 3.3), voir figure 7(b).

(a) Patch Aggr., PSNR = 26.53. (b) Multi-NNF Aggr., PSNR = 32.97.

Fig. 7 Raffinement de la reconstruction avec différentes techniques, (voir Section 3.3).

Une autre situation réelle qui se produit en tenant la caméra pendant l'acquisition est celle des **rotations**. Ici, l'ensemble des photographies (Figure 8) présente une légère rotation dans le sens inverse des aiguilles d'une montre. Il en résulte une perte de structures à l'intérieur de l'image source. De plus, le scénario est très dynamique en arrière-plan.

D'après la reconstruction et la carte des correspondances, figure 9, nous pouvons voir que la version standard produit une très bonne reconstruction. Une fois de plus, l'algorithme est capable de synthétiser une image cohérente avec la référence, même en présence de

(a) Image de référence. (b) Image source

Fig. 8 Ensemble d'images d'entrée avec une rotation.

rotations ou de structures inconnues et d'objets en mouvement. Comme auparavant, l'information est échantillonnée dans plusieurs régions sans cohérence apparente.

(a) Reconstruction. PSNR = 25.60 (b) NNF carte.

(c) Patch Aggr., PSNR = 26.07. (d) Multi-NNF Aggr., PSNR = 31.72.

Fig. 9 Reconstruction d'image et carte des correspondances pour une image avec rotation et (c)-(d) raffinement de la reconstruction avec différentes techniques, (voir Section 3.3).

La troisième expérience comprenait un ensemble d'images qui contiennent principalement des **changements d'illumination**, figure 10. Certains changements géométriques, comme des objets en mouvement, sont également présents.

(a) Image de référence. (b) Image source

(c) Référence normalisée.

Fig. 10 (a-b) Images d'exposition bracketées acquises avec 2 stops de différence. (c) Référence normalisée de couleur en utilisant la spécification d'histogramme. (Section 4.2.1). Images issues de [102].

L'image de reconstruction et la carte des correspondances NNF sont montrées dans la figure 11. En cas de changement de l'éclairage, la version standard échoue fortement. Ceci est attendu parce que le SSD est perdu lorsqu'il est confronté à de fortes variations radiométriques, où des patchs géométriquement identiques mais avec des variations de contraste ne sont pas reconnus comme exacts. Par conséquent, la version standard de Patchmatch ne peut pas reproduire les structures.

Une option pour faire face au changement d'illumination est de normaliser la radiométrie des deux images et de produire des images plus proches en apparence. La figure 10(c), montre la radiométrie de l'image de référence normalisée à la radiométrie de la source via la spécification de l'histogramme. En utilisant l'image couleur normalisée, nous montrons

(a) Reconstruction. PSNR = 22.93.

(b) NNF carte.

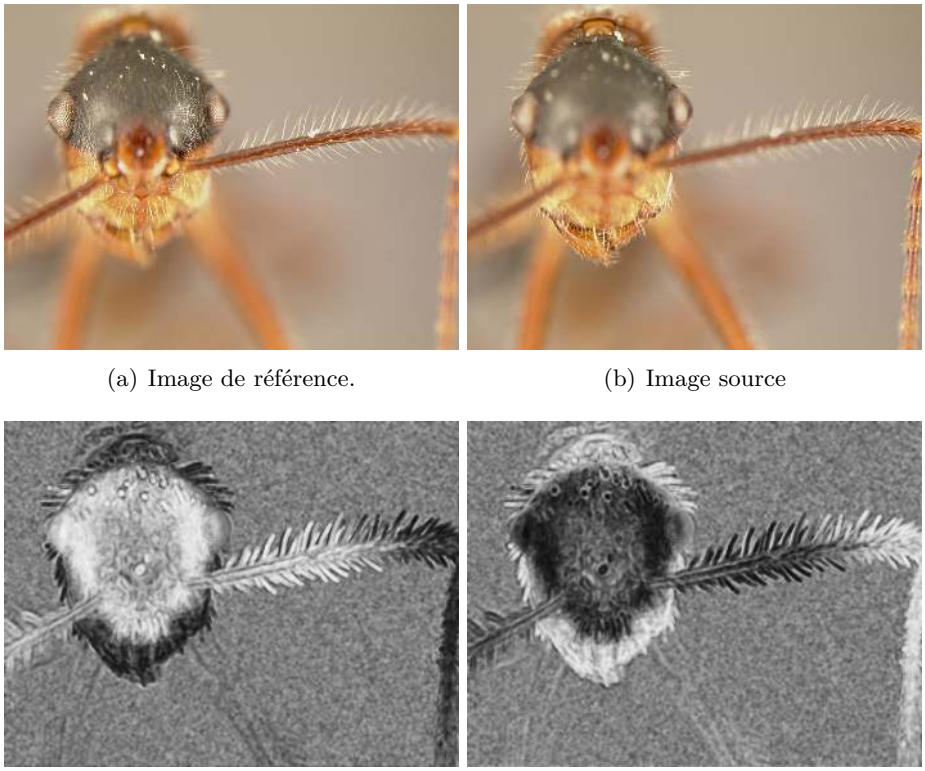
(c) Reconstruction apres specification. PSNR = 27.37.

(d) NNF carte apres specification.

Fig. 11 Reconstruction d'image et carte des correspondances pour une image avec des changements d'illumination.

que la reconstruction est bien meilleure. Pour ce cas, figure 11(c), l'algorithme synthétise une image qui est globalement cohérente avec la référence et l'apparence est transférée avec succès depuis la source.

La dernière expérience concerne le patch matching pour les images qui présentent des **changements de mise au point**. Ici nous utilisons une paire d'images parfaitement alignées avec des niveaux de netteté complémentaires, figure 12. Comme précédemment, l'objectif est de reconstruire la référence à partir du contenu de l'image source.
Comme les images ne présentent pas de déformations géométriques, de translation ou de rotation, nous nous attendons à ce que les meilleures correspondances soient localisées au même lieu, produisant ainsi une carte d'identité. Néanmoins, les résultats, tant pour la reconstruction que pour les NNF, présentent de fortes incohérences. Ces irrégularités sont dues au fait que la mesure de similarité n'associe pas des objets identiques ayant un contenu en fréquence différent.



(a) Image de référence. (b) Image source

(c) Netteté dans la référence. (d) Netteté dans la source

Fig. 12 (a-b) Image de référence et image source sous des niveaux de flou complémentaires. (c-d) Valeurs de netteté normalisées extraites avec la variation totale locale, la netteté est proportionnelle à la luminosité.

En cas de changement de mise au point, la reconstruction avec l'algorithme Patchmatch, figure 13, présente une performance limitée. Cependant, il présente de fortes irrégularités au niveau des pixels et le transfert de netteté de la source n'est pas obtenu.

(a) Reconstruction. (b) NNF Carte.

Fig. 13 Reconstruction d'image en présence de flou.

Fusion d'images multi-exposition pour des scènes dynamiques

Dans le cadre de Exposure Fusion, nous introduisons une méthodologie basée sur des patchs qui minimise l'influence du mouvement sur la fusion des images avec de fortes variations d'illumination. Cette méthode est utilisée pour décrire un procédure général pour la fusion de l'exposition des images capturées dans des scènes dynamiques. La méthodologie proposée résout les problèmes de mouvement de caméra et d'objet de manière très efficace et peut être utilisée sur des RAW Linear Images ou des images non linéaires de 8 bits sans changement significatif.

Notre algorithme consiste à reconstruire les images d'entrée en utilisant la géométrie de l'image de référence et en conservant leur radiométrie originale. Par conséquent, la solution aux mouvements consiste à déplacer des objets en se basant sur l'estimation d'une carte des déplacements. À cause de cela, nous produisons un nouvel ensemble d'images qui est géométriquement cohérent avec la référence mais radiométriquement cohérent avec chaque image source dans l'ensemble. L'étape finale sera l'utilisation simple de l'algorithme de fusion d'exposition [91]. Toutefois, cette étape peut être remplacée par un autre schéma de fusion.
Plus précisément, notre algorithme comporte cinq étapes: Enregistrement, amélioration de la référence, normalisation radiométrique, recherche & Reconstruction et fusion. Ecrivons $S_1, S_2, \ldots, S_N$ pour les images d'entrée et $R = S_{i_0}$ pour la référence.

- *ETAPE I*: Dans un premier temps, l'ensemble des images est enregistré (pas nécessairement avec une précision exacte) par rapport à la référence, en utilisant une homographie et l'algorithme RANSAC.
- *ETAPE II*: Il est impératif que la référence contienne suffisamment de détails pour guider le processus de reconstruction. Ce n'est pas le cas dans les régions saturées. Pour supprimer les saturations, nous améliorons la référence en utilisant les informations de l'image immédiatement la moins exposée, après avoir vérifié la cohérence des deux images dans ces régions. Ceci produit une référence géométrique améliorée.
- *ETAPE III*: La radiométrie de la référence géométrique améliorée est transformée pour correspondre à celle de chaque source (approximativement alignée). Ensuite, un algorithme de correction de couleur est utilisé pour s'assurer que les images résultantes $C_1, C_2, \ldots, C_N$ ne contiennent pas d'artefacts. Ce traitement donne lieu à une série d'images de référence qui peuvent être facilement comparées indépendamment à chaque image source à laquelle elle correspond.
- *ETAPE IV*: Pour chaque i et pour chaque patch de C_i, une recherche par patch dans S_i est effectuée afin de trouver les meilleurs voisinages correspondants dans les

sources. Ces voisinages nous permettent de reconstruire une image L_i qui a la même géométrie que la référence et partage la radiométrie de S_i.
Une étape de raffinement optionnelle peut être incluse pour améliorer la qualité de la reconstruction.

- *ETAPE V*: L'algorithme de fusion d'images est appliqué à la nouvelle collection d'images $L_1, L_2, \ldots, L_N$.

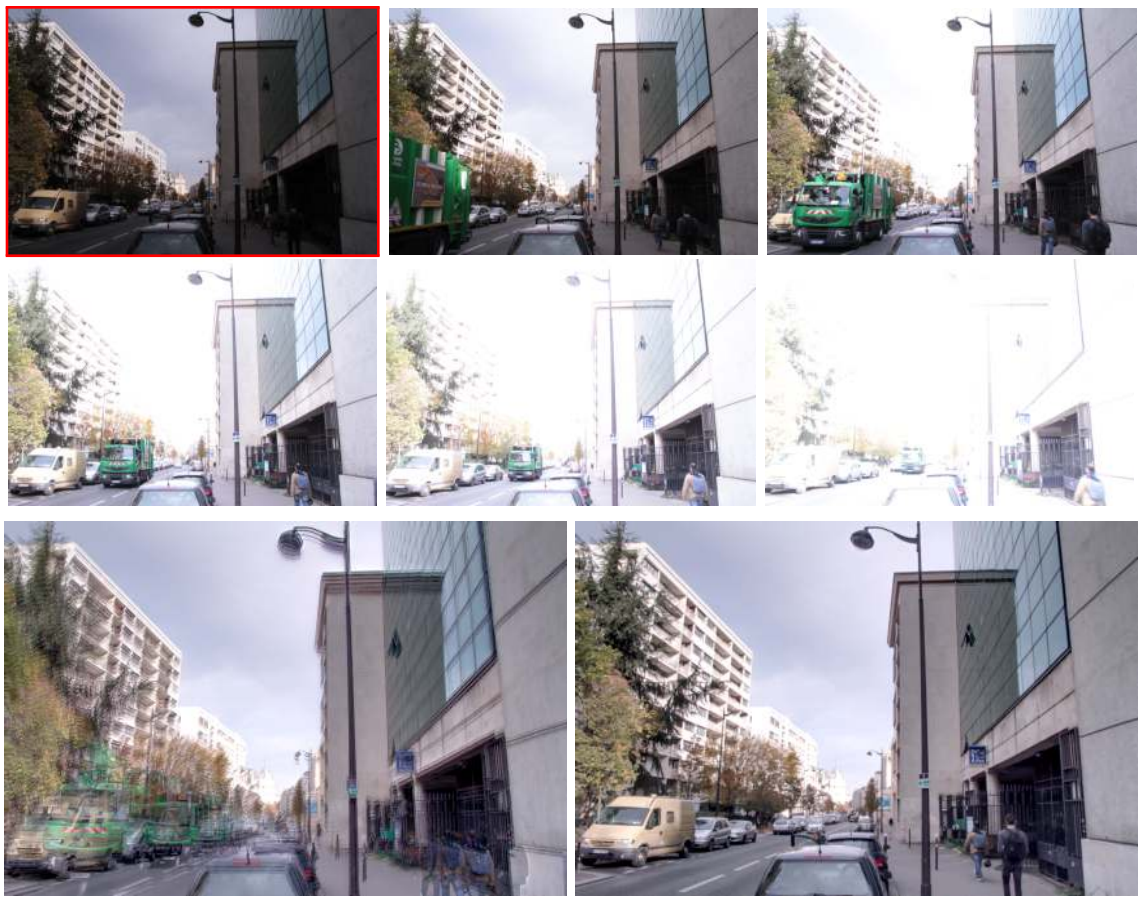
La méthodologie proposée a été évaluée dans plusieurs situations: paramètres statiques, dynamiques et saturations sur l'image de référence. Pour l'expérience sur les scénarios dynamiques, nous avons utilisé plusieurs datasets fournis en ligne et nos propres photographies, contenant différents types de perturbations: motion, occlusion, multiview, etc. Un petit nombre d'expériences est présenté ci-dessous pour le cas statique et dynamique.
Pour une présentation et une analyse complète des expériences, veuillez consulter la Section 4.3.

Dans la figure 14, nous montrons le résultat de la combinaison d'images sur un ensemble d'images *totalement aligné*. Notre méthode est affichée ainsi que Exposure Fusion, nous pouvons voir que le résultat est très proche de celui de Exposure Fusion.

(a) (b)

Fig. 14 Fusion d'images sur des images totalement alignées: (a). Exposure Fusion [91], <1s. (b). Notre méthode, 35s.

La figure 15 montre de grands déplacements d'objets à grande échelle, par exemple le camion vert qui a aussi une taille variable sur les images. La fusion montre une image bien éclairée sans artefacts radiométriques. En particulier, notre méthode fait preuve de plus de cohérence pour reproduire des structures tout en ayant un contraste approprié partout.



(a) (b)

Fig. 15 Fusion d'images sur une scène ayant de grands déplacements de petits et grands objets. **En haut:** Des images d'entrée avec de grands mouvements d'objets. **En bas:** (a) Exposure Fusion et (b) Notre méthode.

Enfin, l'ensemble complet d'expériences (Section 4.3) nous permet de conclure que notre méthode fournit systématiquement une reconstruction appropriée de la référence, même si l'ensemble contient des déformations géométriques significatives ou des changements brusques d'exposition (avec présence de zones saturées ou sombres), étant plus précis avec une version plus sophistiquée de notre algorithme (version *KNN* - voir Chapitre 4). Nous soulignons également que dans le cas statique, notre résultat est presque identique à celui de Exposure Fusion.

Fusion d'images multifocus pour des scènes dynamiques

La procédure standard pour élargir la profondeur de champ limitée des appareils photo numériques est d'utiliser la Fusion d'Images Multifocus (MFIF).
A partir d'une collection d'images acquises avec différents paramètres de mise au point, ces méthodes visent à fusionner le contenu des images de la pile pour produire une image finale qui est nette partout. De telles méthodes peuvent être très efficaces, mais lorsque l'ensemble des images d'entrée ne sont pas alignées avec précision, ou lorsque certains objets sont en mouvement, l'image finale montre des fantômes ou d'autres artefacts.

Dans cette thèse, nous proposons une méthode générique pour surmonter ces limitations. Nous sélectionnons d'abord une image de référence, puis, pour chaque image de la pile, nous reconstruisons une image qui partage la géométrie de la référence et la netteté de l'image d'entrée. La reconstruction est réalisée grâce à une modification spécialement élaborée sur l'algorithme PatchMatch, adaptée aux images floues, et à un post-traitement dédié à la correction des erreurs de reconstruction. Ensuite, à partir de la nouvelle pile d'images, le MFIF est exécuté pour produire un résultat net.

Notre méthode s'inspire des techniques de recherche non locale de patchs, pour résoudre les changements géométriques spatiaux sur les images, ainsi que des mesures de caractéristiques supplémentaires pour réduire l'impact de la dégradation du flou pendant la comparaison de patchs. Pour cela, nous guidons la recherche en utilisant trois caractéristiques distinctives des patches, à savoir (i) la couleur, (ii) la géométrie locale et (iii) la position.

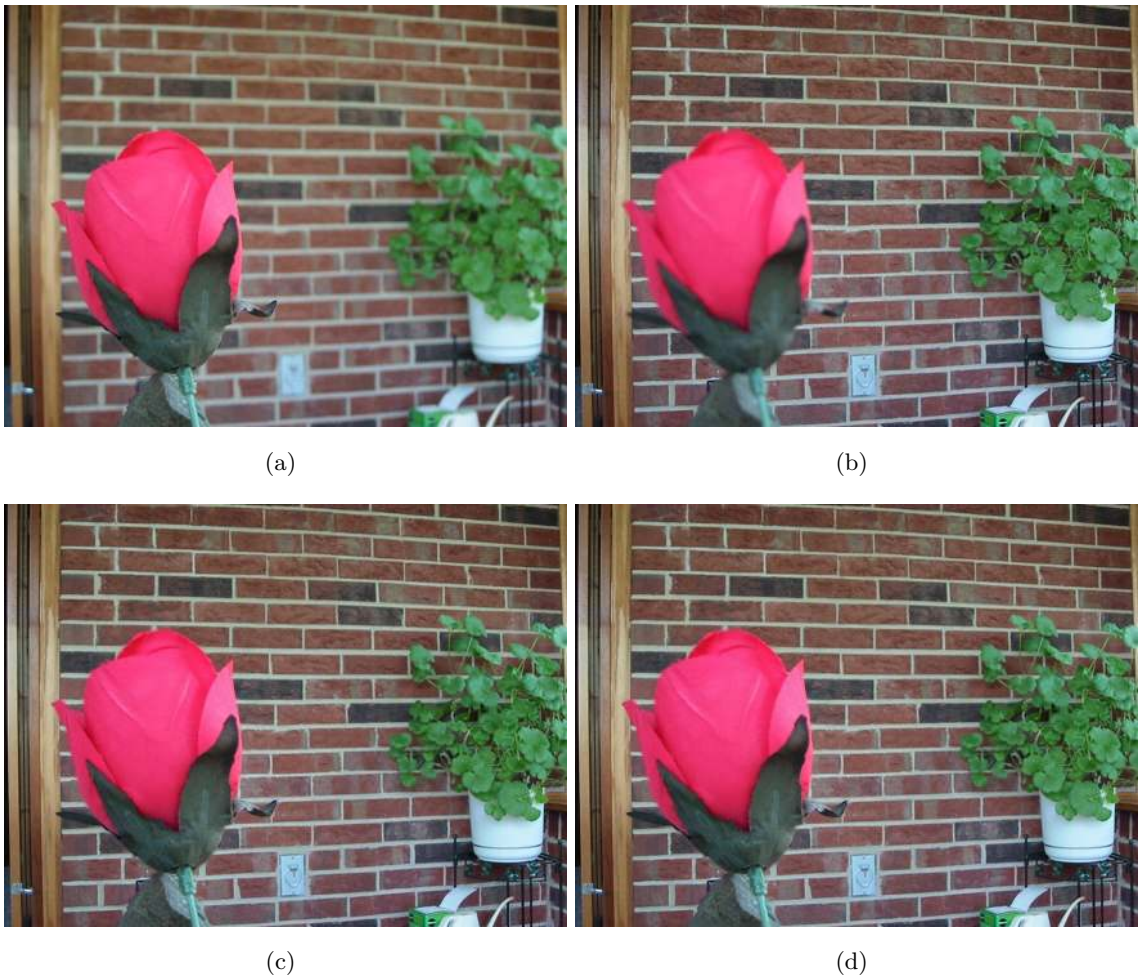
En s'appuyant sur les notions mentionnées, nous construisons une distance qui représente la similarité des patchs tout en minimisant le niveau de distorsion du flou. Pour une image de référence R et une image source S, nous calculons la distance à la position $x \in \Omega_R$ et $x' \in \Omega_S$, comme:

$$\begin{aligned} D(x,x') = \ & \lambda_1 \underbrace{\|\mu(P_R(x)) - \mu(P_S(x'))\|^2}_{\text{Coleur}} \\ +\lambda_2 \underbrace{\|\boldsymbol{\theta}_R(x) - \boldsymbol{\theta}_S(x')\|^2}_{\text{Orientation}} & +\lambda_3 \underbrace{\|D_R(x) - D_S(x')\|^2}_{\text{Descripteurs}} , \end{aligned} \tag{5}$$

où le paramètre $\mu(P_R(x)) \in \mathbb{R}^3$ extrait l'intensité moyenne du patch $P_R(x)$ centré à la position x de l'image R, le paramètre $\boldsymbol{\theta}_R(x) \in \mathbb{R}^{2p^2}$ est le vecteur obtenu à partir du gradient normalisé (gradient divisé par son amplitude) dans un voisinage $p \times p$ autour de x. $D_R(x)$ est un descripteur SIFT extrait de l'image R à la position x, ce qui correspond à l'image S. De plus, les multiplicateurs $\lambda_1,\ \lambda_2$ et λ_3 sont ajustés pour combiner les termes de façon appropriée.

Cette distance est incorporée dans un algorithme des voisins les plus proches, ce qui nous aide à estimer une carte de déplacement qui permet de modifier l'ensemble des images d'entrée pour qu'elles possèdent la même géométrie que l'image de référence. Dans la section expérimentale, nous montrons l'efficacité de notre méthode sur une base de données de cas difficiles d'images contenant des objets en mouvement.
Ci-dessous nous montrons des résultats avec notre méthode sur des images qui présentent un alignement global total et aussi avec des distorsions dynamiques.

Dans la figure 16, nous présentons un ensemble d'images avec différents paramètres focaux. Remarquez qu'après la fusion, notre méthode produit des résultats presque identiques à une méthodologie standard pour le MFIF. Cela nous permet de confirmer que le cadre proposé soutient adéquatement les cas statiques.

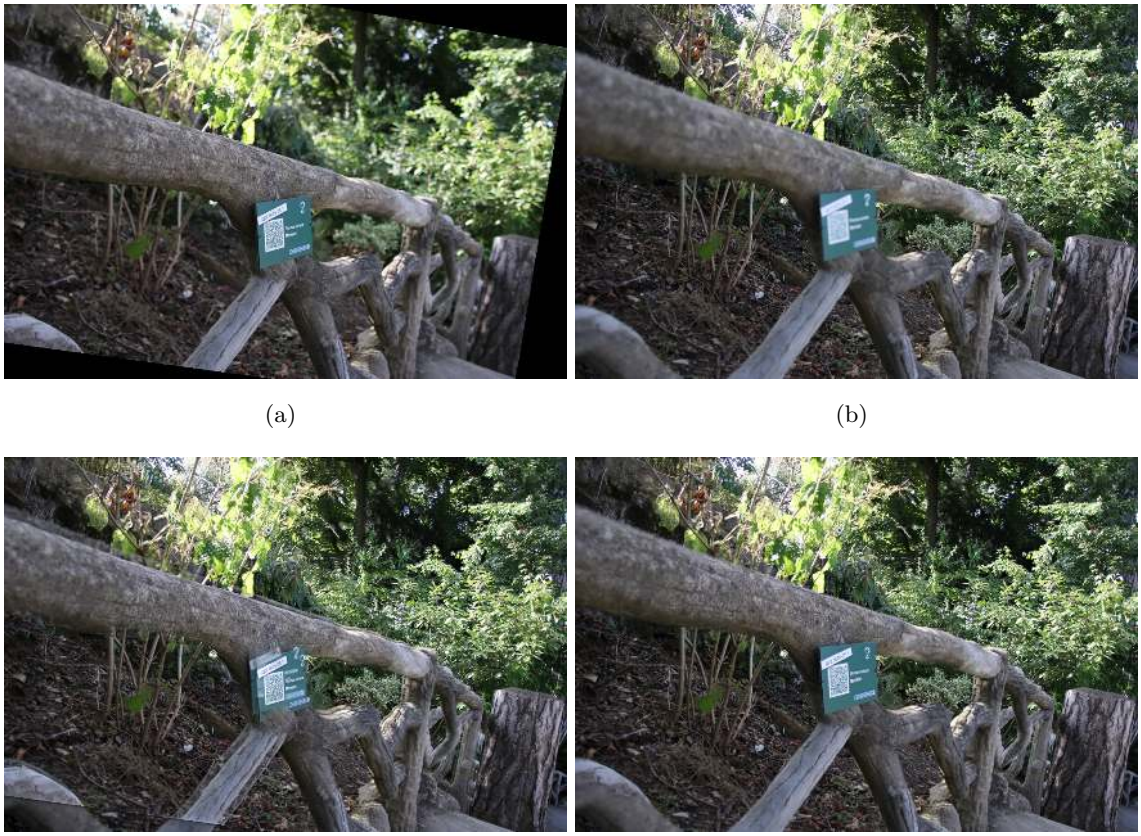


(a) (b)

(c) (d)

Fig. 16 *Expérience sur des images alignées:* (a). Image d'entrée avec un premier plan net. (b) Image de référence d'entrée avec un premier plan flou. (c) Fusion avec le standard algorithme. (d) Notre méthode.

In figure 17, nous présentons le cas où le mouvement de fond est principalement résolu après l'alignement de l'image, mais les objets au premier plan semblent toujours mal placés. Remarquez qu'après la fusion, l'image avec l'algorithme standard présente des artefacts tels que des fantômes et des irrégularités à la transition de régions inconnues dans l'image d'entrée. Au contraire, notre méthode résout correctement ces erreurs après l'enregistrement et les régions inconnues dans les images alignées. La fusion finale est une image sans artefacts qui surpasse l'algorithme standard et qui est capable de rendre la référence plus nette dans les régions floues.



(a) (b)

(c) (d)

Fig. 17 *Correction des désalignements sur les images:* (a). Image d'entrée alignée sur la géométrie de la référence et avec un premier plan net. (b) Image de référence d'entrée avec un premier plan flou. (c) Fusion avec le standard algorithme. (d) Notre méthode.

Les expériences sur des images avec des mouvements complexes (Section 5.3) nous permettent de conclure que notre méthode est généralement robuste aux distorsions géométriques et au flou. Ce système ne résout pas seulement les artefacts dus au mouvement, ou les erreurs après l'enregistrement, mais produit également des résultats presque identiques lorsqu'on les compare aux algorithmes de fusion d'image pour les paramètres statiques.

Avec notre algorithme, la fusion finale est une image sans artefact qui affiche plus de détails et de structures nettes que toute autre image de l'ensemble.

Contributions

Dans ce document, nous présentons un ensemble de techniques pour la fusion d'images Multi-exposition et Multi-focus qui traitent des mouvements de caméra et des mouvements d'objets. Les approches proposées reposent sur une étape de pré-alignement combinée à une analyse de patch qui résout d'autres inexactitudes au niveau des pixels et des incohérences géométriques. De plus, ils s'inspirent de Exposure Fusion [91] en raison de sa praticabilité et de sa simplicité avec peu de réglages de paramètres, d'efficacité de calcul et de réduction des artefacts Halo ou d'inversion de gradient.

Nos méthodes s'appuient sur un ensemble d'images prises indépendamment avec des temps d'exposition ou des réglages de mise au point modifiés. L'acquisition se fait d'une manière qui permet à tous les objets d'être bien éclairés (ou bien focalisés) dans au moins une des images sources. En capturant de manière appropriée les régions qui détiennent une bonne qualité, nos méthodes synthétisent une image qui reproduit de manière cohérente l'illumination; ou la netteté, à partir de la scène réelle.

Les contributions de cette thèse sont énumérées ci-dessous:

- Une étude de la capacité de reconstruction de l'algorithme de Patchmatch en présence de diverses distorsions d'image.
- Un cadre pour la Fusion d'Images Multi-exposition sur des scénarios dynamiques basés sur RAW Linear Images, [98], ou des images non linéaires de 8 bits.
- Un cadre pour la Fusion d'Images Multi-focus sur des scènes dynamiques [99] qui s'appuie sur une nouvelle distance pour la comparaison de patchs avec des distorsions de flou.

Aperçu Général

Cette thèse est organisée comme suit.
Dans le Chapitre 2, nous présentons au lecteur des notions importantes sur l'acquisition d'images numériques standard, ses limites et les outils de calcul actuels pour améliorer la numérisation.
Dans un premier temps, nous expliquons comment l'irradiance d'une scène est reproduite numériquement par l'assemblage d'un objectif photographique, d'un capteur, d'un processeur d'unité et d'autres composants qui caractérisent un appareil photo conventionnel.

Au cours de cette présentation, nous décrirons quelques-unes des raisons pour lesquelles les appareils photo numériques perdent leur précision sur des scènes lumineuses ou lorsque la profondeur de champ est limitée.
Pour ce premier problème, nous introduisons le concept de fonction de réponse de la caméra et expliquons comment son estimation est utilisée pour créer des images HDR. De plus, nous explorons la méthode appelée Exposure Fusion et comment cette méthode simple produit des résultats de type HDR, pour lesquels un état de l'art est offert plus tard. Nous soulignons ici l'ensemble des techniques disponibles qui permettent la fusion d'images avec des perturbations de mouvement, ce qui est précisément le cas que nous allons considérer dans cette thèse.
Le dernier problème est mieux analysé, pour lequel nous décrivons certaines des méthodes actuelles pour compenser le manque de netteté des images. Nous présentons ici l'état de l'art de ces méthodes et les quelques techniques qui traitent du mouvement pour le problème de Focus Stacking.
Nous concluons ce chapitre en présentant certaines des techniques basées sur les patchs qui ont inspiré la méthodologie utilisée dans ce travail.

Dans le Chapitre 3, nous étudions une solution au problème de l'équation (2) afin de corriger les perturbations géométriques dans une paire d'images. Dans ce but, nous étudions les techniques de reconstruction d'images, une approche qui nous permet d'aligner le contenu commun entre les images et qui est basée sur l'estimation d'une carte des déplacements. Dans cette thèse, nous estimons la carte des déplacements par l'extraction d'un ou plusieurs voisins les plus proches qui sont associés à chaque patch possible dans l'image, ce qui dans la pratique est réalisé par l'algorithme de Patchmatch.
A cet égard, nous étudions les capacités de ce cadre (reconstruction d'image avec Patchmatch), pour corriger les distorsions géométriques sur les images avec des conditions d'acquisition variables, y compris les variations de temps d'exposition et les changements de mise au point.

Dans le Chapitre 4, nous proposons une solution à la Fusion d'Images Multiexposition avec du contenu en mouvement. Notre méthodologie repose sur la reconstruction des images normalisées radiométriquement, où la reconstruction s'inspire de la méthodologie du Chapitre 3. Pour cela, nous expliquons les alternatives pour normaliser la radiométrie des images, à la fois dans le domaine linéaire (images RAW) et non linéaire (images 8 bits). Afin d'obtenir un algorithme pleinement opérationnel, nous expliquons comment traiter les régions saturées et régulariser les résultats de reconstruction.
A la fin de ce chapitre, nous analysons la performance de notre méthode à l'aide d'expériences approfondies et comparons nos résultats avec les techniques les plus récentes.

Dans le Chapitre 5, nous proposons une méthode générique pour surmonter les perturbations associées aux objets en mouvement dans le contexte de la Fusion d'Images Multifocus. Nous introduisons d'abord un schéma simple pour la fusion d'images sur des paramètres statiques, puis nous présentons notre solution au cas dynamique. Dans notre méthode, nous sélectionnons une image de référence, puis, pour chaque image de la pile, nous reconstruisons une image qui partage la géométrie de la référence tout en améliorant son contenu de netteté. La reconstruction est réalisée grâce à une nouvelle distance, adaptée aux images floues, qui est incorporée dans l'algorithme PatchMatch. Nous décrivons également une alternative de post-traitement pour corriger les erreurs de reconstruction.
A la fin de ce chapitre, nous montrons l'efficacité de notre algorithme sur une base de données de cas difficiles d'images acquises sur des cas dynamiques et contenant des objets en mouvement. Nos résultats sont comparés à deux méthodes récentes pour les paramètres statiques et dynamiques, respectivement.

Chapter 1

Introduction

During the last decades, technology developments and device miniaturization have been principal factors for the constant growth in the field of digital photography. Today, having access to digital cameras is mainstream and the possibility to register ordinary activities is granted. Virtually everyone is a photographer and not much knowledge is necessary to render good quality images.
This wave, powered either by manufacturers' ambition, consumers demands or scientists curiosity, continuously raises the standards of image quality with increases on resolution, color depth, contrast or detail.
An alternative to hardware reinforcement is to extend the physical limitations of the camera through the development of smarter algorithms. The collection of such digital smart approaches defines what is generally referred to as *Computational Photography*. Among such techniques, image enhancement relates to the computational tools that ameliorate the content of images, where the improvement is over one or several image features. They specialize in tasks as color enhancement, illumination normalization, noise reduction, focus/detail enhancement, etc., therefore, extending the capabilities of standard digital cameras.
The aim of this thesis is to develop patch-based techniques for Computational Photography. Here we concentrate particularly on two applications: *Multiexposure Image Fusion* and *Multifocus Image Fusion* and their extension to dynamic settings. From a stack of images acquired with different settings, the first method aims at capturing the full dynamic range of a scene, while the second one aims at producing an image that is sharp everywhere.

1.1 Motivation

The first problem that we are interested in solving arises with strong spatial variations in illumination. Illumination is a critical factor during image acquisition. On uncontrolled scenarios, it may present variations that modify the visibility and contrast of objects and therefore cause the sensor to render imperfect images. For scenes of very high dynamic

range, the illumination can be up to the order of 20 stops, which surpasses the capabilities of modern commercial sensors only limited to about 15 stops [1]. Even though specialized cameras are more robust to larger dynamics, by integrating more powerful circuitry, they are expensive and mostly unaccessible. As for ordinary cameras, the typical way to overcome this limitation is to use several captures of a scene.

For the acquisition of scenes with high dynamic range, two main currents have emerged: High Dynamic Range (HDR) imaging and Exposure Fusion.
HDR imaging is a technique that has been popularized since the mid-1990's [32], and is capable of rendering vivid images with abundant details on environments that present strong illumination changes, figure 1.1. This approach consists of two steps, HDR map creation and Tone Mapping. The HDR map is a digital reproduction of the irradiance of the objects in the scene. It is estimated by using as input several images that are captured with variable exposure times. The aim of using multiple expositions is to compensate for intensities that either saturate the sensor (very bright intensities) or for objects that are not well illuminated under normal settings of the sensor. Given that natural images frequently exhibit strong contrast changes in the dark-bright spectrum of tones, the HDR map holds a non linear high dynamic nature that exceeds the standard dynamic range of common displays or printing devices. Indeed, although HDR display devices are emerging, typical devices have a dynamic of less than 10 stops. Because of this, an additional treatment, called Tone Mapping [37], is required to appropriately balance color intensities and obtain standard 8 bit images that are easily visualized in commonly commercialized displays.

(a) HDR + tone mapping. (b) Exposure Fusion.

Fig. 1.1 **Top images**: Original stack of bracketed exposure images. (a) Tone mapped image with [42]. (e). Exposure Fusion [91]. Images courtesy of Fattal et al. [42].

In contrast, Exposure Fusion [91] is a much simpler technique that proceeds directly to the fusion stage without the need of estimating the HDR map. As in HDR imaging, the algorithm makes use of several images with varying levels of exposition, and assumes that all objects appear properly illuminated in at least one of the input images. For instance, see the set of bracketed exposure images from figure 1.1. For long expositions, very bright objects appear saturated, while for short expositions they exhibit good intensities. A similar phenomenon occurs for dark objects on longer expositions. Surprisingly, Exposure Fusion manages to employ a simple weighted combination to generate an acceptable image, figure 1.1(b), that is not very different from the HDR result, figure 1.1(a). In this method, the combination of the stack of images I_i is based on the Laplacian components of each image by inspecting the quality of their content in the RGB domain, as follows:

$$\mathcal{L}_l(R) = \sum_{i=1}^{N} \mathcal{G}_l(W_i)\mathcal{L}_l(I_i) \ , \tag{1.1}$$

where $\mathcal{L}_l(I_i)$ is the Laplacian pyramid [20] of image I_i at level l and $\mathcal{G}_l(W_i)$ is the Gaussian pyramid of the quality weight map W_i at level l. The fused image R is then recomposed from the resulting Laplacian pyramid.

Despite considerable advances in both these methods, HDR imaging and Exposure Fusion, they still present some limitations. That is, in spite of its realistic effect HDR imaging is known to present artifacts that come mainly from the Tone Mapping stage [112, 41]. Among them, halo artifacts, gradient reversals or color conditions, like unnatural excess of detail, are the most common. Not to mention that most Tone Mapping operators are scene dependent and that parameter calibration should be performed independently for each image to reach better results.
As for Exposure Fusion, the quality of the fused image depends on how well the input images capture the whole dynamic of the scene. That is because only well exposed intensities, without further processing, are used to synthesize the output.

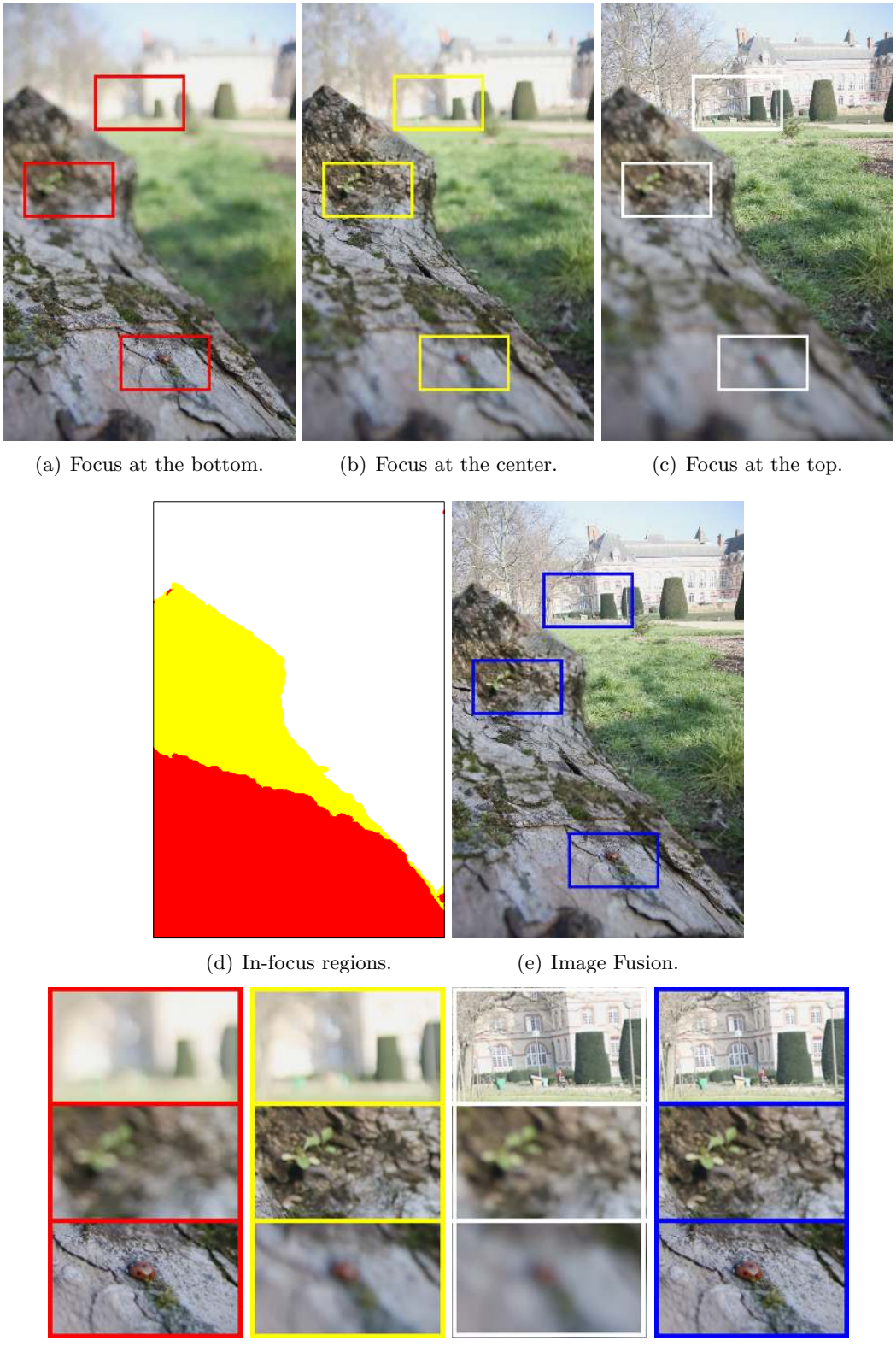
The second problem that we approach in this thesis aims at rendering an image sharp, while relying on images with different focal depth. For that we employ image fusion to combine the sharp regions from each image, a process that is commonly known as *Multifocus Image Fusion* (MFIF). Multifocus image fusion or *Focus Stacking* is a classical technique that seeks to recover and add details to images that suffer from the presence of blur. Such blur can be the result of multiple factors like improper acquisition techniques, environmental conditions or more likely because of objects falling outside the typical depth of field of the sensor. It is also assumed to appear with a spatially varying level of distortion. Following a similar principle as in multiexposure image fusion, focus stacking takes the

best of each image for the composition of a sharper output. It synthesizes an image by blending seamlessly all regions that contain larger presence of detail among the images, see figure 1.2(a)-1.2(c). As a result, the output image (Figure 1.2(e)) displays more details scattered over its spatial domain. Since the process is subject to the content of the source images, a good MFIF method should not only be capable of identifying properly sharp regions but also transferring them without detail reduction.

This is illustrated in figure 1.2. The original stack of images, figures 1.2(a)-1.2(c), contains an excessive amount of blur, except for localized areas that are exclusive to one image in the set. Such sharp areas can best be seen on the map at figure 1.2(d), where red, yellow and white are the color codes assigned to sharp regions on images 1.2(a), 1.2(b) and 1.2(c), respectively. Here, in-focus regions on the map are extracted via the Local Total Variation, for activity level measurement, and fused in a multiresolution fashion with the Laplacian pyramid.

In a general perspective, multiexposure and multifocus image fusion can be broken down into two similar fundamental stages: candidate region selection and fusion. The first task refers to the local identification of pixels whose contrast quality (for MEIF applications) or focus quality (for MFIF applications) is better than that of the corresponding pixel in the remaining images. In the former case, this is commonly carried out by using measures that respond proportionally to illumination, as for instance a combination of Contrast/Saturation and Well-Exposedness (See [91]). In the case of focus stacking, the examination of in-focus objects is popularly performed with measures that analyze the local variation energy which is a reasonable indicator of sharpness. Some focus measures include variance, energy of image gradient, energy of Laplacian, spatial frequency, etc., [57]. Nevertheless no standard consensus has been achieved for a definite blur measure.
The fusion task consists of seamlessly joining all the marked regions, with good quality, so that no discontinuities or artifacts are produced in the process. For fusion purposes, it is common to find schemes where weighted combinations are utilized. Multiresolution approaches like the Laplacian and Gaussian pyramids, as in equation (1.1), and wavelets, with different wavelet families, are also a popular alternative.

Both applications, MEIF and MFIF, can yield artifacts associated to motions which are originated from camera or object displacements while acquiring the input images, figure 1.3. Those artifacts or ghosts effects, appear as vanishing objects and emerge due to local discrepancies during the fusion, where different objects are combined.
Direct solutions to account for motion include image registration or simply the use of tripods while shooting, which could be annoying to carry and restrain from casual scenarios. Whereas such methods ensure global correspondence among still objects, moving objects can still present large displacements and provoke artifacts.



(a) Focus at the bottom. (b) Focus at the center. (c) Focus at the top.

(d) In-focus regions. (e) Image Fusion.

Fig. 1.2 (a-c). Original stack of images with different focus settings. (d) Sharp regions map. (e) Image Fusion. **Bottom images**: zoomed boxes extracted from the input images (red, yellow and white) and the fused image (blue).

Fig. 1.3 Examples of ghost artifacts generated after MEIF. Images courtesy of [3, 102].

Within MEIF and HDR imaging, multiple contributions have been proposed to provide robustness to motion. Some strategies include, deghosting techniques or optimal similarity matching.
As for MFIF, even though it has been extensively studied in the last decades, no motion constraints have been satisfactorily considered. This problem is particularly difficult, because once details are lost due to blurring, there is no geometrical information to be matched with on other images.

In conclusion, MEIF and MFIF, are affected by two types of motions: camera motions and objects motions. The most reliable technique to account for a variety of transformations is image registration via homography. However, due to camera lenses distortions, or because of the acquisition conditions, (namely, if the scene is not made of one single plane and the optical center has moved), the estimation of such transformation can result in mis-registrations. Besides, object motions are not solved.

This thesis examines how to deal with motion under physical variable conditions as illuminations, depth of field and geometric transformations.

1.2 Thesis Problem

Under static settings, MEIF and MFIF can rely on the standard two steps algorithm (candidate region selection and fusion) in order to obtain satisfactory results. Indeed both camera and objects maintain identical spatial positions among the images, allowing for a reliable extraction of local quality measures.
On dynamic settings, however, the problem is more complex because objects can change position or worse geometry, and they are expected to change contrast and focus. These factors make it inviable to rely on the standard procedure, that generates artifacts on the aforementioned conditions. In fact, artifacts due to motion would not be a problem if we knew ahead where the corresponding objects are localized after displacements, or differently, where to extract quality measures associated to matching objects among the images. Instead, we rely on a ***geometric alignment*** to force the objects to share the

same position among the images, producing images that are perfectly registered. For that, we require to localize objects at pixel precision over the images, a problem that is referred to as dense correspondence (for images with same illumination/focus conditions).
For object localization we rely on the hypothesis that even though the content of the images have changed spatially, the internal local content of the objects remains identical or partially unaltered over the images. A toy example can be seen in figures 1.4, where the ball presents spatial changes between the images without substantial internal deformation. This applies as well to structures of smaller scale, that contain enough unique information to be precisely recognized on the images, like for instance, the shoes or hands of the children. It is for that reason that we rely on small image portions or local neighborhoods, from now on called *patches*, to be searched for over the images. Particularly, we use measures that respond proportionally to the similarity content of patches, in a framework that should be significantly robust to illumination or blur.

For MEIF and MFIF, the localization of moving objects can be expressed as an optimization problem where we try to minimize the error between all possible patches from an image and their potential matching candidates in other images, as follows:

$$f(x) = \underset{y \in \Omega_S}{\arg\min} \, \|P_R(x) - \mathcal{T}_R\{P_S(y)\}\|^2, \tag{1.2}$$

where $P_R(x)$ and $P_S(y)$ are the patches centered at position x and y, from a quality perturbed image R and an image S, respectively. Ω_S is the spatial domain of image S, and the operator $\mathcal{T}_R\{S\}$ is a quality transformation that applied to image S calibrates the type of degradation, exposition or blur, to that of image R. In other words, this transformation enhances the patch resemblance between the images, therefore providing invariance to the perturbation. It is worth noting that the nature of the transformation is different for each case, MEIF or MFIF. In practice, estimating the transformation is a complex problem because of unknown variables during the acquisition, like nonlinear illumination changes or blur. Therefore it is estimated differently and may be applied to a patch or the whole

(a) Under-exposed. (b) Normal-exposed. (c) Over-exposed.

Fig. 1.4 (a-c). Bracketed exposure images with moving objects. Images courtesy of [102].

image depending on the case.
In brief, solving equation (1.2) help us find dense matches of local neighborhoods within images R and S. Consequently we recover a map of coordinates $f(x)$ where the objects are identical or similar.
The solution to this problem is presented in the following chapters for each application, where we describe in detail how to deal with multiple images.

The benefits of using patch-based approaches are several. First, patches are highly redundant, meaning that similar patches can be found everywhere in the images, taking advantage of autosimilarity in natural images. This property has shown to be very useful for noise reduction and was capitalized in the seminal work of Buades et al. [16] for image denoising. Here, we extend that approach to correspondence map refinement in a patch aggregated manner. Secondly, thanks to local discriminant content, geometrically similar patches can be coherently found on images that suffer from different degradations. This property is particularly useful for motion estimation problems, where common objects between the images can be tracked, which makes patches suitable for our aim of motion invariance.
In the spirit of the last property, one could enlarge the extend of the patch to add enough additional information for patch matching under blur distortions, as we will do for our second application.

1.3 Contributions

In this document, we present a set of techniques for multiexposure and multifocus image fusion that account for handheld camera motions and object motion. The proposed approaches rely on a prealignment stage combined with a subsequent patch analysis that resolves further inaccuracies at pixel level and geometric inconsistencies. Also, they take inspiration from exposure fusion [91] due to its practicability and simplicity with little parameter tweaking, computational efficiency and avoidance of halo or gradient reversal artifacts.
Our methods rely on a set of images taken independently with modified exposure times or focus settings. The acquisition is done in a way that allows for all objects to be well illuminated (or well focused) in at least one of the source images. By capturing appropriately the regions that hold good quality, our methods synthesize an image that coherently reproduces illumination; or sharpness, from the real scene.

The contributions of this thesis are listed below:

- A study of the reconstruction capacity of the Patchmatch algorithm in the presence of various image distortions.

- A multiexposure image fusion framework for dynamic settings based on raw linear images, [98], or 8-bits non linear images.
- A multifocus image fusion framework for dynamic settings [99] that relies on a new distance for the comparison of patches with blur distortions.

1.4 Overview

This thesis is organized as follows.

In Chapter 2, we introduce the reader with important notions about the standard digital image acquisition, its limitations and current computational tools for enhancing the digitization.
Initially, we explain how the irradiance of a scene is digitally reproduced through the assembly of photographic lens, sensor, unit processor and other components that compose a conventional camera. During this presentation, we will describe some of the reasons why digital cameras lose precision on bright scenes or when the depth of field is limited.
For the former problem, we introduce the concept of camera response function and explain how its estimation is used to create HDR images. Also, we explore the method called Exposure Fusion and how this simple method produces HDR-like results, for which a state of the art is offered later on. Here, we highlight the set of available techniques that allow for the fusion of images with moving perturbations, which is precisely the case we will consider in this thesis.
The latter problem is better analyzed, for which we describe some of the current methods to compensate for the lack of sharpness in the images. Here we present the state of the art for such methods and the very few techniques that deal with motion for the problem of focus stacking.
We conclude this chapter by introducing some of the patch-based techniques that inspired the methodology used in this work.

In Chapter 3, we study a solution to the problem of equation (1.2) in order to correct geometrical perturbations in a pair of images. To that aim, we investigate image reconstruction techniques, an approach that allows us to align the common content between the images and that is based on the estimation of a map of displacements. In this thesis, we estimate the map of displacements by the extraction of either one or several nearest neighbors associated to each possible patch within the image, which in the practice is carried out by the Patchmatch algorithm.
In that regard, we study the capabilities of this setup (image reconstruction with Patchmatch), to correct geometrical distortions on images with variable acquisition conditions

including exposure time and focal changes.

In Chapter 4, we propose a solution to the fusion of multiexposure images with moving content. Our methodology relies on the reconstruction of the radiometrically aligned images, where the reconstruction is inspired by the methodology of Chapter 3. For that we explain the alternatives to normalize the radiometry of the images, both in the linear (RAW images) and non linear (8-bit images) domain. In order to get a fully operational algorithm, we explain how to deal with saturated regions and to regularize the reconstruction results. At the end of this chapter, we analyze the performance of our method with extensive experiments and compare our results with state of the art techniques.

In Chapter 5, we propose a generic method to overcome the perturbations associated to moving objects in the context of multifocus image fusion. We first introduce a simple scheme for image fusion on static settings and proceed to present our solution to the dynamic case. In our method, we select a reference image, and then, for each image of the stack, reconstruct an image that shares the geometry of the reference and the sharpness content of the image at hand. The reconstruction is achieved thanks to a new distance, which is adapted to blurred images, that is incorporated in the PatchMatch algorithm. We also describe a postprocessing alternative for correcting reconstruction errors.
At the end of this chapter, we show the efficiency of our algorithm on a database of challenging cases of hand-held shots containing moving objects. Our results are compared with two recent methods for static and dynamic settings, respectively.

Chapter 2

Background & Previous Works

Multiexposure Image Fusion (MEIF) and *Multifocus Image Fusion* (MFIF) methods have increasingly gained popularity in the last few years.
In this chapter, we present a concise overview and current state of the art for MEIF and MFIF methods.
As both methods will later be approached in a patch-based manner, we also give a short review of patch-based algorithms in *Computer Vision* and *Computational Photography*. As well, some important technical notions about image acquisition are briefly presented.

2.1 Image Acquisition

Digital cameras can be seen as black boxes that digitally reproduce the analog world, that in response to an input produce tri-channelled 8 bit images of a given resolution. Nonetheless, the operations within the camera are highly sophisticated; normally executed within a fraction of a second, and commonly pass unnoticed by the user.
Digital image acquisition comprises the set of operations necessary to capture and save digital images. Current digital cameras are normally manufactured with the following essential components: optical lenses, aperture setting, shutter, *CCD*/*CMOS* sensor, ISO regulator, analog to digital converter and a unit processor, (see [36, Ch. 3] for a complete pipeline of a typical digital camera).
The pipeline can be described as follows. After the shutter is triggered, the light passing through the lenses is concentrated and exposed to the sensor that transforms the electromagnetic energy into an analog signal. The signal is then quantified and converted to a digital value. The resulting array of digital values is the so called RAW image which replicates linearly the radiance reflected by the scene. Finally, embedded software in the unit processor performs a set of treatments over the RAW image in order to turn the data into a displayable image. Figure 2.1 shows a graphic description of all operations inside a digital camera. For a better understanding on the image formation and each of its critical

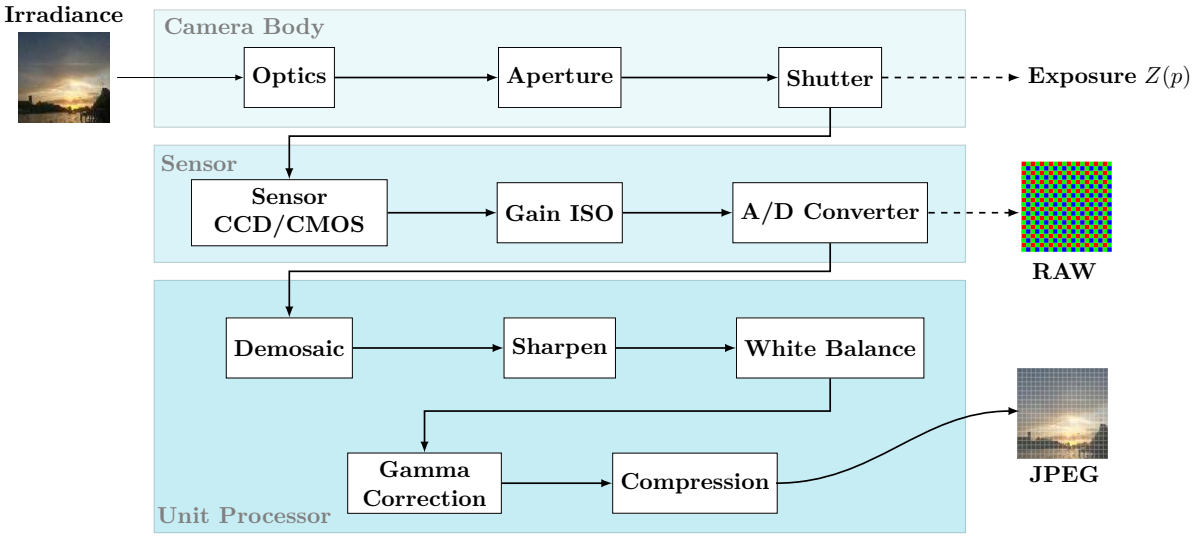


Fig. 2.1 Standard pipeline of a digital camera. Inspired by [120].

stages, the reader is invited to see [120, 36, 2, 90].

By assuming ideal linearity of the sensor, we can formulate the camera acquisition process for an exposed sensor as,

$$Z(p) = E(p) \cdot t \ , \tag{2.1}$$

where E is the scene irradiance, t is the exposure time of the sensor and $Z(p)$ is the observed value at pixel sensor p. The observed irradiance is linearly related to the real irradiance by a factor of time, and other camera related terms that we omit for simplicity. This simple model accounts for image digitization or RAW image creation.
In reality, however, the acquisition process is prone to interferences, as we will see below, and the model of equation (2.1) is only applicable to ideal cameras.

Clearly the most important component in the standard pipeline is the sensor. Nowadays, there are two types of technologies, charged-coupled devices (CCD) and complementary metal-oxide-semiconductor (CMOS), their main difference lying in the way they read and communicate the output charge.
On the one hand, CCD's sensors are silicon based semiconductors that produce an electric charge upon exposition to light, this charge is propagated across the sensor and converted to voltage values as it passes through an amplifier. On the other hand, CMOS sensors are a newer technology of integrated circuits that read and convert the electric charge into voltage at each particular location, where each photo-diode is coupled with an additional

circuitry for such task [120, 2].
However, each configuration is prone to suffer from perturbations. For instance, previous architectures of the CCD sensor are known to suffer from blooming [120]. Blooming is a phenomena that appears when saturation on single photo-diodes is propagated, affecting the readout of neighboring pixels. As for CMOS sensors, their local additional circuitry constitutes an additional source of noise to the digitization [2].

Sources of noise during digitization are not limited to sensor architecture only, indeed multiple perturbations also occur during the general pipeline acquisition. Such perturbations are originated in the real world or within the camera at each stage of the acquisition, consequently adding different types of noise to the rendered image. Common noise types include thermal noise, photon shot noise, dark currents, readout noise, or spatial noise sources [2, 50]. Also, physical conditions on the environment may corrupt the image formation, for example, the amount of incoming light to the sensor or turbulence of the media, which may result in overexposure/underexposure or blur, respectively.

In order to obtain a standard image, RAW images are subject to linear and non linear treatments like demosaicking, white balance, gamma correction or dynamic compression, image compression, sharpening, noise filtering, etc.
In the HDR literature, this subsequent treatment is expressed by abbreviating all operations into a single operator, normally referred to as the camera response function (CRF), which applied to the observed irradiance produce a typical image [36, Ch. 3],

$$I(p) = \mathcal{F}\{\hat{Z}(p)\} \ , \tag{2.2}$$

where the CRF function $\mathcal{F}(\hat{Z})$ is applied to the measured irradiance $\hat{Z}$ (including additional perturbations), yielding image I.

Having a clear understanding of the camera acquisition, is compulsory to elaborate techniques that compensate for them. For instance, the works by [50] and [2] propose a complete camera model, including noise sources, in order to estimate the scene irradiance and to account for the limited dynamic of standard sensors on strong contrast environments.

2.1.1 RAW Images

RAW images are the first digital data obtained by the sensor and they contain information about the real illumination of the scene. The information from RAW images is untreated and therefore is not ready to be displayed in standard monitors. The dynamic range of these images is usually coded on 12 to 14 bits, although their real dynamic is less than 12 bits [1]. Consequently, dark objects will produce very small digital values in comparison to

bright objects that can reach the limits of the sensor, which saturates.

RAW images also provide information (metadata) about the camera configuration and acquisition settings, like focal length, exposure time, shutter speed, white balance multipliers, black level offset, etc. The RAW image information comes in a format exclusive of the manufacturing company (Nikon's *.nef, Canon's *.cr2, Kodak's *.kdc, Epson *.erf, Panasonic's *.rw2, etc.) [119]. In order to access the information inside raw files, various softwares are available, for instance *dcraw, darktable, RawStudio, RawTherapee, Adobe DNG converter*, that are open-source options. A popular commercial software is the *Adobe lightroom*.

Within the RAW image, the data is stored with different color patterns, where each pixel is a response of either a red, green or blue sensor. In technical terms, such distribution of color sensitivity is called Color Filter Array (CFA) and the most popular configuration is the Bayer pattern. After demosaicking, the data is reorganized in a 3D format and each pixel is associated to its RGB components. More additional treatment generally includes: defective pixel removal, white balancing, noise reduction, compression, among others.

2.1.2 Problems due to illumination changes & sources of blur

Standard digital cameras frequently present limitations associated to illumination and depth of field. Understanding such weaknesses help us establishing strategies to reduce those problems as well as the artifacts created by motion.

Illumination changes

Strong illumination variations become a problem when illumination exceeds the operational range of the sensor. For such cases, the sensor gets saturated, resulting in white undetailed regions in the image, that normally should contain visible structures.
Common situations for a sensor to be saturated include direct exposition of the sensor to strong sources of light (figure 2.2 left) or environments where light is non uniformly distributed over the scene, like indoor scenarios, (figure 2.2 right).

As mentioned in the previous chapter, a popular method that solves the contrast change problem consists of combining several images that were captured with different exposure settings. In the next sections (section 2.1.3), a description of these alternatives will be presented.

Fig. 2.2 Common situations where digital cameras render badly exposed images: direct exposition to strong illumination sources (left) or indoor scenarios (right). Images courtesy by [62, 55].

Blur distortions

In the case of blur, distortions of this type may occur for several reasons but mainly for the limited depth of field (DOF) of the camera, which makes the aspect of objects more or less blurred depending on how far they are located from the lenses. By definition, the depth of field of a camera is the distance between the closest and farthest objects that are in-focus. In the case of shallow depth of field, most objects of a particular image appear blurred. The depth of field is modified with changes of the aperture and focal distance, [90, 120]. During the acquisition, in order to compensate for objects falling outside the focal plane, the most simple option is to focus with the lenses or to reduce the aperture size by increasing the *f-number*.

Formally, the focal plane of a lens can be found as:

$$\frac{1}{d} + \frac{1}{d'} = \frac{1}{f} \ , \tag{2.3}$$

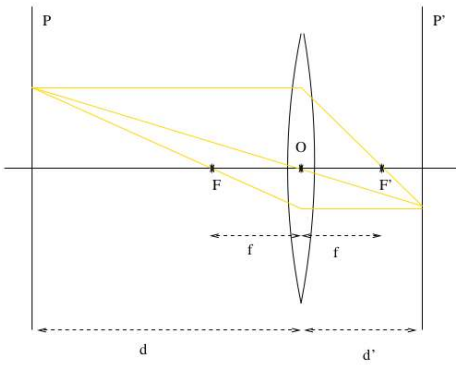


Fig. 2.3 Optical system. Image courtesy by Yann Gousseau.

where f is the focal length, $d > 0$ is the distance to the plane of focus and $d' > 0$ is the distance between the lens and the sensor plane, see figure 2.3.

Other factors, besides the lens, that can provoke blur are: motion, incorrect image acquisition, medium related (like turbulence) and post processing blur [139].
In a similar manner as for strong illumination changes, a common solution to enlarge the limited depth of field consists of fusing images that were captured with different focus settings, a technique named *Multifocus Image Fusion*.

2.1.3 HDR Imaging and Tone Mapping

The aim of High Dynamic Range (HDR) imaging is to enable the capture of a large dynamic of the irradiance of a scene, rather than only a fraction as traditional methods do. Since HDR images are more faithful to reproduce irradiance, the resulting images replicate better the real world in terms of appearance, while being not only rich in colors but also detail. For that, HDR methods have the challenge to provide information even under strong illumination environments, a situation that regularly saturates the sensor (see section 2.1.2).
HDR imaging is normally composed by two main tasks: creation and tone mapping.
The general framework for HDR creation is to estimate the illuminance of the scene by using multiple images, in order to capture the entire dynamic of the scene. More precisely, this methodology creates an irradiance map by relying on several low dynamic range (LDR) images, that individually capture a fraction of the irradiance dynamic. In order to achieve this, two options are possible. One that works with RAW images, for which the access to the irradiance information is relatively straightforward [3, 4]. The second one is by inverting the non linear response of the camera, when working with LDR images [85, 30, 94]. With the latter approach, the inversion of the camera response function (CRF) helps to recover an estimation of the irradiance:

$$\hat{E}_i(p) = \mathcal{F}^{-1}\{I_i(p)\} \cdot \frac{1}{t_i} \ , \tag{2.4}$$

where $\hat{E}_i$ is the estimated irradiance and $\mathcal{F}^{-1}\{I_i(p)\}$ is the inverse of the CRF applied to the input image I_i, acquired with an exposure time t_i.
Provided the inverse of the CRF and assuming that the scene is static, recovering the whole dynamic of the scene is essentially a statistical problem [50, 4], where the estimated maps E_i are properly combined in a weighted combination manner,

$$\hat{E}(p) = \frac{\sum_i^N w_i(p) \cdot \hat{E}_i(p)}{\sum_i^N w_i(p)} \ , \tag{2.5}$$

where the weights w_i can evaluate the quality of exposure for each pixel in image I_i or be chosen in a statistically optimal way [3].
Nonetheless, the estimation of the camera response function is complex, being perturbed by the noise from the 8 bit images.
In practice, what is desired for the inverse of the CRF is a mapping function that receives 8 bit images and returns an image that is proportional to the irradiance of the scene, [36]. In the literature, the most popular option to retrieve both the camera response function and the estimated irradiance is by using the method in [31]. In this work, the authors assume that the CRF is monotonic, thus invertible, and smooth. Their method is formulated as a least squares problem that solves for a function of $\mathcal{F}$ and the irradiance $\hat{E}$:

$$\arg\min_{E,\ g} \sum_{i,p} \|E_i(p) - g(I(p))/t_i\|^2 + \lambda \sum_n g''(n)^2 \ , \tag{2.6}$$

with g = $\mathcal{F}^{-1}$. The first term satisfies equation (2.4) and the second term forces the function to be smooth. The proposed solution to this problem relies on singular value decomposition (SVD) after a matrix formulation.
This methodology represents an important development on the field and multiple researchers have taken inspiration from it. For a deeper comprehensive description on HDR creation methods, the reader is invited to visit dedicated surveys on the topic [114, 36].

Tone Mapping

HDR images typically have a dynamic range that is beyond 8bits images, (up to 20 bits). Therefore, common visualization and printing devices are not well suited for HDR data. *Tone Mapping* is the operation of compressing the dynamic range of HDR images to the 8 bit standard, allowing their visualization on displays or conventional devices.
For tone mapping, the HDR data is non linearly rescaled, such that large values are compressed while the information provided by small intensities is preserved. This non linear rescaling generates a visually appealing image, satisfying human visual system (HVS) criteria [95].

Tone mapping operators can be broadly classified in two categories, global mapping (also named tone reproductive curves) and spatially varying operators (also termed tone reproductive operators).
Among global mapping methods, popular methods are; linearly rescaling the logarithm of the luminance, gamma correction and histogram equalization. In despite of being very practical and fast, these methods are incapable to preserve local contrasts, giving images a flat aspect [42]. Other authors propose methods that replicate the adaptation behavior in the human visual system, such as physiology and perceptually inspired approaches

[111, 87, 9, 60, 124, 44, 59].
Spatially varying operators, on the other hand, allow to compute more accurate image representations by considering contextual information around each pixel. These techniques capitalize the assumption that the human vision is mainly sensitive to local contrast [101]. However the major drawback lies in the fact that these algorithms are computationally expensive and prone to include halo artifacts [112, 92].

Within the last category (spatially varying operators), two classical currents for the tone mapping of HDR images are *gradient based* and *decomposition based* operators.
For gradient based approaches, Fattal et al. [42] proposed to attenuate large magnitude gradients within the HDR map by relying on a weighting function that is extracted in a multiscale manner. Their method results from the observation that drastic changes in the HDR map lead to large gradient responses, whereas detail and texture elements produce gradients of smaller magnitudes. The compressed tone mapped image will be reconstructed based on the new gradient map.
Letting $L(x, y)$ be the logarithm of the HDR image, the compressed gradient map is a response of the product between the gradient of L and a weighting function:

$$C(x, y) = \nabla L(x, y) \cdot \Phi(x, y) \ , \tag{2.7}$$

where $\Phi(x, y)$ penalizes big gradients and enlarges small gradients. More precisely, the weighting function captures the contribution of gradients at different scales k as:

$$\phi_k(x, y) = \frac{\alpha}{\|\nabla L_k(x, y)\|} \left(\frac{\|\nabla L_k(x, y)\|}{\alpha} \right)^{\beta} . \tag{2.8}$$

The parameters, α and β tune the compression. In particular, α establishes the limit where gradient values should be enlarged or reduced. The parameter β regulates the strength of the attenuation or amplification. Once the gradient map is compressed, it is mapped to the spatial domain by solving for an image I that best satisfies $C = \nabla I$, in a least squares sense, as in the following expression:

$$\min_I \iint \|\nabla I - C\|^2 \partial x \partial y \ , \tag{2.9}$$

which is solved with the full multigrid algorithm, after expressing the problem as a Poisson equation.

Regarding *decomposition operators*, one the first works in this category is [37]. Their method uses the bilateral filter to extract two image components from the HDR map, namely, base and detail layers. The tone mapped image is composed by the aggregation of

the detail layer and the compressed base layer:

$$L_{out} = L_{Base} \cdot \kappa + L_{Detail} \ , \tag{2.10}$$

where κ is a compression factor, L_{Detail} contains textures and fine structures, computed as $L - L_{Base}$. The component L_{Base} contains piecewise constant structures and is obtained with the bilateral filter as:

$$L_{Base}(s) = \frac{1}{k(s)} \sum_{p \in \Omega} f(p - s) \cdot g(L(p) - L(s)) \cdot L(p) \ . \tag{2.11}$$

With L corresponding to the logarithm of the HDR map and $k(s)$ a normalization factor. This filter smooths surfaces while preserving edges, for that it considers both spatial and intensity information to compute a weighted average of the input. The function managing the spatial information is the Gaussian kernel f, whereas the function that regulates intensity information is a redescending curve g that decreases the weight for large intensity differences.

Depending on the approach used, the tone mapping operation is susceptible to affect considerably the general aspect of images or to introduce artifacts in the image. In particular, the halo artifact is known to be one of the most annoying to photographers. It is characterized by an aura-like pattern at the borders of darker objects, resulting in a highly unnatural appearance.

2.2 Multiexposure Image Fusion

HDR imaging is a very good technique to surpass dynamic range limitations of digital cameras. A different alternative to the two-step procedure (HDR image creation and tone mapping) presented in the previous section is *Exposure Fusion* [91]. The main idea is to bypass the HDR creation step and to directly create a low dynamic range image (typically made of 8 bits color values) by fusing the input images obtained with different acquisition times. Specific weights are used to ensure that the final result is contrasted enough, has vivid colors and avoid over and under-exposure pixels during the combination. Using classical ideas from computer graphics [100], the images are fused in a multi-scale framework, enabling one to blend images seamlessly.

Exposure Fusion is very efficient and has yielded numerous softwares and plug-ins, e.g. Enfuse for Linux [93] or LR/Enfuse for Lightroom [7]. It has also triggered many methods that propose variants on the original idea. Generally speaking, three types of methods can be defined for multiexposure image fusion (MEIF): blending methods [83, 117, 126], ghost-

free [123, 102, 79, 135, 5] and reconstruction algorithms [115, 55, 137, 47, 98]. Whereas the first category is built on the assumption that no movement exists on the stack of images, the remaining categories aim to solve problems that emerge with motion.

Blending algorithms focus on finding new strategies to combine the image stack in an efficient manner, they can exploit illumination [91] or local features [83]. Similar works involve optimization approaches. In [117] the authors minimize a probabilistic based cost function. In [66] the image entropy is maximized. A different fusion approach was proposed in [109], which makes use of the bilateral filter to preserve details after the fusion. Nevertheless, such approaches have a common drawback: they fail when there is either camera shake or moving objects in the scene. Indeed, they all perform the fusion at pixel level and assume that images are registered and that objects are still. Of course, this is a serious limitation in practice, and several works have tackled this issue.

Anti-ghosting algorithms [123, 48, 5, 102, 138] propose to perform image alignment and possibly to explicitly detect moving objects to prevent using them in the fusion. Their goal is to compute motion maps and mostly all cases require a reference image to rule the process. To that end, intensity mapping functions (IMF) are employed to standardize the color appearance [123, 135, 138] and simple subtraction is used to detect motion. Other methods rely on extraction of general characteristics that are common within the images, as for instance the median threshold bitmap [102], or Dense SIFT descriptors [79] that should be comparable on static objects over the image stack. As for blending, the fusion weights are modified, according to the motion maps, to give priority to images that present no motion. The combination is usually performed on the full scale image resolution. Alongside, other strategies present novel approaches that make the blending richer, as in [74] where a content aware filter is used to enhance detail content.

Reconstruction methods modify the content of the stack so that all images present the same geometry as a selected reference image. They seek to optimize the displacement maps between the input images and for that they rely on the extraction of nearest neighbor fields, [115, 105, 55]. These methods all employ a patch-based optimization process whose purpose is twofold: refine the IMF estimation and improve the similarity search over each cycle [115, 55]. As illustrated in [55], exposure fusion can then be applied to the aligned set, in order to yield a satisfactory final image, even in the case of camera shake and object motions. These sophisticated and efficient methods can be considered as the state of the art for dynamic scenes.
A similar technique is proposed in [47], where the images are registered in a non rigid manner, provided the images present only small displacements.

Since the approaches presented in this thesis are directly related to this last family of methods, we now give a detailed presentation of the most representative works: Non rigid dense correspondence [51], Exposure Stacks for Live Scenes [54], Robust Patch-Based HDR Reconstruction [115] and HDR Deghosting: How to deal with saturation [55].

2.3 Reconstruction techniques for MEIF

Given a set of LDR bracketed exposure images, the aim of these algorithms is to align the content of each source image with respect to a selected image, while preserving as much as possible their original content. These algorithms not only solve global misalignments, as registration methods, but also provide alignment at pixel level on difficult conditions. Such as large displacements and strong contrast changes. They are commonly proposed as optimization problems that iteratively refine geometric correspondences while improving a parametric contrast normalization model. The former task provides geometry consistency on the aligned source by relying on patch-based methods, whereas the latter is in charge of estimating *intensity mapping functions* (IMF) that reproduce the illumination changes between the reference and each source image.

The precursor of those approaches is the work proposed by HaCohen et al. [51]. Even though this work was not particularly created for MEIF, it is shown that the generalized patchmatch algorithm [11] is able to find local transformations that reproduce radiometric and geometric changes, like scale, translation and orientation, on common objects between differently acquired images. Inspired by this, [54] uses their displacement map to estimate a parametric model that normalizes the radiometry between the images in order to check for incoherent regions in the alignment. Such corrupted regions are replaced by gradients transferred from corresponding locations on the reference or the source image after local homography transformation.
A more robust registration is proposed in [115], where the problem is presented as an optimization framework that solves for an HDR map while improving the alignment of the linearized set. They rely on a sophisticated bidirectional mapping to establish relationships over common regions and also on regions that are partially corrupted by saturation. The HDR map is updated so that only good information is transferred to saturated regions in the reference. A similar approach, completely based on LDR images is presented in [55], where the optimization aligns the images by preserving radiance and texture consistency with respect to the reference. The reconstructed images are cleverly combined with the radiometrically normalized reference in order to carry details to the saturated regions. The optimization problem is transformed to the Poisson equation that is solved in the Fourier domain.

Among these methods, the geometrical consistency is provided through a meticulous inspection of patches on their similarity and gradient information [51, 55, 54], or with a strict bidirectional search [115]. This problem is globally solved through modified versions of Patchmatch [10] or its generalized version [11].

The radiometric normalization between the source and the reference is accounted with parametric models for JPEG based methods [51, 55, 54]. Such IMF should reproduce the non linear contrast changes between the differently exposed images and is estimated numerically via least squares by imposing monotonicity constraints and incorporating the RANSAC algorithm to avoid mismatches that may alter the result.

2.3.1 Non Rigid Dense Correspondences (NRDC) [51]

The NRDC is an algorithm that searches for dense correspondences among images captured with very different acquisition parameters. The algorithm is robust to translation, rotation and radiometric transformations by using local transformations over patches. As a result, the algorithm does not restrict the input images to a particular kind of geometric deformation.
As a byproduct, not only a dense displacement map is obtained but also a global parametric color model between the images. Such color model is able to normalize the radiometry even on regions that could not be matched reliably.
The algorithm can be decomposed in 4 principal steps: nearest neighbor search, region aggregation, color model fitting and search range adjustment. Indeed, it is a multiscale method that interlaces the displacement map extraction and color model fitting to provide more stability during the matching search.
It is worth noticing that this complex pipeline is intended to model the global radiometric transformation of images that may present strong geometrical changes. This implies a reduced amount of data during the parametric estimation. In particular, this method is intended to solve color changes that are much more complex than the ones observed in MEIF. Most reconstruction methods in MEIF [55, 54] have been inspired by this work.

Nearest Neighbor Search

Given two input images, S and R, for source and reference, respectively, this step aims to find the nearest neighbor from each patch in S to the reference image, so that for each patch $u \in S$,

$$T(u) = \arg\min_{\widetilde{T}} \|S(u) - R(\widetilde{T}(u))\|_2, \tag{2.12}$$

where $T(u) = [T_x, T_y, T_{rotation}, T_{scale}, T_{gain}, T_{bias}]$ is a transformation that evaluates translation, rotation, scale, color bias and gain inside every patch u. The transformation is expected to provide enough local coherence to be able to account for strong global transformations. Radiometry is adjusted using an affine transform estimated from local mean and variance (gain and bias). Here, the patch $u \in \mathbb{R}^{4\times8\times8} = \{Lab + \|\nabla L(u)\|\}$ contains geometric information from each channel of the *Lab* color space plus the magnitude of the gradient over a 8×8 patch.
The minimization is done by using the Generalized Patchmatch (GPM) algorithm [11] for finding the nearest neighbor for each patch. For the extraction of gain and bias among the images, the patches are previously weighted with a Gaussian kernel in order to provide robustness to rotation.

Aggregating consistent regions

The displacement map is analyzed in search of coherent groups of matches since they offer more reliability of matching than individual matches. Only coherent matches are considered as input for the global color estimation. Aggregation for each match is validated in two steps. First, adjacent consistent matches are evaluated with a consistency error which checks for coherency on the projections with respect to relative displacements. This step is supported by a predefined threshold τ_{local}. Second, in order to avoid an expensive examination over all matches, only random subsets are considered. In this case, coherence region error is obtained by finding the percentage of consistent matches among all matches in a region. This step also helps eliminate small regions.

Global color estimation

The global color estimation serves a two-fold purpose. First it improves the nearest neighbor search and second it generates an image that matches the source to the colors of the reference. The parametric color fitting is able to generate good results even for regions that were not matched. For this purpose the estimation of the color model is a piecewise spline with seven breaks, plus additional constraints on the RGB curves which should be monotonic.

Search Constraints

The stability of the search is provided by imposing constraints on the coherent matches, where they are expected to present only a defined percentage of change in position, scale and orientation. This prevents the algorithm to reproduce only geometric transformations.

2.3.2 Robust Patch-based HDR Reconstruction (RPBR) [115]

The goal of this reconstruction method is to generate an HDR image from LDR images captured on dynamic scenarios. The algorithm relies on a patch-based minimization that is robust to camera/scene and object motions. By relying on the notion that LDR images overlap in the radiance domain, they propose an optimization problem to estimate the HDR map of the scene. In their formulation, each radiance-normalized input images is analyzed in search of matches, and the result is propagated to the remaining images. This combined motion invariance and radiometric normalization scheme, allows the algorithm to manage large and complex motions or fill occlusions.
The main goal of the optimization problem is to create an HDR image containing information from all exposures while being aligned to one of them.

Method

The desired HDR map H is optimized to include well exposed regions and fix errors within the reference, by using properly exposed information within the set of images. For that, an energy functional E is formulated as a weighted combination between three terms: a term that discards bad expositions within L_{ref} and two additional terms that force similarity between the HDR map and each LDR image, in a L^2-norm sense.
This is achieved by the multisource bidirectional similarity measure (MBDS) an extension from the similarity measure proposed in [116]. The MBDS provides consistency and coherence for the color mapping between I_k and the input image set.

$$\begin{aligned} E(H, I_1, \cdots, I_N) = \sum_p \Bigg(& \underbrace{\alpha_{ref}(p) \cdot (\mathbf{h}(L_{ref}(p)) - H(p))^2}_{\text{Good Expositions}} \\ & + (1 - \alpha_{ref}(p)) \sum_k \text{MBDS}(I_k | g^k(L_1), g^k(L_2), \ldots, g^k(L_N)) \\ & + (1 - \alpha_{ref}(p)) \sum_k \Lambda(I_k(p)) \cdot (\mathbf{h}(I_k(p)) - H(p))^2 \Bigg) , \end{aligned} \tag{2.13}$$

where $\mathbf{h}(L_k)$ maps the low dynamic range image L_k to the high dynamic range of H. The map α_{ref} is piecewise trapezoidal function that indicates the quality of exposition for all pixels in the reference. The operator $g^k(L_q)$ sends the radiometry of image L_q to the one of image L_k, $\Lambda(I_k(p))$ is a triangle weighting function used to compose the HDR map as in [31], and I_k is the estimated set of aligned LDR images.
The functional $E(H, I_1, \cdots, I_N)$ is minimized in a multiscale manner, where for every scale the two last terms of $E(H, I_1, \cdots, I_N)$ are minimized iteratively by applying the following two steps until convergence.

First Step - Alignment: Nearest neighbor search by computing the bidirectional displacement map between I_k and the radiometry normalized input.

Second Step - HDR Construction: By relying on the estimated I_k, the algorithm solves for an intermediate HDR map $\hat{H}$:

$$\hat{H}(p) = \frac{\sum_k^N \Lambda(I_k(p)) \cdot \mathbf{h}(I_k(p))}{\sum_k^N \Lambda(I_k(p))} \tag{2.14}$$

A final step is the combination with the reference image, where only well exposed pixels are used:

$$H(p) = \alpha_{ref} \cdot \mathbf{h}(L_ref(p)) + (1 - \alpha_{ref}) \cdot \hat{H}(p) \tag{2.15}$$

Note that to achieve satisfactory results, the algorithm has the requisite for the input images to be linear, e.g. RAW images. If that is not the case, they assume the camera response function (CRF) to be known or that it can be estimated using different alternatives [31, 94], in order to render the LDR images linear in the range [0 1]. For acceleration, they propose to limit the patch similarity search only to regions where the reference is corrupted. Also, to apply the methodology only on neighboring images, rather than the whole set. Finally, the author recognize some negative aspects, like the fact that the reference image may propagate artifacts or that pixel aggregation may introduce some blur, likely corrected at finer scales.

2.3.3 HDR Deghosting: How to deal with saturation? (HDRD) [55]

The HDRD is an iterative algorithm that aims to align a stack of bracketed exposure images, while being guided by a reference image from the set. The reconstructed LDR set is close to the reference in geometry but exposed as each image from the initial set. It takes notions from patch-based optimization algorithms to replace missing information on saturated regions within the reference. Like the previous methods, it is robust to camera and object motions.
The algorithm works iteratively by pairs; every image from the stack versus the reference, at each time building an intermediate image L that is used to propagate geometry to neighboring exposed images. The formulation consists of an energy term that reinforces the geometry and details, and a second term that accounts for illumination with the help of a global photometric transformation:

$$L^* = \arg\min_{L,\tau,\hat{N}} \left(C_r(L, R, \tau) + C_t(S, L, \hat{N}) \right), \tag{2.16}$$

where C_r assesses the radiance consistency between the modified reference R; with the color transformation τ, and the latent image L plus additional gradient information that accounts for exposure changes. The term C_t assesses texture consistency between the latent image and the mapped source with respect to the current nearest neighbor field $\hat{N}$. These terms are defined as follows:

$$C_r(L, R, \tau) = \sum_{i \in \Omega} \left(\|L - \tau(R)\|^2 + \alpha \|\nabla L - \nabla\{\tau(R)\}\|^2 \right) \tag{2.17}$$

$$C_t(L, S, \hat{N}) = \frac{1}{n^2} \sum_{i \in \Omega} \left(\left\| P_L(i) - P_S(\hat{N}(i)) \right\|^2 + \alpha \left\| \nabla P_L(i) - \nabla P_S(\hat{N}(i)) \right\|^2 \right) , \tag{2.18}$$

where $P_v(x)$ is a patch centered at position x from a given image v.

The above optimization problem is decomposed in three subproblems that are solved sequentially at each iteration:

1. Compute an initial color transformation τ based on the histogram of both images. Initialize the algorithm with $L = \tau(R)$ and use the Patchmatch algorithm to extract the map $\hat{N}$ and solve the energy term C_t. The nearest neighbor search is modified such that partially well exposed patches guide the search for a fully well exposed patch on the source. At this step, the algorithm reconstructs and adds detail on regions that were saturated.

2. With the estimated terms $\hat{N}$ and τ, redefine equation (2.16) as:

$$L^* = \arg\min_L (\|L - T\|^2 + \alpha \|\nabla L - \nabla T\|^2) , \tag{2.19}$$

an equation that resembles the screened Poisson equation and is solved in the Fourier domain [14]. The intermediate image T is defined as:

$$T(i) = \frac{1}{w} \left[w_\tau(i) \cdot \tau(R(i)) + \frac{1}{n^2} \sum_{j \in \hat{N}(i)} w_{\hat{N}}(i) \cdot S(\hat{N}(j)) \right] , \tag{2.20}$$

with w_τ and $w_{\hat{N}}$ weights that penalize bad color transformation and erroneous matches in the nearest neighbor field, respectively. The w value is a normalization factor of their combination. The image T induces the algorithm to find a suitable combination between the reference and the sources, so that only reliable (well exposed) pixels are used to synthesize the images L_k.

3. Update the photometric transformation at each iteration by solving:

$$\tau_c = \arg\min_\tau \sum_{i \in \Omega} \|\tau(R_c(i)) - L_c(i)\|_1 , \tag{2.21}$$

with $\tau'(\cdot) \geq 0$, and $\tau(\cdot) \in [0, 1]$. The above problem is solved by using weighted least squares for each color channel $c \in \{\mathcal{R}, \mathcal{G}, \mathcal{B}\}$. Additional constraints are included to find better estimations. For example, hard monotonicity on τ or forcing convexity if the reference is brighter than the source.

A final remark is that the algorithm requires the patch size to be big in order to increase the consistency of matches. Even though it slows the algorithm down, larger support implies strong map refinement, thus solving inaccuracies at pixel level.

2.3.4 Exposure Stacks of live Scenes with Handheld Cameras (ESLS) [54]

This is a systematic image registration algorithm that relies on image reconstruction and radiometric normalizations to account for non rigid deformations. By using a combination of homography estimation and image gradient transfer it solves geometrical inconsistencies on moving regions.

Methodology: first the image with the fewest under or over exposed pixels is selected as the reference. For each pair of reference-source images, a dense correspondence map is obtained between the images, in the spirit of [51]. The matched locations are used to estimate a global color transformation that maps their color distributions. It is used to localize regions that present unreliable matches, due to occlusion/disocclusion, view point changes or non rigid changes. Recovering such regions is obtained with the three following steps:

1. Initial pixel estimation through a color transfer function.
2. Local homography estimation.
3. Geometry transfer from either the source image or the reference image, via Poisson Editing.

Similarly to other methods, the images are processed separately, one after another. However, each aligned images serve as the reference to process immediate images from the stack, in order to reduce possible problems due to radiometric transformations.

Color Transfer Function

A parametric model is used to account for non linear illumination changes among the matched intensities between the reference and source images. This model is robust to noise and mismatches that could alter the color mapping. The color function is found

numerically as the solution of the following problem:

$$\tau_c = \arg\min_{\tau_c} \sum_p \|\tau_c(R_c(p)) - S_c^w(p)\|^2$$
$$\text{s.t } \tau_c'(B) > 0,\ \forall B \in [0,1],\ c \in \{\mathcal{R}, \mathcal{G}, \mathcal{B}\}\ , \tag{2.22}$$

which is solved by using cubic hermite splines, since they preserve monotonicity, in complement with RANSAC to remove outliers. As in [51] the interpolation is guided by monotonically increasing control points, that are sampled strategically from non overlapping intervals within the color domain.
Mismatched pixels are found as changes between the aligned source and the radiometrically transformed reference, controlled by a spatially varying threshold $\delta(p)$, that is a function of the local standard deviation:

$$\frac{|S_c^w(p) - \tau_c(R_c(p))|}{\sqrt{1 + (\tau_c'(R_c(p)))}} < \delta(p)\ . \tag{2.23}$$

Solving for missing correspondences

After identifying reliable matches, the algorithm aims at recovering unmatched regions or simply holes. This is achieved in two steps. First, by transferring radiometrically normalized data from the reference to the image S^w, the partially registered source. Second, by using gradient blending with the corresponding regions in the original source. To guarantee coherent results on the last step, the transfered data should satisfy a motion test and coherence test. The motion test relies on local homography estimation within defined bounding boxes, where enough pixels should be similar. The coherence test is performed with the normalized cross correlation. In case, the tests failed, gradients from the reference are used to synthesize those regions on S^w.

2.3.5 Positive & negative aspects of each reconstruction technique

Non Rigid dense correspondence (NRDC) [51]:

- ✓ Rigorous scheme for patch correspondences even at multiple scales, orientation, contrast changes.
- ✗ The method allows transformations that are too general for the exposure-fusion problem.

Robust Patch-based HDR Reconstruction (RPBR) [115]:

- ✓ Sophisticated scheme for patch correspondences that provides matches for flat saturated regions within the reference. This is a consequence of the bidirectional-search that is modified to deal with partially saturated patches.

- ✗ Even though partial saturations can be solved with a better matching scheme, it relies entirely on multiscale and neighboring propagation of coherent matches to solve for saturated regions that are larger than patches. However, this is mitigated by refining the HDR map estimation and blending only well exposed intensities.

- ✗ The algorithm works in a linear space of intensities, thus finding for correspondences is more accurate since there is not strong influence of contrast changes. This is a disadvantage as non linear JPEG images are common.

Hu et al. (HDRD) [55]:

- ✓ The nearest neighbor search accounts for totally saturated patches. However, it loses stability on big saturated regions.

- ✓ The latent image is enriched on saturated regions through a weighted combination of the reconstructed image, in order to reduce the extension of saturated regions.

- ✗ Radiometric inconsistencies like fake colors, may result due to a badly estimated IMF, which is not corrected during the weighted combination. Saturated regions in the images are not necessarily identical on all channels, thus performing the estimation of the IMF independently may corrupt the quality of the normalization.

Exposure Stacks of Live Scenes (ESLS) [54]:

- ✗ The algorithm does not account for large saturated regions in the reference. In the worse case scenario the method would recover the reference on inconsistent matched regions. Also, the alignment would fail on saturated regions on the reference that are occluded by big objects on the source.

- ✗ As reported in the paper, the process of blending gradients to the reference may be not very effective and would give rise to geometrical inconsistencies.

Our claim is that all previous methods are derived from the approach of [51] which consider general color transforms that are unnecessarily general for the problem of exposure fusion. As a result, and in order to make robust the approach, all these methods include complex and time consuming optimization stages. Instead, and inspired by [3], we propose a patch-based MEIF framework that is robust to motion in the images. It is composed by five fundamental stages: image alignment, reference enhancement, radiometric normalization, search & reconstruction and fusion.

As we will see, this scheme is more simple yet effective, without the often complex and unnecessary optimizations of the previous methods, but without compromising the quality of the resulting image. Our method will be introduced in Chapter 4.

2.4 Multifocus Image Fusion

Imaging devices usually have a limited depth of field, and objects outside this area are rendered out of focus. This is especially problematic with large sensor cameras, large apertures or even on macro photography.
In order to get an image that is sharp everywhere, it is necessary to combine images from the same scene acquired with different focal settings [57]. This task is usually referred to as multifocus image fusion (MFIF). The aim of MFIF is to extract local information that is combined in order to synthesize an image that best describes the scene in terms of sharpness.
Clearly, most tasks involving object recognition will benefit from a technique that renders the whole scene as sharp. Unfortunately, illumination and optical constraints do not always allow for a large depth of field. These constraints motivate the need for some form of combination of various acquisitions.

Multifocus fusion techniques are basically built on two steps: sharp region localization and fusion. They can be categorized based on how the information is extracted, which domain (image domain or transform domain) it is taken from and how it is combined. Classically, a decision map is first computed, either at pixel level [71], using patches [131, 57] or computing regions [73, 82, 29]. In all cases, the identification relies on a sharpness measure, which can be computed through various differential operators [57], through spatial frequency [61, 71, 73], drawing on wavelet decompositions [134, 121], using dense SIFT [78] or through the use of a convolutional neural network [72, 77]. The final category is based on the approach used to combine the images, which usually necessitates some care to avoid artefacts and visible seams. Some popular approaches include a weighted sum of the chosen pixel values or multiresolution-based methods. The general strategy for the latter techniques is based on the rule: decompose-choose max-recompose. Common transforms include the Laplacian pyramid [43] and discrete wavelet transform [134, 21, 127, 121, 52].

Pixel based fusion methods rely on a set of weights to linearly combine the images. In [122], the authors define a focus measure that is based on two parameters: strength and coherence of image gradients. The output of such measure is used to build a weight map that is utilized to combine directly the set of images. Another way is to fuse in a binary fashion by selecting only the information from the image whose focus appears to be the largest among the set, [73]. Thence, the accuracy of the sharpness map is crucial to the final quality. The major contribution of [29] is to build a tree-shaped structure that labels regions as blurred or sharp. For the tree construction, they use an iterative algorithm that adaptively modifies the block size of regions by estimating sharpness in the surrounding area. They evaluate focus as a function of morphologic operators over the magnitude of the

image gradient, called energy of morphologic gradients. The fusion is directly performed as in [73].

In general, wavelet-based fusion schemes compute the fused image by combining the subband components of the set of images in a defined fashion. That is, a final detail component is defined by taking the larger value among the highpass subbands. Some methods explicitly seek to identify the sharp regions based on the wavelet decomposition. A final lowpass component is defined by averaging the corresponding subbands, [134, 21, 127, 121]. In [21, 127] a wavelet transform is performed on the set of images so that lowpass subband and highpass subband are obtained. They use the average of the lowpass to get a fused component, similarly they take the larger value to obtain a highpass subband, then recompose. Oppositely, [121] performs a weighted combination for the lowpass subband, arguing that average combination leads to a contrast reduction. Also, noticing the marginal distribution of wavelet coefficients behave differently for sharp and blurred images, they define a wavelet domain measure to assess depth of field. Such measure is a locally adaptive statistic based on the Laplacian mixture that is parametrized with some standard deviation values σ (large σ means sharp objects). The recognized in-focus regions are used to reconstruct an image.

Other methods approach the MFIF problem differently. In [131], a sparse representation is calculated for each patch inside the set of images. The sharp image is reconstructed by taking the larger coefficient among all representations and relying on a limited set of atoms that define a dictionary. Other approaches build an intermediate fused image that is used to extract a sharpness map which is employed to guide the fusion, [82, 73]. Some methods rely on the camera response to counteract the blurring effect they generate as in [24] where the authors analyze the behavior of a particular case of blurring.

The strongest limitation of all MFIF methods is their limited ability to deal with motion, be it camera shake or motion due to moving objects in the scene. A classical way to limit the influence of camera shake is to register the images before the fusion [136, 49]. However, this approach fails when the scene is not planar or when the camera has moved away from its optical center, which is very common in practice and yields mis-registrations. As a result, ghosts appear in the final fused image. Methods working at pixel level are particularly exposed to these errors. Region based methods are more robust to small mis-registration errors, see e.g. [29], but to a limited extent. Moreover, none of these methods is able to deal with moving objects.
In a different direction, and still to deal with motion, several works have proposed to refine the decision map using spatial coherence, either through image matting techniques [70] or through the use of dense SIFTs [78]. By doing so, these methods attain a better robustness to mis-registration errors or to object motions. However, the decision map they refine is built from all images under the hypothesis that the geometrical content is the same in

these images. For this reason, they cannot deal with non-rigid deformations or strong mis-registrations. More generally, none of the methods in the literature can ensure the global geometrical coherency of the result in case of general objects or camera motion.

2.5 Patch-Based Image Processing

Following the seminal NL-means paper [16], the use of patches has led to the development of very efficient tools within diverse areas, like computer vision, image processing or computational photography. In particular and after the influential Patchmatch algorithm [10], many works have been relying in patch correspondences to synthesize new images or new images parts.

Formally, the term *patch* refers to neighborhoods or image segments of certain shape and size, most generally squared. In this thesis, we denote a patch as $P_I(x)$ a squared window of size $n \times n$ centered at position x within image I. Patches are a powerful tool because of two characteristic properties, redundancy and distinctiveness. First, image patches are highly redundant, presenting strong correlations with other patches at different locations. Second, patches are endowed with discriminant content, providing information about the structures within objects. These two properties combined have been exploited to develop state of the art techniques for texture synthesis [39, 38], image denoising [19, 68] or image edition [10].

The popularization of patches started within the field of texture synthesis. There, patches were successfully used to replicate exemplar image textures [39]. Texture synthesis methods deal with a uniform reproduction of exemplar textures, that are normally, characterized by irregular or periodic patterns. The main idea is to copy small patches within the exemplar image and merge them in the region where the texture should be rendered. This is done under the condition that visual discrepancies should not be created in the process [27, 12]. The seminal work in [39] was later extended by Efros et al. [38] and many papers followed, for instance [75, 22, 128].

Few years later, patches changed the way image denoising was done, this gave rise to a complete new wave of denoising methods producing mainly two currents. One based on non local image denoising [16, 64, 17, 15], and other one relying on sparse methods using image dictionaries [40, 140, 35, 26, 68].

Another field where patches have had tremendous impact is image edition. Techniques like Patchmatch [10, 65, 56, 53] have inspired the creation of very useful tools for image or texture synthesis in a reasonable time. Some important applications are

image retargeting [113], image inpainting [25, 129, 18, 6, 96] or style transfer [13, 81, 88, 45].

In the following chapters we explain how patches can be used to reduce the impact of motions and allow for a local geometric alignment on images acquired under dynamic scenarios.

Chapter 3

Patch-Based Image Reconstruction

For the purpose of digital image edition, artists often face the task of removing undesired objects and replace their gaps with information that should not alter the general coherence of the image. If the desired substitute or target information is contained somewhere within the original image, a copy and paste edition technique [103] would account for it. As for cases where the target information is unknown, one should turn to image synthesis techniques.

Image synthesis is the task of rendering automatically a target image (or region), by using the information provided in supplementary images or the target image itself. This is typically achieved by finding the best regions to fill the gaps. By best regions we refer to small neighborhoods or image patches that hold strong similarity in geometry and color. A patch satisfying those conditions is usually called nearest neighbor or best match. Nearest neighbors are localized by using search techniques that incorporate prior spatial knowledge to ensure precision and avoid computational burden. Some important applications of image synthesis are, texture synthesis and image inpainting for automatic hole filling.

Since the goal of this thesis is to provide robustness to motion in image fusion methods, we suggest to use patch-based image synthesis during the combination. In fact, this idea is similar to the classical way to densely estimate motion, a setting called block matching [104]. Of course, for our purpose it should be done by considering various image distortions: geometric transformations, contrast and color changes and blur.
In this chapter we use the Patchmatch algorithm [10] to efficiently find patch correspondences and synthesize a reference image with the content of a supplementary image, a process that we call image reconstruction. Here, we show that using patches solves local geometric misalignments for images with similar acquisition settings, and explain why

the standard nearest neighbor extraction turns out to be inadequate under illuminations changes or blur perturbations. For such cases, other alternatives are investigated.

3.1 Image Reconstruction

The purpose of image reconstruction is to synthesize a reference image by homogeneously accommodating nearest neighbors sampled from a source image. This results in a new image that replicates the geometry of the reference. We can think of image reconstruction as a technique to guarantee both global and local image alignment.
Image reconstruction algorithms rely on the hypothesis that information synthesis can be achieved through local coherence of multiple fragments of information. Local coherence is satisfied when the rendered information is provided from similar neighborhoods. To that aim, there should exist a measure that provides the degree of resemblance between local neighborhoods or patches. This is normally achieved by the L^2-norm (sum of squared differences (SSD)). Upon a clever strategy of search, this measure is utilized to localize nearest neighbors (ideally found within identical objects between the images), generating a nearest neighbor field.
As we will explain in the next section, prior knowledge about the potential position of the nearest neighbor is utilized to accelerate the process, for instance to confine the search of patches to a neighborhood of certain size, usually called the *search window*.

Given images $R : \Omega_R \to \mathbb{R}^3$ and $S : \Omega_S \to \mathbb{R}^3$, for reference and source, respectively. Let us assume that a nearest neighbor field $M : \Omega_R \to \Omega_S$ is extracted (possibly with the techniques explained below), where for each patch inside R the map returns the coordinate of the nearest neighbor within S, thus the reconstruction is as follows:

$$\widetilde{R}(p) = S(M(p)) \ , \tag{3.1}$$

where $p \in \Omega_R \subset \mathbb{Z}^2$ and $M(p) \in \Omega_S \subset \mathbb{Z}^2$. Basically each pixel on image $\widetilde{R}$ is synthesized with only the pixel linked to the position $M(p)$ in the source.

A graphical description of the notion of image reconstruction is displayed on figure 3.1. Note that for this case, two consecutive pixels (at p and $p+1$) are picked from the same area on the source, which is an indication of regularity within M. That may occur, for example, when two objects are exactly matched. Conversely, irregularly extracted pixels, like pixel $p-1$, may be part of another object.
In fact, depending on the application, the regularity in the map of correspondences M may be more or less important. As we will see, regularity does not always imply a better reconstruction and also inconsistent maps are capable of producing very good results. On

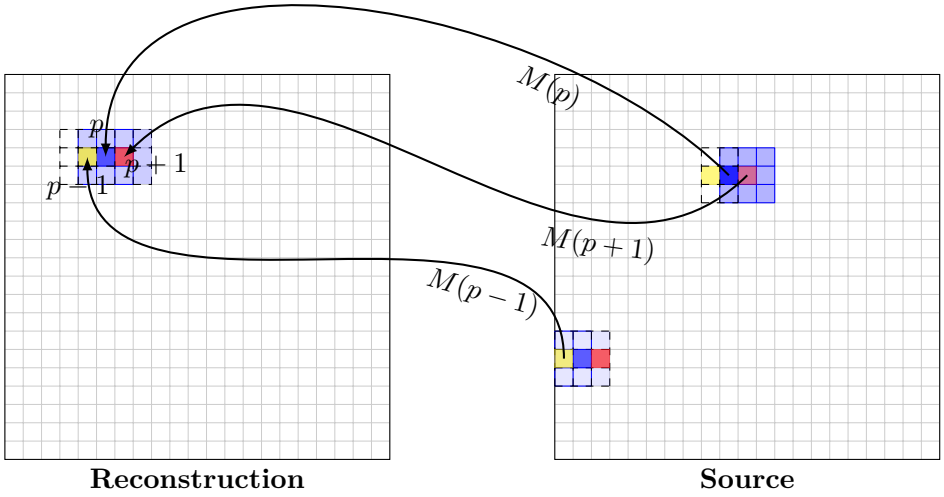


Fig. 3.1 Image reconstruction technique using patches of size 3×3px. The diagram displays the reconstruction of three pixels $p-1,\ p,\ p+1$ which are created by using the central pixels associated to patches from $M(p-1),\ M(p),\ M(p+1)$ on the source image.

this matter, there are methods that sacrifice exactitude on the reconstruction for regularity on the correspondence map or just for faster computational performance. Those methods are called *Approximate Nearest Neighbors* algorithms.

3.1.1 Exhaustive Search Algorithm

The exhaustive search or brute force algorithm, is a straightforward approach to search for nearest neighbors, where essentially, there is no efficient strategy for localization. Here, all possible patches in the reference image are compared with all possible patches in the source image, which results in the identification of the best match. Despite of its simplicity, its computational efficiency is very low, the order of $O(mN^2)$ for images and patches of N and m pixels, respectively.

3.1.2 The Patchmatch Algorithm

The *Patchmatch algorithm* [10] is a highly efficient method for approximate nearest neighbor retrieval. Initially proposed for dense correspondences, it is used for image editing, completion or retargeting. Beginning with a random initialization, the algorithm automatically matches all possible neighborhoods in the reference with the most similar patch in the source. This is done by simply performing two sequential operations that reduce a similarity energy (originally defined through the L^2-norm) after every cycle. More precisely, the Patchmatch algorithm aims at solving the following problem:

$$\underset{M(x)\in\Omega_S}{\text{minimize}}\ E(M) = \sum_{x\in\Omega_R} D(P_R(x), P_S(M(x)))\ , \tag{3.2}$$

with possibly some constraints on $M(x)$, like $M(x)$ being not too far from x. Here Ω_R and Ω_S are the spatial domain of images R and S, respectively, and $M(x)$ is the nearest neighbor field (NNF) at pixel x. The distance D is usually computed as the L^2-distance between the patches $P_R(x)$ and $P_S(y)$ around x and y.

Patchmatch [10] proceeds in three fundamental stages: random initialization, propagation and random search, described hereafter.

i. **Random initialization:** Initialize the NNF with a random coordinate, mapping each patch in the reference to a patch located at certain spatial position inside the source.

ii. **Propagation:** Propagate in scan order. Each pixel x looks for $M(x - v) + v$ as possible better match than the current. Where v is in $V = \{(0, 1), (1, 0), (1, 1)\}$ for odd passes and $-V$ for even ones. If a better match among the three is found, replace $M(x)$ by $M(x - v) + v$.

iii. **Random search:** Search randomly for similar correspondences at different distances around the current candidate match.

The success of this algorithm is due to both computational and structural advantages. First, the algorithm capitalizes coherence within the image by propagating good matches after the random initialization, and therefore minimizing the energy. Indeed, the amount of good matches after initialization is large and the probability for at least one match being well assigned is $(1 - (1 - 1/N)^N)$ for an image of big size [10]. Secondly, the flexible methodology permits additional constraints to better achieve suitable matches, like initialization with predefined NNF or matching over restricted areas within the source image. Additional parameters to increase the precision of the match include: the search window size (affects the maximum displacement of nearest neighbors), the patch size (controls the degree of similarity between patches), the number of iterations or the reduction factor to reduce the space of random search.
In addition, a generalized version of this algorithm [11], is capable of dealing with changes in rotations, scale, translations and accept dense image descriptors or different similarity measures.

Below we present some remarks about the Patchmatch algorithm:

- If the SSD is used, the main constraint to obtain a reliable nearest neighbor field is that both images must have similar appearance. That is, they should share the same palette of colors.

- Although patches are used by default, descriptors can also be used for the extraction of nearest neighbor. Particularly they are well suited for dense matching of common objects within the images.
- The patch size may alter locally the quality of the reconstruction. For smaller patches, the synthesized image will be more accurate on details. Yet, it would be less efficient to transfer local properties from the source.
- Depending on the image content, the algorithm can get stuck in local minima. Another factors for this to happen are the initialization and the patch size. This means that unknown or moving objects can be appropriately synthesized by sampling nearest neighbors from random locations. It also means that exact matches are not guaranteed without additional constraints.

3.2 Patchmatch & Image Perturbations

The Patchmatch algorithm has become a very useful tool for image editing applications and it has served as inspiration for multiple new methods. The purpose of this section is to study this algorithm, for image reconstruction when images undergo various real perturbations.
Particularly, the algorithm is evaluated on images taken from the same scene with a handheld camera under different changes, such as: translations, rotations, illumination and focus changes.

For the experiments we include 3 configurations of the algorithm. The standard Patchmatch algorithm which we denote as (SPM). The descriptor based configuration (DPM), which consists of finding nearest neighbors between dense SIFT descriptors from both images. Lastly, the generalized version of Patchmatch (GPM).
In this study, we used the original implementation of the algorithm [10, 11]. Below, the settings used to configure the Patchmatch algorithm and its variations:

- *Standard Patchmatch* (SPM): patch size 5×5, 4 iterations.
- *Descriptor based Patchmatch* (DPM): 4 iterations. Dense SIFT descriptors, extracted from grayscale images at a single scale with the standard configuration (4×4 cells and 8 bin orientations) [76].
- *Generalized Patchmatch* (GPM): patch size 5×5, 20 iterations. Default parameters for rotation and scale, meaning variations in the range of [0, 2π] radians and scale [0, 2×], respectively.

3.2.1 Translation

Among the set of images (figure 3.2), the perturbation within the source is mainly approximated by a global translation and unnoticeable illumination changes. Moving objects, some of which vary in scale, are also present. Notice that the source image, figure 3.2(b), is displaced to the right, discarding the set of windows located on the left side of the building. Also, we can notice a cyclist disappearing on the bottom left corner.

From the reconstruction and their map of correspondences, figure 3.3, we can easily observe that compared to the SPM, the descriptor based configuration (DPM) generates a uniform offset map, even under image misalignments. However, it is not very precise at detail level. Although the DPM matches densely most of the common regions on both images, it fails to reconstruct accurately regions with scale changes or totally unknown regions, in particular the set of windows and the biker.
The strong coherency displayed by the descriptor based setting indicates that mismatches are highly penalized, which is good for dense match applications but inappropriate for the reconstruction of unknown regions.
On the contrary, the SPM mode builds an offset map that is less coherent, but capable to render accurately objects at detail level, as well as objects that do not appear on the source. This is because the SPM associates patches that do not necessarily belong to the same object in the two images, but rather belong to other similar or repeated regions scattered in the image. For instance, notice that the biker has red shorts on the reference image, but the red color appears only in one region over the source image (street light at the top right), still, Patchmatch localizes those regions and reconstruct the biker's shorts.

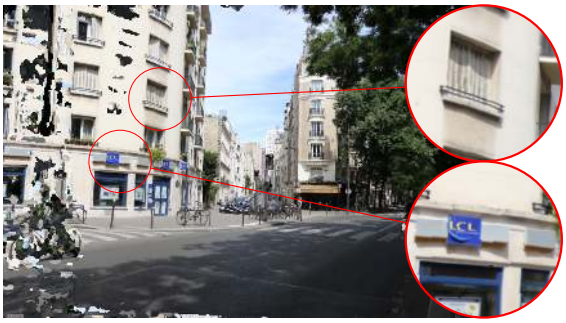
The observations are confirmed by the quality of reconstruction, where the PSNR is much lower for the DPM than for the SPM. However, even though pixel accuracy is higher for the SPM, the reconstructed image (figure 3.3(a)) suffers from visible jitter artifacts that are reduced with a refinement technique (described in Section 3.3), see figure 3.4(b).

(a) Reference. (b) Source.

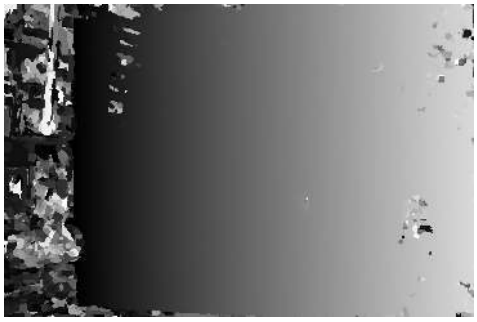
Fig. 3.2 Input image set with visible Translation to the right.

(a) SPM Reconstruction. PSNR = 26.94.

(b) SPM NNF Map.



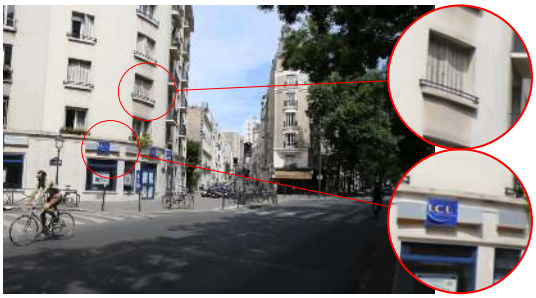
(c) DPM Reconstruction. PSNR = 16.24.



(d) DPM NNF Map.

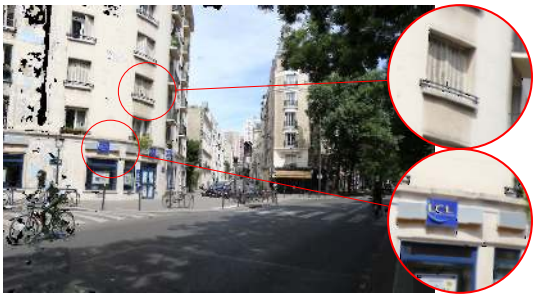
Fig. 3.3 Image reconstruction and index-based correspondence map ($x \times h + y$), for images perturbed with translation.

(a) SPM Patch Aggr., PSNR = 26.53.



(b) SPM Multi-NNF Aggr., PSNR = 32.97.

(c) DPM Patch Aggr., PSNR = 16.95.



(d) DPM Multi-NNF Aggr., PSNR = 18.85.

Fig. 3.4 Refinement of the reconstructions under different modalities, (see Section 3.3).

3.2.2 Rotation

This is another real case occurring due to handheld camera motions. Here the set of photographs (Figure 3.5) presents a slight rotation in the counter-clockwise direction. This results in loss of structures within the source, like the top right building and part of the tree located at the bottom left region. Also, the scenario is highly dynamic on the background. Here, besides the standard mode (SPM) and descriptor mode (DPM), for this experiment, we also include the generalized version of Patchmatch (GPM), which accounts for rotations and scale changes.

(a) Reference (b) Source

Fig. 3.5 Input image set with visible rotation.

From the reconstruction and their map of correspondences, figure 3.6, we can see that the SPM produces the best reconstruction when compared to DPM and GPM. Once again the SPM shows superiority to synthesize an image that is consistent with the reference even in the presence of rotations or unknown structures and moving objects. As before, the information is sampled from several regions with no apparent consistency. It should be noted though, that it presents irregular patterns on and at the transition of unknown structures.

Conversely, the DPM compensates the rotation nearly for all the common structures. This exhibits SIFT capabilities to tolerate rotations [80] . However, visible large patches with strong contrast changes illustrate the inability of SIFT to capture radiometry (SIFT being built from gradient orientations). For those cases, the internal geometry fits but the gray levels are different. Take for instance the white line road marking at the bottom right of the image (figure 3.6(c)). As a result, the final reconstructed image resembles the reference in geometry (to a limited extend) while its illumination and contrast are determined regardless of the reference image. For unknown objects, it is incapable to synthesize coherent structures.

(a) Reconstruction SPM. PSNR = 25.60

(b) SPM NNF Map.

(c) Reconstruction DPM. PSNR = 19.96.

(d) DPM NNF Map.

(e) Reconstruction GPM. PSNR = 19.13.

(f) GPM NNF Map.

Fig. 3.6 Image reconstruction and index-based correspondence map, for images perturbed with rotation.

Concerning the GPM mode, it produced the worst reconstruction with high presence of local deformed regions and inaccuracies at pixel level everywhere on the image. No precise matching of rotated/scaled regions of the images is found.

As explained later on in Section 3.3, the quality of reconstruction can be improved with several refinement techniques. The reconstructions with the above methods SPM, DPM and GPM are refined with two alternatives and presented in figure 3.7.

(a) SPM Patch Aggr., PSNR = 26.07.

(b) SPM Multi-NNF Aggr., PSNR = 31.72.

(c) DPM Patch Aggr., PSNR = 20.92.

(d) DPM Multi-NNF Aggr., PSNR = 21.12.

(e) GPM Patch Aggr., PSNR = 21.

(f) GPM Multi-NNF Aggr., PSNR = 28.03.

Fig. 3.7 Refinement of the reconstructions under different modalities.

3.2.3 Illumination

The set of images included in the experiment contain mainly illumination changes, figure 3.8. Some geometrical changes, like moving objects, are also present.

(a) Reference (b) Source (c) Normalized reference.

Fig. 3.8 (a-b) Bracketed exposure images acquired with 2 stops of difference. (c) Radiometrically normalized reference by using histogram specification (section 4.2.1). Images from [102].

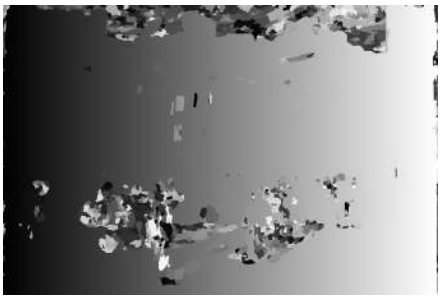
The reconstructions and the corresponding NNF's are shown in figure 3.9. Under illumination changes, the SPM and the DPM decrease their performance. This is expected because the SSD is lost when confronted with strong radiometric changes, where geometrically identical patches with contrast changes are not recognized as exact. Consequently, the resulting image with SPM, fails to reproduce structures. Another reason for the SPM to go wrong is the patch size, that is because small patches may not include enough geometrical landmarks to be matched with other patches on the source image. Both these factors combined make the image irregular, with quantization-like effects, figure 3.9(a).
An additional factor for DPM to fail is the limited robustness of SIFT descriptors to strong illumination changes and their low capabilities to encode textureless regions. This is best observed on flat areas of the reference, where the reconstruction contains only noise. However, since SIFT descriptors excel at encoding geometry, detailed structures are well matched. This yields exact transfer of geometry and color for those regions, and therefore the closer aspect to the source and the reduction in the PSNR measure (because the radiometry of the source and reference are very different).

As both configurations fail to preserve local consistency on the reconstruction, an option is to normalize the radiometry of both images and produce closer images in appearance. Figure 3.8(c), shows the radiometry of the reference image normalized to the source's radiometry via histogram specification. By using the color normalized image, we show that the reconstruction is much better, especially for the standard Patchmatch. For that case, figure 3.9(e), the SPM synthesizes an image that is globally coherent with the reference and the appearance is successfully transfered from the source. This is confirmed in the NNF, where almost exact matches were obtained for textured static objects.

(a) SPM Reconstruction. PSNR = 22.93.

(b) SPM NNF Map.

(c) Reconstruction DPM. PSNR = 9.99.

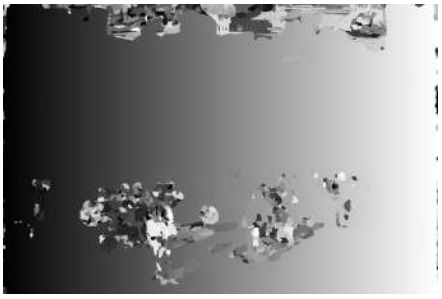


(d) DPM NNF Map.

(e) SPM Reconstruction after hist. Spec., PSNR = 27.37.

(f) SPM NNF Map after hist. Spec.

(g) DPM Reconstruction after hist. Spec., PSNR = 19.70.



(h) DPM NNF Map after hist. Spec.

Fig. 3.9 Image reconstruction and index-based correspondence map ($x \times h + y$), for bracketed exposure images and radiometrically aligned images.

(a) SPM Patch Aggr., PSNR = 22.32.

(b) SPM Multi-NNF Aggr., PSNR = 27.64.

(c) DPM Patch Aggr., PSNR = 10.08.

(d) DPM Multi-NNF Aggr., PSNR = 8.32.

Fig. 3.10 Refinement of the reconstructions under different modalities.

(a) SPM Patch Aggr., PSNR = 26.38.

(b) SPM Multi-NNF Aggr., PSNR = 31.39.

(c) DPM Patch Aggr., PSNR = 20.39

(d) DPM Multi-NNF Aggr., PSNR = 18.61.

Fig. 3.11 Refinement of the reconstructions after radiometric normalization.

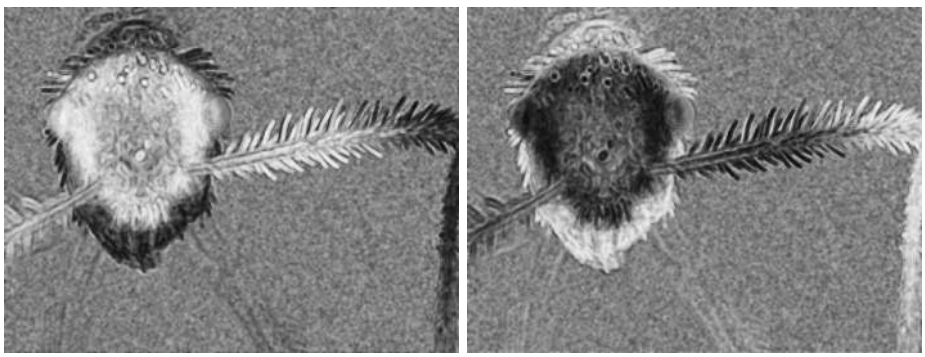
In the case of 'DPM & the normalized reference' figure 3.9(g), slightly better results were obtained than without the normalization, over small flat regions surrounded by textured areas. Regarding large flat regions, the performance remains limited. As shown in figures 3.10 and 3.11 we also tried refinement techniques to improve local consistency.

To conclude, we can say that there are two alternatives to deal with illumination changes. One possibility is to use invariant descriptors, which are commonly not discriminative enough with respect to geometry, therefore many mistakes are produced. Another possibility is to pre-process the image by either applying a global contrast change or by matching the histogram of the images. As we will see in the next chapter, the second option is more effective.

3.2.4 Depth of field

For this experiment we use a pair of perfectly aligned images with complementary levels of sharpness, figure 3.12. As before, the aim is to reconstruct the reference based on the content of the source image.

(a) Reference (b) Source



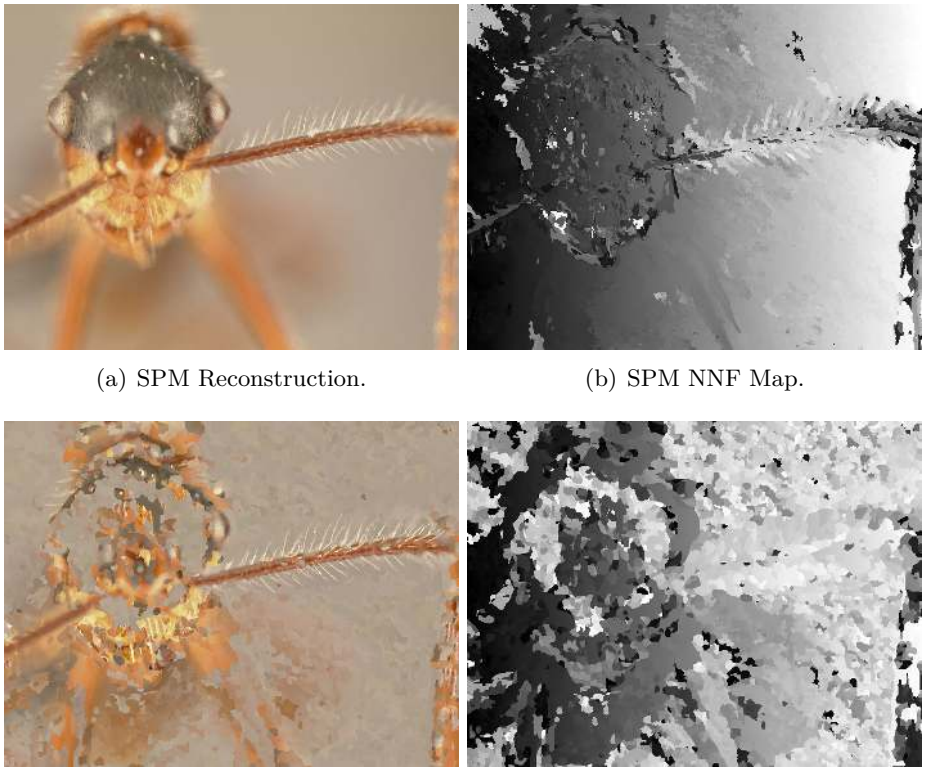
(c) Reference sharpness. (d) Source sharpness.

Fig. 3.12 (a-b) Reference and Source images under complementary levels of blur. (c-d) Normalized sharpness values extracted with the Local Total Variation, with sharpness being proportional to brightness. Images of public domain.

Given that the images do not present geometrical deformations, translation or rotation, we expect for the best matches to be located at the same location, therefore producing an identity map. Nevertheless, the results for both reconstruction and NNF's, display strong incoherences. These irregularities originate because the similarity measure fails to associate identical objects with different frequency content.

Under focus changes, the reconstruction with SPM is a much better approximation than DPM's reconstruction. However, it presents strong irregularities at pixel level and transfer of sharpness from the source is not achieved.

The failure of DPM can be attributed to the strong level of blur that corrupts the content of the images. Since identical objects present different levels of blur, local structures are codified differently by SIFT, leading the Patchmatch algorithm to be easily locked on local minimum areas. Once again, we can see how SIFT descriptors fail to encode flat regions, which results in the matching of flat regions of different colors.



(a) SPM Reconstruction. (b) SPM NNF Map.

(c) DPM Reconstruction. (d) DPM NNF Map.

Fig. 3.13 Image reconstruction and index-based correspondence map ($x \times h + y$).

3.3 Image Reconstruction Refinement

As we just saw, image reconstruction is a task that is highly dependent on the quality of the input images, and for which changes in radiometry or focus can yield very poor results. In fact, even if the input images do not present perturbations of that kind, subpixel accuracy is not guaranteed either, (see Sections 3.2.1 and 3.2.2, for geometrical changes only).
A common practice to alleviate such geometrical errors is to prealign the geometry of the images by using geometric transformations, i.e. homographies. In the same manner, other normalization techniques can be used to counteract illumination changes or blur. These techniques will be reviewed in Chapter 4 and Chapter 5.

We now present alternatives for refining the reconstructed image by exploiting redundancies within the map of correspondences (as *Patch Aggregation* detailed in Section 3.3.1) or relying on different reconstructions from the same scene (as *Multi-NNF Aggregation* detailed in Section 3.3.2).

3.3.1 Patch Aggregation

Patch aggregation or patch voting, is a technique proposed to regularize the image after reconstruction and is based on the map of correspondences, figure 3.14. This simple regularization technique was initially proposed in the original Patchmatch paper [10]. The idea behind this method is that adjacent matches should be coherent in the vicinity of the current match, in order to preserve local coherency.
More precisely, given a coordinate $x \in \Omega_R$ and the map of correspondences $M : \Omega_R \to \Omega_S$, the new aggregated pixel $\widetilde{R}_a(x)$ is obtained by averaging the overlapping pixels, within a local neighborhood $\mathcal{N}$ around x, mapped with the NNF, as follows:

$$\widetilde{R}_a(x) = \frac{1}{n^2} \sum_{p \in \mathcal{N}} S(M(x+p) - p) \ , \tag{3.3}$$

where the local neighborhood $\mathcal{N}$ around 0 is of size $n \times n$.

3.3.2 Multi-NNF Aggregation

This approach consists of reconstructing an image based on multiple maps of correspondences M_i, and was introduced in [11]. Particularly, we use the Patchmatch algorithm in '*knn*' mode, meaning that it extracts m nearest neighbors for each patch inside the reference. This yields m different NNF's that are used for image synthesis in a weighted combination style:

$$\widetilde{R}_{knn}(x) = \frac{1}{\sum_i w_i(x)} \sum_{i=1}^{m} S(M_i(x)) \cdot w_i(x) \ , \tag{3.4}$$

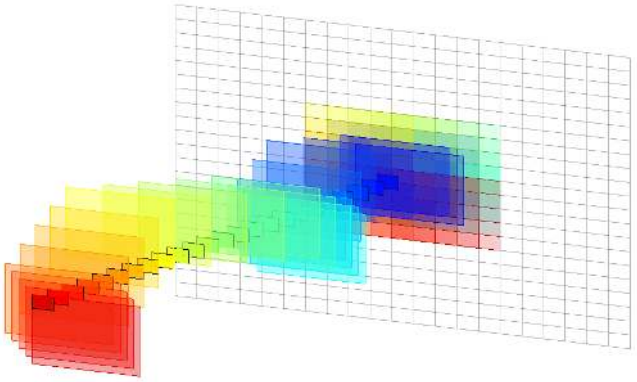


Fig. 3.14 Refinement of image reconstruction through patch aggregation. The idea is to employ pixel redundancies for the estimation of the final pixel. For that we use the pixels generated from the intersection of all possible nearest neighbors that contain the position x. In the diagram we use 5×5 patches, which results in an intersection of 25 pixels to be used in the aggregation.

where $x \in \Omega_R$ and $w_i(x)$ is proportional to the quality of reconstruction for each $i-th$ nearest neighbor,

$$w_i(x) = \exp\left(- \frac{\|R(x) - S(M_i(x))\|^2}{2\sigma^2} \right). \tag{3.5}$$

The σ value is set low, in order to control the maximum error permitted, thus reducing the influence of mismatches.

3.3.3 Yaroslavsky based NNF filtering

Here we propose to use filtering notions to remove spurious matches and enforce regularity within the NNF. The idea of forcing regularity in the displacement map (instead of the reconstructed image), comes from the fact that structures or objects within the image are represented by groups of consecutive coordinates. This statement is true for exact nearest neighbors.
Particularly, we use the *Yaroslavsky* filter because it allows us to include some notions of the type of degradations within the map while being relatively fast. We are interested in solving jitter artifacts with the content of local windows while allowing us to control the strength of refinement. The same idea can be extended to a patch-based filtering, like the Non-local means.

For that, we take the displacement map as a vector field that defines the set of spatial coordinates of nearest neighbors $M(x) \in \Omega_S$ and impose regularity based on the magnitude of the relative displacement around each position $M(x) - x$. This is done by filtering the

vector field, as follows:

$$\widehat{M}(x) = x + \frac{1}{\sum_t w_x(t)} \sum_{|t|<p} \Big(M(x+t) - (x+t) \Big) \cdot w_x(t) \ ,$$

with

$$w_x(t) = \exp \left\{ - \frac{\Big\| M(x+t) - (M(x)+t) \Big\|^2}{2\sigma^2} \right\} \ , \tag{3.6}$$

the σ value is a parameter that defines the maximum relative displacement allowed around each direction. The parameter p defines the extend to which the filter is applied locally. This is exactly the bilateral filtering applied to the vector field $M(x) - x$.

Finally, we use the two dimensional cubic interpolation to map the vector field to discrete positions within the source image S.

Experiments with this refinement technique will be presented in the Chapter 5.

3.4 Discussion

Here we used image reconstruction to align the geometrical content of a source image to the reference. For image reconstruction we employed an efficient extraction of dense correspondences that help us compensate for motions on the source. Several experiments were conducted to study the performance of the algorithms (standard-based SPM, descriptor-based DPM and generalized Patchmatch GPM) for images with geometrical changes. As well, we explored the image reconstruction on bracketed exposure and multifocus images.

The experiments let us conclude that the standard Patchmatch (SPM) configuration is general enough to produce good results on images that have only been perturbed geometrically with rotations or translations, producing accurate reconstructions even when the geometry of the objects is totally different. However, under different degradations like illumination changes or blur, the algorithm loses its ability of reconstruction.
Concerning the SIFT based Patchmatch (DPM), it achieves exact matches for common objects between the images while disregarding translations and moderate variations in rotation and contrast. Other cases like unknown objects or textureless regions within the image, reduce the accuracy of the DPM configuration. But not as much as blur perturbations.

In all cases, the synthesized images through image reconstruction with SPM, DPM and GPM were refined with the techniques described in Section 3.3. We noticed that this

step is necessary to improve the quality of reconstruction and correct errors. In general, the refinement with Multi-NNF aggregation is the best for solving detail inaccuracies and rendering a faithful reconstruction. As for patch-based aggregation, it showed to be very good at removing artifacts to the cost of filtering details when the synthesis was not consistent.

General remarks about the algorithms and experiments can be categorized with respect to the kind of perturbation within the images, namely, geometric changes, illumination changes and blur.

- For geometric changes, we observe that they can be easily removed by using the SPM configuration. Since its precision is good but limited, the synthesized image may exhibit jitter artifacts. For those cases, the Multi-NNF filtering is very efficient.
- For illumination changes, the SPM is not robust to identify nearest neighbors that contain the same geometry. The DPM is more robust to contrast changes but is useless on flat regions or objects that are not present in the source. We found that image pre-processing to normalize the radiometry of the images (e.g. affine contrast changes on RAW images or contrast specification on JPEG images with similar content) is very efficient. These observations are successfully used in Chapter 4.
- Regarding blur distortions, both configurations are very weak. One alternative to deal with nearest neighbor search in the presence of blur, is by conducting a more constrained search, meaning enlarging the patch size and reducing considerably the search window size. This is an option if the images do not present large misalignments, for which image registration is necessary. This alternative is explored in Chapter 5.

3.5 Conclusions

In this chapter we performed extensive experiments on the popular Patchmatch algorithm for the image reconstruction problem by using images acquired with different types of perturbations.
Provided few conditions on the input images (mainly identical sharpness and same palette of colors for shared objects), the standard Patchmatch algorithm performs very well with only simple parameterizations on patch size and number of iterations. Thanks to its non local and patch based nature, the algorithm easily accounts for geometric deformations such as translations, which makes it convenient to deal with motion. Also, the clever manner to capitalize patch repeatability allows for image synthesis on strongly deformed or unknown structures. Nevertheless, less controlled settings like changes in illumination or sharpness may render the patch matching less accurate showing the need for a more sophisticated approach.

The last remark highlights the difficulty to deal with bracketed exposure images or multiple depth of field images, where the L^2-norm is not suited to judge similarity from identical objects. In this regard, techniques like radiometric image normalization or a descriptor based matching give hints about how to deal with perturbations of that kind.
Particularly, the radiometric alignment is a simple way to make the reference and source images more similar, which not only produces better reconstructions but also yields more accurate matches. This strategy though not as precise as DPM, is more stable to the geometry of the reference. In contrast the DPM, is not capable to synthesize unknown objects or reproduce flat regions coherently.

These insights will be used in the following two chapters where adequate local features, pre-processing normalization and/or regularization techniques will be used in the contexts of multiexposure image fusion and multifocus image fusion.

Chapter 4

Multiexposure Image Fusion for Dynamic Scenes

HDR imaging and Multiexposure Image Fusion (MEIF) methods are two techniques capable of producing very realistic images of scenes with high dynamic range. In order to get favorable results, both methods have the requirement for the input images to be perfectly aligned, otherwise, ghosts and blur artifacts appear. A general presentation of the problem, as well as the state of the art in the field, are introduced in Chapter 2 (section 2.2).

In this chapter, we introduce a patch-based methodology that minimizes the influence of motion on the fusion of images with strong illumination variations. Such method is leveraged to describe a general framework for the exposure fusion of images captured under dynamic settings. The proposed methodology solves both issues of camera and object motions in a very efficient manner, and can be used on RAW linear images, or 8bit RGB images with no significant change.
Before describing our method, we explain the standard combination of multiexposure images to deal with static settings. We show how object and camera motions make this algorithm inoperative, and present our solution to the problem. As we will see, we take notions from Chapter 3 to properly correct local geometric deformations by relying on a contrast normalization between the images. Thus, we do not require previous knowledge about the acquisition settings or explicitly estimate the camera response function.
Besides, our method can be potentially adapted to any fusion paradigm. In this thesis, the framework used to apply our method is the classical Exposure Fusion [91], but the approach is general and could be applied to other fusion schemes [83, 117, 126].
Our method was subject to extensive experiments on images with complex motions and illumination changes. We show that for MEIF it is more reliable to perform radiometric normalization at global level than using complex iterative refining methods or locally-based radiometric estimations [51, 55]. Concluding remarks are given at the end of the chapter.

General Description

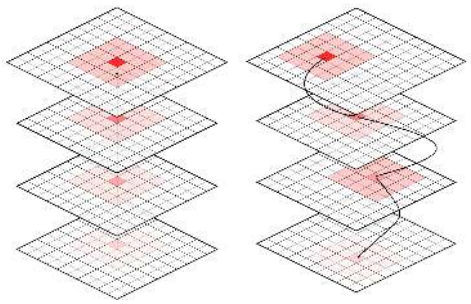
Figure 4.1 presents the classical way to combine multiexposure images (figure 4.1(a)), and also, the principle behind our method (figure 4.1(b)). Basically, instead of fusing values at pixels at the same spatial position, like in [91], we fuse values of pixels having similar neighborhoods in the different images. The method is therefore in the spirit of non-local restoration methods such as the non-local means [16], and also shares similarities with the non-local method introduced in [3] for the creation of HDR images. Our approach has the ability to deal with camera motion (although a previous global image registration is useful to accelerate the process or to enhance the similarity between patches) and with moving objects. Besides, in the absence of motions, our method produces results almost identical to the original exposure fusion algorithm.

4.1 Classical Exposure Fusion

The method introduced in [91] proposes to fuse a series of images $I_1, \ldots, I_N$ acquired with different exposure settings (the knowledge of these is not needed). For each pixel x and index i a weight $W_i(x)$ is defined by taking into account the quality of contrast, color saturation and well-exposedness of the image I_i at this pixel. The idea is then to fuse the values $I_i(x)$ according to the weights. The most straightforward approach would be to define the resulting image R as:

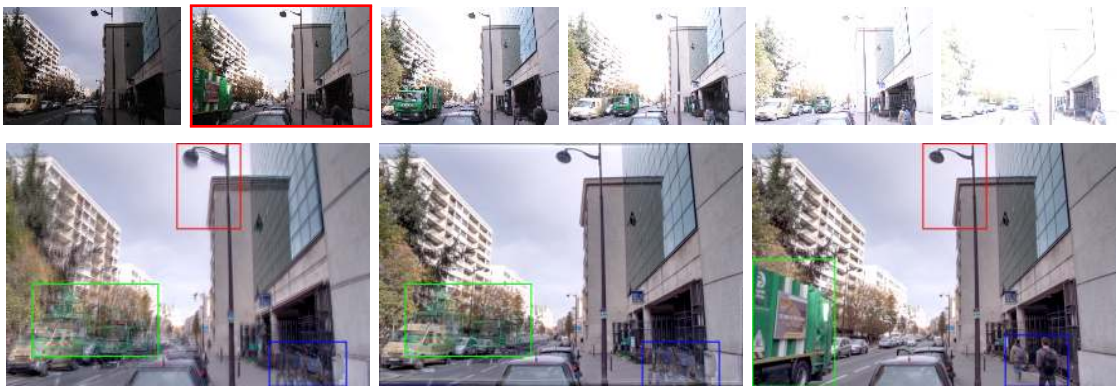
$$R(x) = \sum_{i=1}^{N} W_i(x) I_i(x) \ , \tag{4.1}$$

but this yields incoherences in flat regions and visible seams at slow transitions. In order to achieve seamless fusion, the blending is performed in a multi-scale framework. For each



(a) (b)

Fig. 4.1 Instead of fusing at the same position (a), we use a non-local fusion (b) relying on values from different positions, to account for object motions and camera shake.



(a) (b) (c)

Fig. 4.2 *Top*: Set of misaligned bracketed exposure images displaying large contrast changes. (a). Exposure Fusion. (b). Exposure Fusion on the set of aligned images. (c). Our method.

level l of a Laplacian pyramid [100], the Laplacian pyramid of the resulting image R is computed as:

$$\mathcal{L}_l(R) = \sum_{i=1}^{N} \mathcal{G}_l(W_i)\mathcal{L}_l(I_i) \ ,$$

where $\mathcal{L}_l(I_i)$ is the Laplacian pyramid at level l of image i and $\mathcal{G}_l(W_i)$ is the Gaussian pyramid at level l of the weight map W_i. The final image R is then reconstructed from its Laplacian pyramid.

This algorithm has some desirable properties such as its agnosticism to the different acquisition parameters, and to the camera response (including gamma correction). On the other hand, this (otherwise very efficient) method fails when there is camera shake or object motions. In such cases, the geometry of the different images are not coherent, resulting in ghosting and blur after the fusion. Figure 4.2(a) shows an example of exposure fusion from images taken with a handheld camera and displaying moving objects. The resulting fusion contains visible ghost artifacts, enclosed by color boxes, that strongly impact the quality of the result.

More precisely, when an object location changes between images of the stack, the associated quality weights are mixed together and the sole use of these quality weights cannot prevent different objects to be entangled, resulting in the ghosting effect. For example, a well contrasted object (that moves) will be visible at different locations in the image where his contrast gives him precedence over a flat background.

When a large object moves through a wide range of positions, the result may become catastrophic, figure 4.2(a), green box.

4.2 Exposure Fusion on Dynamic Scenes

To guarantee consistency and avoid repeated objects in the fused image, we propose to enrich the best exposed image with the remaining set of images. To that end, our method selects an image, from now on called *geometric reference* or *reference image*, which is the image with the larger quantity of well exposed pixels, preferably without saturated or underexposed regions, if available (the choice of the reference will be specified afterwards). The main purpose of having a reference image is to impose its geometrical content (including moving objects) to the final result.

Indeed, in all generality it is impossible to render all geometries and objects present in all the images. For example, an object may appear in only one image. What should we do with it? If present in the final result, this object will occlude some other objects that may be present in different images and could have been better reconstructed. The choice of a reference image, though arbitrary, will permit a sound definition of the goal: reconstruct the reference image with the best radiometric precision possible using the rest of the images. The other images will be called *source images* or simply *sources*. This general setup is actually the same as the one used in [55], as detailed in Chapter 2. The comparison to this technique will be detailed and commented further in this chapter.

In short, our algorithm consists in reconstructing the sources using the geometry of the reference and keeping their radiometry. Hence, accounting for motions is a matter of relocating objects based on an estimated map of displacements. As a result, a new set of images that is geometrically consistent with the reference but radiometrically consistent with each source image in the set is produced. The final step will be the straightforward use of the exposure fusion algorithm from [91]. Again, this step can be replaced by another fusion scheme.
More precisely, our algorithm has five stages: Registration, Reference Enrichment, Radiometric Normalization, Search & Reconstruction and Fusion. Let us write $S_1, S_2, \ldots, S_N$ for the input images and $R = S_{i_0}$ for the reference.

- *STAGE I*: Initially, the set of images are loosely registered with respect to the reference, using homography and the RANSAC algorithm.

- *STAGE II*: It is critical for the reference to contain enough details to guide the reconstruction process. This is not the case on saturated regions. To remove saturations, we enhance the reference by using information from the immediately less exposed image, after checking the coherence of the two images in those regions. This produces an enhanced geometric reference.

- *STAGE III*: The radiometry of the enhanced geometric reference is transformed to match the one of each (roughly aligned) source. Then, a color correction algorithm is included to ensure that the resulting images $C_1, C_2, \ldots, C_N$ do not have artifacts. This treatment yields a series of reference images that can be easily compared independently to each source image it corresponds to.
- *STAGE IV*: For each i and for each patch of C_i, a patch-based search in S_i is carried out in order to find the best matching neighborhoods in the sources. These neighborhoods allow us to reconstruct an image L_i that has the same geometry as the reference and shares the radiometry of S_i.
  An optional refinement step (like the ones in Chapter 3, section 3.3) can be included to enhance the quality of the reconstruction.
- *STAGE V*: The fusion algorithm is applied to the stack $L_1, L_2, \ldots, L_N$.

Therefore, our algorithm tries to reorganize the original content of the input images to match the reference's geometry, without losing any distinctive radiometric features.
In the next sections we will detail the different stages and justify the methodological choices that have been made for each of them.

4.2.1 Radiometric Normalization

Given that the images present different exposure settings, objects among the images will present variable illuminations making them difficult to compare with standard similarity measures such as the L^2-norm. Before any comparison between two images, whose goal is to find correspondences, one should normalize the radiometries of the two images to the most possible point (this operation will always be limited by saturations). One can advocate that a similarity measure, say between patches, can be crafted in a way that makes it invariant to affine radiometric transformation, (see for example the affine similarity measure proposed in [34]). On the other hand such a measure may become overly invariant and non specific (**e.g.** correlation is useless in constant regions). This is because, an invariant local measure will implicitly learn the contrast change based only on the patches to be compared, while a global normalization will profit from the whole image information.

In the context of this work, it is reasonable to suppose that a global radiometric normalization will succeed and permit the use of a precise, yet not contrast invariant, similarity measure. Indeed, images are acquired from a given scene with relatively limited framing changes. The goal of this section is to explain how to achieve this normalization. Here we first describe a general approach that does not require any knowledge about the acquisition process and therefore can be applied to non-linear 8bit images. We will also explain how to adopt this to the somehow easier case of RAW images.

Non linear 8bit images

Most acquisition chains output 8bit images whose intensity does not depend linearly on the illuminance. In this case, the contrast transformation between two images of the same scene can be complicated and unknown. A robust way to recover this contrast change is to use the histograms of the images. By comparing the histograms, one can, in favorable cases, exactly recover the contrast change between the two images. When the goal is to match intensities between pixels of two images of the same scene, it is sufficient to match the histograms and many methods have been proposed to this end, midway histogram equalization [33], histogram specification [23, 97], optimal transport [108, 107]. A simple, but limited, possibility is the prescription of a few image statistics such as the standard deviation and mean [28]. We chose to use histogram specification for its flexibility and because in our case (small variations of the scenes) it is enough to preserve the color information of the sources.

Given a pair of images I_0 and I_1, with different levels of exposition, and $H(I)$ denoting the histogram of the image I (for the moment we assume that the image is a gray level image). We search for a contrast transformation h_0 that respects the following requirement:

$$H(h_0(I_1)) = H(I_0) \ ,$$

in other words a histogram specification. The natural way to obtain such h_0 is to apply the inverse cumulative histogram of I_0 to the cumulative histogram of I_1. Due to the quantized nature of images, this procedure can not be carried out exactly, although an approximation can be obtained by using a pseudo-inverse.
In order to find the function h_0 we first compute the cumulative histogram C_0 and C_1, for image I_0 and I_1, respectively. Then, for each intensity $y \in [0,\ 255]$, we search for the intensity $x \in [0,\ 255]$ that satisfies the following expression:

$$h_0(y) = \arg\max_x \ C_0(x) \leqq C_1(y) \ . \tag{4.2}$$

This expression let us find the intensity x that makes the cumulative histogram C_0 more similar to $C_1(y)$. In figure 4.3, this is represented by the dashed lines at graylevels x and y that when evaluated in the cumulative histograms, they map to the same cumulative probability $p = C_1(y) = C_0(x)$. In other words, there is a mapping of intensities from image I_1 to the radiometry of image I_0.

The output image $h_0(I_1)$ has the same geometry as I_1 and the radiometry as I_0, for that reason we refer to I_1 as the geometric image, while I_0 is denoted as the radiometric image. From now on we introduce the notation $h_I(J)$ where I and J are two images, to denote the image that has the same geometry as J and the same (or approximate)

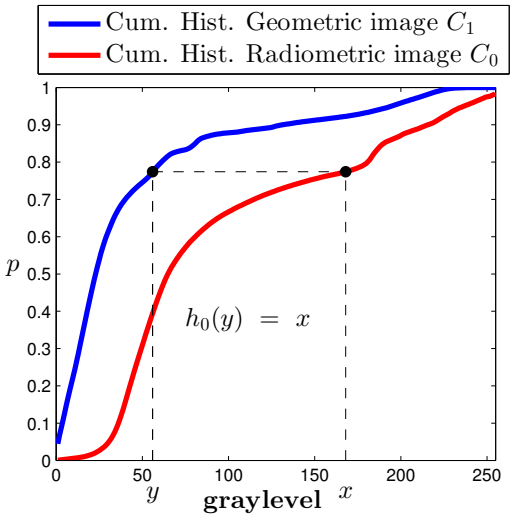


Fig. 4.3 For histogram specification we want to transfer the radiometry of image I_0 to image I_1, where the latter one will conserve its geometry while displaying the colors of the former image. For that we need a function that matches the cumulative histogram of both images at any given intensity y. In the plot we can see that for the intensity y in C_1 there is a match with the cumulative histogram C_0 at graylevel x.

radiometry as I. What we said till now applies to grayscale images. We argue that for our problem, a sufficient extension to the color case is to match the three color channels independently. Indeed, under the assumption that the radiometric difference between the bracketed images comes from changes of white balance, exposure and gamma-correction, then the order of intensities of pixels is preserved on each channel (R, G, or B) between the acquisitions. More precisely, this is true provided the white balance is obtained through a diagonal matrix (Von Kries model) [130], which is true in practice [89]. This approximation is especially accurate when images are aligned and share the same content.

Of course, this procedure is not exact and has to be manipulated with care. Indeed, if image I_0 is saturated (with, say, pixels at 255) then the resulting image will present saturated pixels even if I_1 was saturation free. If this procedure could be exact in the presence of very differently exposed images, then the problem treated here would be solved.

On figure 4.4, the proposed contrast normalization scheme is applied on the best exposed image from the set, red framed picture, to present the same illumination as each image in the set. Notice that despite the misalignment and the presence of some moving objects (pedestrians), no false colors are created, achieving a satisfying color assignment both for dark regions and bright regions on the reference. Again, the result would have been disappointing if the image imposing the geometry was too saturated. Indeed, in bright saturated regions, the geometry is completely lost and details or structures cannot

Fig. 4.4 *Top*: Set of misaligned bracketed exposure images. *Bottom*: Photometrically normalized image with respect to the reference.

be recovered by simple radiometric normalization. For such case, a single intensity (coming from the brightest part of the image imposing the radiometry) is assigned to saturated regions, which results on flat regions after normalization.

Although simple when compared to more elaborated color equalization methods [55, 54], we found that depending only on the histograms brings more robustness to the method. Other works [51, 54] suggest to tune an intensity mapping function (IMF) τ by minimizing functionals of the type:

$$\sum_{(k,l)\in\Omega} \|\tau(I_1(k,l)) - I_0(k,l)\|_p^p \ ,$$

with Ω the set of positions mapping nearest neighbors from I_0 to I_1. This expression forces to track outliers (moving objects and saturations) and remove them from the set Ω. By computing an IMF depending only on the histogram, and since the images have roughly the same content, we obtain a sufficient result for the next steps of our proposed fusion to be carried out successfully. Indeed, in the next sections we correct few inconsistencies left after histogram specification, like saturations or noise enhancement.

For radiometric normalization, and more precisely for histogram matching, we also explored an alternative solution using transportation distances [108] (or Wasserstein distances). In dimension one, this otherwise complex problem boils down to a simple sorting scheme. Even though it works well most of the cases, this alternative is prone to produce artifacts due to inconsistent color assignments, especially when the contrast change between the images is very large. For instance, if the image imposing the geometry I_1 is much brighter (or has too many saturated regions) than the image imposing radiometry

I_0, then flat regions due to saturations will be reproduced with unsaturated pixels, this results in smooth regions with degraded colors. On the contrary, if I_1 is very dark, noise can be enhanced and irregular patterns can be created.

RAW Linear Images

Raw images offer a reliable alternative to obtain a contrast normalized representation of the stack of images. They are the primary data produced by the sensor and store an accurate reproduction of the radiance of the scene, except for noise and saturation. Several models of the raw image formation have been proposed. The main characteristics of all proposed models is that the response is a linear function of the illumination and that the noise level depends on the illumination at each pixel. An offset is added to the image to avoid negative values.

This can be summarized in the following equation:

$$Z_i \sim \mathcal{N}(gaY\tau_i + \mu_R, g^2 aY\tau_i + \sigma_R^2) \ , \tag{4.3}$$

where $\mathcal{N}(\mu, \sigma^2)$ is the Gaussian distribution of mean μ and variance σ^2, Y is the real radiance, Z_i the measured output, μ_R is a fixed offset, g a gain depending on the sensibility setting of the camera and σ_R is the intrinsic readout noise level [2]. Here a is a spatially varying gain, that we will take equal to 1 to simplify the study. Finally, τ_i is the exposure time. Notice that we approximated Poisson noise by a Gaussian distribution, which is a valid assumption, except for extremely dark regions [2].

Under this model, the best least square estimator of Y from Z_i is given by:

$$\widehat{Y}_i = \frac{Z_i - \mu_R}{g\tau_i} \ . \tag{4.4}$$

Of course this is valid within the operational range of the camera (no saturation). As said before, the saturated regions need to be processed differently. Figure 4.5 shows the result on a set of raw images.

4.2.2 Image registration on bracketed exposure images

The image acquisition scenario in which we place our method is a series of images taken with different expositions, potentially using a handheld device. In such a scenario, we cannot assume that the images are perfectly aligned, even when there are no moving objects in the scene. The first step to our fusion algorithm will be a coarse registration that guarantees at least that large portions of the background are aligned. The next steps will deal with moving objects of the scene, as well as registration errors resulting from cases where the homography is not valid. Here, we just want the homography to be a

Fig. 4.5 Contrast normalized raw images (for visualization purposes we applied a global logarithmic tone mapping). The normalization of equation (4.4) is very accurate, except in the saturated regions. The color irregularity on saturated regions, surges because for longer expositions the sensor saturates at lower irradiance values.

rough approximation, mostly aimed at improving the global radiometric correction and at accelerating the forthcoming patch matching, *STAGE IV*.

For the global coarse registration we make the assumption that the optical center of the device does not move much between acquisitions and/or that the scenes are mostly planar. This hypothesis is particularly valid when the objects are far from the camera and the images are taken with a small camera displacement. Under this hypothesis, the geometrical transformation that takes one image of the set to any other is a homography (under pinhole camera model). Consequently we will search for the homography that best aligns the images. We will take one of the images as a geometric reference, and the other images will be transformed to its geometry.

Since, we are interested in a geometrical transformation that explains the dominant movement, using the RANSAC algorithm is particularly indicated. In order to be robust to illumination changes, we use the classical SIFT keypoints as a primary match indicator. Nevertheless, keypoints extraction depends on thresholds which can be untriggered when the region becomes dimmer. When regions of an image are dark in one image and not in the other, keypoints could be missed in the darker image and thus fail to correctly match the geometries of the two images. One way to solve this issue is to adapt the thresholds governing keypoints extraction to accommodate the change in illumination. Another way, is to keep the same thresholds and apply a radiometric normalization to one of the two images. To keep the presentation simple and have a unique keypoint extractor that has not to be tuned to each couple of images, we chose the later approach.

We first apply a radiometric normalization to the geometric reference, according to the method explained earlier. Observe that, strictly speaking, the hypothesis of radiometric normalization is that the two images present the same content, which cannot be the case when a global movement had occurred between the two acquisitions. However, the purpose of the radiometric normalization here is only to permit a faithful extraction of

Algorithm 1 Alignment on Bracketed Exposure Images

Input: Set of images S_i, Reference image S_{ref}
Output: Set of aligned images I_i and common domains Ω_i.

1: **for** image i **do**
2: $\widehat{S}_{ref} \leftarrow h_{S_i}(S_{ref})$ ▷ Radiometry of S_i is transferred to S_{ref}.
3: $SIFT_{ref} \leftarrow$ SIFT features from image $\widehat{S}_{ref}$.
4: $SIFT_i \leftarrow$ SIFT features from image S_i.
5: Find correspondences between $SIFT_i$ and $SIFT_{ref}$
6: RANSAC estimation of a homography H.
7: $I_i = S_i \circ H^{-1}$ (using bilinear interpolation), ▷ Inverse Warping
8: **end for**

SIFT keypoints and does not have to be perfect.
Once these steps are taken our approach is classical. The RANSAC algorithm is applied to the set of matching SIFT descriptors. This provides us with a homography along with a set of matching SIFT couples that agree with this homography. Finally the aligned images are obtained by inverse warping and bilinear interpolation. The whole process is summarized in Algorithm 1. The roughly aligned images I_i and Ω_i the individual common domains with respect to the reference S_{ref}, will serve as input to the rest of the proposed method.

Figure 4.6 presents a set of bracketed exposure images that are affected by a large movement and also a strong illumination change. Even in this rather extreme situation, a correct alignment is computed. At least sufficient enough to carry on the rest of the proposed algorithm, relying on patch matching.

(a) S_1 (b) S_2 (c) S_3 (d) S_4

(e) I_1 (f) I_3 (g) I_4

Fig. 4.6 (a-d). Set of misaligned bracketed exposure images. (e-g). Registered images with respect to the reference S_2.

Indeed, the bridge, the background buildings and all static objects are perfectly aligned even for a much darker exposition than the reference (figure 4.6(e)). As expected, pedestrians and moving objects have moved between the images. The black frames on the aligned images correspond to regions of the reference image that have not been captured in the others due to the relative movement.

4.2.3 Reference Refinement

In *STAGE II*, the content of the reference image is refined on saturated areas, if they exist. Else, textureless regions will be transferred to the final result. For that, the reference I_{ref} (where $ref \in (1, \ldots, N)$) is chosen carefully as the image with less amount of overexposed and underexposed pixels. As explained before, the reference is enhanced by imposing the geometry of the immediately less exposed source image. The reason for choosing this image is because it prevents from large contrast changes and object displacements. We ensure that corresponding regions in this source image are not saturated and are consistent with the surrounding geometry in the reference.
We define that a pixel is saturated if it falls above 95% of the dynamic range. The consistency is checked through a motion mask that is defined as:

$$\|I_{ref} - h_{I_{ref}}(I_j)\| > \beta \ ,$$

where j is the index associated to the closest darker image with respect to the reference I_{ref}, $h_{I_{ref}}(I_j)$ is the contrast normalization of image I_j to the reference's radiometry, and β is a threshold.

4.2.4 Displacement map extraction and reconstruction

Motion during acquisition affects substantially the efficiency of exposure fusion, see figure 4.2(a). Even though image registration solves global misalignment, object motion can still strongly deteriorate the quality of the results, see figure 4.2(b).

In order to solve this further problem of object displacements within the input images, our method estimates local displacements after carrying out a non local search around every neighborhood within the reference image. This gives rise to a displacement map that accounts for the local deformation and displacements between the set of images.

That is, for a series of aligned images $I_1, I_2, \cdots, I_N$, resulting from Algorithm 1, we aim to build a displacement map $f_i(x)$ that maps the spatial deformations between the source I_i and the reference image. The enhanced reference I_{ref}, is intended to guide the estimation of f_i and therefore should have the same radiometry as I_i. For that, we apply the contrast change of section 4.2.1 to the reference, $C_i = h_{I_i}(I_{ref})$, where h_{I_i} is computed

exclusively from $I_{i_{|\Omega_i}}$ the intensities that fall in the common domain Ω_i. The map is defined as a nearest neighbor field that contains the coordinate of the pixel in the source, whose neighborhood (patch) is the closest at x. That is,

$$f_i(x) = \underset{y \in x+\mathcal{N}}{\arg\min}\ d(P_{C_i}(x), P_{I_i}(y)) \ , \tag{4.5}$$

with $\mathcal{N}$ a search window (bigger than the patch), where $P_{C_i}(x)$ is the patch storing the collection of values from C_i, in a $(2s+1) \times (2s+1)$ square window centered at x, and $P_{I_i}(x)$ is defined similarly on I_i. The operator $d(P_A, P_B)$ is a similarity measure between patches.
In our experiments, the similarity measure between the patches is simply the sum of squared differences (SSD):

$$d(P_{C_i}(x), P_{I_i}(y)) = \sum_{k=-s}^{s} (C_i(x+k) - I_i(y+k))^2 \ . \tag{4.6}$$

Equation (4.6) itself is not robust to affine contrast changes between the patches. This drawback makes it unreliable for the comparison of bracketed exposure images, and therefore justifies the importance of the radiometric normalization mentioned before. We also experimented with contrast invariant similarity measures, e.g. cross correlation and the affine measure of [34], but they failed on flat regions larger than the patch size and therefore yield less good results.

Notice that equation (4.5) does not make any assumption about the acquisition parameters of the input images. The only requirement in the extraction of the nearest neighbor field is that both images should be radiometrically normalized. Thus, it can be used on JPEG images or raw images.

Finally, the reconstructed set of bracketed exposure images, consistent with the geometrical reference is obtained by:

$$L_i(x) = I_i(f_i(x)) \ .$$

It is worth noting that image synthesis with patches allows for the transfer of information from the source to the reconstruction. Besides, it is carried out on the original radiometry of the images, meaning that no new intensities are created or transformed. That is because, after reconstruction we want to generate a new image that still preserves initial characteristics from the source.

4.2.5 Algorithm

The complete flowchart of the algorithm is shown in Algorithm 2.
Even though Algorithm 2 is presented for JPEG images, few modifications can be made for using it on raw images. In that case, the luminance based normalization can be performed during the image registration, where the homography transformation is used to remap the luminance images. For that reason, there is not need of the contrast normalization at step 12, and the step at line 18 could be skipped too.

In the algorithm, $h_S(R)$ denotes the operator that normalizes the radiometry between any given image R and S.
For particular conditions between the reference and the source, especially when the reference is darker than the source, it was observed that our radiometric normalization highlights small compression errors from the JPEG images. Such irregularities are solved by including a correction step derived from the Yaroslavsky filter. The radiometric correction (line 18 in our algorithm) is created by imposing identical local similarity weights from the original image R into its contrast normalized version J. This is expressed by $\mathbf{C}(R, J)$, which

Algorithm 2 Non-local Exposure Fusion

Input: Stack of images S_i, with $i \in [1, N]$. Patch size $(2s+1)^2$, search window size $(2m+1)^2$.
Goal: Build a set of aligned images L_i and fuse.

1: Sort images in increasing exposure time. (Still denoted as S_i).
2: Reference selection, $R = S_{i_o}$.

3: **procedure** IMAGE REGISTRATION W.R.T. R
4: **for** each image $i \in [1, N]$ **do**
5: T_i: Homography between $h_{S_i}(R)$ and S_i.
6: Ω_i: Common domain between $T_i(S_i)$ and R.
7: $C_i = h_{T_i(S_i)_{|\Omega_i}}(R)$
8: **end for**
9: **end procedure**

10: **procedure** REPLACEMENT OF SATURATED PIXELS ON R WITH IMAGE $i = i_o - 1$
11: *sat*: saturated pixels $R > \alpha$
12: *mov*: moving pixels $\|R - h_R(T_i(S_i)_{|\Omega_i})\| > \beta$
13: Φ : *sat* and not-*mov*.
14: $\widetilde{R}$: $C_{i|\text{not-}\Phi}$ and $T_i(S_i)_{|\Phi}$. ▷ $\widetilde{R}$ is the new reference.
15: **end procedure**

16: **procedure** PATCH RECONSTRUCTION OF $\widetilde{R}$ FROM $T_i(S_i)$
17: **for** each image $i \in [1, N]$ **do**
18: $\tilde{r} = \mathbf{C}(R, h_{T_i(S_i)}(\widetilde{R}))$ ▷ Yaroslavsky based radiometric correction.
19: $f_i(x) = \text{PatchMatch}(\tilde{r}, T_i(S_i))$ ▷ Displacement Map Extraction.
20: $L_i(x) = T_i(S_i)_{|f_i(x)}$ ▷ Source images Reconstruction.
21: **end for**
22: **end procedure**
23: I_f = ExposureFusion(L_i) ▷ Classical fusion method (section 4.1).

imposes the regularity of image R on image J. This idea was used in [106] to regularize images after color transfer.

Similarly, to improve the quality of reconstruction and noise filtering, we found that using the k-nearest neighbors maps were an excellent choice to refine the reconstructed image, Section 3.3.2. This showed to be the best option in all our experiments to the cost of an additional computational time. For that, the Patchmatch algorithm was tuned to extract $k = 10$ nearest neighbors. By penalizing errors between the reconstructions L_i^k and the reference image $\tilde{r}$, the reconstructed images were combined in a weighted fashion. This optional step can be added after line 20.

4.2.6 A simplified image fusion procedure

In the previous section we proposed an algorithm that captures both geometric and radiometric information from a stack of images in a framework that reduces the impact of moving objects. In this section we introduce a very simplified approach for the same purpose. This time, we do not use patch correspondences to account for displacements or for transferring the geometric content. Instead, the source images are used only to provide radiometric information to the reference.
The idea is that the enhanced reference image is radiometrically normalized to each image from the set. Here, we also include the radiometric correction based on the Yaroslavsky filter, where we impose the local similarity weights of the original reference to its normalized version. The resulting set of images, including the reference, is fused with exposure fusion. In terms of the general pipeline of Algorithm 2, this method suffices to remove *STAGE IV*. In the following, we call this approach the "simplified image fusion".

Notice that this alternative is a safe way to reproduce coherently the geometry within the reference image. Since the search and transfer of geometrical content from source images is avoided, the processing time is notably reduced but the geometry in the final image will not be enriched by the other images. This algorithm's performance depends mostly on the quality of the radiometric normalization, assuming that saturations are completely corrected during the second stage.

4.2.7 Details of Implementation

Here we highlight some important aspects that will ease the replicability of our method and also the quality of the resulting reconstructions.

For the nearest neighbor field extraction, the most important parameter is the patch size. It influences the confidence of matching and the computational efficiency of the search. That is because bigger patches enclose more geometrical information, making them more distinctive but less efficient to process. The smaller they are, the more patches will be

found over the spatial domain, thus, allowing more flexibility on the matching precision. In that sense, a trade-off between efficiency and accuracy should be made.
Without risking precision, we used the patch size to be 3×3 px for all our experiments. For practical reasons, we used the Patchmatch algorithm [10] to extract the nearest neighbor field for each source image. It was set to search for nearest neighbors over the entire image, and to iterate 5 times.

For image registration, we used the *vl-feat* library [125] to extract and match SIFT features, which accounts for the *STAGE I* of our algorithm. The main benefit of including this stage is that it increases geometrical coherence between the images. Particularly, image alignment alleviates strong rotations that can harm the non local similarity search. Also, image misalignments give space to new objects to appear, which may affect the similarity of content between the images. Registering the images would regulate unknown objects appearing at the borders. This case is illustrated in figure 4.6, where the less exposed image contains a larger presence of blue intensities in the sky in comparison to the reference.

4.3 Experiments & Results

About the evaluation of exposure fusion results

Evaluating numerical results in the context of HDR imaging is a difficult subject that has recently received a lot of attention, either in the context of images or videos, see e.g. [36]. For the specific problem of HDR image creation, that is the problem of recovering the true irradiance of a scene from a stack of differently exposed images, the ground truth can be estimated by averaging a large number of acquisitions [50], or performances can be estimated from a given HDR image using a realistic noise model [4]. Evaluation can be performed either through classical image norms or more elaborated and psychologically motivated image quality metrics, often mimicking the human visual system [86]. An extensive study of such quality metrics, in the context of image compression, can be found in [133], see also Chapter 17 of [36]. All these approaches are of course only valid when the scenes are static and captured with motionless camera.

In the case of exposure fusion, the problem is more subtle since the results should primarily be sufficiently contrasted, vivid, or colorful, all notions which are highly subjectives. Several methods have been proposed to compare a tone-mapped image to the original HDR image it has been generated from. For instance, [8] proposes to compare images in a contrast-invariant way. Other approaches have been proposed in [132, 67]. Again, such approaches only work with static scenes and still cameras, and are therefore not adapted to the methods presented in this chapter, whose aim is to deal automatically with object movements, camera shake or optical distortions.

Another approach is to perform subjective user experiments, but these are very time consuming and necessitate an experimental investment that is beyond the scope of this chapter.

For these reasons, we concentrate in this experimental section on different scenarios for which we simply compare the results visually, mostly by identifying artefacts. Before doing so, we also provide a quantitative evaluation of the ability of our method to correctly align the radiometry of images before the fusion.

4.3.1 Contrast Normalization Evaluation

The photometric normalization approach included in our method has been compared with two parametric transformations, namely affine and gamma transforms.

$$\widehat{I}_s = \alpha \cdot I_r + \beta \ , \tag{4.7}$$

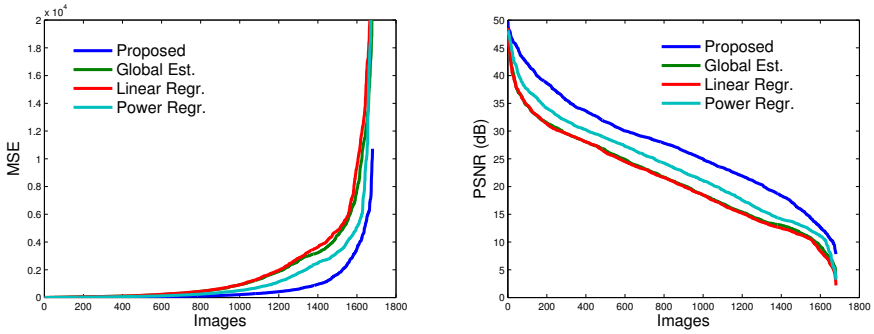
$$\widehat{I}_s = \alpha \cdot I_r^{\beta} \ . \tag{4.8}$$

Where the equations show the linear/power relation between the source image I_s and the reference I_r. Then, the solution to the contrast normalization problem lies on finding the parameters α and β, which is possible with linear regression on both cases, except that for the exponential case (equation (4.8)) regression is performed on the log domain.
Another way to approach the affine case is to impose the same mean intensity and standard deviation as in equation (4.9). We refer to this alternative as the 'Global Estimation' in our experiments.

$$\widehat{I}_s = \frac{\sigma_s}{\sigma_r} I_r + \left[\mu_s - \mu_r \frac{\sigma_s}{\sigma_r} \right] , \tag{4.9}$$

where μ_r and σ_r are the mean and standard deviation of image I_r, correspondingly for image I_s. Even though this model was used in the *Lab* color space by Reinhard et al. [110], we argue that independent *RGB* histogram matching is enough if we assume that the color balance has been done through a diagonal matrix (Von Kries' model) and that non linearities are roughly the same for each channel. For accurate results, the parameter estimation should be performed without saturated pixels.

The above strategies and the proposed contrast normalization for standard 8bit images (section 4.2.1), were evaluated by using the mean squared error (MSE) on different set of images acquired from [84]. The dataset contains 17 sets of images with natural environments, outdoor/indoor features and man-made structures. The images are totally aligned in order to evaluate the color transformation strictly on known pixels, otherwise the accuracy of the results would be corrupted by the inclusion of non aligned images and moving objects. The experiment consisted in taking each image as the geometric reference and normalize it



(a) MSE (b) PSNR

Fig. 4.7 Evaluation of the contrast normalization methods.

to the radiometry of the other images. The normalized image was then compared to the initial source image. This generated a total of 1680 combinations, see figure 4.7.

From the MSE evaluation graph, figure 4.7(a), we can notice that our method (blue line) presents the lowest error, being the best transformation among 81.5% of all combinations. It was followed by power regression with 10.6%, linear regression with 4.7% and 3.2% for the global estimation with image statistics. Besides, our histogram specification method was the most consistent of all with a total of 989 pairs having an error below to 200 units, compared to 582, 563, 744 for global estimation with image statistics, linear regression and power regression, respectively. The PSNR curve is also presented in figure 4.7(b).
The increasing deterioration of the contrast normalization depends on several factors, namely, large exposure changes between the pair of images or noise in dark regions for a underexposed reference image. Notice that when mapping the color from a very dark image to a brighter image, the method will perform poorly. That is because underexposed images generally present noise or quantization defects that are enhanced after contrast normalization.

4.3.2 On Multi-Exposure Fusion

The proposed methodology was evaluated under several situations: static settings, dynamic settings and saturations on the reference. For the experiments we used the datasets provided in [63, 102] and our own photographs, containing different types of perturbations: motion, occlusion, multiview, etc. A selection of experiments with their corresponding analysis is presented below.

Our method from Algorithm 2, that we called 'Non-Local Exposure Fusion' is denoted by *NLEF* in the experiments. The version that includes the optional regularization with the $k-$nearest neighbors in the final step, is called *NLEF-KNN*. Both alternatives are

compared with the state of the art image reconstruction algorithm for HDR-like image rendering proposed by *Hu et al. [55]* and the deghosting algorithm (*BMD*) [102].
For Hu's method [55] we employed their implementation. We used a patch size of 3, but also a patch size of 10, the size used in [55], to be more fair with their configuration. As before, the whole image is used as a search window. The BMD algorithm [102], implemented by us, is configured with the parameters used in [102], a circular structural element for the morphological erosion and dilation operations, with sizes 3 and 17, respectively.

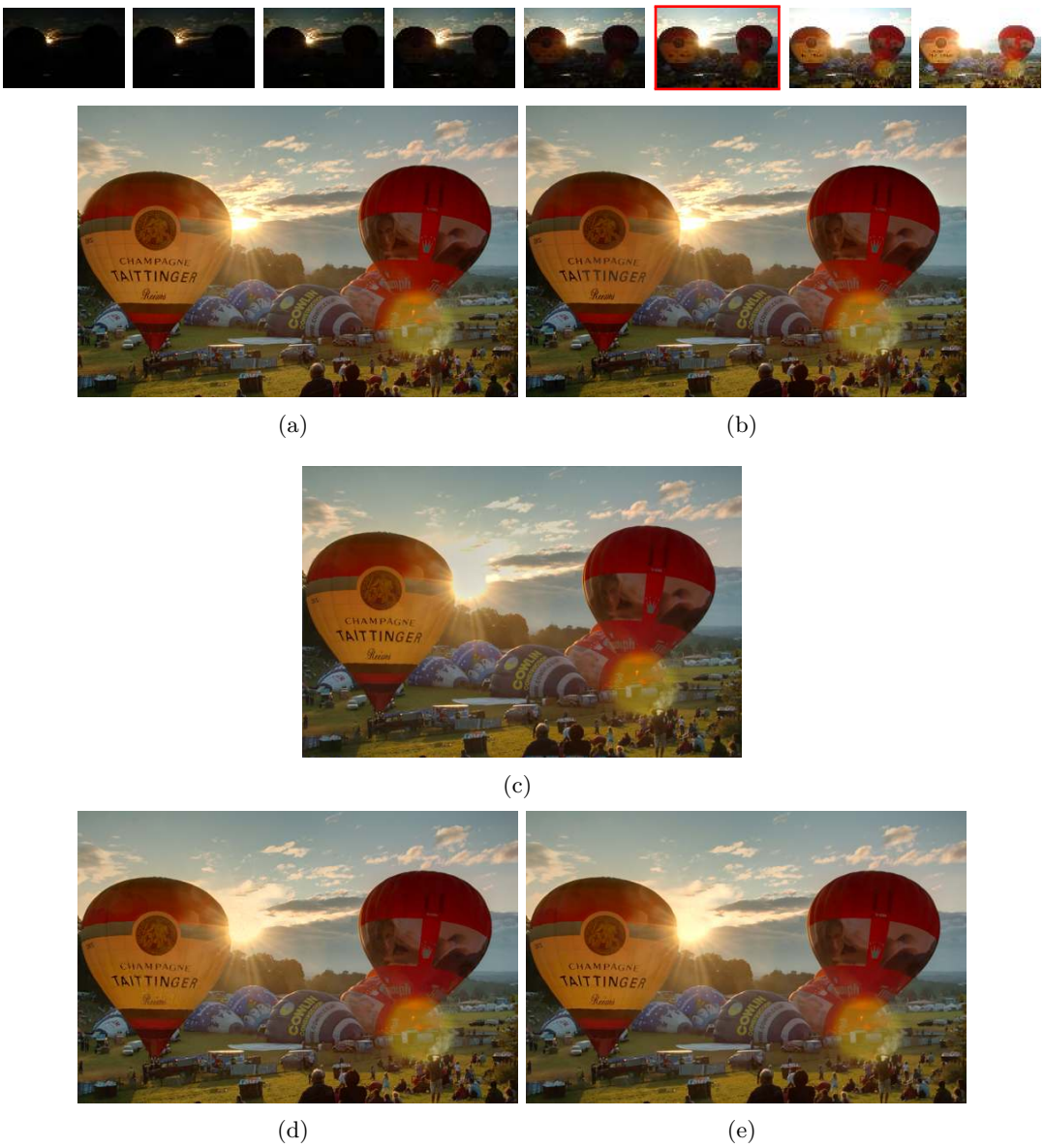
The methods are analyzed on the following aspects:

1. Totally aligned images, section 4.3.3.
2. Non aligned images: Handheld camera motions & object motions, section 4.3.4.
3. Robustness to reference change, section 4.3.5.
4. Patch size change, section 4.3.6.
5. Geometric distortions in the fusion, section 4.3.7.
6. False colors in the fusion due to contrast normalization failure, section 4.3.8.
7. Simplified image fusion, section 4.3.9.
8. Correcting errors within the simplified approach, section 4.3.10.
9. Standard non rigid dense correspondence for exposure fusion, section 4.3.11.

4.3.3 Totally aligned images

Figure 4.8, shows the result of exposure fusion on a set of *totally aligned images*. The BMD algorithm presents some halo-like effects around the balloons and also exhibits a darker appearance than exposure fusion. The method from Hu et al. [55], accurately reproduces details but some regions appear opaque, and there is a loss of information in the saturated regions. As for our method, the standard version presents some subpixel imperfections at detail level, like the letters of the yellow balloon. Nevertheless, such inaccuracies are adequately corrected with the use of the KNN maps in the combination. Regarding saturations, our method is better than Hu et al. [55], but not as good as the original Exposure Fusion. That is because the supporting image for the reference enhancement was also affected by saturations over corresponding areas. For static settings, this can be improved by using a darker image than the I_{ref-1} image. For dynamic settings, that alternative is risky as objects can be moving and the useful content would not be very valuable for enhancement.

Figure 4.9 is another example of image fusion with static scenes. For this set the fusion with all methods is almost as good as with Exposure Fusion. This is a result of the fact that the chosen reference does not present large saturations and the exposure change of



(a) (b)

(c)

(d) (e)

Fig. 4.8 On totally aligned images: (a). Exposure Fusion [91], <1sec. (b). Exposure Fusion with BMD [102], <1sec. (c). Hu et al. [55], 176secs (d). NLEF, 17secs. (e). NLEF-KNN, 35secs.

the input images did not present a big challenge to the contrast changes approaches for Hu et al. [55] and NLEF.

In figure 4.10, we present an experiment with little moving objects and small exposure changes between the images. For such controlled settings, the reference has no saturations and the remaining images have almost identical colors, thence all methods present very good results for correcting imperfections due to motion, except of course, the original Exposure fusion [91]. Notice that Exposure fusion presents ghost artifacts, see for instance

the walking people and the flying birds.

Another example where Exposure Fusion fails on aligned settings with motion is presented in figure 4.11. For this experiment, the walking person in the foreground partially appears in the final result. Notice that even though, BMD produces a geometrically consistent image, the fusion is darker than the other methods. That is because it only utilizes the best exposed image to account for corresponding moving regions.

Regarding the reconstruction methods, both alternatives deliver a geometrically consistent image with respect to the reference and the colors are identical to Exposure Fusion.

(a) (b)

(c)

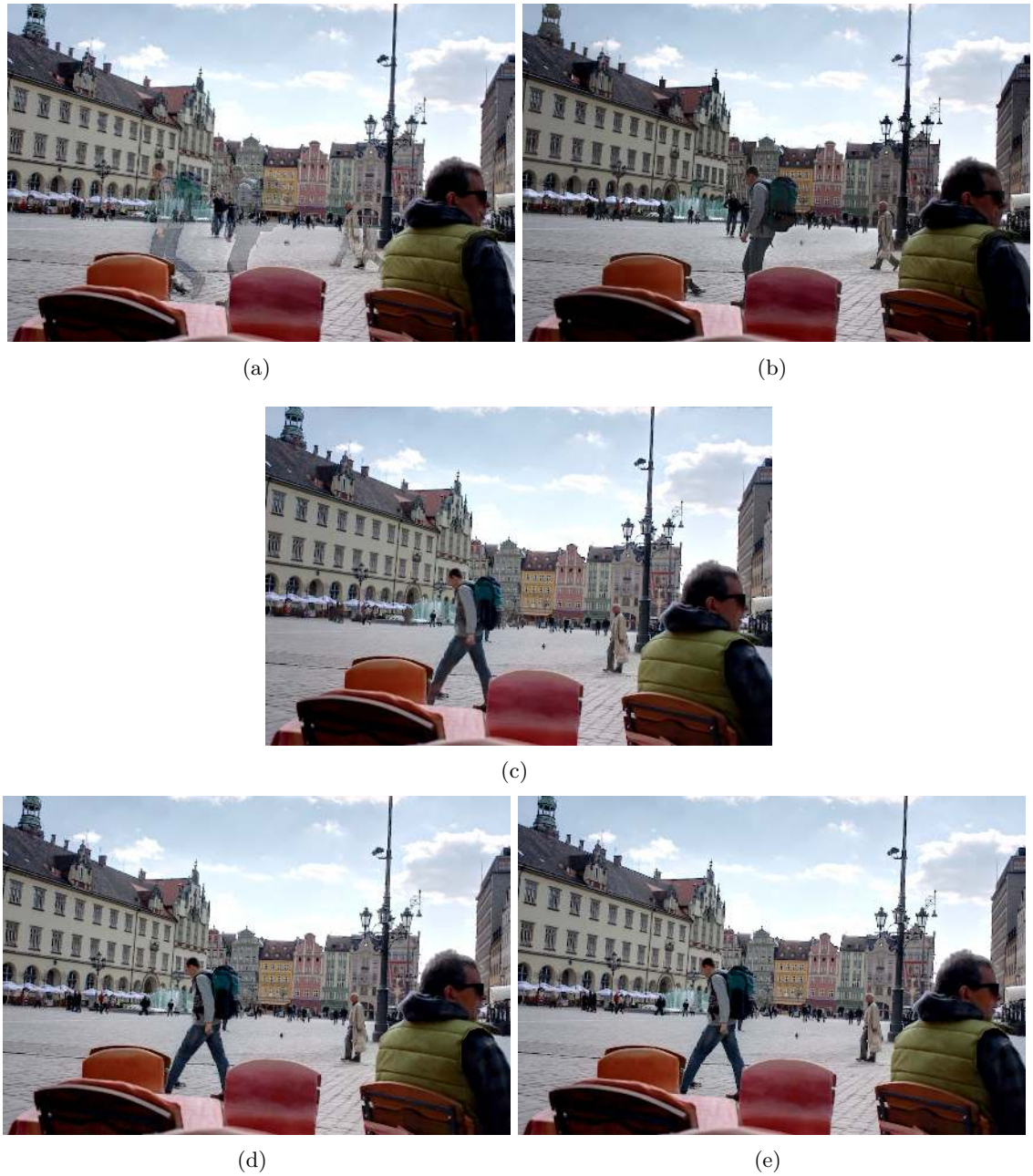
(d) (e)

Fig. 4.9 On totally aligned images: (a). Exposure Fusion [91], <1sec. (b). Exposure Fusion with BMD [102], <1sec. (c). Hu et al. [55], 39secs (d). NLEF, 7secs. (e). NLEF-KNN, 8secs.

However, the result from Hu et al. [55] shows a reduction of contrast in dark regions, e.g. at the left leg of the walking person and his backpack. As before, NLEF-KNN is more precise than NLEF.

(a) (b) (c) (d) (e)

Fig. 4.10 Global alignment with small translations: (a). Exposure Fusion [91] on aligned set, <1sec. (b). Exposure Fusion with BMD [102], <1sec. (c). Hu et al. [55], 38secs. (d). NLEF, 7secs. (e). NLEF-KNN, 16secs.



(a) (b)

(c)

(d) (e)

Fig. 4.11 Global alignment with object motion: (a). Exposure Fusion [91] on aligned set, <1sec. (b). Exposure Fusion with BMD [102], <1sec. (c). Hu et al. [55], 185secs. (d). NLEF, 17secs. (e). NLEF-KNN, 43secs.

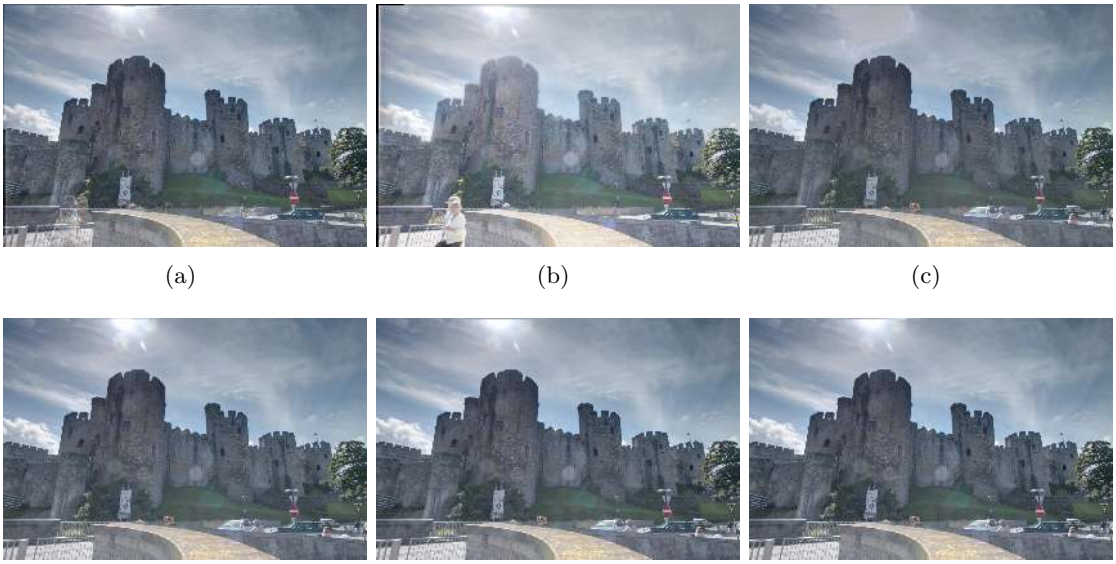
4.3.4 Misaligned images: camera motions & object motions

Figure 4.12, presents the fusion for a set of images with *global and local misalignments*, and saturations on the reference image. There, we can see that exposure fusion does not present artifacts around edges of static structures. But again, moving objects appear as ghosts. For BMD, the fusion presents a brighter appearance with respect to the others. Even though there is no ghost, some regions have a strong level of blur. This behavior is again due to failures of the moving mask estimation and fusion, which gives priority to only one image that happens to be badly exposed on those regions.

Although the result from Hu et al. [55] still presents very good detail reproduction, it shows presence of an opaque veil on dark detailed areas. A different kind of artifact (sparse colored dots) appears over regions where the reference image is saturated. The performance is corrupted due to an imprecise estimation of the intensity mapping function (IMF), which is affected by strong illumination changes and improper substitution of saturated regions. Nevertheless, a significant reduction of those negative aspects was noticed when enlarging the patch size to 10×10px which provided a better image reconstruction to the cost of higher computational demand.
Also, it was observed that their implementation regularly gets locked in an infinite loop when performing weighted least squares for the IMF.

As for our method, it produces superior results than BMD and Hu et al. [55] with patch size of 3×3, with not artifact creation around saturated regions or moving objects, lower dependency to the patch size and with identical quality to Exposure Fusion for static and undeformed regions.

Based on experiments with small (e.g. figure 4.13) and large displacements (e.g. figure 4.14) we can say that both reconstruction methods (NLEF and Hu et. al [55]) are robust to geometric transformations, like rotations or translations. Generally, what changes between them is the sensitivity to additional factors, like saturations for Hu et al. [55], or the ability to reproduce accurate details, which for NLEF-KNN is higher for more KNN maps.
Since Hu et al. [55] does a patch-based regularization, using a patch size of 10 is enough to correct perturbations but sometimes yields over blurring. Regarding the BMD deghosting method we observed that it is very unstable, failing under the assumption that no motion is present above or below any threshold. This approach produces critical failures specially on moving objects that share colors above and below the estimated threshold. For that reason, we skip the BMD method in future experiments and concentrate on evaluating



(a) (b) (c)

(d) (e) (f)

Fig. 4.12 Global and local misalignments: (a). Exposure Fusion on aligned set [91], 4sec. (b). Exposure Fusion with BMD [102], 6sec. (c). Hu et al. [55] patch size 3×3, 211secs. (d). Hu et al. [55] patch size 10×10, 884secs. (e) NLEF, 190secs. (f). NLEF-KNN, 423secs.

(a) Fusion NLEF-KNN. (b) Fusion Hu et al. [55].

Fig. 4.13 Image fusion on a scene having large displacement of small objects. **First row:** Source images with large object motion. **Second row:** Fused images for NLEF-KNN and Hu et al. [55] using the as reference the image 1.

further the reconstruction methods.

Figure 4.13 is an example of small scale objects with big displacements. In this case, both reconstruction methods present good results with well balanced colors and no artifacts. Only slight contrast changes in the background are visible.

Figure 4.14 shows large displacements from large scale objects, e.g. the green truck which also has a variable size. The fusion shows a well illuminated image with no radiometric artifacts. Particularly, NLEF-KNN exhibits more contrast in some regions, e.g. the building gate at the bottom right. The corresponding regions with the Hu et al. [55] fusion appear blurred.

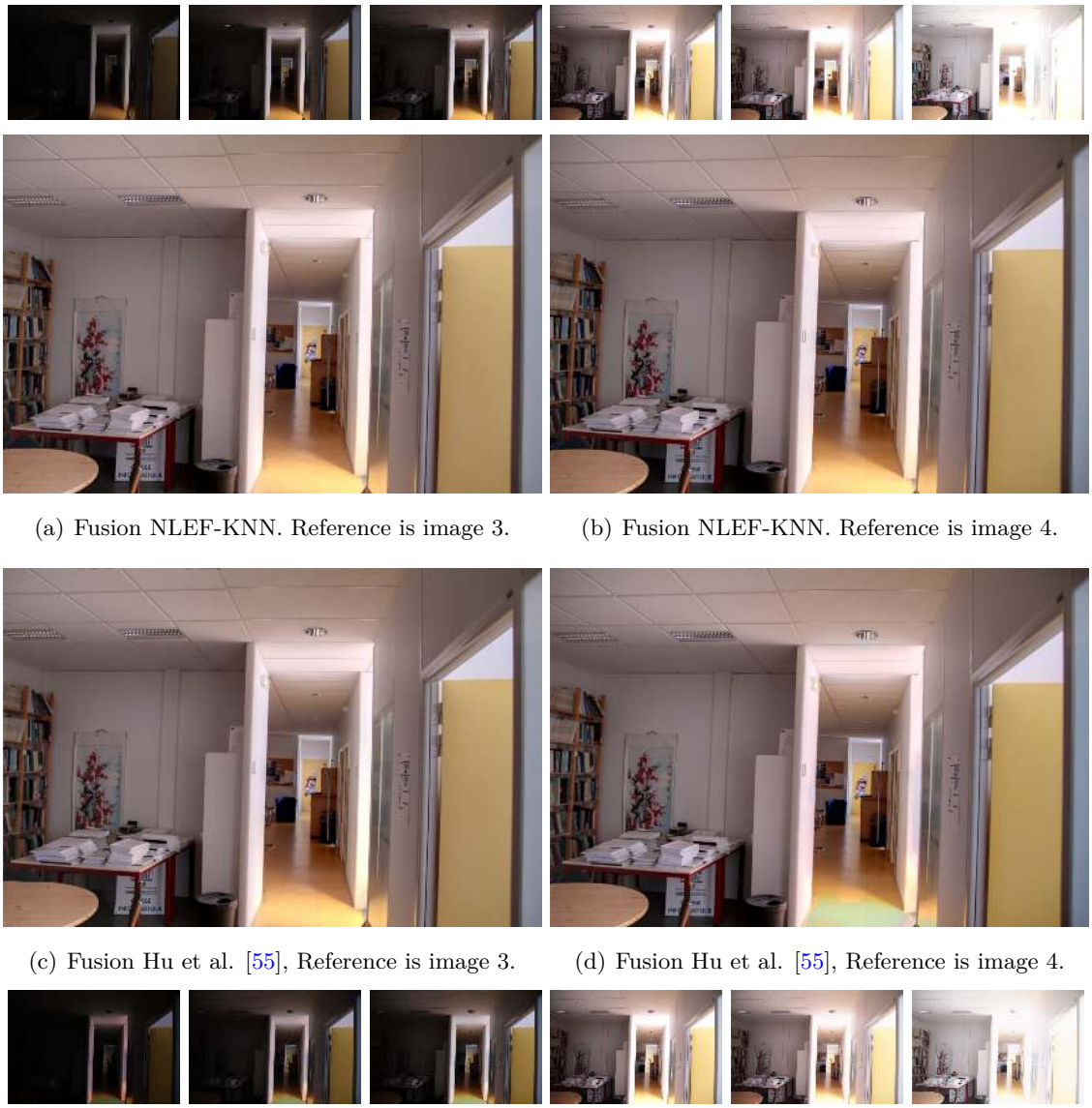
(a) Fusion NLEF-KNN. (b) Fusion Hu et al. [55].

Fig. 4.14 Image fusion on a scene having large displacement of small and large objects. **First row:** Source images with large object motion. **Second row:** Fused images for NLEF-KNN and Hu et al. [55] using the as reference the image 1.

4.3.5 Robustness to reference change

The input images from figure 4.15 present global misalignments due to camera motions.

By taking image 3 as a reference, noise gets partially enhanced with NLEF-KNN. That is because the reference contained localized noise that was highlighted during the contrast normalization and the 10 KNN's were not enough to get rid of it in totality. Nevertheless, contrast changes were suitably corrected.



(a) Fusion NLEF-KNN. Reference is image 3.

(b) Fusion NLEF-KNN. Reference is image 4.

(c) Fusion Hu et al. [55], Reference is image 3.

(d) Fusion Hu et al. [55], Reference is image 4.

Fig. 4.15 **Top row:** Source images. (a)-(d) Fused images. **Bottom row:** Reconstructed images with Hu et al. [55] using image 4 as reference.

When using a longer exposed reference (image 4) the illumination was consistently good, and no geometrical artifacts, like noise or false colors appeared during the fusion, see figure 4.15(b).

Regarding Hu et al. [55], using a darker reference yields well balance of colors and robustness to noise. However, even in the absence of motion, the under exposure of the reference (input image 3) provokes geometrical distortions on dark regions, see the dark regions under the square table, figure 4.15(c), where one of the table's leg has disappeared. For a brighter exposed reference (image 4), the strategy of [55] fails badly due to large

saturated regions on the reference, resulting in false colors. Specifically, those colors appear on all reconstructed images, bottom set figure 4.15, that have lower exposure than the reference, see bottom row.

This experiment shows that NLEF is less sensitive to the choice of the reference image, in contrast to Hu et al. [55] that is prone to reduce its performance because of enlarged saturated regions from higher exposures.
However, this also highlights the limits of NLEF, which is able to denoise a dark reference image upon comprehensible levels of noise. Regarding saturations, its robustness depends on a successful reference enhancement (*STAGE II*), specifically when transferring well exposed regions, with no motion, from a lower exposed image.

4.3.6 Patch size change

In some cases, the fusion with Hu et al. [55] gets harmed by the tunning of the patch size, which may result in geometrical errors (figure 4.16, upper right corner) or false colors (figure 4.19).
As reported by the authors, Hu et al. [55] method is affected by small patch sizes. A small patch size, say 5×5px, will reduce the matching precision, which results in badly estimated radiometries. On the contrary, larger patches improve the matching strategy and the possibility to fail reduces, but the computational cost increases. This behavior was observed on all our experiments for which we adjust the patch size to 10×10px to increase the efficiency of [55].
Particularly, Hu et al. [55] method appears in disadvantage when it has to reconstruct flat areas much larger than the patch size, like in figure 4.16(a). For that case, an erroneous structure appears in the sky. Indeed, there we can see that the reconstruction did a good job in the vicinity of unsaturated areas, but it lost precision at the center when there was not geometry guiding the matching.

In the case of NLEF-KNN the patch size change does not have a negative impact in the global appearance of the image, as it does for Hu et al. [55]. As we can see on figures 4.16 and 4.17, the colors of the final image are not corrupted and the geometry of objects around and at saturated regions is not deformed. Nevertheless, it does present a reduction in the precision to reproduce details, a situation that is alleviated by adding more KNN maps in the refinement of the reconstruction. In fact, this is not an unfair practice knowing that Hu et al. [55] method refines the reconstructed image with a patch-based approach. That means that when patches of 5×5 and 10×10 are used, the estimation of each pixel will be determined by using 25 and 100 pixels, respectively. In comparison, our refinement (see

Section 3.3.2) only uses 10 pixels that result from each estimated nearest neighbor fields by using Patchmatch with KNN=10, but for this experiment we tuned it to KNN=30.

Our method does not show a dependence between the dimension of saturations and the patch size. For a suitable compromise between precision and computational efficiency, we can tune the patch size to 3×3 without risking precision.

(a) Patch 5×5. (b) Patch 10×10.

(c) Patch 5×5. (d) Patch 10×10.

Fig. 4.16 Final fusion after image reconstruction with different patch sizes. **Top:** Hu et al. [55]. **Bottom:** NLEF-KNN.

(a) Patch 5×5. (b) Patch 10×10.

(c) Patch 5×5. (d) Patch 10×10.

Fig. 4.17 Final fusion after image reconstruction with different patch sizes. **Top:** Hu et al. [55]. **Bottom:** NLEF-KNN.

4.3.7 Geometric distortions within the fused image

A particular case showed to be difficult for NLEF-KNN, when the reference is not properly enhanced and the contrast normalization fails. The combined two factors may create inconsistent geometrical structures that are not totally corrected during the reconstruction and refinement.

Figure 4.18 shows an example of such a situation. There, saturated zones in the reference are removed with the help of the input image 1. During the radiometric normalization, inconsistent blue intensities replace a flat gray cloud in the upper central part.

(a) Fusion NLEF-KNN. (b) Fusion Hu et al. [55].

Fig. 4.18 Geometrical distortions in NLEF-KNN.

4.3.8 False colors due to contrast normalization failures

Failure due to the creation of false colors is an additional case observed with Hu et al. [55]. This situation tends to appear within or around large saturated zones in the reference image. On such cases, an increasing error appears and propagates during the reconstruction of neighboring images. In particular, the intensity mapping function (IMF) estimation is prone to fail after the imprecise image reconstruction. To compensate for these situations, the reference image should be better chosen and the patch should be larger.

(a) Fusion Hu et al. [55]. (b) Fusion NLEF-KNN.

(c) Fusion Hu et al. [55]. (d) Fusion NLEF-KNN.

(e) Fusion Hu et al. [55]. (f) Fusion NLEF-KNN.

Fig. 4.19 False colors in the fusion with Hu et al. [55].

Figure 4.19, shows the fusion with Hu et al. [55] and the fusion with NLEF-KNN. For the first two set of images, figures 4.19(a) and 4.19(c), the fusion appear unnaturally more green and yellow in the background when using Hu et al. [55].

4.3.9 Simplified image fusion

In this section, we compare the best configuration of our method, NLEF-KNN, with the simplified approach from section 4.2.6.
Figure 4.20 shows the results with this approach. In our experiments, we noticed that the simplified approach showed to be very good for cases where the difference of expositions between the images is small, with little presence of underexposed or overexposed areas on the reference. However, such conditions are often not satisfied in a regular basis and the fusion would generally contain presence of noise around dark zones. Noise is enhanced as a result of normalizing the reference to brighter images, and it appears in the final image because Exposure fusion prefers better exposed structures. This is noticeable in the zoomed insets from figure 4.21 (right column) that enclose dark regions. Contrarily, the NLEF-KNN method does not present these problems (figure 4.21, left column), because the reconstruction of brighter expositions does not present noise since the original images do not include noise. Moreover, further inconsistencies are wiped out with the KNN refinement.

Fig. 4.20 Fusion of the simplified approach.

It is worth emphasizing that combining only the mapped reference excludes valuable information from other images that could enrich the fusion. This highlights the need of a patch-based similarity search, where the impact of noise will be reduced as the patch size increases. Moreover, the patch-based reconstruction also neglects problems on the registration stage, and the color normalization by searching for neighborhoods that resemble the reference.

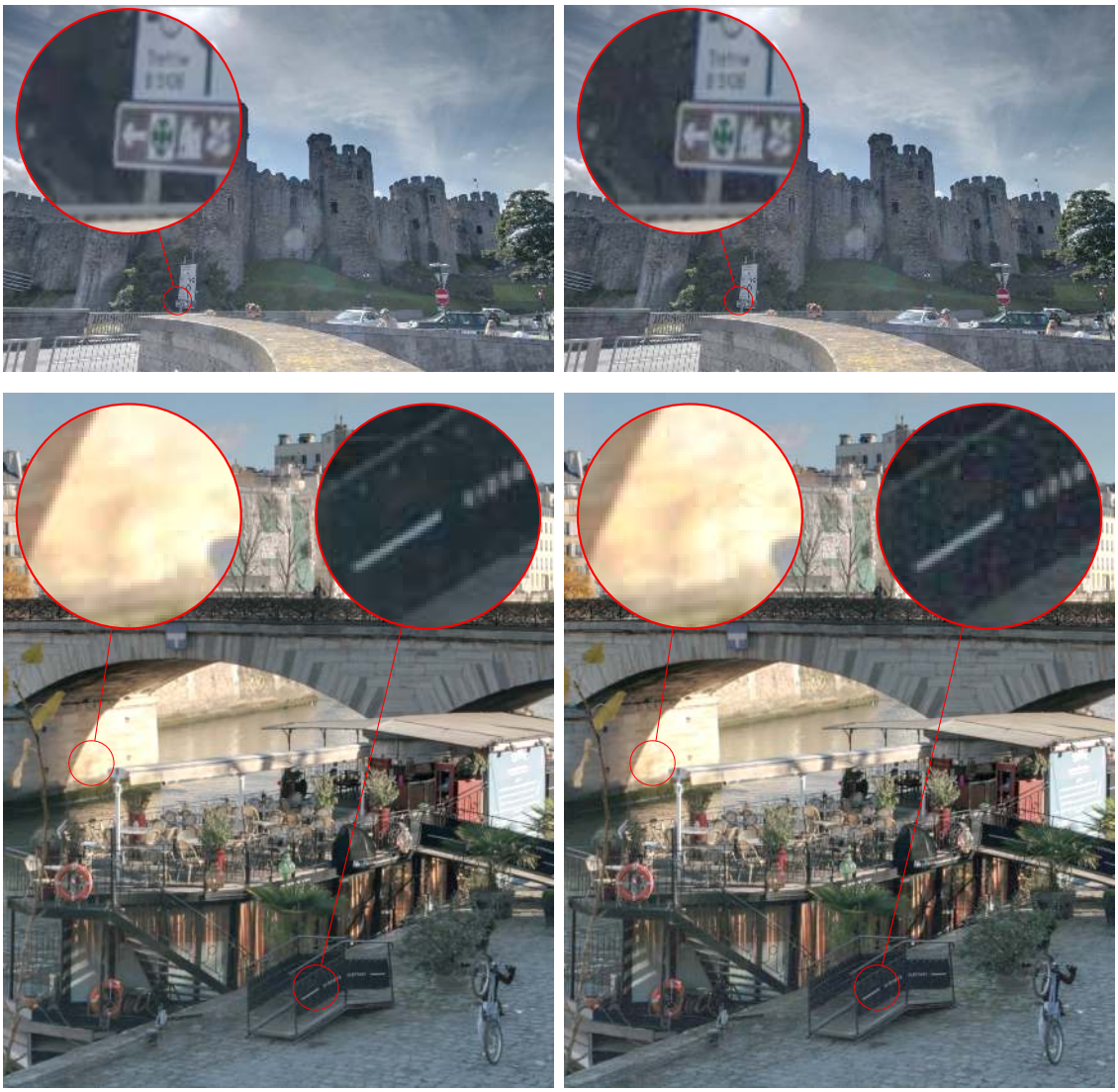


Fig. 4.21 Comparison of the fusion with the reconstruction approach NLEF-KNN (left column) and the simplified method (right column).

We also experimented with the simplified method but using the less exposed image as reference. On such setting, the fusion is prone to undesired artifacts that come from the darker image, like quantization errors and noise augmentation. This is because the shorter exposure does not preserve details and contains strong levels of noise. However, the simplified approach works very well for controlled situations where the less exposed image was not very underexposed and its exposure change was not very big with respect to other images from the set.
Figure 4.22 shows two cases where this simple strategy fails (top row). As well as two cases where the fusion shows good quality and almost no artifacts (bottom row).

Fig. 4.22 Exposure fusion with the less exposed image as reference in the simplified approach. **Top:** The radiometry is preserved but the fusion presents strong noise in dark areas (left image) or quantization errors (right image) that come from inaccurate reproduction of dark objects after acquisition. **Bottom:** For cases when the reference image is not saturated and the exposure level is not too different from the other images, the fusion is very good.

4.3.10 Correcting errors within the simplified approach

The previous experiment is interesting in the sense that the simplified approach is very straightforward and has the potential to produce good results with very few computations. In order to explore further this simple setting, we reduce the noise by denoising the contrast normalized image.

For that we use the simple Yaroslavsky based correction and the state of the art "Noise Clinic filter" [69], with its default parameters.

The results from figure 4.23 show that, in this setting none of the methods reduce completely the amount of noise from the images in the dark regions, but the impact of noise is strongly reduced. For both methods the contrast in the images remains the same.

(a) Fusion - Yaroslavsky Based. (b) Fusion - Noise Clinic denoise filter.

Fig. 4.23 Filtering the simplified fusion.

4.3.11 Non rigid dense correspondences for exposure fusion

The work by Hacohen et al. [51] has the aim of finding dense correspondences among images that present extreme radiometric and geometric transformations. Consequently, it has inspired most of the state of the art reconstruction methods presented in Section 2.2. The experiments below, show the limitations of [51], for multiexposure images with non rigid changes.

With their method, the produced NNF map is used to reconstruct an image that presents strong geometrical inconsistencies. Particularly, the fusion fails on regions where there was a displacement or small deformations, figure 4.24, left column. A reason for this algorithm to fail is because it is proposed for more drastic radiometric changes than those displayed by multiexposure image fusion images, and because too much invariance yield a poor ability to identify similar patches. On the contrary, excellent results are achieved with the fusion of the contrast normalized images, figures 4.24, right column.

Fig. 4.24 Fusion of the reconstructed images with NRDC (left column) and with the contrast normalized images (right column).

4.4 Conclusions

This chapter presents a methodology for the fusion of bracketed exposure images under dynamic settings. Our method, especially with the KNN version, has proven to produce accurate results under the presence of motions and strong illumination changes. Colors appear as vivid as in the exposure fusion method.

Our method combines a carefully crafted contrast normalization and a sophisticated non local combination to create a simple but efficient methodology. Experimental results let us conclude that our method is more flexible than its counterparts, by making very weak assumptions about the acquisition settings, **e.g.** exposure time. We do not explicitly estimate the camera response function, or use any tone mapping operators. The main contribution of the proposed scheme, is a more clear and well structured methodology for multiexposure image fusion (MEIF) without the often complex optimizations from reconstruction algorithms such as those in [115, 55, 54, 51]. We show that rather than focusing only on a sophisticated framework for detail reconstruction, it is crucial to ensure the synergy between three key components: radiometric normalization between the images, displacement map estimation and an additional geometric refinement, like the multi-NNF aggregation. This framework achieves good results without compromising the quality of colors and details, and also yields additional robustness to noise.

We also confirmed that in the motionless case, our results are almost identical to Exposure Fusion.

Chapter 5

Multifocus Image Fusion

When a photograph of a scene is taken, the geometry of the optics imposes a certain depth of field. Outside the in-focus region, objects are blurred. A simple solution to this problem is to acquire the scene at multiple depths of field and combine the images to produce one blur-free image. Most existing multifocus image fusion (MFIF) algorithms are restricted to the case where no movement of the objects or camera has occurred. In this case, MFIF boils down to the detection of the sharpest image among the stack, for each pixel of the scene, and a mixing algorithm to seamlessly fuse the images.

When the objects or the camera move, the problem becomes much more difficult. If the movement is not taken into account, ghosting artifacts appear after the fusion and objects are either duplicated or distorted. Whereas sole camera movement can be, to some extent, corrected by the computation of homographies linking the different images (for a small movement around the optical center, or planar scenes) to a general movement including objects displacements is challenging.

We propose a patch-based correction that accounts for global and objects misalignments, and can be used as an initial step to any standard MFIF methods.

Before detailing our method in section 5.2, we propose to solve MFIF for static scenes by including a measure of blur based on the local total variation (LTV), and show its multi-scale extension in section 5.1. Extensive results and comparisons are given in section 5.3.

5.1 MFIF for Static Settings

We place ourselves in the case where images are perfectly aligned and denote by I_i the stack of N partially out of focus images. The general approach to perform MFIF consists of a candidate region selection and fusion. Basically the sharpest regions among all images are localized, for each individual pixel, and combined to build a sharp scene. The fusion can give priority to some images using a weighted combination, or be performed at multiple scales or stitched like a collage, 5.1(d).

A simple possible fusion is presented below:

$$F(x) = I_k(x) \ , \qquad \text{where} \quad k = L(x) \in [1,\ N] \ , \tag{5.1}$$

where $L(x)$ is the index of the sharpest image at position x, and F is the synthesized image whose sharpness is expected to be better than any of the I_i. Notice that for this model, only one pixel among the stack is used for synthesis, the remaining pixels are discarded. L is a decision map that labels the sharpest pixel among the images and is estimated via region, pixel or patch-based measures. The latter one being the most popular choice for sharpness estimation. Please refer to Chapter 2 for a state of the art on the field.
Figure 5.1 presents a synthetic example where the image 5.1(c) has been blurred at different regions with several low pass filters (Gaussian Filter) of increasing blurring effect, $\sigma_{blur} = [1.5, 1.8, 2.1, 2.5, 3]$. Figure 5.1(a) shows the sharp regions for each filtered image.

Fig. 5.1 **Top row:** Set of synthetic multifocus images I_i. (a). Ground truth sharp regions. (b). Estimated sharpness by the LTV, with $\sigma = 5$. (c). Original Image. (d). Direct fusion output, PSNR = 33.39 dB. (e) Multiscale Fusion, PSNR = 34.66 dB.

The direct combination method of equation (5.1) is a simple alternative to synthesize a sharp image. However, it is prone to enhance seams at the frontier of pixels sampled from different images. The resulting image, even thought close to the original image (Figure 5.1(c)), presents artifacts that corrupt the quality of synthesis. For instance notice the zoomed boxes within figure 5.1(d).
In fact, this is the same problem tackled by Exposure Fusion [91], and is solved in a multiresolution fashion by taking notions from Ogden et al. [100]. In the spirit of [91] and [100], we describe a MFIF framework guided by a decision map that is computed via the local total variation.

5.1.1 Focus Measure with the LTV

To assess if an object falls in the focal plane of the camera, there are multiple alternatives proposed either in the spatial domain [61, 73] or in the frequency domain [134, 121]. Some popular methods analyze the variance, topology or contrast of a block. Among them, energy of image gradient and energy of Laplacian, are widely used. In [57], the authors listed the desired properties of a useful focus measure, it should be independent of image content, monotonic with respect to blur, have minimal computational complexity and be robust to noise. In the case of aligned images, the fact that the underlying scene is the same for all images of the stack simplifies the challenge that a trustable measure of blur represents. Indeed, one only needs a local measure that can serve as a comparison guide and not necessarily an absolute measure of blur. In other words the independence with respect to the image content is no longer needed. The main idea that we follow is that when the same image feature is subjected to various amounts of blur, its variations (derivatives) tend to decrease. The smaller the energy of a derivative is, the blurrier the image. The monotony is therefore respected. Finally, an integral of the energy over a small patch serves as regularization that brings robustness to noise. We therefore, can present local measure of (inverse) blur that we call LTV_σ.

Let $u(\mathbf{x})$ be an image with $\mathbf{x} \in \mathbb{R}^2$. We express the local total variation as follows:

$$LTV_\sigma(u)(x) = \left(\|\nabla u\| * K_\sigma\right)(x) \ , \tag{5.2}$$

where K_σ is a constant kernel of width σ (scale parameter). The operator $\{*\}$ is the convolution.
By using equation (5.2), we can define the decision map L as:

$$L(x) = \arg\max_i \ LTV_\sigma(I_i)(x) \ , \tag{5.3}$$

which produces the map displayed in figure 5.1(b). The LTV captures accurately sharpness within the images, producing well defined frontiers of the different levels of blur, and thus is a close approximation to the groundtruth map.
Nonetheless, notice that for strong discontinuities within the images, the LTV has a powerful response when the blur is strong, $\sigma_{blur} = 3$. This happens in figure 5.1(b) where strokes corresponding to strong edges indicate that the "sharpest" image is in fact the one with larger blur.

This spurious decisions are in fact explainable as a scale artifact. Indeed, when a strong edge is subjected to a strong blur, its total variation is spread through a window roughly of the size of the blur itself. When width σ of the window over which we aggregate the gradient's norm is smaller than the spread then, in regions in the vicinity of the edge (situated farther than σ but within the blur width), the sharp image will have a zero LTV measure whereas the blurry image will have a non zero measure because of the spread.

To solve this issue we propose a multi-scale approach that we detail below.

5.1.2 Multiscale Fusion

As a matter of fact, the LTV weakness to strong blur can be solved if the scale parameter properly encloses the local content. Instead of being constant it should be inversely proportional to the blur strength. A simple technique to guarantee that, is by using a multiscale approach. As in [91, 100], we use the Laplacian pyramid to solve the problem.

In order to perform a multiscale combination for totally registered images I_i we propose the Algorithm 3. First, we extract the Laplacian pyramids and the LTV from each image. Then for each level l of the pyramid, we combine the Laplacian responses by giving priority to the largest LTV, as follows:

$$\mathcal{L}_l(F) = \sum_{i=1}^{N} \mathcal{G}_l(\mathbf{W}_i)\mathcal{L}_l(I_i) \ , \tag{5.4}$$

where $\mathcal{L}(u)$ and $\mathcal{G}(u)$ are the Laplacian and Gaussian pyramids of image u, respectively. The matrix $\mathbf{W}_i(x) \in [0, 1]$ is an indicator function that is '1' only for the image that encloses the largest LTV at position x.

$$\mathbf{W}_k(x) = \begin{cases} 1, & \text{for } k = L(x) \ , \\ 0, & \text{otherwise.} \end{cases} \tag{5.5}$$

where L is the decision map from equation (5.3).
Finally, the Laplacian pyramid obtained from equation (5.4) is recomposed, yielding the final image F. This strategy reduces visible artifacts, thus improving the quality of synthesis, see figure 5.1(e).

Algorithm 3 MFIF on Static Settings

Input: Aligned stack of multifocus images I_i, with $i \in [1, N]$.
Goal: Produce a sharp image F.

1: Decision map estimation $L(x)$.
2: Computation of $\mathbf{W}_i$.
3: Laplacian pyramid extraction $\mathcal{L}_l(I_i)$.
4: Gaussian pyramid extraction $\mathcal{G}_l(\mathbf{W}_i)$.
5: **for** each level l **do**
6: Combine the Laplacian pyramids $\mathcal{L}_l(I_i)$.
7: **end for**
8: Recompose $\mathcal{L}_l(F)$.

For real static settings the Algorithm 3 also presents positive results. Figure 5.2, displays a set of 13 images with varying depth of field.

The artifacts that appear with the naive direct fusion (figure 5.2(c)) are corrected when applying the multiscale combination (figure 5.2(d)). The fusion is a natural-looking image with sharp regions all over its spatial domain. Besides the simplicity of this method, the scale of LTV has little influence.

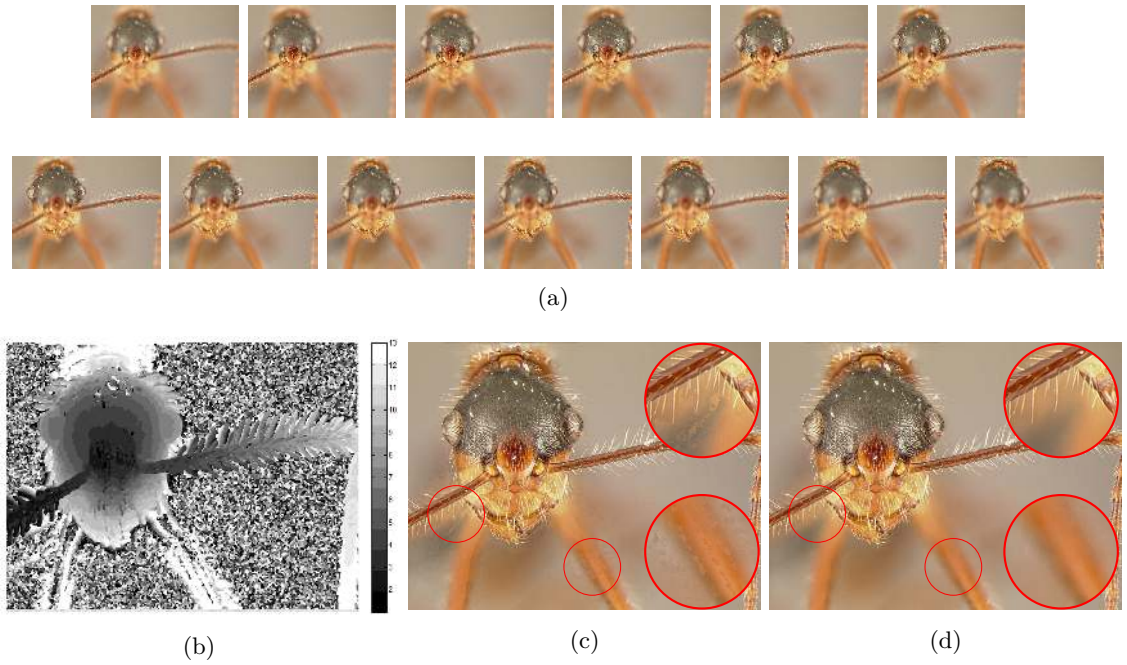


Fig. 5.2 **Top row:** Set of multifocus images. **Bottom row:** (b) Sharpness map with LTV. (c) Direct fusion. (d) Multiscale multifocus fusion.

5.2 MFIF For Dynamic Settings

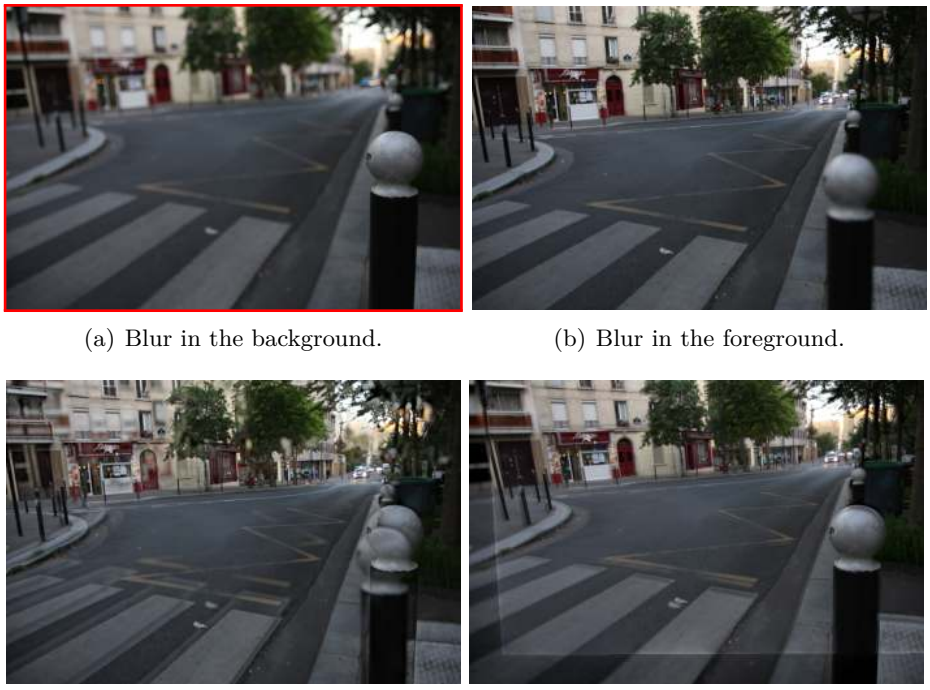
On dynamic settings, the images are captured with object or camera motions. For these cases, the method presented above (Section 5.1) is likely to be flawed and produce ghosts effects or artifacts, such as double edges. As of today, very few MFIF methods [136, 118, 49, 70] have approached the challenging problem of moving acquisitions. In most cases, only rotations and translations are solved by registration. However, alignment methods are not a sufficient solution to correct object motion. Besides, their performance is restricted to moderate blur levels or geometrical deformations.
Below, we expose the standard approach to solve global misalignments and present a new method to solve object motions.

5.2.1 Image Registration

Image registration of out of focus images is generally performed by estimating an image transformation that corrects the geometric distortion between the images. Such transformation is designed to reproduce precisely the spatial change of common landmarks. Where such landmarks should be localized in a blur invariant fashion.

Some strategies to align out of focus images include homography estimation [49], affine estimation based on multiscale optical flow [136], or a MAP formulation to correct translations and fuse the images at the same time [118]. Another approach is to render the fusion robust to small deformations [70]. In this method the authors define alpha matte functions to better refine boundaries of sharp regions.
Under suitable conditions of the input images, these algorithms work well. Nevertheless, their efficiency is reduced mainly in two cases, whenever blur is very strong or the images contain complex deformations. In the first case, landmarks, usually defined as lines or corners may be difficult to match, thus reducing the amount of data to estimate correctly the geometric transformation. In the second case, the image settings may be complex and solving only for translations [118] or small displacements [70] will not deal with the problem. Besides, deformations of moving objects or non rigid changes are not considered.

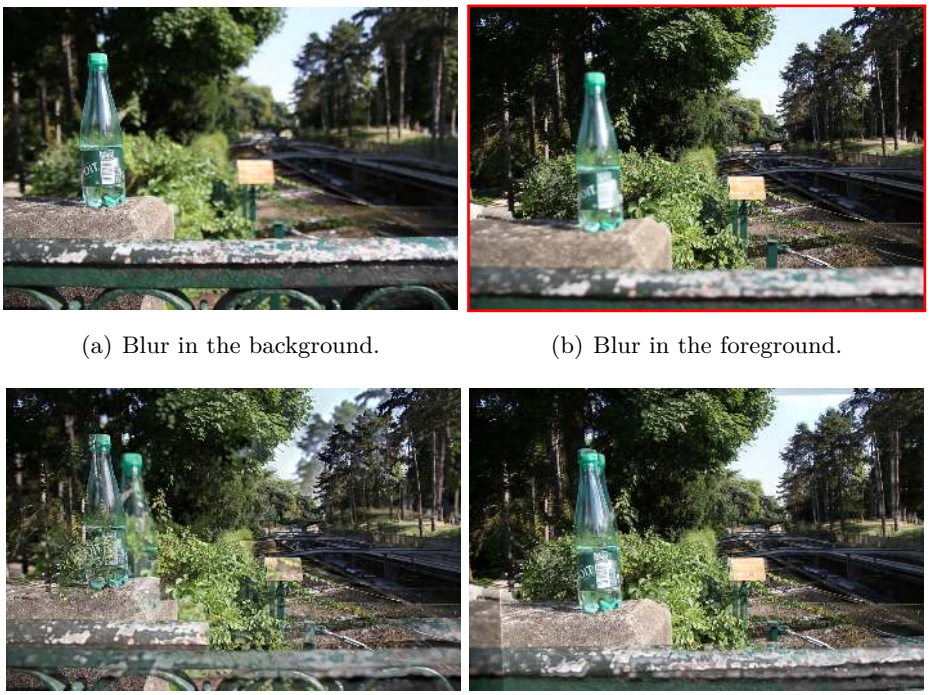
With that in mind, we aim to use a registration method that does not need to be perfect, especially in the presence of moving objects, but that does a good job for global alignment. We have found that the classical image registration framework (SIFT keypoints, RANSAC and homography estimation) satisfies our requirements, while allowing for a variety of deformations. In our case, we apply this framework on the second level of the Gaussian pyramid and scale the estimated homography to the full resolution, for initialization. This is done to prevent small blur from affecting the estimation.



(a) Blur in the background. (b) Blur in the foreground.

(c) (d)

Fig. 5.3 **Top row:** (a-b) Set of multifocus images. The reference is framed in a red box. **Bottom row:** (c) Fusion of the input images. (d) Fusion of the aligned images.



(a) Blur in the background. (b) Blur in the foreground.

(c) (d)

Fig. 5.4 **Top row:** (a-b) Set of multifocus images. The reference is framed in a red box. **Bottom row:** (c) Fusion of the input images. (d) Fusion of the aligned images.

For MFIF, we register the set of images to the sharper image in the set, then fuse with the Algorithm 3 from Section 5.1. See figures 5.3-5.4 for image fusion on handheld camera images.

As expected, the fusion of images that include motion, is perturbed by artifacts in the final image, figure 5.4(c). Notice that the fusion after image alignment produces more visually appealing results. However, even with no moving objects, the homography does not produce perfect registration. In this case, because the scene is not planar and the optical center has moved.

5.2.2 Feature Based Geometric Alignment

In the previous section, we stressed out the importance of image registration for dynamic cases. Existing methods solve global misalignments to some extent, thus reducing substantially the creation of artifacts during fusion. Conversely, moving objects move independently, thence no single estimation of global transformation can correct them. Instead moving objects should be treated independently to fully remove artifacts associated to their movement.
In this section we propose a technique that not only solves residual errors after image registration, but also corrects artifacts due to moving objects.
The idea is to use patches to find local correspondences, in a similar framework to the non local exposure fusion described in Chapter 4. As seen before, scanning a local neighborhood for matching patches among the images, can provide independence to motion or complex deformations. However a different approach should be taken this time, because standard similarity measures fail to associate blurred and sharp patches from the same object. In essence, we use a reference-based approach to avoid the influence of moving objects and correct its blurred regions by using the sharp content from the remaining images. Then the fusion can be performed with the procedure explained in the previous section.

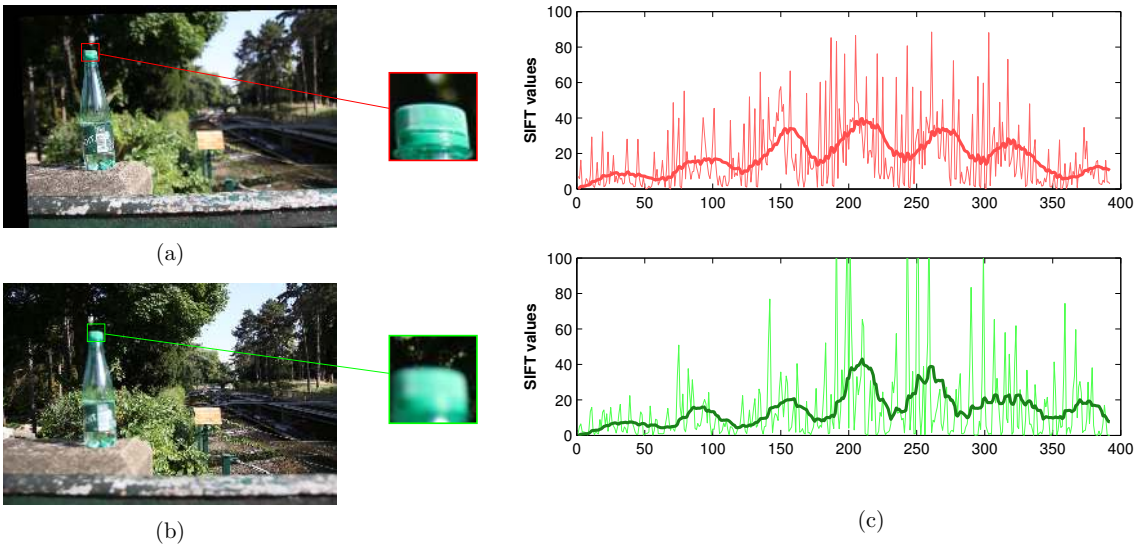
Description

Given a set of images with various in-focus regions, the purpose of our method is to sharpen one of the images by combining with the remaining images. Logically, sharp regions on the reference should be preserved and blurred regions should, ideally, disappear by using their sharp version from other images. For that we rely on a reference image, selected as the sharpest image in the set, which in practice is the image that has the larger total variation. This reference image is used to guide a search of corresponding regions among the set of images, and therefore identify where displacements occurred. The other images will be called sources.

By comparing each source image against the reference, the algorithm automatically identifies identical regions among the pairs. Such common regions are later inspected for sharpness and used to synthesize an image that is sharp all over its spatial domain; provided that all objects are sharp in at least one of the images.

Our method is inspired by non local patch search techniques, to account for spatial changes on the images, and additional feature measures to reduce the impact of blur degradation during the patch comparison. For that we guide the search using three distinctive features on (*i*) color, (*ii*) local geometry and (*iii*) space.

The first two constraints can be formalized by using several measures that are constant for sharp patches and their corresponding degraded versions, see figure 5.5. That is, given a sharp patch and its associated blurred version, a simple estimator for the color on both patches is the mean intensity. Second, to capture similar geometrical features we rely on local gradient orientations, since they tend to behave similarly over characteristic structures, like edges or corners (subject to the degree of blur). We also rely on SIFT descriptors in order to add robustness to blur and geometrical distortions. The third criterion is satisfied by limiting the space of search around each position.



(a)

(b)

(c)

Fig. 5.5 Shared characteristics within blurred and sharp patches extracted from identical objects: *average intensity, gradient orientations* and *SIFT descriptors*. The average intensity around the central position of each box is $\mu_{src} = [116, 228, 203]$, $\mu_{ref} = [124, 240, 209]$, for the source and reference, respectively. (a) Source image aligned to the reference. (b) Reference image. (c) SIFT descriptor for each zoomed box. Notice that SIFT descriptors are larger because they are extracted from a larger space, with a spatial grid of 7×7 while keeping the standard 8 orientations.

Figure 5.5, provides an example of the previous measures for two corresponding patches taken from a blurred (reference) and sharp (source) image. We can see that the average intensity is not strongly affected by the presence of blur. Likewise, SIFT descriptors have a similar behavior, which can be confirmed by their average curves (only used for illustration purposes). In this approach, SIFT descriptors are extracted densely at the original scale and over a larger spatial domain around each pixel.

We build a distance that represents the similarity of patches and is robust to blur to some extent.
That is, for a reference image R and a source image S, we compute the distance at $x \in \Omega_R$ and $x' \in \Omega_S$, as:

$$\begin{aligned} D(x,x') = \; & \lambda_1 \underbrace{\|\mu(P_R(x)) - \mu(P_S(x'))\|^2}_{\text{Color}} \\ +\lambda_2 \underbrace{\|\boldsymbol{\theta}_R(x) - \boldsymbol{\theta}_S(x')\|^2}_{\text{Orientation}} & +\lambda_3 \underbrace{\|D_R(x) - D_S(x')\|^2}_{\text{Descriptors}} , \end{aligned} \tag{5.6}$$

where the parameter $\mu(P_R(x)) \in \mathbb{R}^3$ extracts the average intensity from the patch $P_R(x)$ centered at position x from image R, the parameter $\boldsymbol{\theta}_R(x) \in \mathbb{R}^{2p^2}$ is the vector made of the unit normalized gradient (gradient divided by its amplitude) in an $p \times p$ neighborhood around x. $D_R(x)$ is a SIFT descriptor extracted from image R at position x, correspondingly for image S. In addition, the multipliers λ_1, λ_2 and λ_3 are tuned to combine the terms appropriately.

This distance is composed by three terms that combine color and geometry. The main purpose of this distance is to let us localize reliable patch correspondences, while being substantially invariant to blur. In essence, we intend for the distance to be proficient in determining whether a pair of patches are similar or not, while disregarding their blur perturbation. That is, the functional should be large for different patches, especially when they present large variations in geometry or radiometry.

Equation (5.6) is integrated into a minimization framework that is similar to the non local method from Chapter 4. We want to pair the coordinates (x, x') of patches so to minimize $E(x, x')$, thus correcting for motion within a local neighborhood. For that, we prealign the images to reduce geometric deformations and constrain the patch search to a limited spatial domain around each potential candidate.

In particular, blur invariance is directly proportional to patch size, where robustness is achieved with larger patches. This is because larger patches would enclose more discriminant information compared to small patches that contain fine details easily wiped out by blur.

Therefore, in the presence of blur, larger patches preserve more geometrical content, this enlarges the possibility to find suitable nearest neighbors.

Algorithm

In this section we explain how to integrate our distance of equation (5.6) in a focus stacking framework.
As in Chapter 4, the geometric alignment is performed independently for each source image S within the input set.
For each patch in the reference image R, we find a nearest neighbor in the source that best resembles the geometrical content from the reference's patch. The patch search is confined to a local neighborhood of predefined size around a candidate region. More precisely, we incorporate our distance from equation (5.6) in a multiscale Patchmatch framework for obtaining a nearest neighbor field M in an efficient manner. That is to say, we obtain a displacement map for a coarse version of the images, we interpolate it to a finer scale and take this as a seed for the PatchMatch at the finer scale (by doing so, we constrain the search in PatchMatch at a finer scale not to deviate too much from the coarser displacement map). We repeat this process until the finest scale is attained. This multiscale approach enforces coherence of the final result.

The generated displacement maps are employed to reconstruct the source images and produce a set that is geometrically consistent with the reference, therefore reducing the chances for artifacts to appear during the fusion. The general overview of our method is presented in Algorithm 4.

The operation $\mathcal{G}_l(R)$ denotes the Gaussian pyramid of image R at level l. The notation EPatchMatch defines the energy modified Patchmatch with map initialization M_l and constrained search window v around each potential candidate. Lastly, the operator $\mathbf{C}(M)$ corresponds to the Yaroslavsky based correction that forces the map M to be piece-wise regular (for a detailed description, see Section 3.3.3).
In fact, besides the refined map $\widehat{M}_i$, the Yaroslavsky correction also offers a measure $\phi(x)$ for the confidence of refinement at pixel x, and is obtained by $\phi(x) = \sum_i^N w(i)/N$. Where w (see equation (3.6)) is the local regularization term which indicates how big or small is the relative displacement of the nearest neighbors located in the vicinity of x. In other words, it defines the amount of local matches that satisfy the condition for the relative displacement to be smaller than σ, (see Section 3.3.3), and therefore gives us a hint about the piecewise regularity around x.

Finally, the image fusion is performed in a weighted fashion while giving priority to pixels that were well synthesized (relying on the confidence function ϕ). This is easily introduced in Algorithm 3, by modifying the weight function $\mathbf{W}_i$ from line 2 as below.

Algorithm 4 Non-local Multifocus Image Fusion

Input: Set of multifocus images S_i, with $i \in [1, N]$. Patch size p, search window size v, number of levels n.
Makes use of: EPatchMatch(A,B,M,p,v) : returns the displacement map between images A and B, using an initial displacement map M.
Goal: Build a set of aligned images L_i and fuse into final result I_f.

$i_o = \arg\max_{i \in [1,N]} \|\nabla S_i\|$
Reference selection, $R = S_{i_o}$.

procedure IMAGE REGISTRATION W.R.T. R
  for each image $i \in [1, N] \setminus i_o$ **do**
    T_i: Homography between R and S_i.
    $I_i = T_i(S_i)$ ▷ I_i is globally aligned with R.
  end for
end procedure

procedure PATCH RECONSTRUCTION OF R FROM I_i
  for each image $i \in [1, N] \setminus i_o$ **do**
    M_n: Map initialization with identity.
    for level $l = n$ to 1 **do**
      $r_l = \mathcal{G}_l(R)$ ▷ Downsample operation at level l.
      $s_l = \mathcal{G}_l(I_i)$
      M_l = EPatchMatch(r_l, s_l, M_l, p, v) ▷ Displacement Map Extraction.
      M_{l-1} = Upsample(M_l) ▷ Except when $l = 1$
    end for
    $[\widehat{M}_i(x),\ \phi_i(x)] = \mathbf{C}(M)$ ▷ Map regularization.
    $L_i(x) = I_i(\widehat{M}_i(x))$ ▷ Source images Reconstruction.
  end for
end procedure
I_f = Weighted-Fusion(L_i, ϕ_i) ▷ MFIF for the static case.

$$\mathbf{W}_k(x) = \begin{cases} 1, & \text{if} \quad \phi_k(x) > \epsilon, \text{ with } k = L(x) \ . \\ 0, & \text{otherwise.} \end{cases} \tag{5.7}$$

If $\mathbf{W}(x)$ is zero over all k images, meaning that no sharp image was well reconstructed, then we prioritize the reference by making $\mathbf{W}_{i_o}(x) = 1$, where i_o is the subscript associated to the reference image. This is a fall-back solution: we keep the information of the reference image when we are not sure of the reconstruction L_i.

5.3 Experiments and Results

For evaluation we conduct experiments on real scenarios rather than on artificially generated images. For that, we captured 25 scenes with an average of 3 images each. Within these sets, we included common perturbations of handheld camera acquisitions, namely, translations, rotations, moving objects or even non rigid deformations. In the analysis, we also use 9 static image stacks that are classically used in the MFIF literature.
The acquisition was done with the camera *Canon 7D EOS* on manual mode. The level of blur was altered by using the auto-focus point selection to select the focused object.
In all our experiments, we use a patch size $p = 5$ to compute the mean intensity, a search

window $v = 5$, and a threshold $\epsilon = 0.5$ for computing $\mathbf{W}$ (equation (5.7)). The number of levels in the Gaussian pyramid was set to $n = 4$ for images of 667×1001 pixels. As mentioned earlier, SIFT descriptors are computed from larger regions $7 \times 7 \times 8$ in order to enclose more geometric information. The λ terms of equation (5.6) were tuned by applying the distance to a set of pairs of perfectly aligned images. The values of λ_i were chosen as the inverse of the mean contribution of each of the three terms of equation (5.6) over a fixed set of stacks. For all our experiments we fixed $\lambda_{1,2,3} = [0.0845,\ 0.0533,\ 8.7266]$.

Experimental framework The classical way to evaluate MFIF methods is to measure the discrepancy between results and a given ground truth, see e.g. [57]. No-reference evaluations have also been proposed, basically by evaluating the ability of the fusion method to accurately account for the different images in the stack, see e.g. [78]. In both cases, many alternatives to the classical and limited L^2 norm can be used. However, all such evaluations are made under the hypothesis that the scenes are registered and still (with no moving objects). Now, the purpose of the present thesis is precisely to deal with mis-registration errors and moving objects. To the best of our knowledge, no database of dynamic images provides a ground truth. For this reason, results are evaluated visually on a database of challenging scenes. We provide comparisons of our method (NL-MFIF) with the static case, as well as with two state-of-the art methods [78] and [77]. Where the former method is proposed to fuse static scenes and the later one, aims at correcting image displacements with the local refinement of the sharpness map.

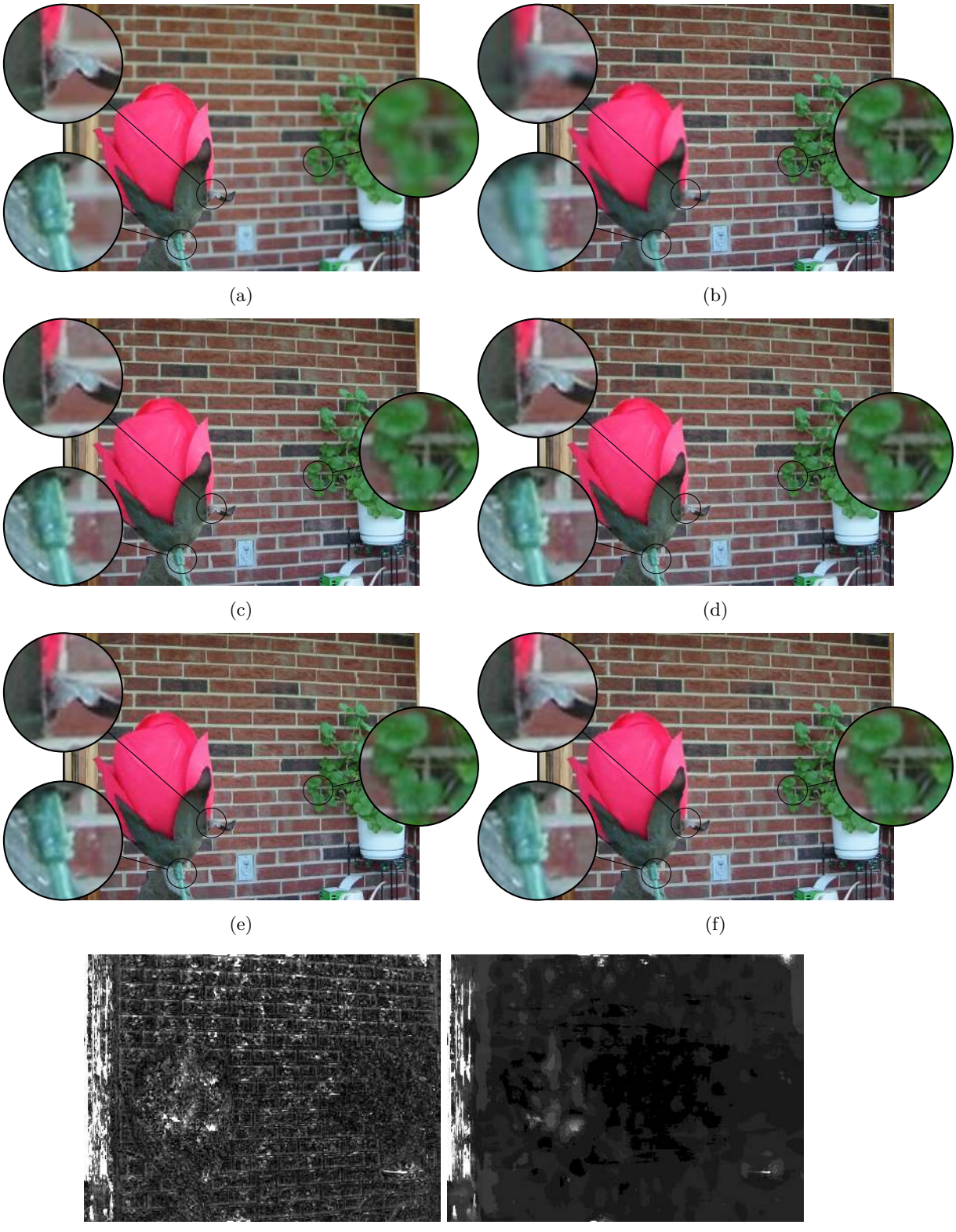
5.3.1 On Static Scenes

As a sanity check, here we are interested in testing NL-MFIF on static images where the only change between the images is the depth of field. For these cases, the estimated nearest neighbor field should resemble the identity map and the fusion should not contain artifacts. Besides the methods in [77] and [78] (originally proposed for the fusion of static and dynamic scenes, respectively), we compare the output of our Algorithm 4, with the proposed method for static cases, Algorithm 3. Results for two scenes are presented in figures 5.6 and 5.7, but the observations are consistent over all the sets.

Visually, the fused images for all four methods are almost identical, with sharp structures everywhere. This indicates that NL-MFIF, does not produce distortions in the fusion. For NL-MFIF, the accuracy in the fusion comes from the fact that the local geometric alignment chose almost the exact nearest neighbors in most of the cases. This is visualized in the relative displacement maps, figures 5.6(h) and 5.12(h). Notice that the maps are not zero everywhere because of geometrical redundancies over local neighborhoods. This can be seen particularly around edges or uniform flat regions, like the wall and certain parts of the book cover for figure 5.6, and around the bricks for figure 5.7.

(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 5.6 *Aligned Images:* (a). Aligned source image with sharp background. (b) Reference image with blurred background. (c) Fusion with [77]. (d). Fusion with [78]. (e). Fusion of the registered images. (f) NL-MFIF's output. (g). Relative displacement or offset map, obtained by the modified Patchmatch. (h) Refined offset map after Yaroslavsky correction.



(a) (b) (c) (d) (e) (f) (g) (h)

Fig. 5.7 *Aligned Images:* (a). Aligned source image with sharp foreground. (b) Reference image with blurred foreground. (c) Fusion with [77]. (d). Fusion with [78]. (e). Fusion of the registered images. (f) NL-MFIF's output. (g). Relative displacement or offset map, obtained by the modified Patchmatch. (h) Refined offset map after Yaroslavsky correction.

We also observed that depending on the number of input images and their variable level of blur, the results from [78] and [77] may differ slightly from our method. That is because very strong levels of blur may displace edges further than at sharp images. Also, because our method relies on a single image to reproduce the other images, this results in the preservation of the geometry of such image, the reference. This favorably prevents from blurry shadows around edges.

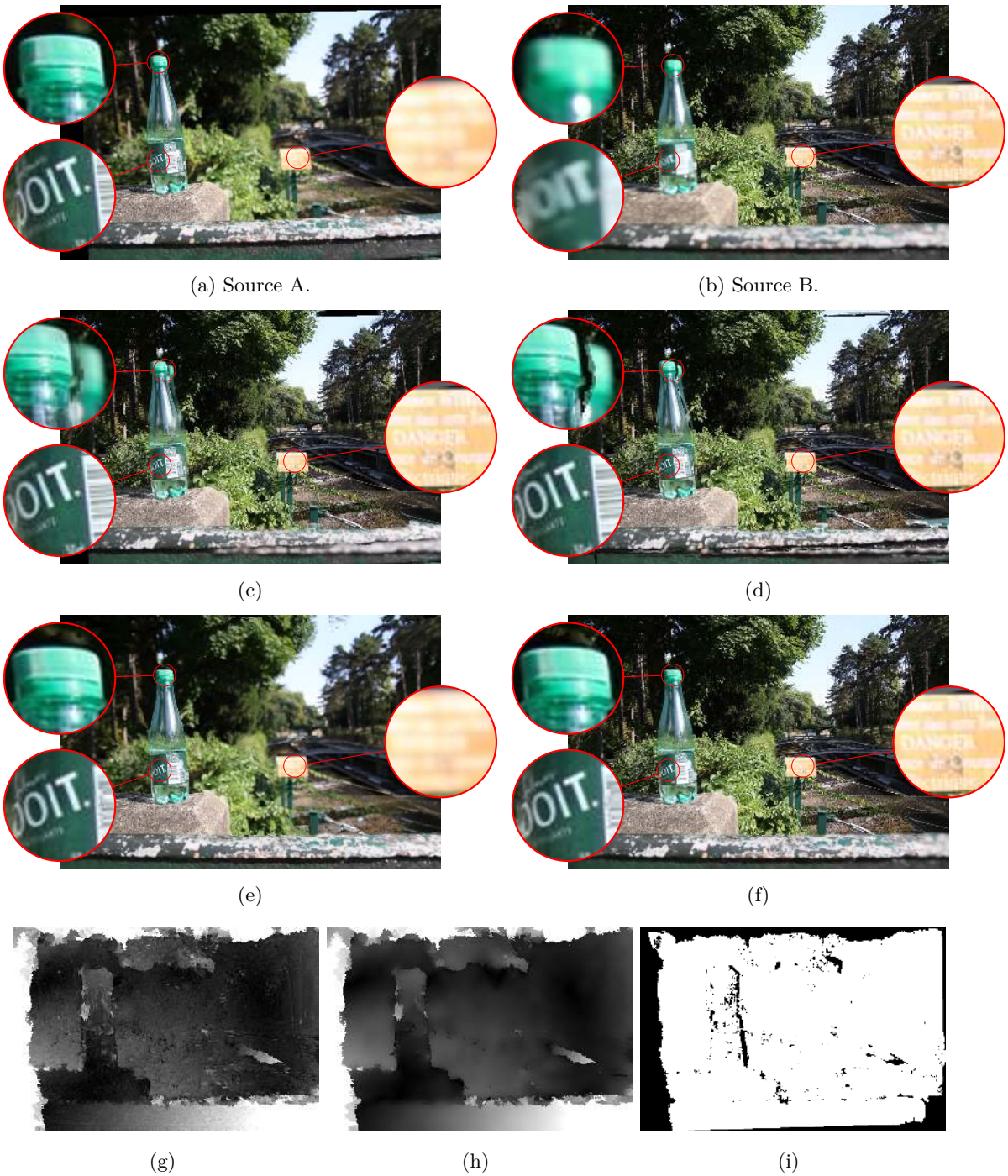
5.3.2 Correcting Image Registration Errors

There are several reasons for the image alignment to produce inaccurate results, e.g. strong deformations, non-planar scenes, viewpoint change, etc., these cases are analyzed in our experiments. For instance, figures 5.8 and 5.9 present the case where the background motion is mostly reproduced by the homography, but objects in the foreground are still misplaced, e.g. bottle and handrail.
For the sets in figures 5.8 and 5.9, notice that after registration, the input source images 'A' are slightly rotated and displaced when compared to the reference, and present unknown regions at the borders.

In the case of [77] (figure 5.8(c)), the method coherently renders sharp regions from the background of the sharper image. Here, this method for static scenes benefits from the fact that even if there was motion in one of the images, the background clearly appears sharp everywhere in the other. Yet, it fails where the misalignment occurs, e.g. the bottle and the unknown flat regions of the source image 'A', (also figure 5.9(c)).

Even though the method in [78] proposes a strategy to deal with misalignments, the fusion presents artifacts (figure 5.8(d) and figure 5.9(d)), such as ghosts and irregularities at the transition of unknown regions in the images.

On the contrary, NL-MFIF properly deals with such errors after registration and the unknown regions within the aligned images. The final fusion is an artifact free image that outperforms the other methods, figure 5.8(f). Observe that the proposed method renders the reference sharper in blurred regions, like the bottle, without corrupting sharp regions, like the warning sign. Moreover, in the worst case scenario when the reconstruction fails, the weighted fusion prevents from including untrusted regions ($\phi < \epsilon$) in the combination. This last aspect can be seen in the sky of figure 5.8(e), where the reconstructed image presents few irregularities that were successfully excluded from the final fusion.
This proves the importance of utilizing the confidence map ϕ, figure 5.8(i). Without such a map, high frequency errors would be tagged as sharp and used in the fusion. In this regard, empty holes within the map refer to structures whose nearest neighbors are not found coherently, and are commonly associated to regions that are not present on both images due to occlusions or strong deformations. See for instance, the border's map or the disoccluded region at the right side of the bottle.



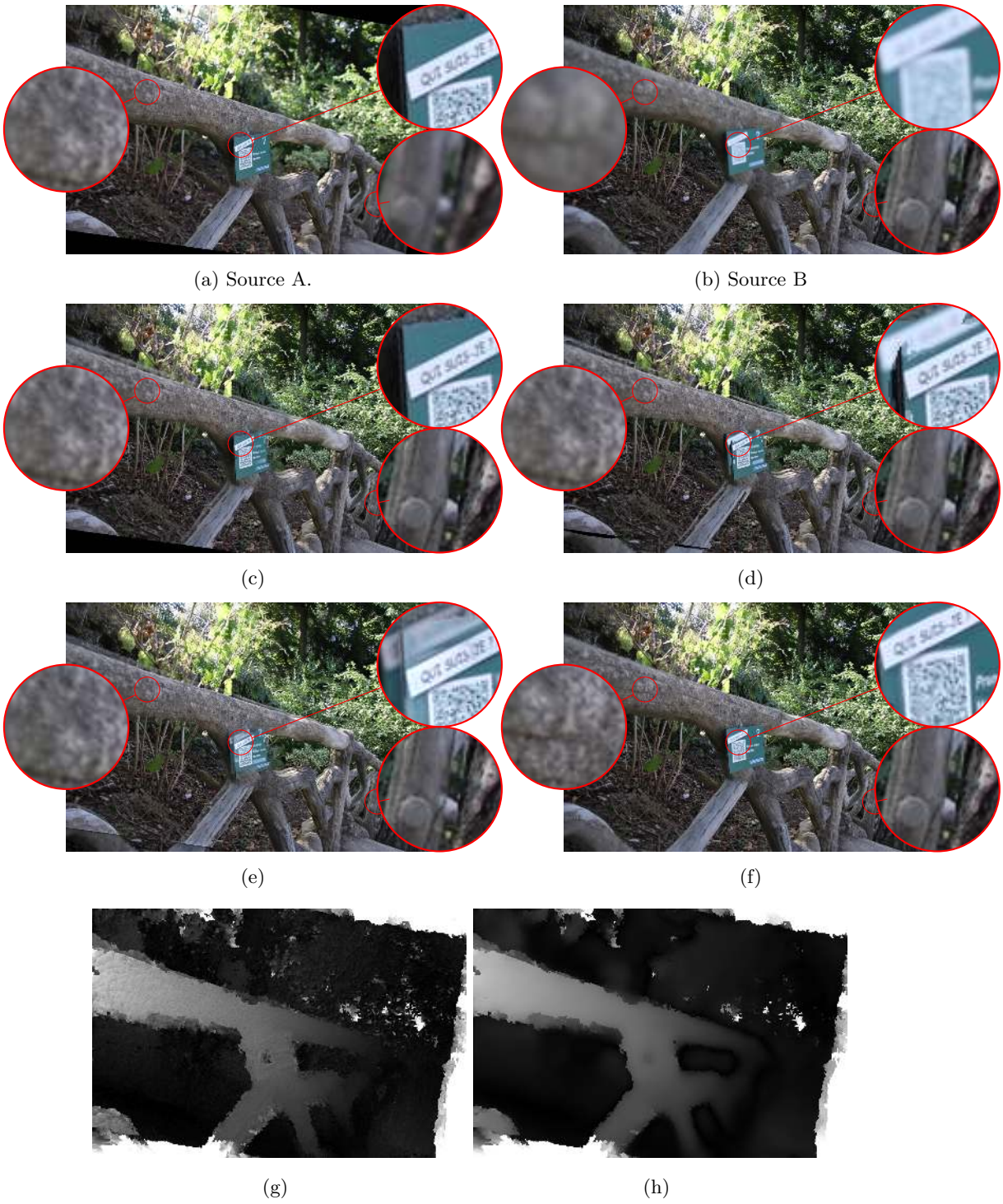
(a) Source A. (b) Source B.

(c) (d)

(e) (f)

(g) (h) (i)

Fig. 5.8 *Misalignment corrections:* (a). Aligned source image with sharp foreground. (b) Reference image with blurred foreground. (c) Fusion with [77]. (d). Fusion with [78]. (e) Geometrical reconstruction of the source image to the reference. (f) NL-MFIF's output. (g). Relative displacement or offset map, obtained by the modified Patchmatch. (h) Refined offset map after Yaroslavsky correction. (i) Confidence map after thresholding.



(a) Source A. (b) Source B

(c) (d)

(e) (f)

(g) (h)

Fig. 5.9 *Misalignment corrections:* (a). Aligned source image with sharp foreground. (b) Reference image with blurred foreground. (c) Fusion with [77]. (d). Fusion with [78]. (e) Fusion of the reconstructed scenes with the multiscale static method. (f) NL-MFIF's output. (g). Relative displacement or offset map, obtained by the modified Patchmatch. (h) Refined offset map after Yaroslavsky correction.

Notice that unknown regions are synthesized by taking pieces of information sampled from distinct regions, but only if they hold similar characteristics in geometry and color. This is noted in the relative displacement maps, figures 5.8(h) and 5.9(h), where the intensity encodes the distance from where the nearest neighbor was sampled. In general, nearest neighbors were sampled from within an average of 10 pixels, which indicates the common distortions after the image alignment. The best match for unknown structures, was sampled from farther positions, more than 20 pixels away.
It should be highlighted, however, that gaps in the confidence map are not necessarily associated to bad reconstructions. But since they are sampled more irregularly, they are more vulnerable to suffer from jitter or subpixel innacuracies, which may not be totally corrected by the regularization stage (figures 5.8(f) and 5.9(f)), thus we prefer to take precautions and prioritize the reference.

A common observation for NL-MFIF is that the fusion generally shares three favorable characteristics: blur is coherently reduced from the reference, sharp regions are not deteriorated and unknown regions within the source are reconstructed in a natural manner.

5.3.3 Correcting Object Motion

For well aligned images, object motions are the only cause of artifacts during fusion. On this regard, we evaluate the capacity of our algorithm to correct artifacts due to different kind of motions. We consider a variety of motion distortions, like rigid and non rigid motions within the images, figures 5.10, 5.12, 5.14 and 5.15.

In figure 5.10, the scene presents motion in the foreground and background. In the first case, flowers move independently due to wind. In the background we can see people moving. Figure 5.12 also presents complex geometrical changes, especially in the background where all the activity is concentrated, e.g. moving cars and people.

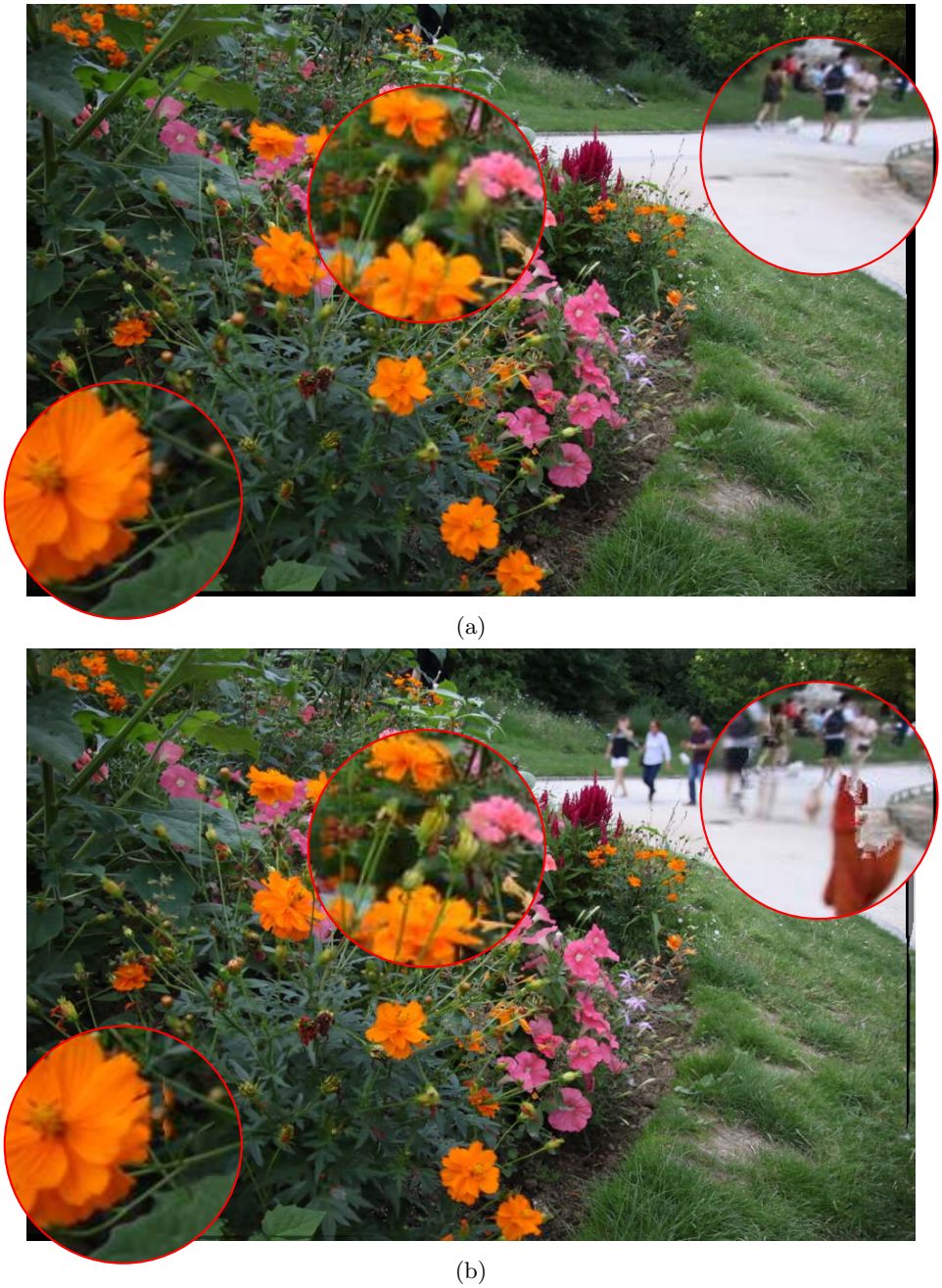
All these movements produce disocclusions of sharp regions, that consequently appear in the fusion as they are preferred over blurred regions, figures 5.10(a), 5.12(a) and 5.13(a). In spite of such challenging setup, the output of NL-MFIF is a natural looking image. Again, blurred regions in the reference are enriched with details by using the common content within the images.

A less frequent case was the appearance of small irregularities due to fine objects with small displacements. Artifacts of this type may occur when one of the reconstructed and refined images is not completely aligned to the reference at the region where the object moved. It is not necessarily because of reconstruction errors but rather because of the strength of regularization. That is, the aim of the proposed regularization is to make the vector field piecewise regular so that the boundaries of moving objects are not corrupted. For very small objects, a strong regularization of the vector field might reduce the precision in the reconstruction of those objects, which could yield small artifacs in the fusion.

(a)

(b)

Fig. 5.10 *Object corrections:* The top row images are the input set of out of focus images after alignment, the reference is the image 3. (a) Fusion of the registered images. (b) NL-MFIF's output.



(a)

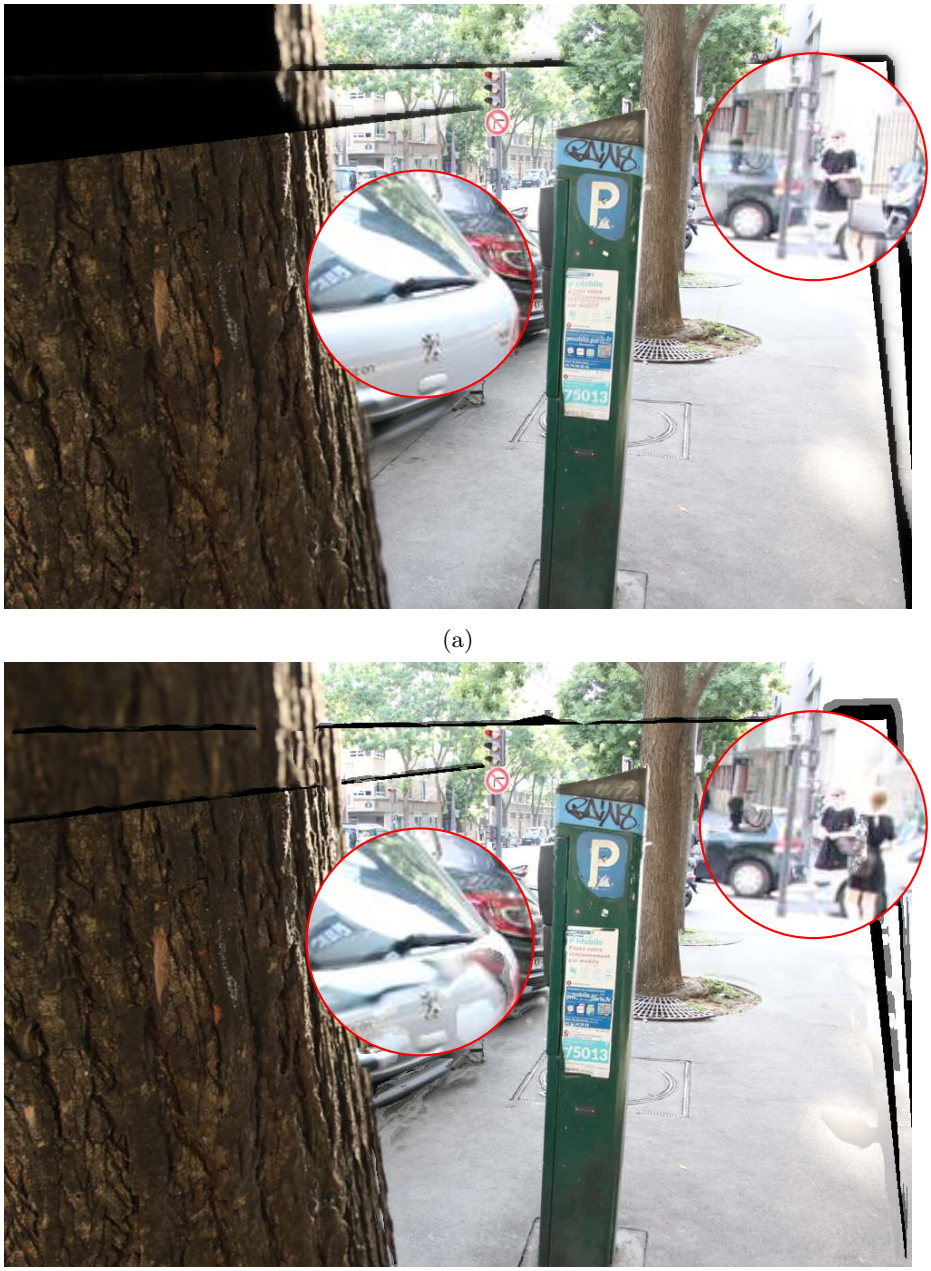
(b)

Fig. 5.11 *Object corrections:* (a) Fusion with [77]. (b). Fusion with [78].

(a)

(b)

Fig. 5.12 *Object corrections:* The top row images are the input set of out of focus images after alignment, the reference is the image 3. (a) Fusion of the registered images. (b) NL-MFIF's output.



(a)

(b)

Fig. 5.13 *Object corrections:* (a) Fusion with [77]. (b). Fusion with [78].

(a) Source A.

(b) Source B.

(c) Static MFIF.

(d) PatchMatch with L^2 norm.

(e) NL-MFIF without Eq.(5.7).

(f) NL-MFIF.

(g) Fusion from [77].

(h) Fusion from [78].

Fig. 5.14 MFIF on dynamic cases with moving objects and non-planar scenes. We present the fusion with the MFIF method for static cases, the dynamic MFIF setting using the PatchMatch with L^2 norm, our method, NL-MFIF, without and with the Yaroslavsky filtering (Eq.(5.7)) and the methods by Liu's et al. with CNN's [77] and Dense SIFT [78].

(a) Source A. (b) Source B. (c) Source C.

(d) Fusion static MFIF. (e) PatchMatch with L^2 norm.

(f) NL-MFIF without Eq.(5.7). (g) NL-MFIF

(h) Fusion from [77]. (i) Fusion from [78].

Fig. 5.15 MFIF on dynamic cases with moving objects and non-planar scenes. We present the fusion with the MFIF method for static cases, the dynamic MFIF setting using the PatchMatch with L^2 norm, our method, NL-MFIF, without and with the Yaroslavsky filtering (Eq.(5.7)) and the methods by Liu's et al. with CNN's [77] and Dense SIFT [78].

5.4 Conclusions

In this chapter we presented a framework for multifocus image fusion (MFIF) on images with moving alterations. At the core of our methodology, we proposed a patch-based algorithm that corrects local geometric deformations by relying on three characteristics, namely, color, gradient orientations and SIFT descriptors.
Experiments on images with complex motions let us conclude that our method is generally robust to geometric distortions and blur. This framework, not only solves artifacts due to motion, or errors after image registration, but also produces almost identical results when compared to image fusion algorithms for static settings. With our algorithm, the final fusion is an artifact-free image that displays more details and sharp structures than any image from the set. Consequently, our method is more general than other MFIF methods that only correct small displacements.
These results confirm that, despite the absence of works in the literature, MFIF of dynamic scenes is possible by relying on suitable tools. In our case, a constrained patch-based search while using the proposed similarity measure is robust enough.

There are challenging situations that limit the accuracy of this work. Particularly on two cases, the accuracy of reconstruction of small scale objects and the strength of blur. The first case emerges from the need to remove strong jitter, which forces us to apply a strong regularization of the displacement map. This situation could be improved by using a more sophisticated approach that preserves piece-wise regularity of fine variations. Second, because depending on how strong is the blur distortion all geometrical features could be removed, therefore making inviable the creation of an artifact free reconstruction.

Chapter 6

Conclusions & Perspectives

In this thesis we explored patch-based techniques and their integration into multiexposure and multifocus image fusion to deal with motions. Our methodology showed to be robust to object and camera motions and is able to correct ghosts and irregular structures after the fusion.

In Chapter 3 we studied how patch correspondences are a reliable tool to solve image distortions in a pair of images. We investigated how to estimate displacements within the images by relying on the extraction of nearest neighbors, and how to use a map of nearest neighbors to synthesize a new image that is geometrically consistent with the geometry of a given reference. This geometric alignment framework constitutes what we called image reconstruction.
In this thesis, we used the Patchmatch algorithm to extract the map of nearest neighbors. Originally introduced to compute approximate nearest neighbors, this algorithm offers the possibility to include some prior knowledge about the displacements and to control the locality of the search in a very efficient manner. We noticed that this scheme was very useful at correcting rotations and translations for images with similar radiometry, but fails for images that present different exposure or focal settings. This is a consequence of the L^2-norm which is not suited to radiometric or frequency variations within similar patches.

A solution to the radiometric change between the images consists of a contrast normalization before the extraction of the nearest neighbor field. This was explored in Chapter 4 where a general methodology for multiexposure image fusion is proposed. At the core of our method we rely on an image reconstruction of the radiometrically normalized images, a suitable correction of saturations within the reference and a regularization scheme to correct reconstruction errors.

We found that it is not necessary to rely on very complex parametric models to obtain a global contrast change that reproduces the illumination variations within the images.

Instead, the estimation of such variations can be simplified by using global approaches, like histogram specification, either with optimal transport or pseudo-inversion of the histogram.

When visually compared to deghosting algorithms and other image reconstruction frameworks our method appeared to be superior, consistently yielding better or similar results. Deghosting algorithms have the serious challenge of detecting moving objects that are composed by dark and bright intensities. This typically results in cutting them into incoherent pieces, and yields distorted reconstructions in the final combination. Most reconstruction methods deal badly with saturation in the reference image. They either restrain from using bright intensities [51, 54] or rely on the partially unsaturated geometry within the patch to find perfect matches [115, 55]. This last alternative is strongly affected by the patch size and the extent of saturations within the image. In contrast, our method yields very few artifacts due to saturation or reconstruction errors.

In Chapter 5 we extended the approach from Chapter 4 in order to develop a motion-aware MFIF method. For that, we rely on a reference image that imposes its geometry to the final result. As before, each source image is modified, using patch-based image reconstruction, to present the same geometry as the reference, without losing its sharpness properties. The reconstruction is performed using the map of nearest neighbors of Chapter 3. Since this scheme yields a new stack of completely aligned images, potentially any MFIF procedure can be used for the fusion.
Thanks to a new distance and adequate regularization scheme, we overcome the main difficulties for dealing with moving objects in multifocus images. First, a constrained multiscale patch matching helps us during the difficult comparison of local elements, that are extracted from images with different levels of blur. Second, the proposed regularization and fall-back strategy prevent from visual artifacts that could result from inaccurate patch comparisons. Our method clearly outperforms the two recent methods [78, 77], respectively based on SIFTs and CNNs, as far as the avoidance of ghosts and artifacts is concerned. To the best of our knowledge, the proposed method is the first non-local approach to multi-focus image fusion.

Perspectives

A first immediate follow up of the thesis is the publishing of the codes of both methods proposed in this document. In particular, both applications appear especially suited for the online demo publication IPOL [58].

As explained earlier, dealing with moving scenes makes it intractable to use conventional quantitative evaluation methods, because of the lack of ground-truth. Therefore a subjective user evaluation of the methods presented in this thesis would be most welcome

and would necessitate the development of a rigorous experimental set up which was beyond the scope of this thesis.

In Chapter 4 we showed the required stages to make MEIF methods robust to motions. Among those stages, reference enhancement and contrast normalization are the most important for controlling the visual quality of the final image. If the reference enhancement fails then irregular structures may appear on very bright regions. If the contrast normalization fails, then radiometric inconsistencies will most certainly be present in the fusion. Concerning the latter case, we found that histogram matching was good enough to normalize the radiometry of multiexposure images. As for the former case, it is solved by removing saturations from the reference image, a very challenging problem.
As we proposed, an alternative is to use the immediately less exposed image with respect to the reference, which solves saturations only when such image does not contain overexposed pixels. Besides, we argue that most of the methods in the literature have mainly focused on patch correspondences or motion detection, rather than including a robust strategy for transferring unsaturated content to affected areas. To that aim we propose two potential directions. One that better selects the image whose content is going to be used. This implies to better inspect the content within multiple images: how useful is the information, or how much the corresponding content is affected by motion or even saturations. A second direction could be the transfer of unsaturated regions in a methodology that is inspired by a global multi-image patch-based inpainting.

The motion aware method from Chapter 5 relies on a dense correspondence method. This strategy should be constrained in order to provide robustness to the different levels of blur within the images. For that reason we used 3 strategies, namely, (*i*) image prealignment, (*ii*) multiscale initialization and (*iii*) using a specially crafted distance. In the experiments, this approach showed to be well suited to reduce the impact of blur, however, it is computationally expensive and the precision may probably be enhanced, particularly for cases when a fine position cannot be found because one of the images has a very low frequency content. In this regard, it would be interesting to explore other metrics within our framework, for instance the invariant moments from [46] or its more recent extensions.
A different direction is to investigate other techniques that better regularize the displacement map without corrupting the quality of the reconstruction of small objects. Indeed, for very blurred images, we are forced to increase the strength of the refinement.
Let us mention a possible application of our MFIF framework, macro-photography, where solving misalignments is an extremely time-consuming process that is usually done manually. Other applications include dense correspondence frameworks, like optical flow or dense stereo matching.

References

[1] DxO Mark. https://www.dxomark.com/Cameras/Ratings/Landscape. Accessed: November 2017.

[2] Aguerrebere, C., Delon, J., Gousseau, Y., and Musé, P. Study of the digital camera acquisition process and statistical modeling of the sensor raw data. *Preprint HAL* (2012).

[3] Aguerrebere, C., Delon, J., Gousseau, Y., and Muse, P. Simultaneous HDR image reconstruction and denoising for dynamic scenes. In *Computational Photography (ICCP), 2013 IEEE International Conference on* (2013), IEEE, pp. 1–11.

[4] Aguerrebere, C., Delon, J., Gousseau, Y., and Musé, P. Best algorithms for HDR image generation. a study of performance bounds. *SIAM Journal on Imaging Sciences 7*, 1 (2014), 1–34.

[5] An, J., Lee, S. H., Kuk, J. G., and Cho, N. I. A multi-exposure image fusion algorithm without ghost effect. In *Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference on* (2011), IEEE, pp. 1565–1568.

[6] Arias, P., Facciolo, G., Caselles, V., and Sapiro, G. A variational framework for exemplar-based image inpainting. *International journal of computer vision 93*, 3 (2011), 319–347.

[7] Armes, T. LR/Enfuse. https://www.photographers-toolbox.com/products/lrenfuse.php. Accessed: December 2017.

[8] Aydin, T. O., Mantiuk, R., Myszkowski, K., and Seidel, H.-P. Dynamic range independent image quality assessment. *ACM Transactions on Graphics (TOG) 27*, 3 (2008), 69.

[9] Barkowsky, M., and Le Callet, P. On the perceptual similarity of realistic looking tone mapped high dynamic range images. In *Image Processing (ICIP), 2010 17th IEEE International Conference on* (2010), IEEE, pp. 3245–3248.

[10] Barnes, C., Shechtman, E., Finkelstein, A., and Goldman, D. Patchmatch: A randomized correspondence algorithm for structural image editing. *ACM Transactions on Graphics-TOG 28*, 3 (2009), 24.

[11] Barnes, C., Shechtman, E., Goldman, D. B., and Finkelstein, A. The generalized patchmatch correspondence algorithm. In *European Conference on Computer Vision* (2010), Springer, pp. 29–43.

[12] Barnes, C., and Zhang, F.-L. A survey of the state-of-the-art in patch-based synthesis. *Computational Visual Media* (2016), 1–18.

[13] Bénard, P., Cole, F., Kass, M., Mordatch, I., Hegarty, J., Senn, M. S., Fleischer, K., Pesare, D., and Breeden, K. Stylizing animation by example. *ACM Transactions on Graphics (TOG) 32*, 4 (2013), 119.

[14] Bhat, P., Curless, B., Cohen, M., and Zitnick, C. L. Fourier analysis of the 2D screened poisson equation for gradient domain problems. In *European Conference on Computer Vision* (2008), Springer, pp. 114–128.

[15] Brox, T., Kleinschmidt, O., and Cremers, D. Efficient nonlocal means for denoising of textural patterns. *IEEE Transactions on Image Processing 17*, 7 (2008), 1083–1092.

[16] Buades, A., Coll, B., and Morel, J.-M. A non-local algorithm for image denoising. In *Computer Vision and Pattern Recognition, 2005. IEEE Computer Society Conference on* (2005), vol. 2, IEEE, pp. 60–65.

[17] Buades, A., Coll, B., and Morel, J.-M. Nonlocal image and movie denoising. *International journal of computer vision 76*, 2 (2008), 123–139.

[18] Bugeau, A., Bertalmío, M., Caselles, V., and Sapiro, G. A comprehensive framework for image inpainting. *IEEE Transactions on Image Processing 19*, 10 (2010), 2634–2645.

[19] Burger, H. C., Schuler, C. J., and Harmeling, S. Image denoising: Can plain neural networks compete with BM3D? In *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on* (2012), IEEE, pp. 2392–2399.

[20] Burt, P., and Adelson, E. The laplacian pyramid as a compact image code. *IEEE Transactions on communications 31*, 4 (1983), 532–540.

[21] Chai, Y., Li, H., and Li, Z. Multifocus image fusion scheme using focused region detection and multiresolution. *Optics Communications 284*, 19 (2011), 4376–4389.

[22] Cohen, M. F., Shade, J., Hiller, S., and Deussen, O. *Wang tiles for image and texture generation*, vol. 22. ACM, 2003.

[23] Coltuc, D., Bolon, P., and Chassery, J.-M. Exact histogram specification. *IEEE Transactions on Image Processing 15*, 5 (2006), 1143–1152.

[24] Cossairt, O., Zhou, C., and Nayar, S. Diffusion coded photography for extended depth of field. In *ACM Transactions on Graphics (TOG)* (2010), vol. 29, ACM, p. 31.

[25] Criminisi, A., Pérez, P., and Toyama, K. Region filling and object removal by exemplar-based image inpainting. *IEEE Transactions on image processing 13*, 9 (2004), 1200–1212.

[26] Dabov, K., Foi, A., Katkovnik, V., and Egiazarian, K. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on image processing 16*, 8 (2007), 2080–2095.

[27] D'angelo, E. *Patch-based methods for variational image processing problems*. PhD thesis, École Polytechnique Fédérale de Lausanne, 2013.

[28] Darabi, S., Shechtman, E., Barnes, C., Goldman, D. B., and Sen, P. Image melding: combining inconsistent images using patch-based synthesis. *ACM Transactions on Graphics (TOG) 31*, 4 (2012), 82.

[29] De, I., and Chanda, B. Multi-focus image fusion using a morphology-based focus measure in a quad-tree structure. *Information Fusion 14*, 2 (2013), 136–146.

[30] Debevec, P. E., and Malik, J. Recovering high dynamic range radiance maps from photographs. In *SIGGRAPH* (1997), ACM.

[31] Debevec, P. E., and Malik, J. Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes* (2008), ACM, p. 31.

[32] Debevec, P. E., Taylor, C. J., and Malik, J. Modeling and rendering architecture from photographs: A hybrid geometry-and image-based approach. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (1996), ACM, pp. 11–20.

[33] Delon, J. Midway image equalization. *Journal of Mathematical Imaging and Vision 21*, 2 (2004), 119–134.

[34] Delon, J., and Desolneux, A. Stabilization of flicker like effects in image sequences through local contrast correction. *SIAM Journal on Imaging Sciences* (2010), 703–734.

[35] Dong, W., Li, X., Zhang, L., and Shi, G. Sparsity-based image denoising via dictionary learning and structural clustering. In *Computer Vision and Pattern Recognition (CVPR), 2011 IEEE Conference on* (2011), IEEE, pp. 457–464.

[36] Dufaux, F., Le Callet, P., Mantiuk, R., and Mrak, M. *High dynamic range video: from acquisition, to display and applications*. Academic Press, 2016.

[37] Durand, F., and Dorsey, J. Fast bilateral filtering for the display of high-dynamic-range images. In *ACM Transactions on Graphics (TOG)* (2002), vol. 21, ACM, pp. 257–266.

[38] Efros, A. A., and Freeman, W. T. Image quilting for texture synthesis and transfer. In *Proceedings of the 28th annual conference on Computer graphics and interactive techniques* (2001), ACM, pp. 341–346.

[39] Efros, A. A., and Leung, T. K. Texture synthesis by non-parametric sampling. In *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on* (1999), vol. 2, IEEE, pp. 1033–1038.

[40] Elad, M., and Aharon, M. Image denoising via sparse and redundant representations over learned dictionaries. *IEEE Transactions on Image processing 15*, 12 (2006), 3736–3745.

[41] Farbman, Z., Fattal, R., Lischinski, D., and Szeliski, R. Edge-preserving decompositions for multi-scale tone and detail manipulation. In *ACM Transactions on Graphics (TOG)* (2008), vol. 27, ACM, p. 67.

[42] Fattal, R., Lischinski, D., and Werman, M. Gradient domain high dynamic range compression. In *ACM Transactions on Graphics (TOG)* (2002), vol. 21, ACM, pp. 249–256.

[43] Fedorov, D., Sumengen, B., and Manjunath, B. Multi-focus imaging using local focus estimation and mosaicking. In *Image Processing, 2006 IEEE International Conference on* (2006), IEEE, pp. 2093–2096.

[44] Ferradans, S., Bertalmio, M., Provenzi, E., and Caselles, V. An analysis of visual adaptation and contrast perception for tone mapping. *Pattern Analysis and Machine Intelligence, IEEE Transactions on 33*, 10 (2011), 2002–2012.

[45] Fišer, J., Jamriška, O., Lukáč, M., Shechtman, E., Asente, P., Lu, J., and Sỳkora, D. Stylit: illumination-guided example-based stylization of 3D renderings. *ACM Transactions on Graphics (TOG) 35*, 4 (2016), 92.

[46] Flusser, J., Suk, T., and Saic, S. Recognition of blurred images by the method of moments. *IEEE Transactions on Image Processing 5*, 3 (1996), 533–538.

[47] Gallo, O., Troccoli, A., Hu, J., Pulli, K., and Kautz, J. Locally non-rigid registration for mobile HDR photography. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (2015), pp. 49–56.

[48] Gomez, A. L., Saravi, S., and Edirisinghe, E. A. Multiexposure and multifocus image fusion with multidimensional camera shake compensation. *Optical Engineering 52*, 10 (2013), 102007–102007.

[49] Goshtasby, A. A. Fusion of multifocus images to maximize image information. In *Society of Photo-Optical Instrumentation Engineers (SPIE) Conference Series* (2006), vol. 6229.

[50] Granados, M., Ajdin, B., Wand, M., Theobalt, C., Seidel, H.-P., and Lensch, H. Optimal HDR reconstruction with linear digital cameras. In *IEEE Conference on Computer Vision and Pattern Recognition* (2010), IEEE, pp. 215–222.

[51] HaCohen, Y., Shechtman, E., Goldman, D. B., and Lischinski, D. Non-rigid dense correspondence with applications for image enhancement. *ACM transactions on graphics (TOG) 30*, 4 (2011), 70.

[52] Hariharan, H. Extending depth of field via multifocus fusion.

[53] Ho Lee, J., Choi, I., and Kim, M. H. Laplacian patch-based image synthesis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 2727–2735.

[54] Hu, J., Gallo, O., and Pulli, K. Exposure stacks of live scenes with hand-held cameras. *Computer Vision–ECCV 2012* (2012), 499–512.

[55] Hu, J., Gallo, O., Pulli, K., and Sun, X. HDR deghosting: How to deal with saturation? In *2013 IEEE Conference on Computer Vision and Pattern Recognition* (2013), IEEE, pp. 1163–1170.

[56] Hu, Y., Song, R., and Li, Y. Efficient coarse-to-fine patchmatch for large displacement optical flow. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 5704–5712.

[57] Huang, W., and Jing, Z. Evaluation of focus measures in multi-focus image fusion. *Pattern recognition letters 28*, 4 (2007), 493–500.

[58] Image Processing On Line. http://www.ipol.im/. ISSN:2105-1232, http://dx.doi.org/10.5201/ipol.

[59] Jobson, D. J., Rahman, Z.-U., and Woodell, G. A. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *Image Processing, IEEE Transactions on 6*, 7 (1997), 965–976.

[60] Johnson, G. M., and Fairchild, M. D. Rendering HDR images. In *Color and Imaging Conference* (2003), vol. 2003, Society for Imaging Science and Technology, pp. 36–41.

[61] Kaneda, K., Ishida, S., Ishida, A., and Nakamae, E. Image processing and synthesis for extended depth of field of optical microscopes. *The Visual Computer 8*, 5-6 (1992), 351–360.

[62] Karaduzovic-Hadziabdic, K. HDR imaging website, international university of sarajevo. http://projects.ius.edu.ba/ComputerGraphics/HDR/, 2016. Accessed: February 2016.

[63] Karaduzovic-Hadziabdic, K., Telalovic, J. H., and Mantiuk, R. Expert evaluation of deghosting algorithms for multi-exposure high dynamic range imaging. In *2th International Conference and SME Workshop on HDR imaging* (2014).

[64] Kervrann, C., and Boulanger, J. Optimal spatial adaptation for patch-based image denoising. *IEEE Transactions on Image Processing 15*, 10 (2006), 2866–2878.

[65] Kim, J., Liu, C., Sha, F., and Grauman, K. Deformable spatial pyramid matching for fast dense correspondences. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2013), pp. 2307–2314.

[66] Kotwal, K., and Chaudhuri, S. An optimization-based approach to fusion of multi-exposure, low dynamic range images. In *Information Fusion (FUSION), 2011 Proceedings of the 14th International Conference on* (2011), IEEE, pp. 1–7.

[67] Krasula, L., Narwaria, M., Fliegel, K., and Le Callet, P. Rendering of HDR content on LDR displays: An objective approach. In *Applications of Digital Image Processing XXXVIII* (2015), vol. 9599, International Society for Optics and Photonics, p. 95990X.

[68] Lebrun, M., Buades, A., and Morel, J.-M. A nonlocal bayesian image denoising algorithm. *SIAM Journal on Imaging Sciences 6*, 3 (2013), 1665–1688.

[69] Lebrun, M., Colom, M., and Morel, J.-M. The noise clinic: a blind image denoising algorithm. *Image Processing On Line 5* (2015), 1–54.

[70] Li, S., Kang, X., Hu, J., and Yang, B. Image matting for fusion of multi-focus images in dynamic scenes. *Information Fusion 14*, 2 (2013), 147–162.

[71] Li, S., Kwok, J. T., and Wang, Y. Combination of images with diverse focuses using the spatial frequency. *Information fusion 2*, 3 (2001), 169–176.

[72] Li, S., Kwok, J. T., and Wang, Y. Multifocus image fusion using artificial neural networks. *Pattern Recognition Letters 23*, 8 (2002), 985–997.

[73] Li, S., and Yang, B. Multifocus image fusion using region segmentation and spatial frequency. *Image and Vision Computing 26*, 7 (2008), 971–979.

[74] Li, Z., Zheng, J., Zhu, Z., and Wu, S. Selectively detail-enhanced fusion of differently exposed images with moving objects. *IEEE Transactions on Image Processing 23*, 10 (2014), 4372–4382.

[75] Liang, L., Liu, C., Xu, Y.-Q., Guo, B., and Shum, H.-Y. Real-time texture synthesis by patch-based sampling. *ACM Transactions on Graphics (ToG) 20*, 3 (2001), 127–150.

[76] Liu, C., Yuen, J., Torralba, A., Sivic, J., and Freeman, W. T. SIFT flow: Dense correspondence across different scenes. In *European conference on computer vision* (2008), Springer, pp. 28–42.

[77] Liu, Y., Chen, X., Peng, H., and Wang, Z. Multi-focus image fusion with a deep convolutional neural network. *Information Fusion 36* (2017), 191–207.

[78] Liu, Y., Liu, S., and Wang, Z. Multi-focus image fusion with dense SIFT. *Information Fusion 23* (2015), 139–155.

[79] Liu, Y., and Wang, Z. Dense SIFT for ghost-free multi-exposure fusion. *Journal of Visual Communication and Image Representation 31* (2015), 208–224.

[80] Lowe, D. G. Distinctive image features from scale-invariant keypoints. *International journal of computer vision 60*, 2 (2004), 91–110.

[81] Lukáč, M., Fišer, J., Bazin, J.-C., Jamriška, O., Sorkine-Hornung, A., and Sỳkora, D. Painting by feature: texture boundaries for example-based image creation. *ACM Transactions on Graphics (TOG) 32*, 4 (2013), 116.

[82] Luo, X., Zhang, J., and Dai, Q. A regional image fusion based on similarity characteristics. *Signal processing 92*, 5 (2012), 1268–1280.

[83] Ma, K., and Wang, Z. Multi-exposure image fusion: A patch-wise approach. In *IEEE International Conference on Image Processing (ICIP)* (2015.).

[84] Ma, K., Zeng, K., and Wang, Z. Perceptual quality assessment for multi-exposure image fusion. *IEEE Transactions on Image Processing 24*, 11 (2015), 3345–3356.

[85] Mann, S., and Piccard, R. On being 'undigital' with digital cameras: Extending dynamic range by combining differently exposed pictures. In *IS&T* (1995), pp. 442–448.

[86] Mantiuk, R., Kim, K. J., Rempel, A. G., and Heidrich, W. HDR-VDP-2: a calibrated visual metric for visibility and quality predictions in all luminance conditions. In *ACM Transactions on graphics (TOG)* (2011), vol. 30, ACM, p. 40.

[87] Mantiuk, R., Myszkowski, K., and Seidel, H.-P. A perceptual framework for contrast processing of high dynamic range images. *ACM Transactions on Applied Perception (TAP) 3*, 3 (2006), 286–308.

[88] Martínez, J., Dumas, J., Lefebvre, S., and Wei, L.-Y. Structure and appearance optimization for controllable shape design. *ACM Transactions on Graphics (TOG) 34*, 6 (2015), 229.

[89] Mazin, B. *Robust Methods for Illuminant Estimation and Color Images Matching*. PhD thesis, Telecom ParisTech, 2014.

[90] Maître, H. *From Photon to Pixel: The Digital Camera Handbook*. Wiley-ISTE, 2015.

[91] Mertens, T., Kautz, J., and Van Reeth, F. Exposure fusion. In *Computer Graphics and Applications, 2007. PG'07. 15th Pacific Conference on* (2007), IEEE, pp. 382–390.

[92] Meylan, L. *Tone mapping for high dynamic range images*. PhD thesis, Ecole Polytechnique Federale De Lausanne, 2006.

[93] Mihal, A. Enfuse. http://enblend.sourceforge.net/, 2009. Accessed: December 2017.

[94] Mitsunaga, T., and Nayar, S. K. Radiometric self calibration. In *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.* (1999), vol. 1, IEEE, pp. 374–380.

[95] Narwaria, M., Da Silva, M. P., Le Callet, P., Valenzise, G., De Simone, F., and Dufaux, F. Quality of experience and HDR: concepts and how to measure it. In *High Dynamic Range Video.* Elsevier, 2016, pp. 431–454.

[96] Newson, A., Almansa, A., Fradet, M., Gousseau, Y., and Pérez, P. Video inpainting of complex scenes. *SIAM Journal on Imaging Sciences 7*, 4 (2014), 1993–2019.

[97] Nikolova, M., Wen, Y.-W., and Chan, R. Exact histogram specification for digital images using a variational approach. *Journal of Mathematical Imaging and Vision 46*, 3 (2013), 309–325.

[98] Ocampo-Blandon, C., and Gousseau, Y. Non-local exposure fusion. In *Iberoamerican Congress on Pattern Recognition* (2016), Springer, pp. 484–492.

[99] Ocampo-Blandon, C., Gousseau, Y., and ladjal, S. Non local multifocus image fusion scheme for dynamic scenes. In *2018 25th IEEE International Conference on Image Processing (ICIP), Submitted.* (2018), IEEE.

[100] Ogden, J. M., Adelson, E. H., Bergen, J. R., and Burt, P. J. Pyramid-based computer graphics. *RCA Engineer 30*, 5 (1985), 4–15.

[101] Palmer, S. E. *Vision Science: Photons to Phenomenology*, vol. 1. MIT press Cambridge, MA, 1999.

[102] Pece, F., and Kautz, J. Bitmap movement detection: HDR for dynamic scenes. In *Visual Media Production (CVMP), 2010 Conference on* (2010), IEEE, pp. 1–8.

[103] Pérez, P., Gangnet, M., and Blake, A. Poisson image editing. In *ACM Transactions on Graphics (TOG)* (2003), vol. 22, ACM, pp. 313–318.

[104] Po, L.-M., and Ma, W.-C. A novel four-step search algorithm for fast block motion estimation. *IEEE transactions on circuits and systems for video technology 6*, 3 (1996), 313–317.

[105] Qin, X., Shen, J., Mao, X., Li, X., and Jia, Y. Robust match fusion using optimization. *IEEE Transactions on Cybernetics* (2015).

[106] Rabin, J., Delon, J., and Gousseau, Y. Removing artefacts from color and contrast modifications. *IEEE Transactions on Image Processing 20*, 11 (2011), 3073–3085.

[107] Rabin, J., Ferradans, S., and Papadakis, N. Adaptive color transfer with relaxed optimal transport. In *Image Processing (ICIP), 2014 IEEE International Conference on* (2014), IEEE, pp. 4852–4856.

[108] Rabin, J., Peyré, G., Delon, J., and Bernot, M. Wasserstein barycenter and its application to texture mixing. In *International Conference on Scale Space and Variational Methods in Computer Vision* (2011), Springer, pp. 435–446.

[109] Raman, S., and Chaudhuri, S. Bilateral filter based compositing for variable exposure photography. In *Proceedings of Eurographics* (2009).

[110] Reinhard, E., Adhikhmin, M., Gooch, B., and Shirley, P. Color transfer between images. *IEEE Computer graphics and applications 21*, 5 (2001), 34–41.

[111] Reinhard, E., and Devlin, K. Dynamic range reduction inspired by photoreceptor physiology. *Visualization and Computer Graphics, IEEE Transactions on 11*, 1 (2005), 13–24.

[112] Reinhard, E., Heidrich, W., Debevec, P., Pattanaik, S., Ward, G., and Myszkowski, K. *High dynamic range imaging: acquisition, display, and image-based lighting*. Morgan Kaufmann, 2010.

[113] Rubinstein, M., Gutierrez, D., Sorkine, O., and Shamir, A. A comparative study of image retargeting. In *ACM transactions on graphics (TOG)* (2010), vol. 29, ACM, p. 160.

[114] Sen, P., and Aguerrebere, C. Practical high dynamic range imaging of everyday scenes: Photographing the world as we see it with our own eyes. *IEEE Signal Processing Magazine 33*, 5 (2016), 36–44.

[115] Sen, P., Kalantari, N. K., Yaesoubi, M., Darabi, S., Goldman, D. B., and Shechtman, E. Robust patch-based HDR reconstruction of dynamic scenes. *ACM Trans. Graph. 31*, 6 (2012), 203.

[116] Simakov, D., Caspi, Y., Shechtman, E., and Irani, M. Summarizing visual data using bidirectional similarity. In *Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on* (2008), IEEE, pp. 1–8.

[117] Song, M., Tao, D., Chen, C., Bu, J., Luo, J., and Zhang, C. Probabilistic exposure fusion. *Image Processing, IEEE Transactions on 21*, 1 (2012), 341–357.

[118] Sroubek, F., and Flusser, J. Registration and fusion of blurred images. *Image Analysis and Recognition* (2004), 122–129.

[119] Sumner, R. Processing raw images in matlab. https://users.soe.ucsc.edu/~rcsumner/rawguide/RAWguide.pdf, 2014.

[120] Szeliski, R. *Computer vision: algorithms and applications*. Springer Science & Business Media, 2010.

[121] Tian, J., and Chen, L. Adaptive multi-focus image fusion using a wavelet-based statistical sharpness measure. *Signal Processing 92*, 9 (2012), 2137–2146.

[122] Tian, J., Chen, L., Ma, L., and Yu, W. Multi-focus image fusion using a bilateral gradient-based sharpness criterion. *Optics communications 284*, 1 (2011), 80–87.

[123] Tico, M., Gelfand, N., and Pulli, K. Motion-blur-free exposure fusion. In *Image Processing (ICIP), 2010 17th IEEE International Conference on* (2010), IEEE, pp. 3321–3324.

[124] Tumblin, J., and Rushmeier, H. Tone reproduction for realistic images. *Computer Graphics and Applications, IEEE 13*, 6 (1993), 42–48.

[125] Vedaldi, A., and Fulkerson, B. VLFeat: An open and portable library of computer vision algorithms. In *Proceedings of the 18th ACM international conference on Multimedia* (2010), ACM, pp. 1469–1472.

[126] Vonikakis, V., Bouzos, O., and Andreadis, I. Multiexposure image fusion based on illumination estimation. In *Proceedings of pacific graphics* (2007).

[127] Wang, W.-W., Shui, P.-L., and Song, G.-X. Multifocus image fusion in wavelet domain. In *Machine Learning and Cybernetics, 2003 International Conference on* (2003), vol. 5, IEEE, pp. 2887–2890.

[128] Wei, L.-Y., Lefebvre, S., Kwatra, V., and Turk, G. State of the art in example-based texture synthesis. In *Eurographics 2009, State of the Art Report, EG-STAR* (2009), Eurographics Association, pp. 93–117.

[129] Wexler, Y., Shechtman, E., and Irani, M. Space-time completion of video. *IEEE Transactions on pattern analysis and machine intelligence 29*, 3 (2007).

[130] Wyszecki, G., and Stiles, W. S. *Color science*, vol. 8. Wiley New York, 1982.

[131] Yang, B., and Li, S. Multifocus image fusion and restoration with sparse representation. *Instrumentation and Measurement, IEEE Transactions on 59*, 4 (2010), 884–892.

[132] Yeganeh, H., and Wang, Z. Objective quality assessment of tone-mapped images. *IEEE Transactions on Image Processing 22*, 2 (2013), 657–667.

[133] Zerman, E., Valenzise, G., and Dufaux, F. An extensive performance evaluation of full-reference HDR image quality metrics. *Quality and User Experience 2*, 1 (2017), 5.

[134] Zhang, Q., and Guo, B.-l. Multifocus image fusion using the nonsubsampled contourlet transform. *Signal Processing 89*, 7 (2009), 1334–1346.

[135] Zhang, W., Hu, S., Liu, K., and Yao, J. Motion-free exposure fusion based on inter-consistency and intra-consistency. *Information Sciences 376* (2017), 190–201.

[136] Zhang, Z., and Blum, R. S. Image registration for multifocus image fusion. In *Battlespace Digitization and Network-Centric Warfare* (2001), vol. 4396, pp. 279–290.

[137] Zheng, J., and Li, Z. Superpixel based patch match for differently exposed images with moving objects and camera movements. In *Image Processing (ICIP), 2015 IEEE International Conference on* (2015), IEEE, pp. 4516–4520.

[138] Zheng, J., Li, Z., Zhu, Z., Wu, S., and Rahardja, S. Hybrid patching for a sequence of differently exposed images with moving objects. *Image Processing, IEEE Transactions on 22*, 12 (2013), 5190–5201.

[139] Zhu, X. *Measuring spatially varying blur and its application in digital image restoration.* PhD thesis, University of California, Santa Cruz, 2013.

[140] Zoran, D., and Weiss, Y. From learning models of natural image patches to whole image restoration. In *Computer Vision (ICCV), 2011 IEEE International Conference on* (2011), IEEE, pp. 479–486.