Highly accurate computer vision and image comparison

Pascal Monasse

École Normale Supérieure de Cachan

Mémoire d'Habilitation à Diriger des Recherches
(spécialité : mathématiques)

Highly accurate computer vision and image comparison

Pascal Monasse

soutenue publiquement le 10 décembre 2013
devant le jury composé de

Rapporteurs :	Adrien Bartoli	Université d'Auvergne
	Yann Gousseau	Télécom ParisTech
	Philippe Salembier	Université Polytechnique de Catalogne
Examinateurs :	Antonin Chambolle	École Polytechnique
	Laurent Najman	Université Paris-Est
	Marc Pierrot-Desseilligny	École Nationale des Sciences Géographiques
Garant :	Jean-Michel Morel	ENS Cachan

To the memory of Vicent Caselles (1960-2013),
outstanding mathematician and friend

Contents

Chapitre 1

Résumé des travaux de recherche

Mes recherches portent sur la comparaison d'images et les sujets connexes. Ma thèse de doctorat (Monasse, 2000) se concentrait sur le recalage d'images (Monasse, 1999; Dibos et al., 2003a) dans un sens restreint : seul le recalage par similitude était considéré. Cela a été étendu au recalage par homographie (Moisan et al., 2012), ce qui est plus ou moins la limite du recalage rigide. Dans presque tous les cas, la comparaison d'images implique une détection de zones ou points d'intérêt et leur mise en correspondance. Une construction utile fournissant de telles zones d'intérêt a été introduite dans ma thèse et nommée l'arbre des formes (Ballester et al., 2003), qui peut être vu comme le résultat de la fusion des deux arbres des composantes (Caselles et al., 2008). Cet arbre met en évidence de riches propriétés concernant la structure géométrique de l'image, et une description de la structure de l'image à la manière de la théorie de Morse est possible sous des conditions assez générales. Ma collaboration avec Vicent Caselles a exploré ces aspects (Caselles and Monasse, 2010) et en particulier une notion équivalente aux points selles a été généralisée en régions selles pour des fonctions continues, de la même manière que les extrema se généralisent en régions extrêmes. Cela a conduit à une définition cohérente des lignes de niveau avec des applications utiles : des opérateurs invariants par contraste peuvent être implémentés par des schémas géométriques et numériques de haute précision (Moisan, 1998) ; l'analyse des filtres auto-duaux de la morphologie mathématique en est facilitée (Heijmans and Keshet, 2002) ; des contours contrastés de régions peuvent être extraits (Desolneux et al., 2001; Cao et al., 2005) ; le mouvement par courbure moyenne peut être calculé avec une bonne stabilité numérique (Ciomaga et al., 2010). Pour la plupart de telles applications, une extraction de lignes de niveau moins pixelisées est nécessaire. J'ai développé une variante de la *Fast Level Set Transform* (FLST) issue de ma thèse pour traiter de l'image interpolée en bilinéaire. En explorant ces sujets, mon attention était toujours dirigée vers la comparaison d'images.

Le premier usage des arbres de composantes pour la comparaison d'images a été exposé en 2002 par Matas et al. avec les fameux *Maximally Stable Extremal Regions* (MSER) (Matas et al., 2004). Leur interprétation en termes d'arbres de composantes n'était pas explicite dans l'article original, et leur définition est

restée ambiguë. Par conséquent, il y a au moins deux implémentations publiques d'extraction de MSER et leurs définitions diffèrent. Bien que la robustesse des MSER en termes d'invariance et de répétabilité ait été reconnue (Mikolajczyk et al., 2005), leur défaut principal par rapport au désormais standard *Scale Invariant Feature Transform* (SIFT) de David Lowe (Lowe, 2004) est leur faible nombre. Il peut y avoir un ordre de grandeur de différence entre le nombre de descripteurs MSER et SIFT. Ce défaut a été corrigé dans une récente collaboration avec le doctorant Yongchao Xu et ses encadrants, Laurent Najman et Thierry Géraud : nous avons proposé d'utiliser toutes les régions de Morse (extrêmes et selles) (Xu et al., 2013). Les résultats prouvent une amélioration de l'état de l'art en recalage d'images et dans quelques domaines de la reconstruction 3D.

La reconstruction de la géométrie 3D à partir d'images stéréo, que ce soit des paires ou des images multiples, peut être considérée comme la tâche de comparaison d'images par excellence. Motivé en particulier par les applications en imagerie spatiale stéréo, telles qu'explorées par le groupe de recherche MISS, conduit par le CNES, l'agence spatiale française, et l'ENS Cachan, groupe dans lequel je suis un participant actif, mon attention s'est tournée vers la reconstruction 3D et sujets proches. Lorsque des mesures précises sont visées, toute la chaîne stéréo doit être revisitée. Cela commence par la calibration interne des caméras. L'expérience a montré que corriger la distorsion due à l'objectif tout en estimant les paramètres internes de la caméra peut conduire à des compensations numériques malvenues, où l'apparence d'une texture observée peut être expliquée par de la distorsion ou par un changement de point de vue. Curieusement, aucune évaluation quantitative satisfaisante de la correction de distorsion n'a pu être trouvée dans la littérature. La thèse de doctorat de Zhongwei Tang (Tang, 2011) répondait à ce problème en utilisant une "harpe de calibration", un montage assez facile avec des fils opaques et flexibles tendus à travers un cadre rigide. Des photographies de cette harpe permettent de mesurer la distorsion résiduelle après correction (Tang et al., 2012). Il s'avère que des niveaux de corrections sans précédent, de l'ordre de 0.03 pixel, peuvent être atteints avec cette harpe. L'inconvénient est qu'une composante homographique ne peut pas être retrouvée par ce processus, et les images corrigées correspondent à une caméra sténopée *virtuelle*, dont les paramètres internes n'ont pas de raison de correspondre à une caméra réelle (Grompone von Gioi et al., 2010). Cela se traduit par une calibration externe déformée par une transformation globale rigide de l'espace, ce qui n'est pas une poblème dans la plupart des applications car cela peut être corrigé par quelques points de contrôle au sol si nécessaire. Dans le cas de paires stéréo, la rectification épipolaire doit aussi être effectuée. Un nouvel algorithme (Monasse et al., 2010) améliorant la méthode de référence de Fusiello et Irsara (Fusiello and Irsara, 2008) a été proposé ; il évite le risque de tomber dans un minimum local de l'énergie minimisée. Tout comme dans l'algorithme original de Fusiello et Irsara, une rectification épipolaire quasi-euclidienne est appliquée, donc la distorsion est minimale.

L'autre composante de la calibration stéréo multi-vues est la calibration externe. Elle suppose généralement que les paramètres internes sont déjà connus, bien que ce ne soit pas strictement nécessaire. La méthode consiste en l'estimation des positions et orientations relatives de la caméra dans les différentes vues. La première étape cherche des points correspondants dans des paires d'images, le plus souvent par la méthode SIFT. De fausses correspondances sont écartées en

utilisant RANSAC ou une variante (Fischler and Bolles, 1981). Les trajectoires à travers les images sont alors construites par suivi des points correspondants par transitivité. Une solution élégante et économique de calculer ces trajectoires a été exposée dans un travail conjoint avec Pierre Moulon (Moulon et al., 2012). Après cela, la manière la plus standard d'effectuer la calibration externe est la méthode incrémentale : une paire d'images avec suffisamment de correspondances est choisie, sa matrice essentielle calculée (Longuet-Higgins, 1981; Nistér, 2004; Li and Hartley, 2006) et les rotation et translation relatives déduites. Puis des vues avec des points correspondants sont ajoutés de façon incrémentale. Chaque nouvelle vue est positonnée par rapport à la scène en utilisation des correspondes de points 3D-2D (Haralick et al., 1989). Régulièrement on applique quelques itérations d'une optimisation non-convexe impliquant toutes les variables (positions de caméra, orientations et points 3D) (Triggs et al., 2000). Le logiciel le plus connu suivant cette procédure est Bundler. Dans cette chaîne, modifier tous les paramètres d'estimation de modèle par la méthodologie *a contrario* (Desolneux et al., 2008) a permis d'obtenir des améliorations significatives de précision dans un travail conjoint avec Pierre Moulon et Renaud Marlet (Moulon et al., 2013a). Le problème de la méthode incrémentale est sa sensibilité à l'ordre d'ajout des images, avec des erreurs qui s'accumulent et provoquent une dérive. Quelques méthodes globales de calibration ont été introduites dans les quelques dernières années. La calibration globale évite la dérive des méthodes incrémentales mais se heurte souvent à un problème d'optimisation trop difficile à résoudre directement, donc des solutions sous-optimales sont recherchées. Nous avons developpé notre propre méthode (Moulon et al., 2013b), qui passe à l'échelle et donne une calibration plus précise que l'état de l'art.

Je suis également impliqué dans le journal en ligne IPOL (Image Processing On Line), en particulier en ce qui concerne la comparaison d'images, le recalage et la reconstruction 3D. Le but est de fournir une implémentation de référence en logiciel libre d'algorithmes d'état de l'art, en particulier dans le domaine du calcul de cartes de disparités à partir de paires d'images et de publier leur description algorithmique exacte, facilement reproductible dans n'importe quel langage.

Chapter 1

Research Summary

My central research area revolves around image comparison. My PhD dissertation (Monasse, 2000) focused on image registration (Monasse, 1999; Dibos et al., 2003a) in a restricted sense: only similarity registration was investigated. This has been extended to homographic registration (Moisan et al., 2012) and this is roughly as far as rigid registration can lead. In almost all cases, image comparison involves a detection of interest areas or points and their correspondence. A useful construction yielding such interest areas was introduced during my PhD as the tree of shapes (Ballester et al., 2003), which can be seen as the result of merging the two component trees (Caselles et al., 2008). This tree has rich properties concerning the geometry of the image, and a description of the image structure in the manner of Morse theory is possible under mild conditions. This has been investigated in detail in collaboration with Vicent Caselles (Caselles and Monasse, 2010) and in particular a notion equivalent to saddle points was generalized to saddle regions for continuous functions, in the same manner as extrema generalize to extremal regions. This has led to a consistent definition of level lines with useful applications: contrast-invariant operators can be implemented with high accuracy geometric and numerical schemes (Moisan, 1998); analysis of self-dual filters of mathematical morphology is facilitated (Heijmans and Keshet, 2002); contrasted contour regions can be extracted (Desolneux et al., 2001; Cao et al., 2005); mean curvature map can be computed with numerical stability (Ciomaga et al., 2010). For many such applications, an extraction of less pixelized level lines is necessary. I developed a variant of the Fast Level Set Transform (FLST) of my PhD work to deal with the bilinear image interpolation. While investigating these topics, my main focus was still directed toward image comparison.

The first use of the component trees for image comparison has been demonstrated by Matas et al. in 2002 with the celebrated Maximally Stable Extremal Regions (MSER) (Matas et al., 2004). Their interpretation in terms of component trees was not quite clear in the original paper, and their definition remained ambiguous. As a consequence, there are at least two public implementations of MSER detectors and their definitions differ. While the robustness of MSER in terms of invariance and repeatability was recognized (Mikolajczyk et al., 2005), their main defect with respect to the now standard Scale Invariant Feature Transform (SIFT) of David Lowe (Lowe, 2004) is their sparsity. There can be an order of magnitude difference in the number of MSER descriptors and SIFT

descriptors. This has been remedied, in a recent collaboration with the PhD student Yongchao Xu and his advisors, Laurent Najman and Thierry Géraud: we proposed to use all Morse regions (extremal and saddle regions) (Xu et al., 2013). Results show that it improves the state of the art in image registration and in some areas of 3D reconstruction.

The reconstruction of 3D geometry from stereo images, whether pairs or multiple images, can be considered the image comparison task *par excellence*. Motivated in particular by applications to spatial stereo imagery, as investigated by the MISS research group, led by CNES, the French space agency, and ENS Cachan, group in which I am an active participant, my focus turned to 3D reconstruction and all related aspects. When precise measurements are aimed at, the whole stereo pipeline needs to be revisited. It starts with the internal camera calibration. It was observed that correcting lens distortion while estimating internal camera parameters can lead to detrimental numerical compensations, where the apparence of an observed pattern can be explained by distortion or by change of point of view. Curiously, no satisfying quantitative assessment of distortion correction could be found in the literature. The PhD thesis of Zhongwei Tang (Tang, 2011) answered that by using the so called "calibration harp", a device fairly easy to build with opaque and flexible strings tensely stretched across a rigid frame. Photographs of this harp allow measuring the residual distortion after correction (Tang et al., 2012). It turns out that unprecedented levels of correction accuracy, in the order of 0.03 pixel, can be reached with this harp. The drawback is that a homographic component cannot be recovered by this process, and the corrected images correspond to a *virtual* pinhole camera, whose internal parameters have no reason to correspond to the real camera (Grompone von Gioi et al., 2010). This is reflected in the external calibration as a global rigid transform of the space, which is not a problem in most applications because it can be corrected by a few ground control points if necessary. In the case of stereo pairs, the epipolar rectification step must also be performed. A new algorithm (Monasse et al., 2010) improving a state of the art method of Fusiello and Irsara (Fusiello and Irsara, 2008) was proposed and avoids being trapped in a local minimum of the optimized energy. As in the original algorithm of Fusiello and Irsara, a quasi-Euclidean epipolar rectification is applied, so the distortion is minimal.

The other component of multiple view stereo calibration is the external calibration. It usually assumes that the internal parameters are already known, though not strictly necessary. The method consists in estimating the relative positions and orientations of the camera in the different views. The first step is to find corresponding points or features in image pairs, usually achieved through the SIFT method. False correspondences are discarded using RANSAC or a variant (Fischler and Bolles, 1981). Tracks across images are then built by following corresponding points by transitivity. An elegant and economical way to compute the tracks was exposed in a joint work with Pierre Moulon (Moulon et al., 2012). Then, the most standard way to achieve external calibration is the incremental method: an image pair with enough correspondences is chosen, its essential matrix computed (Longuet-Higgins, 1981; Nistér, 2004; Li and Hartley, 2006) and the relative rotation and translation deduced. Then views with corresponding points are incrementally added. Each new view is positioned with respect to the scene using pose estimation from 3D-2D point matches (Haralick et al., 1989). Regularly one applies a few iterations of a non-

convex optimization involving all variables (camera positions, orientations and 3D points) (Triggs et al., 2000). The popular software following this procedure is Bundler. In this pipeline, modifying all model estimation parameters using the *a contrario* methodology (Desolneux et al., 2008) was proved to yield significant calibration precision improvements in a joint work with Pierre Moulon and Renaud Marlet (Moulon et al., 2013a). The problem of the incremental method is its sensitivity to the order in which images are added, with accumulating errors provoking drift. A few global calibration methods were presented in the last few years. Global calibration avoids the drift of incremental methods but often has a problem of optimization that is too difficult to solve directly, so suboptimal solutions are sought. We developed our own method (Moulon et al., 2013b), which is scalable and yields more precise calibration than state of the art methods.

I am also involved in the online journal IPOL (Image Processing On Line), especially in the areas of image comparison, registration, and 3D reconstruction. The goal is to provide reference open source implementations of state of the art algorithms, in particular in the area of disparity map computation from a stereo pair and to publish their accurate algorithmic description, easily reproducible in any language.

Chapter 2

Research Details

2.1 Tree of shapes

The component trees, namely the min- and max-trees of connected components of level sets, were introduced by Salembier et al. (Salembier et al., 1998) in order to study connected filters of mathematical morphology. A more efficient algorithm for their extraction in the case of high bit-depth images was proposed by Najman and Couprie (Najman and Couprie, 2006) using the union-find data structure (Tarjan, 1975). Noticing the rather low performance of the latter algorithm on 8-bit images, Nister proposed a new algorithm (Nistér and Stewénius, 2008). It happens that Nister's algorithm is exactly a non-recursive reimplementation of the original Salembier et al. flooding method. Meanwhile, a unique tree built from min- and max-components was introduced (Monasse and Guichard, 2000; Ballester et al., 2003). The proposed 2D bottom-up algorithm, while fast in general, could be made much faster by a top-down approach (Song, 2007), but using the unusual hexagonal connectivity. The higher dimensional extraction of the tree of shapes was made possible by fusion of the min- and max-trees (Caselles et al., 2008). However, it was shown that a slight modification of the digital image interpretation leads to a very efficient algorithm in any dimension (Géraud et al., 2013) using again the union-find data structure. The algorithm interprets the image as a set-valued function, where a modification of the definition of level set is necessary (Najman and Géraud, 2013).

Using the components as features for image comparison was pioneered in rigid registration (Monasse, 1999). Each component is described by a few invariant moments, which are used for feature correspondence. The same principle was rediscovered in the highly influencial work of Matas et al. and applied to estimation of fundamental matrix (Matas et al., 2004). On top of that, a stability criterion to select distinctive components was introduced, hence the name *maximally stable extremal regions* (MSER). The robustness of MSER in comparison to linear scale-space based features as SIFT was checked (Mikolajczyk et al., 2005). The low number of MSERs is their main drawback. However, a different criterion that is truly contrast invariant could be used instead, leading to a much higher number of features (Xu et al., 2013).

The geometric analysis of a function by means of its level lines can be traced back to Marston Morse, though preliminary works in the field of topography

already appeared in the 19th century. This is known as Morse theory (Milnor, 1963). In the search of a generalization of total variation to functions defined on a multi-dimensional domain, the tree structure associated to the level lines was first noticed by Kronrod in the 1940's (Kronrod, 1950). Using connected components of iso-level lines as points and deriving a topology from the topology of the definition domain, he showed the resulting connected dendrite structure. It presents many similarities to the graph developed by Reeb (Reeb, 1946). An equivalence relation is established between points of the domain: two points are equivalent if they belong to the same connected component of iso-level set. The quotient topology by this equivalence relation gives the connection between equivalence classes, which are connected components of iso-level sets. A computational equivalent is known as *digital Morse theory* (Cox et al., 2003). The nestedness structure of the connected components of iso-level sets was used in computer graphics for data visualization (Bajaj et al., 1996; van Kreveld et al., 1997). The topological analysis using level lines was also investigated in the context of robotics (Kweon and Kanade, 1994). In image processing, the level sets hierarchy was proposed as a fundamental contrast invariant representation of image (Caselles et al., 1999a; Caselles et al., 1999b). This provides the adequate framework for the analysis of connected operators of mathematical morphology (Salembier and Serra, 1995).

2.1.1 Theoretical justification of the tree of shapes

The tree of shapes was introduced in my PhD thesis (Monasse, 2000) as a fusion of the component trees, namely the min- and max-trees. Given an image $u : \Omega \subset \mathbb{R}^n \to \mathbb{R}$, we define its

- Upper level sets $[u \geq \lambda] := \{x : u(x) \geq \lambda]\}$;
- Lower level sets $[u < \lambda] := \{x : u(x) < \lambda]\}$,

where $\lambda \in \mathbb{R}$. The maps $\lambda \to [u \geq \lambda]$ and $\lambda \to [u < \lambda]$ are respectively decreasing and increasing for the inclusion relation among subsets of Ω. When switching to connected components (assuming most visual "objects" should appear connected), this translates to an inclusion tree structure. This easy to prove result depends neither on any regularity of u nor on the topology of Ω. Each one of these trees is sufficient to reconstruct the full image, so they are redundant. A unique tree containing both upper and lower components is interesting, but depends on two assumptions:

1. u is upper semicontinuous;
2. Ω is unicoherent [1].

Then components must also be modified with a saturation operator sat, an increasing and idempotent set operator such that for all $A \subset \Omega$ connected:

1. $\Omega \setminus \operatorname{sat} A \in \{\emptyset\} \cup \mathcal{CC}(\Omega \setminus A)$,
2. $\operatorname{sat}(\Omega \setminus \operatorname{sat} A) \in \{\emptyset, \Omega\}$.

In other words, a saturation operator adds to a set all connected components of its complement (called the holes) except at most one (called the exterior), and the saturation of the exterior is the full domain Ω. For example, we can fix a

1. A notion akin but different from simple connectedness: for any U, V connected open sets such that $\Omega = U \cup V$, $U \cap V$ is connected.

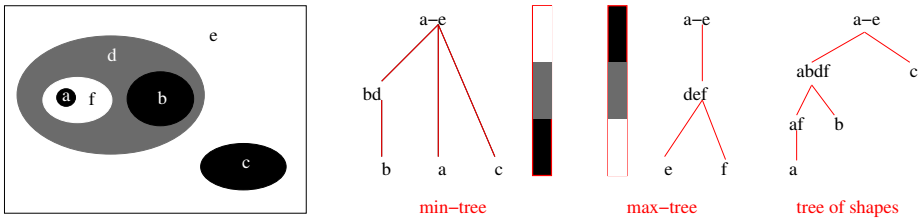


Figure 2.1: Trees of components and tree of shapes. In the min-tree, nodes are connected components of lower level sets. In the max-tree, nodes are connected components of upper level sets. For the tree of shapes, the point p_∞ was chosen in e. Notice the nodes of the tree of shapes that are not components, but *saturations* of components: $af = \operatorname{sat} f$ and $abdf = \operatorname{sat} bd$.

point $p_\infty \in \Omega$ and define as exterior of $A \in \Omega$ the component $\operatorname{cc}(\Omega \setminus A, p_\infty)$ and the saturation as the complement of the exterior.

A saturation operator preserves the topology: to a connected/closed/open set A it associates a connected/closed/open set $\operatorname{sat} A$. An important property is that for a closed set, $\operatorname{sat} A = \operatorname{sat} \partial A$.

Under these conditions, it was shown that the saturations of all upper *and* lower components (called shapes) of u can be organized in a *single* inclusion tree structure (Ballester et al., 2003), see Figure 2.1. The elements of this tree are thus not necessarily components of u but saturated components of u. As with component trees, the tree of shapes is sufficient to recover the image u. The proof of the tree structure relies on the important fact that shapes are either closed or open, as the components they are built from. This is ensured by the definition of large/strict inequality for upper/lower level sets in conjunction with the upper semicontinuity hypothesis.

2.1.2 Fusion of component trees

As the tree of shapes was presented as a mix of the component trees, it is logical that an appropriate fusion of the component trees would yield the tree of shapes (Caselles et al., 2008). For that, the notion of limit node of a tree is important. An interval $[A, B]$ of an inclusion tree is defined as the set of all elements of the tree between two sets A and B: they all contain A and are contained in B, where A and B can be *any* subsets of Ω, not necessarily elements of the tree itself. A limit node is defined as the intersection or union of all subsets of an interval. Elements of the tree are limit nodes, but they are not alone. In the same manner as Dedekind's construction of the real numbers from rational ones by using "cuts", the "irrational" elements of an inclusion tree are the limit nodes which are not in the original tree: the completion of the rational numbers using their order relation can be transposed in the same manner to the tree of subsets with its *partial* order relation of inclusion.

The irrational elements of the max-tree $\mathcal{U}(u)$ are sets of $\mathcal{CC}([u > \lambda])$ and the irrational elements of the min-tree $\mathcal{L}(u)$ are sets of $\mathcal{CC}([u \leq \lambda])$. Notice that they are not components, so a limit node in a tree is not found in the dual tree. Also, the hole of an element of one tree is the saturation of a component in the dual tree.

A branch of an inclusion tree $\mathcal{T}$ is an interval $[A, B]$ with $A \in \mathcal{T}$ such that

$$\forall A' \in \mathcal{T}, A' \subset B \Rightarrow A \cap A' \neq \emptyset. \tag{2.1}$$

The inclusion relation is a total order when restricted to a branch. It happens that the set $\{\operatorname{sat} C : C \in [A, B]\}$ is an interval of the tree of shapes when $[A, B]$ is an interval of $\mathcal{U}(u)$ or $\mathcal{L}(u)$. That means that branches of the component trees are not interrupted in the tree of shapes. Therefore, the fusion algorithm can begin with segmentations of $\mathcal{U}(u)$ and $\mathcal{L}(u)$ into their maximal branches. The question is how to reconnect the branches in the tree of shapes, which we detail in the next paragraph. It looks like we put components into the tree of shapes, not their saturation. Actually a post-processing step should happen, where each element A is replaced by the union of all elements in the subtree rooted at A, which is actually $\operatorname{sat} A$.

The connection of segmented maximal branches in the trees happens as follows. Considering a maximal branch $[A, B]$ of $\mathcal{U}(u)$ with B its upper limit node, $B \in \{\Omega\} \cup \mathcal{CC}([u > \lambda])$. In the first case, we attach the branch to the root. In the second case, we define $B' = \operatorname{cc}([u \geq \lambda], B) \in \mathcal{U}(u)$. B' is the lower end of a maximal branch of $\mathcal{U}(u)$. But $\operatorname{sat} B$ is also a hole of some $N \in \mathcal{L}(u)$ at level λ. $[A, B]$ should be attached either to N or to B'. Actually $\operatorname{sat} N$ and $\operatorname{sat} B'$ are either nested or both Ω. We must attach to the smaller of both sets. Two questions should be answered to translate these results into an actual algorithm:

1. How to find the component N in the dual tree?
2. How to compare $\operatorname{sat} N$ and $\operatorname{sat} B'$, whereas these saturated sets are not yet computed?

To answer Question 1, we can record one point p in the external boundary of B. We find the smallest component in $\mathcal{L}(u)$ containing p and go up this tree and stop just before reaching level λ. To know if p is in the *external* boundary of B as opposed to an *internal* boundary, this is solved by the answer to Question 2: we follow each connected component of the boundary of B and compute the enclosed region by Green's formula. A negative value indicates an internal boundary (and is ignored), a positive value indicates the external boundary and is the area of its saturation. Just looking at the areas, we can compare $\operatorname{sat} N$ and $\operatorname{sat} B'$. An analogous argument applied to reconnection of segmented maximal branches of $\mathcal{L}(u)$. Figure 2.2 illustrates the process.

2.1.3 Grain filters

The tree of shapes is very well suited for the study of connected operators of mathematical morphology. A connected filter is such that the connected components of level sets of the resulting image are all connected components of level sets at the same level (Salembier and Serra, 1995). A characteristic property of these filters is that the level lines are not smoothed: they are either present or absent in the filtered image, but never modified. Typical connected filters are the area opening and closing of Vincent (Vincent, 1993). They are defined with respect to a parameter $\epsilon \geq 0$ as:

$$M_\epsilon^+ u = \sup_{B \in x + B_\epsilon} \inf_{y \in B} u(y) \text{ and} \tag{2.2}$$

$$M_\epsilon^- u = \inf_{B \in x + B'_\epsilon} \sup_{y \in B} u(y), \tag{2.3}$$

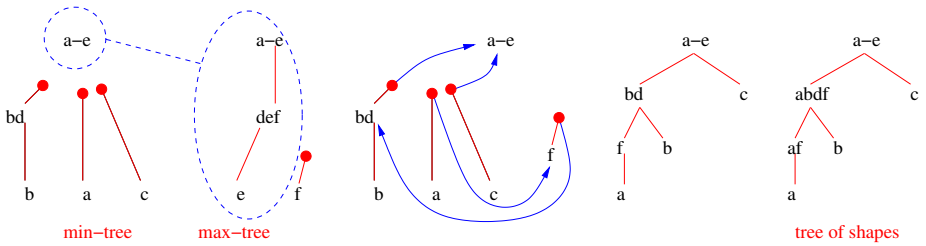


Figure 2.2: Fusion of component trees to produce the tree of shapes. Refer to Figure 2.1. First, components that contain $p_\infty \in e$ are isolated to constitute the root. Subtrees become orphans. Each free end is connected to the adequate node. This creates a new tree. To obtain the tree of shapes, each node must be saturated, which consists of augmenting the node with all its descendents.

with

$$B_\epsilon = \{B \text{ Lebesgue-measurable and connected}, 0 \in B, |B| \geq \epsilon\} \tag{2.4}$$
$$B'_\epsilon = \{B \text{ Lebesgue-measurable and connected}, 0 \in B, |B| > \epsilon\}, \tag{2.5}$$

$|B|$ being the Lebesgue measure of the set B. Area opening and closing are increasing and idempotent (required conditions to be called "filters" in mathematical morphology). Typically they remove small details in the image, and are adequate for removing salt and peper noise. Actually, they should be used in tandem for such a task, so that we could use: $M^+_\epsilon M^-_\epsilon$ or $M^-_\epsilon M^+_\epsilon$, which are actually different idempotent filters. Because of this non commutativity, they should be used as iterated sequential filters, for example:

$$M^+_{\epsilon_n} M^-_{\epsilon_n} \cdots M^+_{\epsilon_1} M^-_{\epsilon_1}, \tag{2.6}$$

with $\epsilon_1 < \cdots < \epsilon_n$ (for digital images, we would use $\epsilon_i = i$ and n the expected size of the connected components of salt-and-pepper noise). This is an efficient filter but it is rather slow to compute. An alternative is what we call the grain filter (Caselles and Monasse, 2002). It is defined indirectly through the associated set operator:

$$G_\epsilon X = \bigcup_{C \in \mathcal{CC}(X), |\operatorname{sat} C| \geq \epsilon} \{\operatorname{sat} C \setminus \bigcup_i C'_i\}, \tag{2.7}$$

where C'_i are internal holes of C of Lebesgue measure larger than ϵ. The grain filter is defined as the stack filter:

$$G_\epsilon u(x) = \sup\{\lambda : x \in G_\epsilon[u \geq \lambda]\}. \tag{2.8}$$

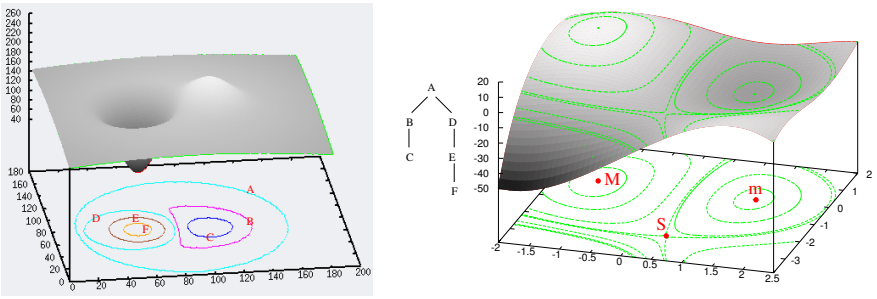
The key property of the grain filter is that it is self-dual on continuous functions: it commutes with the negative operator. This amounts to say that a single application of G_{ϵ_n} is sufficient, the intermediate G_{ϵ_i} having no influence on the final result. In practice, the grain filter is almost identical to the iterated sequential application of area openings and closings. All have the nice property that the limit shapes of the filtered image are limit shapes of the original image, so that they act as simplifications of the tree of shapes. Another natural

requirement that is satisfied is that $G_\epsilon u$ converges uniformly to the continuous function u when ϵ goes to 0 (the definition domain must be a continuum). It can be shown that the grain filter is a paradigm for all increasing and self-dual connected filters such that the set operator acts additively on connected components. The only degree of liberty is the increasing criterion saying if we should preserve a saturated set. The simplest such criterion is the comparison of the area with the threshold ϵ, leading to the grain filter. However, the astute reader should have noticed the last assumption: additive behavior of the set operator on connected components. That means that connected components are treated independently. This seems a natural assumption, but we could imagine some interesting filters that do not handle components independently, but for example would rely on distance between components. For example, close components with similar shape could be assumed a texture and preserved, while individually each component could be construed as noise.

These operators decrease the total variation of the image by removing some connected components of level sets. Another way to decrease the total variation of the image is to change the levels of these components in order to reduce the contrast with their parent and children. This is easily seen via the coarea formula for the total variation. This is not anymore a contrast invariant operator, since the levels change, but it shares the "good" properties of the grain filters, though it is naturally not idempotent. A good way to change the level is to change the level of a component towards the level of its parent at a speed proportional to the ratio of its perimeter by its measure until they disappear by reaching the level of their parent (Dibos et al., 2003b). For example, a disk or radius r changes level with a speed proportional to $2\pi r/\pi r^2 = 2/r$. Therefore small disks disappear before large disks, which is reasonable assuming that they correpond to small details and more likely to be due to noise. The only reasonable way to compute this self-dual operator is with the tree of shapes.

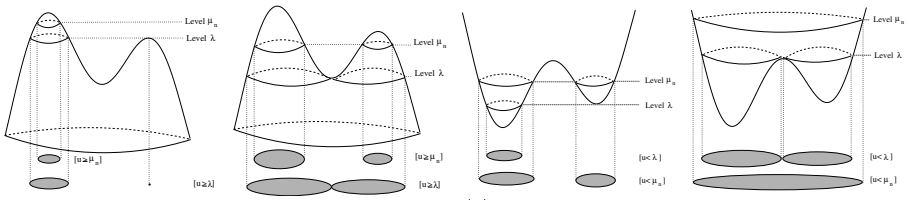
2.1.4 Geometric description of images

Assuming the level lines $\mathrm{cc}([u = \lambda])$ of the image were Jordan curves, the tree structure of level lines would become very natural: each Jordan curve separates the plane in two connected components, the bounded one being called the interior of the curve. Level lines being disjoint, a level line L is considered an ancestor in the tree of another level line L' if the interior of L contains L', see Figure 2.3. However, the presence of saddle points prevents such an interpretation, with typically some level lines in the shape of an 8. Moreover, the topological description of the image level lines is difficult to establish without further hypotheses on u. In Morse theory, the function is assumed to be twice differentiable and to have nondegenerate Hessian at critical points. The former condition may be too strong for images, while the latter prevents the presence of plateaux in particular. Such a description, valid for a continuous image u with no differentiability requirement, seems desirable. Even though such a continuous function can be approached by a sequence of Morse functions, there is no canonical approximation and the topology of the approaching Morse functions does not inform on the behavior of u. Actually, our analysis requires another assumption that seems perfectly reasonable for digital images: it is said to be *weakly oscillating*, meaning that it has a finite number of regional



(a) Tree of level lines (b) Singular points

Figure 2.3: Tree of some level lines of a Morse function, each of them being a Jordan curve. Some "level lines" are not Jordan curve, specifically those going through singular points, such as minimum m, maximum M and saddle point S.



(a) Appearance (b) Merge (c) Disappearance (d) Split

Figure 2.4: Topological changes of level sets at a singular level λ. When the sequence μ_n converges from above to λ, topological changes occur at the limit. These functions have three maximal monotone sections.

extrema[2]. The extrema and saddle points of Morse theory translate in this new framework into extremal regions and saddle regions. The singularities, or topology changes, are described by purely geometric arguments.

A monotone section of u is defined as a connected component of a bilevel set of u where no "bifurcation" happens: from an interval $I \subset \mathbb{R}$, $X_I \in \mathcal{CC}([u \in I])$ such that $\forall \alpha \in I$, $X_I \cap [u = \alpha]$ is connected. The union of a family of monotone sections having a common point x is still a monotone section. This allows to define the maximal monotone sections of u[3], see Figure 2.4. The assumption of weak oscillation is necessary for this result, though it can be relaxed. At a point x, we can consider the maximal monotone section containing x, associated to an interval $I(x)$. The end points of this interval are noted $\eta_+(x) = \sup I(x)$ and $\eta_-(x) = \inf I(x)$. Such values are called *singular* values of u. It can be shown that a weakly oscillating function has a fundamentally finite structure: there is a finite number of maximal monotone sections, and therefore a finite number of singular values (Caselles and Monasse, 2010). It may seem obvious since u has a finite number of regional extrema, but a maximal monotone section may not contain any regional extremum at all. Moreover, a complete characterization of the limit nodes of the tree of shapes can be achieved. As expected, these are

2. An application of the grain filter G_ϵ for any small $\epsilon > 0$ ensures that property and is the natural way to achieve that, since it corresponds just to a pruning of the tree of shapes.

3. Notice the analogy with connected sets and the definition of connected components, though the proof of our result involves more sophisticated arguments

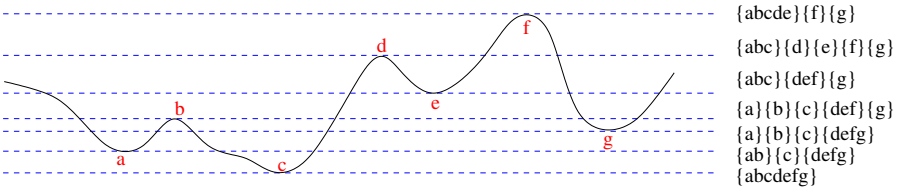


Figure 2.5: Signature of a signal and critical values. In 1D, all are values of extrema. In 2D, saddle values are also critical.

the sets of form $\operatorname{sat} C$ with

$$C \in \mathcal{CC}([u \geq \lambda]) \cup \mathcal{CC}([u > \lambda]) \cup \mathcal{CC}([u \leq \lambda]) \cup \mathcal{CC}([u < \lambda]) \tag{2.9}$$

where λ is a singular value.

Another way to characterize the topology change is to look at the separation of regional extrema operated by thresholds at levels λ. At a level λ, $\mathcal{CC}([u \geq \lambda]) \cup \mathcal{CC}([u < \lambda])$ is a partition of Ω. A regional extremum of u is contained in a single element of this partition. So we can also consider that at level λ, we have a partition of the regional extrema. Looking at the evolution of this partition when λ changes, we define the *critical* values of u as the values λ such that the partition is not constant in any neighborhood of λ, see Figure 2.5 [4]. Actually, the critical and singular values of u coincide. They coincide also with the notion of singular value of tree of shapes of u: that is the level λ of a limit node of a maximal monotone section of u.

If we come back to the correct definition of a level line of a continuous function u, we would like to keep the natural properties of a Jordan curve, even though it may not be one. We have several candidates:

1. $\partial \operatorname{cc}([u \geq \lambda])$;
2. $\operatorname{cc}(\partial \operatorname{cc}([u \geq \lambda]))$;
3. $\partial \operatorname{sat} \operatorname{cc}([u \geq \lambda])$;

with cc selecting one arbitrary connected component. All definitions are actually parts of the isolevel set $[u = \lambda]$. The problem of the first definition is that it may not even be connected. This is the defect corrected by the second definition, but such curve may still present self-crossing. The third definition defines the strict boundary necessary to recover the "interior": if L is such a set, we have $\operatorname{sat}(L)$ as closed interior of L, and it is equal to $\operatorname{sat} \operatorname{cc}([u \geq \lambda])$. This represents a connected set, and its complement is also connected.

2.1.5 Bilinear level lines

The algorithm for extraction of the tree of shapes for a pixel-constant interpretation of the digital pixel values, the fast level set transform (FLST) (Monasse and Guichard, 2000), yields pixelized level lines. This pixelization may be detrimental to further processing. The problem is that the image is interpreted as

4. Figure 2.5 illustrates the 1D case. Notice that in 1D, there could be "critical" values corresponding to inflection points. However, inflection points do not produce any topology change and behave like regular points. The presence of an inflection point implies a null second derivative, which means the function is not a Morse function.

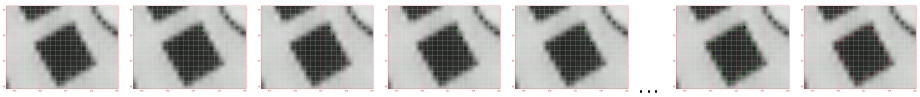


Figure 2.6: Bilinear level line extraction. The algorithm starts from an initial point between two adjacent pixels, at half-integer ordinate, and proceed by following the level line from dual pixel to dual pixel. It proceeds into a new dual pixel when it reaches a point where x or y is half-integer. The algorithm stops when the curve closes the loop.

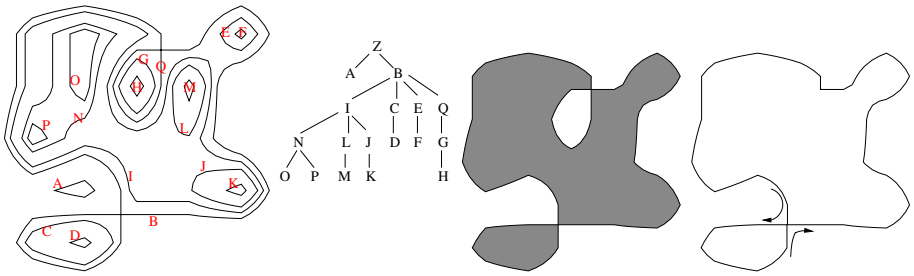


Figure 2.7: Level lines and saddle points. The level line B goes through two saddle points. According to definition of level line at the end of Section 2.1.4, the true level line B is on the right, but our algorithm following the level lines detects only the left curve because it turns right at a saddle point. This choice is correct for the part around C, but not for Q.

piecewise constant and discontinuous. An option is to consider rather the bilinear interpolation of the pixel values, that is the interpolation with splines of order 1. This yields a continuous image, whose extrema are at original local extrema of the pixel values. Also, the level line can be locally explicitly parameterized, permitting to sample it with no difficulty. However, there are some difficulties with singular levels, but since these are discrete, they can be avoided (Caselles and Monasse, 2010).

Since there are infinitely many different level lines in the continuous case (on the contrary to the pixelized case), some predefined levels of extraction should be chosen. Two algorithms are proposed. The first one is more natural, following level lines (see Figure 2.6) but must avoid initial levels of the digital image, since at such levels singularities can occur, in particular plateaux. It can accept levels of saddle points and give consistent results, but there are some ambiguities left in the inclusion relation, see Figure 2.7. The second algorithm deals with arbitrary levels but is more complex, akin to the FLST.

Inside the square $[0,1]^2$, assuming prescribed values at the four vertices, we have the interpolation equation:

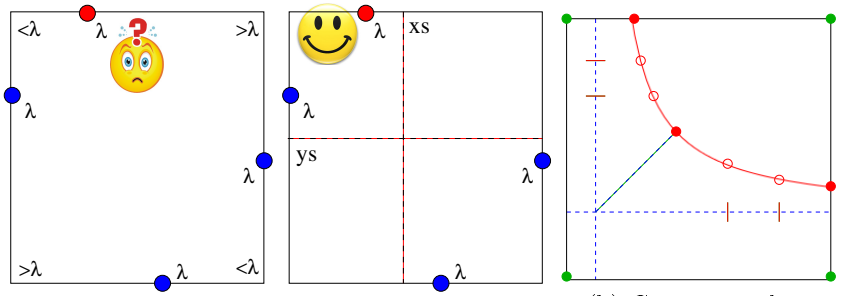
$$u(x,y) = u_{00}(1-x)(1-y) + u_{10}x(1-y) + u_{01}(1-x)y + u_{11}xy. \tag{2.10}$$

The behavior is determined by the quantity

$$a = u_{00} + u_{11} - u_{10} - u_{01}. \tag{2.11}$$

If $a = 0$, we get

$$u(x,y) = (u_{10} - u_{00})x + (u_{01} - u_{00})y + u_{00}. \tag{2.12}$$



(a) Finding exit point (b) Curve sampling

Figure 2.8: Following a level line inside four adjacent prescribed points and sampling the hyperbola branch. (a) When the level line at level λ gets inside the square from the top side (at red point) and the saddle point (x_s, y_s) is inside, there is one possible exit point on each remaining side. To disambiguate, we compare the position of entry point relative to x_s. Since the level line does not cross the asymptotes, the exit point is on the left hand side here. (b) We sample the branch of hyperbola with the following point: the entry and exit points inside the square and the maximum curvature point if inside the square (it is on the angle bisector of the asymptotes). On each side, we sample uniformly in x or in y, depending on whether $|x'(y)| < 1$ or $|y'(x)| < 1$.

If the factors in front of x and y both vanish, we have all four corners at same value u_{00} and there is no level line through the square (since we avoid prescribed values of u). Otherwise, we get a line equation for the level line. The other case $a \neq 0$ is more interesting because we can rewrite

$$u(x, y) = a(x - x_s)(y - y_s) + \lambda_s, \tag{2.13}$$

with $x_s = (u_{00} - u_{01})/a$, $y_s = (u_{00} - u_{10})/a$ and $\lambda_s = a_{00} - a x_s y_s$. This formulation shows that level lines are equilateral hyperbolae of center $S = (x_s, y_s)$, the saddle point of u, and horizontal and vertical asymptotes. The exception is for the level line $u(x, y) = \lambda_s$, in which case we get the horizontal and vertical lines through (x_s, y_s). For $\lambda \neq \lambda_s$, the level line $u(x, y) = \lambda$ is the intersection of a hyperbola with the square $[0, 1]^2$. It can represent a single branch (S is outside the square) or two branches (S is inside the square). The latter case is a bit more difficult to handle, because we must take care of not jumping from one branch to the other while following the fragment of level line, see Figure 2.8.

This algorithm extracts level lines but not the tree structure. However, a simple post-processing can recover the inclusion information. It consists in finding intersections of level lines with regularly spaced horizontal lines. These are discrete points, their number is even. Ordering these points by x values among each line, interior/exterior of a level line is determined by odd/even number of intersections up to that abscissa.

The second algorithm is able to deal with arbitrary levels. It computes first the "fundamental" tree of bilinear level lines, which consists of the level lines at critical values. From this fundamental tree, the inclusion tree of any family of level lines can be computed.

Higher order of interpolation would be even more desirable, for example splines of order 3 or 5. Unfortunately, the equation of the level line in a square

whose vertices are four prescribed points is implicit in the form $u(x, y) = \lambda$, with u a higher order bivariate polynomial. An explicit expression of the form $x = f(y)$ or $y = f(x)$ is difficult to get, and following the level line by any other means seems also difficult: its behavior may be complex, and present cusps for example.

2.1.6 Curvature map computation

The mean curvature at a regular point x of a smooth function u coincides with the curvature of the level line through x (with a sign depending on the orientation of the level line). Local extrema of curvature of a curve provide in many cases a compact and faithful idea of the curve itself. More generally, a curvature map of the image may be interesting. There are two general ways to compute it. The direct way is to use the explicit formula of mean curvature:

$$\operatorname{curv}(u)(x) = \frac{u_{xx}u_y^2 - 2u_{xy}u_x u_y + u_{yy}u_x^2}{(u_x^2 + u_y^2)^{3/2}}(x), \tag{2.14}$$

and use a finite difference scheme to approximate the partial derivatives. The indirect way is much more complex: compute the second derivative of the arc-length parameterized level line through x.

Actually, the Euclidean covariance of the curvature map is impossible to achieve with a finite difference scheme (Mondelli and Ciomaga, 2011). The only solution left is the second method. As described, it is very sensitive to the pixelization, even when using bilinear level lines. The solution is to smooth the image before, or to smooth the level lines themselves. A good smoothing preserving the Euclidean invariance while preserving the level line structure is to apply the PDE:

$$\frac{\partial u}{\partial t} = |Du| \operatorname{curv}(x)^\alpha, \tag{2.15}$$

with $\alpha = 1$ (mean curvature motion) or $\alpha = 1/3$ (affine smoothing). This is applied up to a time evolution t at which the pixelization effect disappears, typically the scale at which a disk of 1 pixel diameter disappears through application of the equation. Again, multiple numerical schemes exist to simulate those, the best ones using a stack filter. However, they all create unwanted diffusion. Geometric schemes applied to the level lines directly are preferable (Ciomaga et al., 2011).

For mean curvature motion, a solution is to parameterize the curve by arc-length $(x(s), y(s))$. Applying a Gaussian convolution to each of these coordinate signals indepedently, $(G_\sigma * x(s), G_\sigma * y(s))$, we get a new curve. If Gaussian standard deviation σ is small enough, it should not self-intersect. We then have to reparameterize by arc-length before iterating the process (Mokhtarian and Mackworth, 1992). Requiring small Gaussian smoothing σ implies many iterations to simulate the result of mean curvature motion at time t.

A geometrical scheme for the affine curvature motion (Alvarez et al., 1993) was proposed by Moisan (Moisan, 1998). It consists of these steps (Ciomaga et al., 2010):

1. Segment the curve by cutting it at inflection points, which remain fixed.
2. Change each part of the curve by taking the middle points of σ-chords.

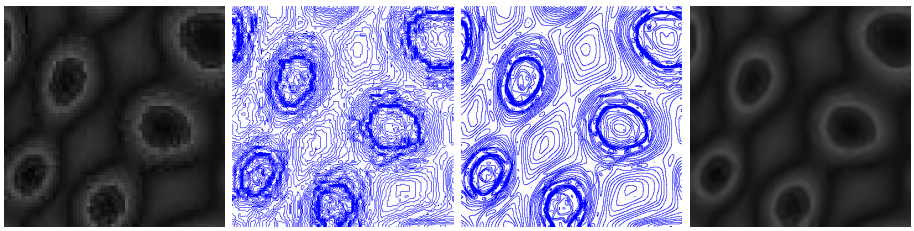
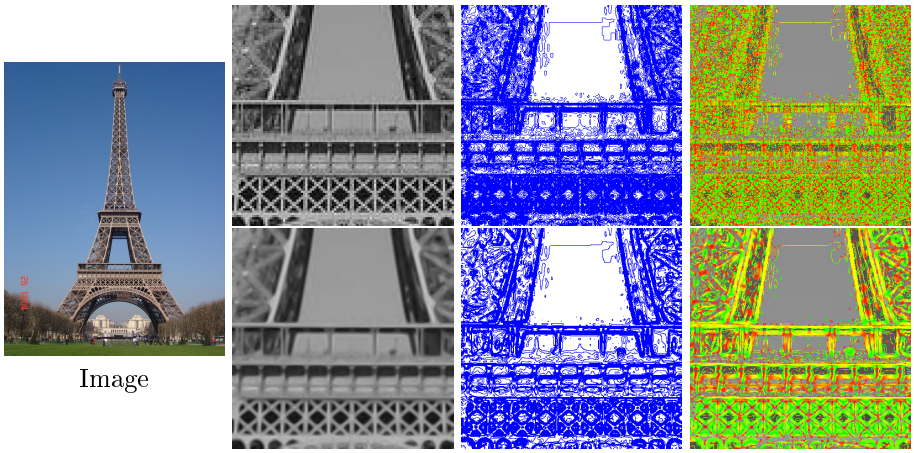


Figure 2.9: Geometric scheme for affine curvature motion. From left to right: original image, its bilinear level lines, evolution of the level lines, reconstructed image.



Image

Figure 2.10: Curvature map. Left: original image. Top row: crop of image, bilinear level lines and curvature map. Bottom row: the same after affine curvature smoothing.

3. Reconnect the curve at the former inflection points.

4. Iterate.

The erosion parameter σ represents an area. σ-chords are chords of the curve that delimit a surface of area σ with the convex curve. Applying this scheme to the level lines of an image, we get results of Figure 2.9.

An advantage of such a scheme is that the curvature computations are not pixelized, that is, they are measured at floating point coordinates. Therefore, we can zoom in the original image and still see the curvature map at any scale, hence the name of curvature microscope. To visualize the curvatures as color-coded pixels, we take all level lines going through the pixels and compute an average of their curvature, which is translated to a chromatic value.

Examples of curvature map computations on photographs of monument (Figure 2.10), of bacteria (Figure 2.11) and of fingerprint (Figure 2.12) show that the smoothing is necessary to have a readable map removing the pixelization effect (Ciomaga et al., 2013).

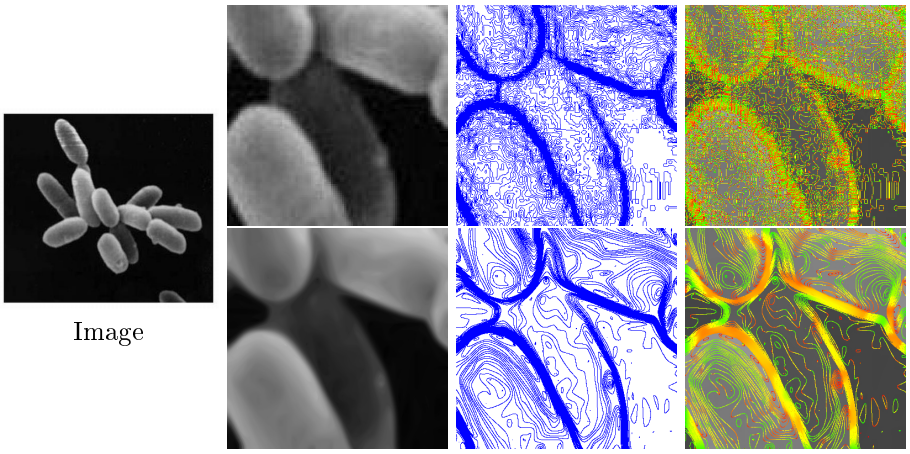


Figure 2.11: Curvature map. Left: original image. Top row: crop of image, bilinear level lines and curvature map. Bottom row: the same after affine curvature smoothing.

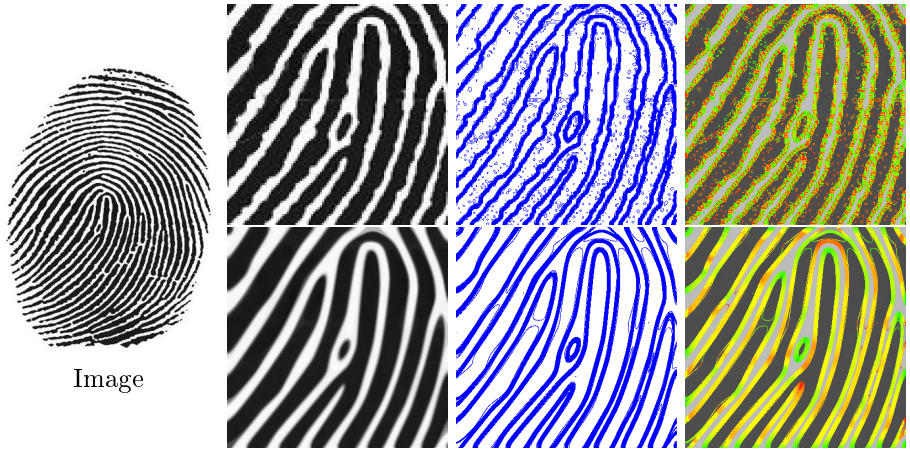


Figure 2.12: Curvature map. Left: original image. Top row: crop of image, bilinear level lines and curvature map. Bottom row: the same after affine curvature smoothing.

2.2 Calibration for 3D stereo reconstruction

Having worked on image registration during my PhD, and using as feature points the centroids of monotone sections of the tree of shapes, my interest turned to 3D reconstruction, whether from stereo pairs or multiple images. The ultimate goal is to use the camera as a real photogrammetric device, allowing precise quantitative measurements. While considered a solved problem since the geometry aspects are well understood, it is rather striking that so few quantitative assessments of common methods are found in the literature. This can also be observed by the poor quantity of publicly available photogaphs of a 3D scene with ground truth geometry.

Whereas some visually impressive results of 3D reconstruction are frequently advertised, including with large scale data, the early parts of such pipelines are of paramount importance, and especially the calibration step is essential. There are two sides to calibration, internal calibration, obtaining the intrinsic parameters of cameras, and external calibration, recovering the relative orientations and postions of the cameras in the different views. Some methods propose to solve both at once with no calibration rig, but it seems logical to validate independently each step with measurable precision. Most of my work on these topics was supported by ANR project Callisto (ANR-09-CORD-003) [5] that I coordinated, involving IMAGINE, the LTCI (Telecom ParisTech), the MAP5 (Université Paris-Descartes) and the CMLA (Ecole Normale Supérieure de Cachan).

The first notice of the geometric constraint between two pinhole views is due to Longuet-Higgins (Longuet-Higgins, 1981). This is expressed by the *essential matrix*, a 3×3 matrix that restrains corresponding points to lay on corresponding lines, the epipolar lines. There are several reasons why the *essential matrix* is difficult to use: the constraint is expressed in terms of points in real-world coordinates, which can only be computed from images after *internal calibration*, the process of estimating the projection parameters of the camera; the computation of the essential matrix is complex due to polynomial but nonlinear constraints between the nine coefficients of the matrix (Nistér, 2004). The real three-dimensional computer vision development can be traced back to the independent discovery of the *fundamental matrix* in the middle of the 1990's by Bill Triggs (Triggs, 1995) and by Olivier Faugeras and his collaborators at INRIA (Luong and Faugeras, 1996). This represented a significant progress in the sense that the only constraint of the matrix is having rank two. Ignoring this constraint during the estimation, enforced *a posteriori* by projection, the problem can be formulated as least squares and easily solved, even though numerical difficulties must be handled (Hartley, 1997a). The equations are derived from the correspondence of a few points (more rarely lines), which must take into account the presence of some outliers. This requires a robust estimation method, the most popular in this context being RANSAC (random sample consensus) (Fischler and Bolles, 1981) or one of its multiple variants. Of course, the process can only lead to *projective* 3D reconstruction but minimal external Euclidean data, such as known distances or orthogonality, can be sufficient to lift the reconstruction to Euclidean. The usual next step is to rectify the stereo pair to simulate a common projection plane and a displacement of the focal

5. `http://imagine.enpc.fr/~monasse/Callisto/`

points along the horizontal axis, which consists in applying adequate homographies to the images (Hartley, 1999). Then, the dense correspondence problem can be addressed. The apparent motion of each pixel is horizontal, its amplitude is called *disparity*. Among the numerous propositions to solve this problem (at least 150 in the popular Middlebury benchmark), two classes of methods can be distinguished:

- Global methods: a global energy combining a term of data fidelity (typically conservation of color during apparent motion) and a term of regularity of the motion field make these methods similar to *optical flow* estimation (Barron et al., 1994) with the additional constraint that the motion field is purely horizontal. The main drawback of these methods is the difficulty of minimizing the energy, almost always non-convex, and the important computational cost. A notable method in this category based on graph cuts (Kolmogorov and Zabih, 2001) guarantees a sub-optimal energy mimimum is reached. However, even the global optimum may suffer from the modelization bias.
- Local methods: small to moderate size patches of one image are considered and searched in the other image, using a distance or dissimarility function. Though these methods can be fast (the few real-time techniques are based on local methods), they usually suffer from the fact that the patch shapes are not adapted to the underlying geometry of the scene, which is unknown. This leads to the infamous *fattening effect*, yielding a dilation of foreground object shapes in the disparity map (Scharstein and Szeliski, 2002).

All methods rely on numerous parameters, usually without any automatic procedure to fix them. The inverse of the depth of points is an affine function of their disparity, whose two parameters depend on camera parameters and the rectification transforms.

Using the camera as a precise photogrammetric device, as a low cost alternative to costly time of flight or difference of phase laser systems, requires an accurate calibration. Most notably the lens geometric distortion must be corrected, otherwise the whole theory collapses because of the violation of the fundamental pinhole camera model. The very delicate auto-calibration process (Hartley, 1994), based on the observation of several images of a scene with unknown geometry and trying to solve the Kruppa equations, assumes a simplified camera model and has not yet reached an adequate level of precision. Efficient and simple calibration methods require the use of a planar rig (Zhang, 2000). A distortion model with few parameters is assumed. It is typically radial with unknown center or radial with fixed center but supplemented by a tangential distortion. The most widely used software is a Matlab toolbox written by Jean-Yves Bouguet [6]. A method from Lavest et al. (Lavest et al., 1998) is notable for its care of accuracy. The method estimates simultaneously the distortion parameters, the intrinsics (principal point, focal length, skew) and the rig geometry if not perfectly planar. It is based on a global *bundle adjustment*, which minimizes a non convex least square energy in a high dimensional space.

In the multiple view stereo case, the generalization of the fundamental matrix is the multi-view tensor (Hartley and Zisserman, 2000). While the trifocal tensor (Hartley, 1997b) gathers more constraints that the addition of its two-view

6. `http://www.vision.caltech.edu/bouguetj/calib_doc/`

constraints, the multiple view constraints stop at three views: all constraints involving more than three views are combinations of trifocal constraints (Ma et al., 2004). The most widely used software for external calibration of the views is Bundler (Snavely et al., 2010), an incremental method relying on frequent bundle adjustments to reduce drift. A more capable but less easy to use it the open source APERO software (Pierrot-Deseilligny and Cléry, 2011) Global methods are more recent and were pioneered by Martinec and Pajdla (Martinec and Pajdla, 2007). Another category of methods for global calibration relies on the factorization of the structure and motion matrix into a low rank product (Sturm and Triggs, 1996). The large dimension matrix contains many unknown coefficients since point correspondences cannot be observed in all views, connecting the problem to the very active field of low-rank matrix completion.

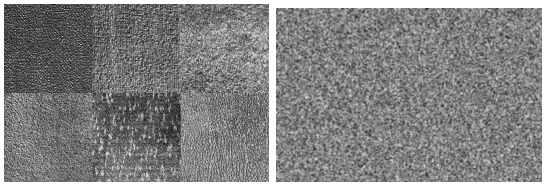
Once the external calibration is estimated, two categories of methods exist:

- Plane sweep methods: dense multi-view correspondences are recovered by assuming planar patches in 3D space and testing the reprojection in the different images (Collins, 1996). This is the mutli-view equivalent of the disparity map estimation. It then provides a point cloud, which then must be meshed and optimized. A state of the art pipeline according to a popular benchmark (Strecha et al., 2008) was demonstrated by Vu and al. (Vu et al., 2009).
- Visual hull methods: a mesh is directly constructed from the silhouette of the observed object in the different images (Lazebnik et al., 2001). This category of methods is limited to cases where a central object is seen from multiple angles and the background can be easily removed.

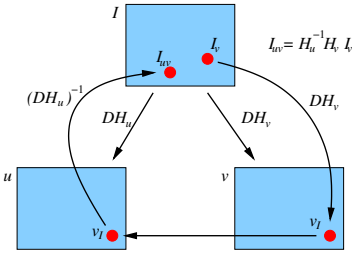
Although remarkable feats of engineering resulting in very large scale urban reconstructions were demonstrated in recent years (Agarwal et al., 2009; Frahm et al., 2010), the large scale and *accurate* 3D reconstruction from multiple views is still a fairly open problem.

An area of research with potential high impact for the movie and entertainment industry is the multiple movie stereo reconstruction, also known as 3D+time markerless stereo. We can cite a tentative approach (Courchay et al., 2009).

Finally, let us mention a less researched area of research, probably because it requires a special aparatus for data acquisition: *epipolar plane imagery* (Bolles et al., 1987). The camera motion is controlled by a linear stage and uniform. After rectification, an $x - t$ cut along each horizontal epipolar line yields an epipolar plane image, exhibiting straight segments of line whose slope is linked to depth by a simple function. Detection of these segments gives an evaluation of the depth. The advantage of the technique is that it combines the precision of large baseline stereo (considering images of the sequence far apart) without compromising the feasible correspondence problem (intermediate frames build a segment in the epipolar plane image). A sophisticated method of Criminisi et al. (Criminisi et al., 2005) is often cited. An approach relying on extension of epipolar plane segments into lines, intersection of these lines resulting in a tesselation of the epipolar plane image and interpolation of the disparity in each segment is demonstrated by Monasse et al. (Monasse et al., 2007; Rudin et al., 2011). Very good-looking results with a different method were recently demonstrated (Kim et al., 2013).



Possible patterns (texture and noise images)



Loop validation

Figure 2.13: Distortion estimation using a calibration pattern. On the right is illustrated the loop validation for discarding pattern to image correspondences, using a third image to create a loop.

2.2.1 Internal calibration

Distortion correction

Some tests with structure-from-motion software in controlled environment, using a flat calibration pattern, suggest that estimating the camera parameters while simultaneously recovering structure is an unstable process: some displacements of feature points at fine scale in the images can be attributed to camera motion, to optical defects or to 3D geometry itself. Even with a carefully laid out calibration object and fixing its 3D geometry, variations of camera position and distortion estimations are still important, even though the residual reprojected pixel error is a fraction of one pixel. The most significant problem lies with the geometric distortion of the lens. This deviation to a pinhole camera model can be very detrimental to accuracy. The PhD work of Zhongwei Tang was for a big part focused on this problem (Tang, 2011).

The use of a highly textured flat pattern was investigated (Grompone von Gioi et al., 2010). This can be a real pattern from a photograph or a synthetic one obtained by generating a white noise image, smoothing it slightly by a Gaussian filter and printing the image on a paper sheet, see Figure 2.13. The advantage is that it creates a multitude of feature points (several thousands SIFT points) which can be matched between the ground truth pattern and a photograph of it. It is important to remove outlier correspondences between the image of the pattern and the ground-truth pattern itself. The best way to validate the correspondences is to use a second photograph of the pattern and check the loop consistency. Indeed, following matching points to come back to pattern image, the distortion does not play any role, and there is just a homography between points of the pattern image and their resulting position after loop closing, see Figure 2.13. This homography can be estimated with a RANSAC algorithm and outliers discarded. Moreover, the standard model of distortion comprising a radial distortion supplemented by a tangential component and involving just a few parameters was found insufficient to model the complex distortions that can occur. A model of vector field with two bivariate polynomials of higher order improves notably the results. However, checking the results on photographs of stretched fine threads shows that the rectification of these curve lines by the distortion correction is not quite straight, with an average error of 0.08 pixel. This is attributed to the fact that the pattern object

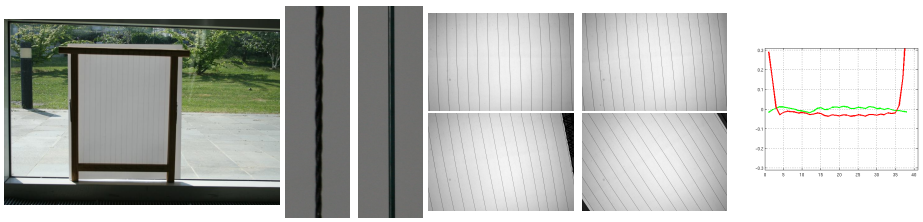


Figure 2.14: The calibration harp, two types of thread (sewing and fishing, notice the fishing thread is straighter) and photographs of the harp (notice some curved lines due to distortion). On the right is the residual error of one thread after correction (two sides).

was not perfectly flat, even though care has been taken to fix it to quite planar surfaces, like a mirror or an aluminium plate.

Extending this idea of using a plumb-line method, we designed a *calibration harp*, consisting of a rigid frame with opaque flexible threads stretched across the frame. This provides a series of ground truth straight lines, which can be used to measure the distortion and its correction (Tang et al., 2012). This device is used to evaluate the residual distortion of the lens: multiple photographs of the harp under different orientations are taken. The line segments are extracted and refined to yield largely subpixel precision, following a method proposed by F. Devernay (Devernay, 1995). Some average and maximum measures of error of each line with respect to its straight regression line give quantitative assessments of the distortion. Best results were obtained with our texture image and non-parametric distortion model, compared with classical bundle adjustment and commercial software. The average distance was measured as 0.04 pixel on average with a maximum of 0.16. These are half the best errors obtained by the competition.

When the distortion is corrected by a plumb-line method, the only insurance is that straight lines in space are projected on straight lines in the image (Grompone von Gioi et al., 2011). But any homography applied subsequently to the corrected image has the same property. Therefore, when applying the correction, there is no guaranty that some artificial homography H was not introduced in the correction. In that case, the point-to-camera projection equation becomes

$$x = HK \begin{pmatrix} R & T \end{pmatrix} X \tag{2.16}$$

instead of $x = K \begin{pmatrix} R & T \end{pmatrix} X$. But the RQ factorization of HK can be written $HK = K'R'$. This has two consequences:

1. The corrected images correspond to projections through a virtual pinhole camera of matrix of instrinsics $K' \neq K$.

2. The translation and rotation (R, T) between a world-coordinate frame and the camera become $(R'R, R'T)$ with the virtual camera. As a consequence, the 3D position of the virtual camera $-R^{-1}T$ remains the same, but the rotation becomes $R'R$. Notice that the relative rotation between two views becomes: $R'R_2R_1^{-1}R'^{-1} \neq R_2R_1^{-1}$.

Figure 2.15: Simulated lateral chromatic aberration on part of a black disk with different shifts. The color fringes are hardly noticeable below 0.05 pixel shift.

Chromatic abberration correction

The techniques used in distortion correction, in particular the correction model, can be used for lateral chromatic aberration correction (Rudakova and Monasse, 2014). This is due to the fact that the lens has a different refraction depending on wavelength and is visible as a shift between the three color planes of the image. For example, color fringes at the boundary of gray level spots may be visible. Even a shift of some fraction of pixel may be noticeable. This requires a very fine non-rigid registration of the color planes. The algorithm developed with Victoria Rudakova uses a calibration pattern made of multiple black disks on white background. This is printed on a white sheet of paper and photographed. Notice that it is not necessary to have a perfectly flat pattern and the non-planarity should not affect much the precision. The circular regions are first detected by a simple thresholding of in each channel image. Their shape is then precisely localized using a model of linear transition of the intensity at the boundary, and constant levels inside the disk and outside. A Levenberg-Marquardt minimization of the model fitting involving position and attitude of the circular shape and its intensity is performed. Only the center of each circle is then used as interest point. The green channel is then used as reference, and the interest points of blue and red channels are matchted to their nearest interest point in the green channel. A dense registration map is obtained by fitting polynomial models of high order to the green-red and green-blue correspondences.

Experiments show that the detection of disk centers is quite precise. Synthetic tests show a good resistance to noise, with less than 0.02 pixel error for additive Gaussian noise of standard deviation 1 in an 8-bit image. This is valid even for small disks (radius 10 pixels). This allows to pack many disks on the pattern and thus having numerous interest points. At least as important, a good robustness with respect to aliasing is measured, with less than 0.05 pixel localization error for highly aliased images. This is noteworthy, since the red and blue channels of the Bayer pattern produce half-size images that are notably aliased. The green channel is kept at the original resolution; the half of all image pixels with only red or blue information are interpolated by average of known neighbor green pixels.

A good registration of maximum shift of 3 pixels down to 0.1 pixel is observed for different cameras and focal lengths. This is the limit at which the aberration becomes hardly visible, see Figure 2.15. Comparison with commercial software shows that our correction outperforms all others by a comfortable margin, see Figure 2.16. This is due to

1. precise detection of interest points;
2. high order model of registration.

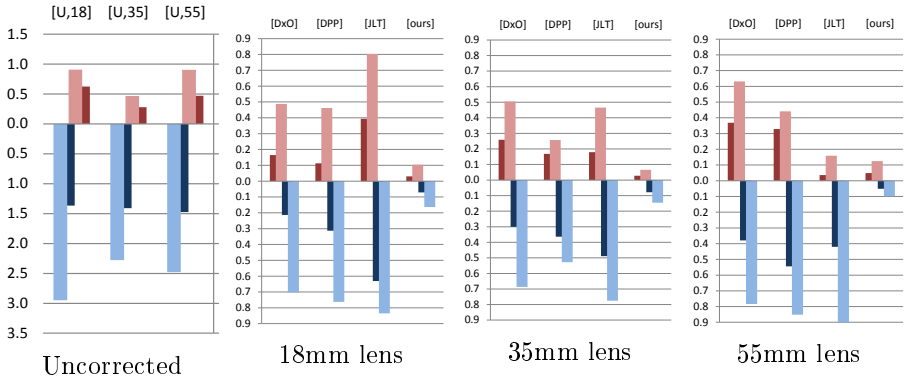


Figure 2.16: Performance comparison of lateral chromatic aberration correction (mean and maximum residual error after disk center registration between green and red/blue channels) for different commercial software programmesand our method. Camera: Canon EOS 40D with zoom lens, experiments at different focal lengths. The left graph shows uncorrected aberration (notice the different scale).

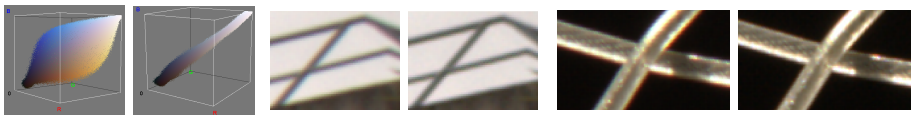


Figure 2.17: Lateral chromatic correction on real data with Canon EOS40D. We show original (left) and corrected (right) data.

Results of the estimated correction on real data are shown in Figure 2.17. Distribution of pixel colors in RGB space of a photograph of the black and white pattern is much more concentrated after correction. Notice also the disparition of color fringes on different parts of images.

2.2.2 External calibration

The external calibration is a solved problem for a pair of cameras. Assuming the internal calibration matrices are known and the essential matrix computed, a simple decomposition via SVD of the essential matrix provides the relative rotation and translation (up to a 4-fold indeterminacy, which is easily solved by the cheirality constraint, namely imposing interest points to be in front of each camera). The problem arises with loops: estimated relative transforms between coordinate frames in a loop must compose to identity. Considering that each estimate has a certain amount of error, balancing equally the error is difficult. Incremental methods rely on bundle adjustment to "close" the loops. Global calibration methods try to deal earlier with the loops, at higher computational cost.

Epipolar rectification

Once the fundamental matrix between two pinhole camera views has been computed from point correspondences, based most of the time on point correspondences and outlier elimination through some form of RANSAC algorithm, putting the images into epipolar rectified geometry is generally a necessary step before disparity estimation. The rigidity constraint being written with the fundamental matrix F

$$x'^T F x = 0 \tag{2.17}$$

with x and x' corresponding points in their respective image in homogeneous coordinates, the rectification is achieved by applying homographies of matrix H and H' to the images such that

$$H'^T [i]_\times H = \lambda F \tag{2.18}$$

with $i = \begin{pmatrix} 1 & 0 & 0 \end{pmatrix}^T$. Indeed, in that case the rigidity constraint becomes

$$(H'x')^T [i]_\times (Hx) = 0 \tag{2.19}$$

which amounts to $(Hx)_2 = (H'x')_2$, meaning that points Hx and $H'x'$ are on the same horizontal line in both images. Multiple solution pairs (H, H') satisfy (2.18) and there is some leeway in the choice of solutions. Additional constraints may be imposed, mostly to distort as little as possible the original images (with different measures of distortion among the methods). The method of Fusiello and Irsara (Fusiello and Irsara, 2008; Monasse, 2011) tries to simulate physical zoom and rotations of the camera, hence its name "quasi-Euclidean". For that, it has to assume a calibration matrix

$$K = \begin{pmatrix} f & 0 & w/2 \\ 0 & f & h/2 \\ 0 & 0 & 1 \end{pmatrix}. \tag{2.20}$$

f is the unknown focal length but the principal point is fixed at $(w/2, h/2)$, the center of image. The rotations of the two cameras are paremeterized by 5 angles, 3 for each camera but an arbitrary common rotation around optical axis leaves the pair rectified. Supplemented with f, a vector v of 6 unknowns is then searched minimizing the error measure

$$\sum_{i=1}^{N} \epsilon(x'_i, H'(v)[i]_\times H(v), x_i)^2 \tag{2.21}$$

with ϵ the error function, preferably having geometric meaning rather than purely algebraic. Fusiello and Irsara choose the Sampson error. This minimization is based on Levenberg-Marquardt, so may fail to converge to the local minimum.

With Jean-Michel Morel and Zhongwei Tang, an improvement was proposed (Monasse et al., 2010), relying on the single unknown f. Having only one parameter instead of 6, the minimization is much simpler and even a global minimization could be sought, even though we still relied on Levenberg-Marquardt iterative minimization. The method is based on the observation that assuming f known, we can decompose the rectification in three elementary steps, see Figure 2.18:

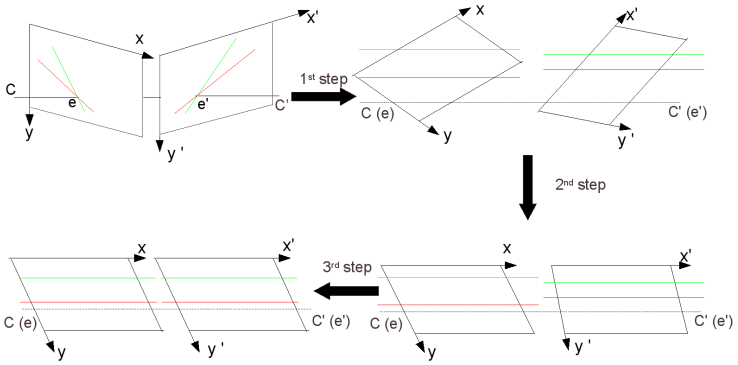
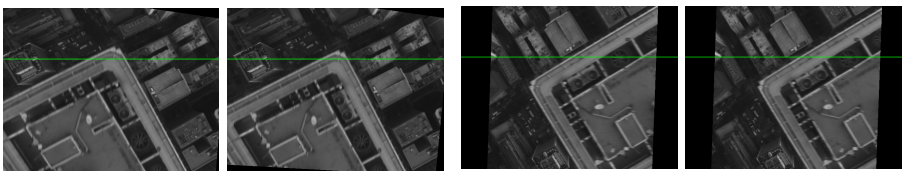


Figure 2.18: Three-step epipolar rectification.



Fusiello and Irsara's method Three-step method

Figure 2.19: Epipolar rectification of a pair of aerial images, with superimposed green line to help visual comparison. Notice the residual vertical motion in Fusiello-Irsara's method.

- Apply minimal rotations such that epipoles $a = (e_x, e_y, 1)$ and $a = (e'_x, e'_y, 1)$ are mapped to infinity $b = (e_x, e_y, 0)$ and $b' = (e'_x, e'_y, 0)$. Writing $H = KRK^{-1}$, we can write $RK^{-1}a = K^{-1}b$ and find R as the closest rotation to identity aligning the 3D vectors $K^{-1}a$ and $K^{-1}b$.
- Rotate each image to map b and b' to i with similar method.
- The new fundamental matrix may be written $F' = K^{-T}[i]_\times RK^{-1}$ and R, rotation of one camera around optical axis, can be recovered from $K^T F' K$.

The three steps depend on K, which is parameterized by the unknown f, but are computable with closed form formulae. Instead of the Sampson error, the symmetric transfer error is used:

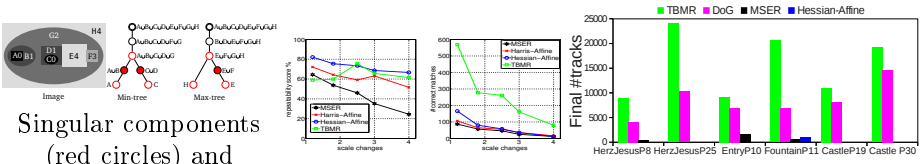
$$\epsilon(x'_i, H'(f)[i]_\times H(f), x_i)^2 = d(x'_i, Fx_i)^2 + d(x_i, F^T x'_i)^2 \tag{2.22}$$

with d the point-to-line geometric distance based on points and lines equations in homogeneous coordinates.

It turns out that due to the lower dimensionality of the search space, this algorithm is less susceptible to get stuck in a local minimum. This is particularly the case when the initial apparent motion of corresponding points is mostly vertical, as in Figure 2.19.

Tree-Based Morse Regions

Feature points and their correspondences are at the basis of many crucial steps in 3D computer vision. The most popular interest points are based on



Singular components (red circles) and TBMR (red disks)

Score on 'Bark' images

Number of tracks in multi-view stereo

Figure 2.20: Tree Based Morse Regions. Left: definition. Middle: comparison with other methods in terms of repeatability and number of correct matches in Bark image (notice comparable repeatibility but a much higher number of matches for TBMR). Right: comparison of the number of tracks in datasets of multi-view stereo with different interest point dectectors.

local extrema of the derivative of Gaussian scale-space (DoG), with each point associated by a fixed-length descriptor computed by the SIFT method (Lowe, 2004). Fast implementations exist and DoG based interest points have the avantage of being numerous, especially if one starts from octave -1 (doubled image size). However, it is to be noted that the method depends on multiple parameters, even though default values are appropriate in general. Also, the method is not contrast invariant: Gaussian convolution does not commute with general contrast change, and even though the descriptor construction is contrast invariant, relying on direction of gradient, it is computed on a convolved image.

In a celebrated benchmark (Mikolajczyk et al., 2005), MSER (Matas et al., 2004) was shown to outperform DoG and variants in the most important cases: more robustness with respect to geometric transformation and with contrast changes. The cases where MSER perform poorly are when the images have undergone blur, strong JPEG compression, or an important change of scale. The first two are just related to image quality and should be easy to avoid. The real limitation is the sensitivity to scale changes. Nevertheless, MSER suffers from two defects: MSERs are usually much less numerous than DoG interest points, and they depend on a few parameters, though much fewer than SIFT. Moreover, their precise definition is not quite clear. Given a connected component C_λ of $[u \geq \lambda]$, the stability function is defined as:

$$\frac{|C_{\lambda-\delta}| - |C_{\lambda+\delta}|}{|C_\lambda|}, \tag{2.23}$$

with $\delta > 0$, where $C_{\lambda-\delta} = \mathrm{cc}([u \geq \lambda - \delta], C_\lambda)$. The ambiguity arises in the term $C_{\lambda+\delta}$: when there is a bifurcation between levels λ and $\lambda + \delta$, should it be the greatest connected component of $[u \geq \lambda]$ inside C_λ, or the union of all such connected components? The original paper does not precise this point leading to slightly different interpretations of MSER. Local minima in the component trees of this stability function are the MSERs. The associate descriptor is usually SIFT, even though MSER being a true region of the image, more specific descriptors should be possible. The parameter δ implies that MSER are not quite contrast invariant.

A simpler, contrast invariant variant of MSER was called TBMR for tree based Morse regions (Xu et al., 2013). They are the connected components of level sets just before a bifurcation, see Figure 2.20. Actually, there is still

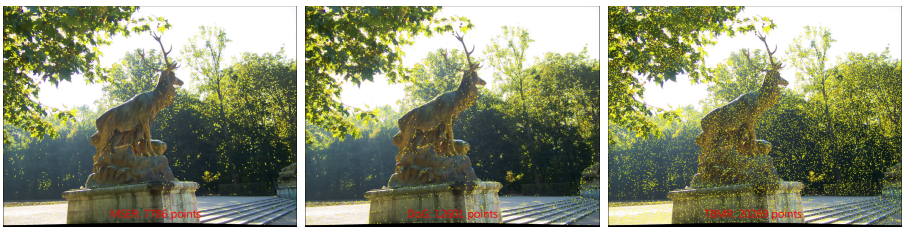


Figure 2.21: Repartition of interest points in an image. Notice that TBMR are more numerous and more uniformly distributed than others. MSER has very few detections on the object of interest (the sculpture) due to low contrast.

Using DoG Using TBMR

Figure 2.22: Three dimensional reconstruction from multiple view stereo images, using DoG or TBMR as interest points. Notice that the back facade of the castle is lost using DoG because of lack of correct correspondences, while TBMR recovers a more complete 3D.

a parameter, the min area of the grain filter to apply. This not only removes small regions, but also modifies the tree structure: some bifurcations become regular components as all children but one had an area smaller than the grain filter parameter.

TBMR have good repeatibility score on standard benchmarks, but with much higher number of correspondences, see Figure 2.20. Also, their repartition is more uniform, see Figure 2.21. This has important consequences for image registration and 3D reconstruction from stereo, see Figure 2.22. These results are obtained by incremental calibration using different interest points, followed by PMVS algorithm (Furukawa and Ponce, 2010).

Incremental calibration

Incremental calibration is the paradigm employed by the popular open-source software Bundler (Snavely et al., 2010). Pairs of corresponding points between images are computed and filtered to remove outliers (via a RANSAC algorithm computing the fundamental matrix). Ensuring transitivity of correspondences, tracks of points are constructed, a track corresponding to a 3D point (Moulon et al., 2012). Then an initial pair of images with enough point correspondences is chosen. Knowing the fundamental matrix F, the essential

matrix E can be deduced by

$$E = K'^T F K, \tag{2.24}$$

with K and K' the intrinsics matrices. From the formula $E = [T]_\times R$ with R and T the relative rotation matrix and translation vector between cameras, R and T can be recovered. Once done, the triangulation of matching points recovers 3D points in the fixed coordinate frame linked to the reference camera of the image pair. A third view having maximum tracks in common is appended. This gives a set of 3D-2D correspondences. From this, pose estimation recovers the relative position and orientation of this camera. Adding incrementally views is susceptible to drift: errors accumulate and views added late in the process are badly positioned with respect to the first views. To compensate that, a bundle adjustment happens after adding each new view:

$$\arg \min_{\{R_i\},\{T_i\},\{X_j\}} \sum_{i,j} d(x_{ij}, K_i \begin{pmatrix} R_i & T_i \end{pmatrix} X_j)^2, \tag{2.25}$$

where i is the view index, j the 3D point index, X_j a 3D point and d the Euclidean distance in the image plane. Image point x_{ij} is the observed 3D projection of track of index j in view i. This optimization is typically slow because it is in a space of large dimension, non-convex and likely to get trapped in a local minimum.

The whole procedure depends on several kinds of model estimations:

- Homography estimation;
- Fundamental or essential matrix estimation;
- Pose estimation.

Of course, each of these steps can be contaminated by outliers. The robust algorithm of choice for model estimation is RANSAC (Random Sample Consensus) (Fischler and Bolles, 1981). It depends on a parameter of precision σ that should be fixed in accordance with the noise level. It discriminates between noisy inlier points and outliers. The right value of σ being unknown, a fixed threshold is usually used. With Pierre Moulon and Renaud Marlet, it was checked that using a fixed uniform threshold for all images is not optimal. A better solution is to estimate σ at the same time as the model.

This is achieved using the *a contrario* framework, according to which observed unlikely events are significant. Given a model M and supposing k inliers, the expectation of the event "having k inliers or more", or number of false alarms, is given by (Moisan et al., 2012):

$$NFA(M,k) = N_{\text{out}}(n - N_{\text{sample}}) \binom{n}{k} \binom{k}{N_{\text{sample}}} (e_k(M)^d \alpha_0)^{k - N_{\text{sample}}} \tag{2.26}$$

where N_{out} is the number of models estimated by a minimum size of N_{sample} correspondences among n (usually $N_{\text{out}} = 1$, but 3 for fundamental matrix estimation based on 7 correspondences), $e_k(M)$ is the k-th lowest error to M, α_0 is the probability of a random correspondence having error at most 1 pixel and d the error dimension: 1 for distance point-line, 2 for point-point. This relies on the background model according to which a random correspondence is a pair of uniformly distributed points in their respective image. The number of inliers $k > N_{\text{sample}}$ is not known so a minimization happens:

$$NFA(M) = \min_k NFA(M,k). \tag{2.27}$$

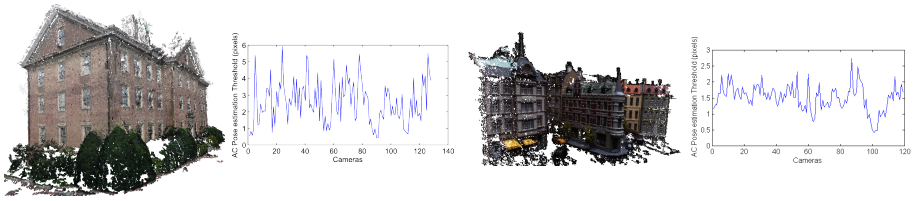


Figure 2.23: Reconstruction and estimated pose estimation threshold σ depending on image number, as estimated by *a contrario* RANSAC algorithm. Notice the significant variation of the threshold.

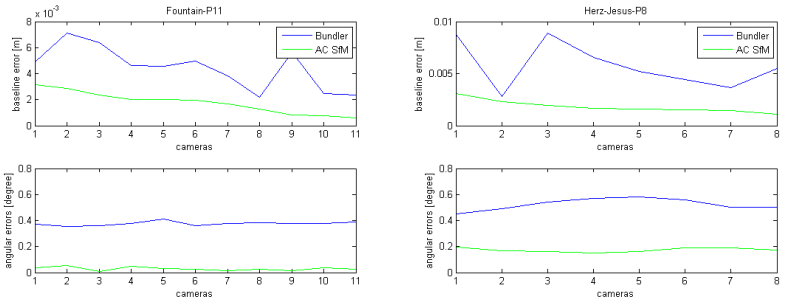


Figure 2.24: External calibration accuracy in position and angle with classical Bundler (and its standard parameters) and with *a contrario* structure from motion (which uses adaptive thresholds).

The optimal k_0 gives the error threshold $\sigma = e_{k_0}(M)$. Figure 2.23 shows some 3D reconstruction from multiple-view stereo with estimated σ in pose estimation of each image. Notice the variation of this threshold depending on the image.

Experiments on data with available ground truth show that replacing all model estimations by their *a contrario* equivalent yields better results than any uniform thresholds could achieve (Moulon et al., 2013a). This is especially true of the rotation accuracy, while the position accuracy is better but by a smaller margin. This can be observed in Figure 2.24.

The background model (null hypothesis) for *a contrario* estimation assumes that point matches are uniformly distributed in their respective image. A deviation from this model can be deemed as significant simply because feature points are not uniformly distributed. Improved results in *a contrario* fundamental matrix estimation relying on a more adpative background model have since been demonstrated (Espuny et al., 2014).

Scalable global calibration

As mentioned, the incremental calibration can be slow, since multiple bundle adjustments must be performed to avoid drift. Also the result depends on the initial pair chosen and on the order in which views are appended, on frequency of bundle adjustment, etc. The most prominent problem is that the bundle adjustments happening in high dimension spaces, a lot of local minima can be expected and a good initialization is necessary. A few methods have been proposed to deal globally with all views and estimate rotations and translations.

Loops in camera trajectory are handled. The two problems are usually separated: rotations are estimated, then translations. Different techniques are used to handle both problems.

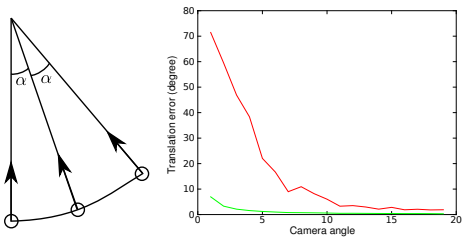
All pairs of images (i, j) from which an essential matrix could be computed yield an estimated relative motion (R_{ij}, t_{ij}), except that only the direction of t_{ij} is known, not its amplitude. A set of global rotations R_i with respect to a fixed coordinate frame is sought:

$$\forall i, j \quad R_j = R_{ij} R_i. \tag{2.28}$$

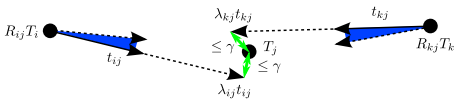
The difficulty is that the parameterization of rotation matrices cannot be achieved linearly. The solution proposed by Govindu (Govindu, 2001) and independently by Martinec (Martinec and Pajdla, 2007) parameterizes each rotation by a quaternion. This yields a set of linear equations that is easily solved. However, the unit length of each quaternion is not guaranteed by this minimization, and, since we are unable to impose the constraints in the minimization, each is normalized *a posteriori*. Though some variants exist, no public method is yet able to deal exactly with global rotations.

However, the first task is to make sure that all R_{ij} are plausible. Indeed, in artificial environments it is not rare to have sufficient structural coherence between two images that are actually observing different objects. For example two different facades can match because of coherent shape and organization of features, such as windows. Zach et al. (Zach et al., 2010) proposed to find loops in the visibility graph (whose vertices represent views and an edge between views i and j is present whenever E_{ij} could be computed), cycle errors to identity are computed and errors above a threshold provoke the rejection of the edges. Only cycle lengths up to 6 are considered, because of lower complexity and because longer cycles should allow more leeway in their error. We supplement the algorithm with the normalization factor $\sqrt{l}$ for the error, where l is the cycle length. As observed by Enqvist et al. (Enqvist et al., 2011), an error proportional to $\sqrt{l}$ is the expected behavior of a normal cycle. To avoid limitation of the check to small cycles, we iterate the algorithm until no more edge is removed. Finally, we check all triplets forming loops in the graph and reject the ones with an error greater than $2°$.

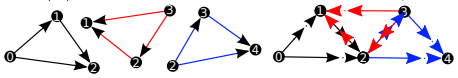
To estimate translations, we proceed in two steps: trifocal tensors are computed on length 3 loops to estimate more precise relative translations using *a contrario* RANSAC, then translation registration is performed. The trifocal tensor is estimated with known rotations, so four point correspondences in the three images are enough to have a solution. The minimal solver is based on solving the feasibility of a linear programme involving the l_∞ reprojection error, see Figure 2.25. This is shown to be more accurate and quite faster than a straight l_∞ minimization using slack variables to handle outliers, as proposed by Sim and Hartley (Sim and Hartley, 2006). The method is also shown to yield much more precise translation direction than the one based on pairs, especially with small baselines. The translation registration minimizes the maximum error between the vectors $\lambda_{ij} t_{ij}$ (scaled relative translation vector) and $T_j - R_{ij} T_i$ (relative translation computed from absolute translations). The unknowns are the T_i and λ_{ij}. This is solved by a linear program, faster than the SOCP formulation of Sim and Hartley, which is based on angular errors instead of Euclidean distances.



(a) Accuracy of translation direction



(b) Translation error measure γ.

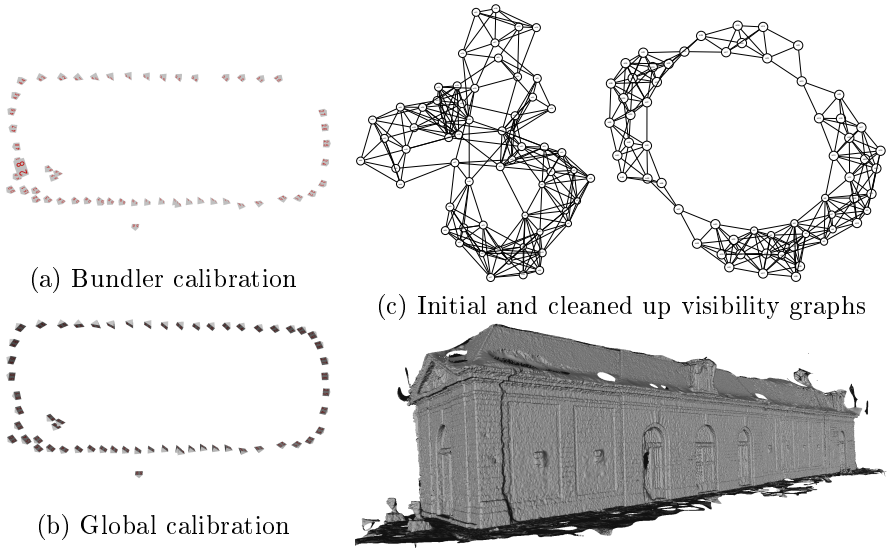


(c) Translation registration of trifocal tensors.

Figure 2.25: Translation registration in global external calibration. (a) Accuracy of translation direction estimation in bifocal (red) and trifocal (green) tensor estimation, as a function of scene viewing direction. Test performed on synthetic images with Gaussian noise perturbing 2D point projections. (b) The error measure γ we use in estimation of translation registration. (c) 3 local tensors and merged translations.

Finally, our pipeline (Moulon et al., 2013b) operates quite faster (by a factor of about five) than the concurrent global pipeline of Olsson and Enqvist (Olsson and Enqvist, 2011), with comparable precision as the best incremental and global methods on Strecha's dataset. With larger scale datasets, the time efficiency of the pipeline is even more notable, with success where Bundler was unable to calibrate precisely and with late bundle adjustments taking very long to compute (because of higher number of variables).

We show a 3D reconstruction based on a dataset of 61 images of the Sceaux Orangerie in Figure 2.26. A larger dataset of 161 images of Opera Garnier yields the reconstruction of Figure 2.27.



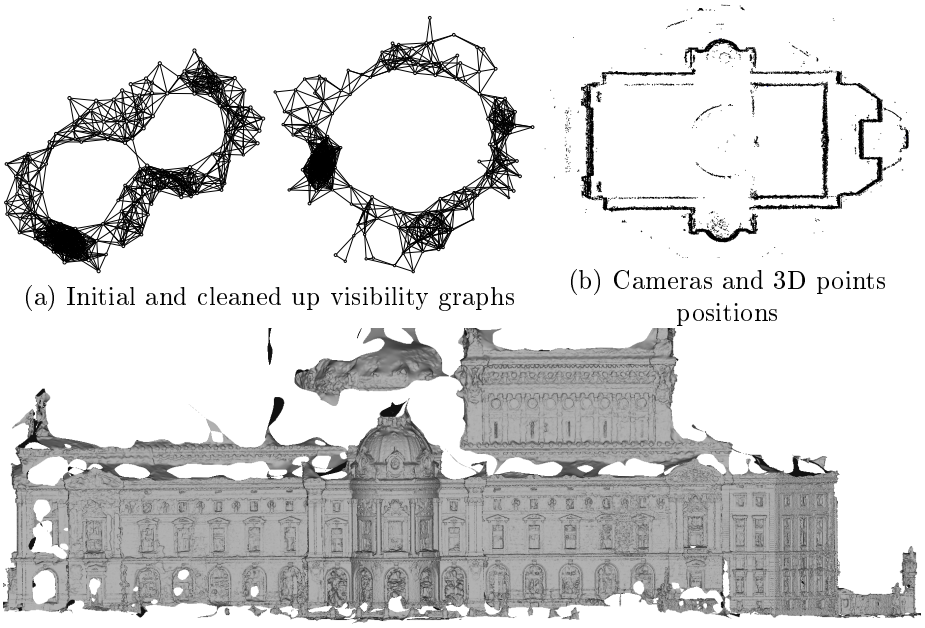
(a) Bundler calibration

(b) Global calibration

(c) Initial and cleaned up visibility graphs

(d) Reconstruction

Figure 2.26: Reconstruction based on global registration. Notice the missing views in upper-left part of (a) with incremental calibration of Bundler. The visibility graph has some wrong edges in (c), which are removed by cycle analysis.



(a) Initial and cleaned up visibility graphs

(b) Cameras and 3D points positions

(c) Reconstruction

Figure 2.27: Reconstruction based on global registration.

Chapter 3

Perspectives

3.1 Disparity map computation

When a pair of images are rectified, the computation of the disparity map is a challenge, with over 150 methods tested in the Middlebury benchmark [1]. Whereas the benchmark offers a glimpse of the state of the art, many of the methods have no publicly available implementation and therefore their results on different datasets are unknown. In the Middlebury benchmark, the underlying geometry of the scene is mainly a collection of planes, the ground truth data is questionable with its limited precision, most scenes are fronto-parallel. This does not give a good indication of the performance of the methods in satellite imagery for example. One strength of the IPOL journal (Image Processing On Line) is that each article provides an open source reference implementation of the algorithm, the article itself serving as documentation. It is unnecessary and probably impossible to reimplement all methods tested in the Middlebury benchmark: many published papers do not provide all necessary information to reproduce the results. But online demonstrations of a few representative methods would permit a real comparison. This would open the way to a true understanding of the strengths and weaknesses of the methods. This work has already begun with several landmark algorithms already submitted to IPOL or in preparation.

There are typically two categories of methods: local and global methods. Local methods rely on limited size patches around each pixel and try to locate them in the search image using some distance measure. Many times, some post-processing is done: filters to eliminate wrong correspondences and recomputation based on reliable points. Global methods define an energy, similarly to optical flow methods, exhibiting a data fidelity term and a smoothness term. These present other kinds of problem, with the difficulty of minimizing the energy, the optimization can have a high computation cost, and the parameters are often numerous and hard to tune. Local methods typically suffer from the adherence (or fattening) effect. This is an artifact created by the patches that are not adapted to the local geometry. Some recent methods take large square patches but overlay a weight map that tries to follow the image geometry, with the assumption that discontinuity of depth is usually accompanied

1. `http://vision.middlebury.edu/stereo/`

by a discontinuity of color. Several representatives of these evolved local methods are in preparation for publication in IPOL, like weighted patches (Yoon and Kweon, 2006) and cost aggregation based on guided filter (Rhemann et al., 2011). Concerning global methods, the celebrated algorithm of Kolmogorov and Zabih (Kolmogorov and Zabih, 2001) based on graph cuts has been submitted for publication (Kolmogorov et al., 2013).

Another research track concerns the precision of the disparity computation. It has been recently shown that the recovered height precision can be estimated by the formula

$$\delta h(x) = \frac{1}{(b/h)nt^{3/2}} \frac{\overline{\sqrt{v(x)}}}{\overline{(v)}_x^2}, \tag{3.1}$$

with n the patch size, t the exposure time, and v the image. This is valid under the assumption of a Poisson noise, which is reasonable. The role of b/h (ratio baseline-depth) and n is well-known. What has been overlooked is the role of the exposure time t of the camera. Increasing t reduces the noise and therefore the error with a very favorable exponent, 3/2. In other terms, with almost noise free images, it should be possible to consider a low b/h, which is favorable since it allows to have more similar images and facilitates the matching, and a low n, which reduces the adherence. Investigating the effect of noise on disparity precision and trying to reach the theoretical bounds given by the above formula are the goals of the new ANR project STEREO (programme ASTRID 2012), involving CMLA and IMAGINE. This involves the creation of *perfect images*, that is, images with no aliasing and no noise created by subpixel registration, super-resolution, accumulation and resampling of a burst of images. This is better than long exposure time since there is less risk of saturation. The quality of distortion correction with the calibration harp must be verified (experiments suggest an average rectification error of $1/30th$ pixel could be reached). Subpixel block matching with small blocks could then be performed (the target is of the order of $1/100th$ pixel precision). Methods based on optical flow should also be tested. Their advantage is that the subpixel character is automatic and does not require any interpolation of the image.

3.2 Multiple view stereo

In multiple view stereo, numerous problems remain, especially concerning global calibration. Probably the most difficult part concerns the global rotation registration. The fact that the rotation constraints on matrices are nonlinear makes any optimization procedure difficult. So far, few methods have been proposed, and none is really satisfactory. In comparison, the translation registration seems a much easier problem, with several existing methods. Other problems in the pipeline still exist, such as how to distribute the error evenly in loops, how to remove outliers in the visibility graph, identifying degenerate situations for the fundamental matrix, which are frequent in artificial environments because of mostly planar surfaces... The multiple view stereo calibration is still an active area of research, especially with respect to global methods and when handling large datasets.

Other problems in multiple view stereo include the fusion of disparity maps. For example, given three images, three pairs can be formed and each one gives

rise to disparity maps. These disparity maps may not be dense, due to occlusions and false negative disparities. The question is how to merge these maps to obtain denser and more precise maps. Actually, a disparity map is specific to a pair of rectified images, so the problem may be better reformulated as the fusion of point clouds. Getting the point cloud from a disparity map requires the use of calibration parameters. Initial experiments performed in the framework of the MISS project show that the internal and external calibrations given as metadata with some Pleiades satellite imagery are not precise enough to significantly merge such point clouds.

Another area of interest is linked to poorly textured scenes, with some applications having a high industrial impact. For example, the reconstruction of indoor 3D geometry faces multiple problems: non-uniform lighting, reflections, poor textures, close range photography... For such scenes, the number of interest points may be too low. Other features may be used, most notably segments, and *a priori* hypotheses about the 3D scene must be assumed, such as planar surfaces. Convenience of the data capture may require the use of wide angle cameras, meaning that disortion correction is necessary and that non-uniform resolution of the corrected images must be taken into account.

Bibliography

Agarwal, S., Snavely, N., Simon, I., Seitz, S. M., and Szeliski, R. (2009). Building Rome in a day. In *Proceeding of the 12th International Conference on Computer Vision (ICCV)*, pages 72–79. IEEE.

Alvarez, L., Guichard, F., Lions, P.-L., and Morel, J.-M. (1993). Axioms and fundamental equations of image processing. *Archive for Rational Mechanics and Analysis*, 123(3):199–257.

Bajaj, C., Pascucci, V., and Schikore, D. (1996). Fast isocontouring for improved interactivity. In *In Proceedings of IEEE Symposium on Volume Visualization*, pages 39–46.

Ballester, C., Caselles, V., and Monasse, P. (2003). The tree of shapes of an image. *ESAIM: Control, Optimisation and Calculus of Variations*, 9:1–18.

Barron, J. L., Fleet, D. J., and Beauchemin, S. S. (1994). Performance of optical flow techniques. *International Journal of Computer Vision*, 12(1):43–77.

Bolles, R. C., Baker, H. H., and Marimont, D. H. (1987). Epipolar-plane image analysis: An approach to determining structure from motion. *International Journal of Computer Vision*, 1(1):7–55.

Cao, F., Musé, P., and Sur, F. (2005). Extracting meaningful curves from images. *Journal of Mathematical Imaging and Vision*, 22(2-3):159–181.

Caselles, V., Coll, B., and Morel, J.-M. (1999a). Topographic maps and local contrast changes in natural images. *International Journal of Computer Vision*, 33(1):5–27.

Caselles, V., Lisani, J.-L., Morel, J.-M., and Sapiro, G. (1999b). Shape preserving local histogram modification. *IEEE Transactions on Image Processing*, 8(2):220–230.

Caselles, V., Meinhardt, E., and Monasse, P. (2008). Constructing the tree of shapes of an image by fusion of the trees of connected components of upper and lower level sets. *Positivity*, 12(1):55–73.

Caselles, V. and Monasse, P. (2002). Grain filters. *Journal of Mathematical Imaging and Vision*, 17(3):249–270.

Caselles, V. and Monasse, P. (2010). *Geometric Description of Images as Topographic Maps*. Number 1984 in Lecture Notes in Mathematics. Springer.

Ciomaga, A., Moisan, L., Monasse, P., and Morel, J.-M. (2013). Image curvature microscope. Preprint Image Processing On Line (IPOL), `http://www.ipol.im/pub/pre/H4/`.

Ciomaga, A., Monasse, P., and Morel, J.-M. (2010). Level lines shortening yields an image curvature microscope. In *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP)*, pages 4129–4132. IEEE.

Ciomaga, A., Monasse, P., and Morel, J.-M. (2011). Image visualization and restoration by curvature motions. *SIAM Multiscale Modeling and Simulation*, 9(2):834–871.

Collins, R. T. (1996). A space-sweep approach to true multi-image matching. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 358–363. IEEE.

Courchay, J., Pons, J.-P., Monasse, P., and Keriven, R. (2009). Dense and accurate spatio-temporal multi-view stereovision. In *Proceedings of the 9th Asian conference on Computer Vision (ACCV)*, volume 2, pages 11–22. Springer-Verlag.

Cox, J., Karron, D., and Ferdous, N. (2003). Topological zone organization of scalar volume data. *Journal of Mathematical Imaging and Vision*, 18(2):95–117.

Criminisi, A., Kang, S. B., Swaminathan, R., Szeliski, R., and Anandan, P. (2005). Extracting layers and analyzing their specular properties using epipolar-plane-image analysis. *Computer Vision and Image Understanding*, 97(1):51–85.

Desolneux, A., Moisan, L., and Morel, J.-M. (2001). Edge detection by helmholtz principle. *Journal of Mathematical Imaging and Vision*, 14(3):271–284.

Desolneux, A., Moisan, L., and Morel, J.-M. (2008). *From Gestalt Theory to Image Analysis: A Probabilistic Approach*, volume 34 of *Interdisciplinary Applied Mathematics*. Springer.

Devernay, F. (1995). A Non-Maxima Suppression Method for Edge Detection with Sub-Pixel Accuracy. Technical Report RR-2724, INRIA.

Dibos, F., Koepfler, G., and Monasse, P. (2003a). *Geometric Level Set Methods in Imaging, Vision, and Graphics*, chapter Image Alignment, pages 271–295. Springer. S. Osher and N. Paragios, eds.

Dibos, F., Koepfler, G., and Monasse, P. (2003b). *Geometric Level Set Methods in Imaging, Vision, and Graphics*, chapter Total Variation Minimization for Scalar/Vector Regularization, pages 121–140. Springer. S. Osher and N. Paragios, eds.

Enqvist, O., Kahl, F., and Olsson, C. (2011). Non-sequential structure from motion. In *IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pages 264–271. IEEE.

Espuny, F., Monasse, P., and Moisan, L. (2014). A new a contrario approach for the robust determination of the fundamental matrix. In Huang, F. and Sugimoto, A., editors, *Proceedings of the PSIVT 2013 Workshop on Geometric Computation for Computer Vision (GCCV)*, number 8334 in Lecture Notes In Computer Science, pages 181–192. Springer.

Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.

Frahm, J.-M., Fite-Georgel, P., Gallup, D., Johnson, T., Raguram, R., Wu, C., Jen, Y.-H., Dunn, E., Clipp, B., Lazebnik, S., et al. (2010). Building Rome on a cloudless day. In *Proceedings of the IEEE European Conference on Computer Vision (ECCV)*, pages 368–381. Springer.

Furukawa, Y. and Ponce, J. (2010). Accurate, dense, and robust multiview stereopsis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(8):1362–1376.

Fusiello, A. and Irsara, L. (2008). Quasi-Euclidean uncalibrated epipolar rectification. In *Proceedings of the 19th International Conference on Pattern Recognition (ICPR)*, pages 1–4. IEEE.

Géraud, T., Carlinet, E., Crozet, S., and Najman, L. (2013). A quasi-linear algorithm to compute the tree of shapes of nD images. In *Mathematical Morphology and Its Applications to Signal and Image Processing*, pages 98–110. Springer.

Govindu, V. M. (2001). Combining two-view constraints for motion estimation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 2, pages 218–225. IEEE.

Grompone von Gioi, R., Monasse, P., Morel, J.-M., and Tang, Z. (2010). Towards high-precision lens distortion correction. In *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP)*, pages 4237–4240. IEEE.

Grompone von Gioi, R., Monasse, P., Morel, J.-M., and Tang, Z. (2011). Lens distortion correction with a calibration harp. In *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP)*, pages 617–620. IEEE.

Haralick, R. M., Joo, H., Lee, C.-N., Zhuang, X., Vaidya, V. G., and Kim, M. B. (1989). Pose estimation from corresponding point data. *IEEE Transactions on Systems, Man and Cybernetics*, 19(6):1426–1446.

Hartley, R. and Zisserman, A. (2000). *Multiple View Geometry in Computer Vision*. Cambridge University Press.

Hartley, R. I. (1994). Euclidean reconstruction from uncalibrated views. In *Applications of Invariance in Computer Vision*, pages 235–256. Springer.

Hartley, R. I. (1997a). In defense of the eight-point algorithm. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(6):580–593.

Hartley, R. I. (1997b). Lines and points in three views and the trifocal tensor. *International Journal of Computer Vision*, 22(2):125–140.

Hartley, R. I. (1999). Theory and practice of projective rectification. *International Journal of Computer Vision*, 35(2):115–127.

Heijmans, H. J. and Keshet, R. (2002). Inf-semilattice approach to self-dual morphology. *Journal of Mathematical Imaging and Vision*, 17(1):55–80.

Kim, C., Zimmer, H., Pritch, Y., Sorkine-Hornung, A., and Gross, M. (2013). Scene reconstruction from high spatio-angular resolution light fields. In *ACM Transactions on graphics (proceedings of SIGGRAPH)*.

Kolmogorov, V., Monasse, P., and Tan, P. (2013). Kolmogorov and Zabih's graph cuts stereo matching algorithm. Preprint Image Processing On Line (IPOL) `http://www.ipol.im/pub/pre/97/`.

Kolmogorov, V. and Zabih, R. (2001). Computing visual correspondence with occlusions using graph cuts. In *Proceedings of the 8th International Conference on Computer Vision (ICCV)*, volume 2, pages 508–515. IEEE.

Kronrod, A. (1950). On functions of two variables. *Uspehi Mathematical Sciences (NS)*, 35(5):24–134.

Kweon, I. S. and Kanade, T. (1994). Extracting topographic terrain features from elevation maps. *CVGIP: Image Understanding*, 59(2):171–182.

Lavest, J.-M., Viala, M., and Dhome, M. (1998). Do we really need an accurate calibration pattern to achieve a reliable camera calibration? In *Proceedings of the IEEE European Conference on Computer Vision (ECCV)*, pages 158–174. Springer.

Lazebnik, S., Boyer, E., and Ponce, J. (2001). On computing exact visual hulls of solids bounded by smooth surfaces. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 151–161. IEEE.

Li, H. and Hartley, R. (2006). Five-point motion estimation made easy. In *Proceedings of the 18th International Conference on Pattern Recognition (ICPR)*, volume 1, pages 630–633. IEEE.

Longuet-Higgins, H. (1981). A computer algorithm for reconstructing a scene from two projections. *Nature*, 293:133–135.

Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2):91–110.

Luong, Q.-T. and Faugeras, O. D. (1996). The fundamental matrix: Theory, algorithms, and stability analysis. *International Journal of Computer Vision*, 17(1):43–75.

Ma, Y., Huang, K., Vidal, R., Košecká, J., and Sastry, S. (2004). Rank conditions on the multiple-view matrix. *International Journal of Computer Vision*, 59(2):115–137.

Martinec, D. and Pajdla, T. (2007). Robust rotation and translation estimation in multiview reconstruction. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE.

Matas, J., Chum, O., Urban, M., and Pajdla, T. (2004). Robust wide-baseline stereo from maximally stable extremal regions. *Image and Vision Computing*, 22(10):761–767.

Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., and Van Gool, L. (2005). A comparison of affine region detectors. *International Journal of Computer Vision*, 65(1-2):43–72.

Milnor, J. W. (1963). *Morse Theory*. Number 51 in Annals of Mathematics Studies. Princeton University Press.

Moisan, L. (1998). Affine plane curve evolution: A fully consistent scheme. *IEEE Transactions on Image Processing*, 7(3):411–420.

Moisan, L., Moulon, P., and Monasse, P. (2012). Automatic homographic registration of a pair of images, with a contrario elimination of outliers. *Image Processing On Line (IPOL)*, 2012. `http://dx.doi.org/10.5201/ipol.2012.mmm-oh`.

Mokhtarian, F. and Mackworth, A. K. (1992). A theory of multiscale, curvature-based shape representation for planar curves. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(8):789–805.

Monasse, P. (1999). Contrast invariant registration of images. In *Proceedings of IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, volume 6, pages 3221–3224. IEEE.

Monasse, P. (2000). *Contrast invariant representation of digital images and application to registration*. PhD thesis, Université Paris IX-Dauphine.

Monasse, P. (2011). Quasi-Euclidean epipolar rectification. *Image Processing On Line (IPOL)*, 2011. `http://dx.doi.org/10.5201/ipol.2011.m_qer`.

Monasse, P. and Guichard, F. (2000). Fast computation of a contrast-invariant image representation. *IEEE Transactions on Image Processing*, 9(5):860–872.

Monasse, P., Morel, J.-M., and Tang, Z. (2010). Three-step image rectification. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 89.1–89.10. BMVA Press.

Monasse, P., Rudin, L., and Cao, F. (2007). Super-dense digital terrain elevation reconstruction through method of epipolar characteristics. In *Proceedings of the Annual Conference of the American Society of Photogrammetry and Remote Sensing (ASPRS)*, volume 2, pages 429–439. Curran Associates, Inc.

Mondelli, M. and Ciomaga, A. (2011). Finite difference schemes for MCM and AMSS. *Image Processing On Line (IPOL)*, 2011. `http://dx.doi.org/10.5201/ipol.2011.cm_fds`.

Moulon, P., Monasse, P., et al. (2012). Unordered feature tracking made fast and easy. In *European Conference on Visual Media Production (CVMP)*.

Moulon, P., Monasse, P., and Marlet, R. (2013a). Adaptive structure from motion with a contrario model estimation. In *Proceedings of the IEEE Asian Conference on Computer Vision (ACCV)*, pages 257–270. Springer.

Moulon, P., Monasse, P., and Marlet, R. (2013b). Global fusion of relative motions for robust, accurate and scalable structure from motion. In *Proceedings of the International Conference on Computer Vision (ICCV)*, to appear. IEEE.

Najman, L. and Couprie, M. (2006). Building the component tree in quasi-linear time. *IEEE Transactions on Image Processing*, 15(11):3531–3539.

Najman, L. and Géraud, T. (2013). Discrete set-valued continuity and interpolation. In *Mathematical Morphology and Its Applications to Signal and Image Processing*, pages 37–48. Springer.

Nistér, D. (2004). An efficient solution to the five-point relative pose problem. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 26(6):756–770.

Nistér, D. and Stewénius, H. (2008). Linear time maximally stable extremal regions. In *Proceedings of the IEEE European Conference on Computer Vision (ECCV)*, pages 183–196. Springer.

Olsson, C. and Enqvist, O. (2011). Stable structure from motion for unordered image collections. In *Image Analysis*, pages 524–535. Springer.

Pierrot-Deseilligny, M. and Cléry, I. (2011). APERO, an open source bundle adjustment software for automatic calibration and orientation of set of images. In *Proceedings of the ISPRS Commission V Symposium, Image Engineering and Vision Metrology, Trento, Italy*, pages 2–4.

Reeb, G. (1946). Sur les points singuliers d'une forme de Pfaff complètement intégrable ou d'une fonction numérique. *Comptes Rendus de L'Académie ses Sciences, Paris*, 222:847–849.

Rhemann, C., Hosni, A., Bleyer, M., Rother, C., and Gelautz, M. (2011). Fast cost-volume filtering for visual correspondence and beyond. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3017–3024.

Rudakova, V. and Monasse, P. (2014). Precise correction of lateral chromatic aberration in images. In Klette, R., Rivera, M., and Satoh, S., editors, *Proceedings of the 6th Pacific-Rim Symposium on Image and Video Technology (PSIVT 2013)*, volume 8333 of *Lecture Notes in Computer Science*, pages 12–22. Springer.

Rudin, L. I., Morel, J.-M., Monasse, P., and Cao, F. (2011). System and method for three-dimensional estimation based on image data. Patent. US 8014588.

Salembier, P., Oliveras, A., and Garrido, L. (1998). Antiextensive connected operators for image and sequence processing. *IEEE Transactions on Image Processing*, 7(4):555–570.

Salembier, P. and Serra, J. (1995). Flat zones filtering, connected operators, and filters by reconstruction. *IEEE Transactions on Image Processing*, 4(8):1153–1160.

Scharstein, D. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International Journal of Computer Vision*, 47(1-3):7–42.

Sim, K. and Hartley, R. (2006). Recovering camera motion using linfty minimization. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, volume 1, pages 1230–1237. IEEE.

Snavely, N. et al. (2010). Bundler: Structure from motion (sfm) for unordered image collections. *Code available at: http://phototour. cs. washington. edu/bundler*.

Song, Y. (2007). A topdown algorithm for computation of level line trees. *IEEE Transactions on Image Processing*, 16(8):2107–2116.

Strecha, C., von Hansen, W., Van Gool, L., Fua, P., and Thoennessen, U. (2008). On benchmarking camera calibration and multi-view stereo for high resolution imagery. In *Proceedings of the Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1–8. IEEE.

Sturm, P. and Triggs, B. (1996). A factorization based algorithm for multi-image projective structure and motion. In *Proceedings of the IEEE European Conference on Computer Vision (ECCV)*, pages 709–720. Springer.

Tang, Z. (2011). *Calibration de caméra à haute précision*. PhD thesis, CMLA, ENS Cachan.

Tang, Z., Grompone von Gioi, R., Monasse, P., and Morel, J.-M. (2012). High-precision camera distortion measurements with a "calibration harp". *Journal of the Optical Society of America A*, 29(10):2134–2143.

Tarjan, R. E. (1975). Efficiency of a good but not linear set union algorithm. *Journal of the ACM*, 22(2):215–225.

Triggs, B. (1995). Matching constraints and the joint image. In *Proceedings of the 5th International Conference on Computer Vision (ICCV)*, pages 338–343. IEEE.

Triggs, B., McLauchlan, P. F., Hartley, R. I., and Fitzgibbon, A. W. (2000). Bundle adjustment—a modern synthesis. In *Vision Algorithms: Theory and Practice*, pages 298–372. Springer.

van Kreveld, M., van Oostrum, R., Bajaj, V., Pascucci, V., and Schikore, D. (1997). Contour trees and small seed sets for isosurface traversal. In *Proceedings of the 13th Symposium on Computational Geometry*, pages 212–220. ACM.

Vincent, L. (1993). Grayscale area openings and closings, their efficient implementation and applications. In *First Workshop on Mathematical Morphology and its Applications to Signal Processing*, pages 22–27.

Vu, H. H., Keriven, R., Labatut, P., and Pons, J.-P. (2009). Towards high-resolution large-scale multi-view stereo. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1430–1437. IEEE.

Xu, Y., Monasse, P., Najman, L., and Géraud, T. (2013). Tree-based Morse regions: a topological approach to local feature detection. Preprint.

Yoon, K.-J. and Kweon, S. (2006). Adaptive support-weight approach for correspondence search. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(4):650–656.

Zach, C., Klopschitz, M., and Pollefeys, M. (2010). Disambiguating visual relations using loop constraints. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1426–1433. IEEE.

Zhang, Z. (2000). A flexible new technique for camera calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(11):1330–1334.

Detailed Résumé

Pascal Monasse, PhD

Contact information

IMAGINE, École des Ponts ParisTech
6-8, avenue Blaise Pascal – Cité Descartes
Champs-sur-Marne
77455 Marne-la-Vallée cedex 2, France

Phone: (+33/0) 1 64 15 21 76
Fax: (+33/0) 1 64 15 21 86

E-mail: monasse@imagine.enpc.fr
Web: `http://imagine.enpc.fr/~monasse/`

Education

PhD in Mathematics
University Paris IX, September 1996 - June 2000

Master degree of mathematics (numerical analysis)
University Paris VI, Septembre 1994 - June 1995

Engineer degree
École Nationale des Ponts et Chaussées, Septembre 1991 - June 1995

Professional experience

[Sept. 2008 - Present] Research scientist, IMAGINE/LIGM, École des Ponts Paris-Tech, Université Paris-Est, France
- Scientific adviser to three PhD students in computer vision.
- Work on computer vision, all aspects of 3-D stereo reconstruction
- Coordinator of ANR grant (ANR-09-CORD-003), project Callisto (326 k€)

[Sept. 2007 - Aug. 2008] Research scientist, CMLA, Ecole Normale Supérieure de Cachan, France
- Scientic advisor to two PhD students in image processing

- Participant of CNES project MISS, *high resolution satellite stereo imaging concerning all aspects of 3D reconstruction*

[Sept. 2001 - July 2007] Junior researcher (until 2004) then Senior researcher (leader of a team of three PhD researchers) at Cognitech Inc., Pasadena, California

- *Accurate, georeferenced topographic mosaic reconstruction from airborne video*: Principal Investigator with Dr. Lenny Rudin for this DARPA contract (GeoSpatial Representation and Analysis), phase I and II completed (contract of 3 M$).
- Technical leader for development of *Cognitech GeoMeasure software, allowing automatic construction of planar mosaics from high-altitude airborne video (with optional use of GPS and inertial data)*, leadership of three programmers. This software was developped for the National Geospatial intelligence Agency (NGA), the US federal mapping agency, responsible among other tasks for technical maintenance of GPS (contract of 200 k$).
- *Localization and tracking of airborne video in large database image*: scientific leader for this contract (400 k$) for the US Navy (Naval Airfare Weapons Center, China Lake).
- *Real-time photogrammetric method for landing of UAV from video*, DARPA contract reserved for small business, phases I and II completed (contract of 1 M$).
- *Photogrammetric 3-D reconstruction from stereo pairs*, US Navy contract (Office of Naval Research).
- Internal research for *Cognitech VideoInvestigator* software in image and video registration, photogrammetry, camera calibration, detection... This software is used by FBI, Scotland Yard, French Gendarmerie Nationale, the Hong Kong police, LAPD, NASA, US Navy, National Geospatial intelligence Agency...

[July 2000 - Aug. 2001] Research scientist at CMLA, ENS Cachan, France

- Development of a fast algorithm for extraction of continuous level lines in an image.
- Research in motion estimation in video with techniques of maximal flows in graphs.

[Sept. 1995 – June 1996] Military service as scientist at Matra Cap Systèmes, France. Development of a software kernel for image processing applications.
[Apr. 1995 – Sept. 1995] MSc internship at Laboratoire de Météorologie Dynamique, ENS Paris, France. Construction of wavelet bases on the interval with prescribed boundary conditions for solving PDE.
[Apr. 1990 - June 1990] Engineering internship class at Credome, Paris (group Publicis) on statistical optimization of commercials.

Research interests

- Computer vision
- Image processing

Prizes and awards

- Best student short paper for Pierre Moulon at the 10th European Conference on Visual Media Producation (2013).
- Winner with IMAGINE group of ProVisG Mars 3D Challenge (2011), "testing and improving the state of the art in visual odometry and 3D terrain reconstruction in planetary exploration", `http://www.provisg.eu/news/PRoVisG-Mars-3D-Challenge`
- Cognitech's PixL2GPS software (for dense GPS tagging of video frames) is the winner in the Aerospace, Defense and Security category for the American Technology Awards "The Termans" (2010) delivered by the TechAmerica Foundation. I was the scientific leader and main developer in this software product.
- "Top performer" at DARPA workshop (2007) ahead of teams of UCLA, Rensslaer Polytechnic Institute, University of Santa Barbara, University of North Carolina, University of South Carolina...
- Best paper award (1999) at the Second International Conference on Scale-Space Theories in Computer Vision, Corfu, Greece

Scientific publications

Book

1. Caselles, V. and Monasse, P. (2010). *Geometric description of images as topographic maps.* Number 1984 in Lecture Notes in Mathematics. Springer.

Book chapters

2. Dibos, F., Koepfler, G., and Monasse, P. (2003). *Total variation minimization for scalar/vector regularization.* In *Geometric Level Set Methods in Imaging, Vision and Graphics.* S. Osher and N. Paragios, eds., pages 121–140, Springer
3. Dibos, F., Koepfler, G., and Monasse, P. (2003). *Image Alignment.* In *Geometric Level Set Methods in Imaging, Vision and Graphics.* S. Osher and N. Paragios, eds., pages 271–295, Springer

Patents

4. Rudin, L.I., Musé, P., and Monasse, P. (2012). *System and method for pattern detection and camera calibration.* USPTO patent #8,106,968
5. Rudin, L.I., Morel, J.M., Monasse, P., and Cao, F. (2011). *System and method for three-dimensional estimation based on image data.* USPTO patent #8,014,588
6. Rudin, L.I., Lisani, J.L., Monasse, P., and Morel, J.M. (2009). *Object recognition based on 2D images and 3D models.* USPTO patent #7,587,082

Peer-reviewed journals

7. Tang, Z., Grompone von Gioi, R., Monasse, P., and Morel, J.-M. (2012). High-precision camera distortion measurements with a "calibration harp". *Journal of the Optical Society of America A*, 29(10):2134–2143
8. Moisan, L., Moulon, P., and Monasse, P. (2012). Automatic homographic registration of a pair of images, with a contrario elimination of outliers. *Image Processing On Line* DOI: http://dx.doi.org/10.5201/ipol.2012.mmm-oh
9. Ciomaga, A., Monasse, P., and Morel, J.-M. (2011). Image visualization and restoration by curvature motions. *Multiscale Modeling and Simulation*, 9:834.
10. Monasse, P. (2011). Quasi-Euclidean epipolar rectification. *Image Processing On Line* DOI: http://dx.doi.org/10.5201/ipol.2011.m_qer
11. Caselles, V., Meinhardt, E., and Monasse, P. (2007). Constructing the tree of shapes of an image by fusion of the trees of connected components of upper and lower level sets. *Positivity*, Birkhäuser, 12(1):55–73
12. Ballester, C., Caselles, V., and Monasse, P. (2003) The Tree of Shapes of an Image. *ESAIM: Control, Optimisation and Calculus and Variations*, 9:1–18
13. Caselles, V., and Monasse, P. (2003). Grain filters. *Journal of Mathematical Imaging and Vision*, 17(3):249–270
14. Monasse, P., and Guichard, F. (2000). Scale-space from a level lines tree. *Journal of Visual Communication and Image Representation*, 11:224–236
15. Monasse, P., and Guichard, F. (2000). Fast computation of a contrast invariant image representation. *IEEE Transactions on Image Processing*, 9(5):860–872
16. Monasse, P., and Perrier, V. (1998). Orthonormal wavelet bases adapted for partial differential equations with boundary conditions. *SIAM Journal of Mathematical Analysis*, 29(4):1040–1065

Proceedings in refereed conferences

17. Espuny, F., Monasse, P., and Moisan, L. (2014). A new a contrario approach for the robust determination of the fundamental matrix. In Huang, F. and Sugimoto, A., editors, *Proceedings of the PSIVT 2013 Workshop on Geometric Computation for Computer Vision (GCCV)*, number 8334 in Lecture Notes In Computer Science, pages 181–192. Springer.
18. Rudakova, V. and Monasse, P. (2014). Precise correction of lateral chromatic aberration in images. In Klette, R., Rivera, M., and Satoh, S., editors, *Proceedings of the 6th Pacific-Rim Symposium on Image and Video Technology (PSIVT 2013)*, volume 8333 of *Lecture Notes in Computer Science*, pages 12–22. Springer.
19. Espuny, F. and Monasse, P. (2014). Singular vector methods for fundamental matrix computation. In Klette, R., Rivera, M., and Satoh, S., editors, *Proceedings of the 6th Pacific-Rim Symposium on Image and*

Video Technology (PSIVT 2013), volume 8333 of *Lecture Notes in Computer Science*, pages 290–301. Springer.

20. Moulon, P., Monasse, P., and Marlet, R. (2013). Global fusion of relative motions for robust, accurate and scalable structure from motion. In *Proceedings of the International Conference on Computer Vision (ICCV)*, to appear.
21. Moulon, P., Monasse, P., and Marlet, R. (2012). Adaptive structure from motion with *a contrario* model estimation. In *Proceedings of the IEEE Asian Conference on Computer Vision (ACCV)*, pages 257–270. Springer.
22. Grompone von Gioi, R., Monasse, P., Morel, J.-M., and Tang, Z. (2011). Lens distortion correction with a calibration harp. In *Proceedings of the 18th IEEE International Conference on Image Processing (ICIP)*, pages 617–620. IEEE.
23. Grompone von Gioi, R., Monasse, P., Morel, J.-M., and Tang, Z. (2010). Towards high-precision lens distortion correction. In *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP)*, pages 4237–4240. IEEE.
24. Ciomaga, A., Monasse, P., and Morel, J.-M. (2010). Level lines shortening yields an image curvature microscope. In *Proceedings of the 17th IEEE International Conference on Image Processing (ICIP)*, pages 4129–4132. IEEE.
25. Monasse, P., Morel, J.-M., and Tang, Z. (2010). Three-step image rectification. In *Proceedings of the British Machine Vision Conference (BMVC)*, pages 89.1–89.10. BMVA Press.
26. Aganj, E. and Monasse, P. (2009). Multi-view texturing of imprecise mesh. In *Proceedings of Asian Conference on Computer Vision (ACCV)*, pages 468–476. Springer.
27. Courchay, J., Keriven, R., Monasse, P., and Pons, J.P. (2009). Dense and accurate spatio-temporal multi-view stereovision. In *Proceedings of Asian Conference on Computer Vision (ACCV)*, pages 11–22. Springer.
28. Monasse, P. and Rudin, L., and Cao, F. (2007). Super-dense digital terrain elevation reconstruction through method of epipolar characteristics. In *Proceedings of Annual Conference of the American Society of Photogrammetry and Remote Sensing (ASPRS)*, volume 2, pages 429–439. Curran Associates, Inc.
29. Rudin, L., Monasse, P., and Yu, P. (2005). Geometrical methods for accurate forensic videogrammetry, part II. Reducing complexity of Cartesian scene measurements via epipolar registration. In *SPIE Conference on Image and Video Communications and Processing*, volume 3, pages 272–283. International Society for Optics and Photonics.
30. Rudin, L., Monasse, P., and Yu, P. (2005). Epipolar photogrammetry: a novel method for forensic image comparison and measurement. In *Proceedings of IEEE International Conference on Image Processing (ICIP)*, volume 3, pages 385–388.
31. Lisani, J.L., Rudin, L., Monasse, P., Morel, J.M., and Yu, P. (2005). Meaningful automatic video demultiplexing with unknown number of cameras, contrast changes, and motion. In *Proceedings of IEEE Conference*

on Advanced Video and Signal Based Surveillance (AVSS), pages 604–608. IEEE.

32. Lisani, J.L., Moisan, L., Monasse, P., and Morel, J.-M. (2000). Affine invariant mathematical morphology applied to a generic shape recognition algorithm. In *Proceedings of International Symposium of Mathematical Morphology (ISMM)*, volume 18, pages 91–98. Springer.
33. Monasse, P. (1999). Contrast invariant registration of images. In *Proceedings of International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, volume 6, pages 3221–3224. IEEE.

Other publications

34. Moulon, P., Duisit, B., and Monasse, P. (2013). Global multiple view color consistency. *The 10th European Conference on Visual Media Production (CVMP)*.
35. Moulon, P. and Monasse, P. (2012). Unordered feature tracking made fast and easy. *The 9th European Conference on Visual Media Production (CVMP)*.
36. Monasse, P. and Perrier, V. (1995). Ondelettes sur l'intervalle pour la prise en compte de conditions aux limites. In *Comptes Rendus de l'Académie des Sciences*, Paris, I(312), pages 405-410.

Invited talks

37. *"Geometric description of images as topographic maps"*, CEREMADE seminar, University Paris Dauphine, July 2013
38. *"Analyse topologique et visualisation des ensembles de niveau d'image"*, colloquium Mathématiques pour l'image, Orléans, France, June 2012
39. *"A geometric scheme for affine shortening and application to image curvature microscope"*, workshop Geometric flows in finite or infinite dimension, CIRM, Luminy, France, March 2011
40. *"Les images vues comme cartes topographiques"*, seminary Méthodes mathématiques du traitement d'images, University Pierre et Marie Curie, November 2010
41. *"Les mathématiques dans le traitement d'image et la vision tridimensionnelle"*, Maths Club, University Paris Diderot, April 2010
42. *"Représentation auto-duale d'une image en lignes de niveau"*, colloquium Morphologie mathématique: structures et connexions, in honor of Jean Serra's 70*th* anniversary, ESIEE, Marne-la-Vallée, France, April 2010
43. *"Filtrage des images par mouvements par courbure et filtrages non réguliers*, colloquium Image, EDP et Géométrie, Institut Fourier, Grenoble, France, December 2009
44. *"Geometric Description of Images as Topographic Maps"*, Lecture Series in Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, Beijing, September 2009

45. *"Advances in algorithms and software tools for automatic geo-mosaic/geo-TIFF creation from video sensors"*, closed DARPA workshop on Geospatial Representation and Analysis, Arlington, Virginia, October 2007

46. *"Automatic 2-D and 3-D Georeferencing and Mosaicing from Video"*, closed DARPA workshop on Geospatial Representation and Analysis, Coeur d'Alene, Idaho, April 2007

47. *"Automatic 2-D and 3-D Georeferencing and Mosaicing from Video"*, closed DARPA workshop on Geospatial Representation and Analysis, Boothbay Harbor, Maine, August 2006

48. *Remote video-sensing of 3-D geometry through method of epipolar characteristics with cognitech's sky-scanner*, closed DARPA workshop on Geospatial Representation and Analysis, Snoqualmie Falls, Washington, April 2006

49. *Computation theory for dense meaningful topographic mosaics"*, closed DARPA workshop on Geospatial Representation and Analysis, San Diego, California, November 2005

50. Closed DARPA workshop on Geospatial Representation and Analysis, Savannah, Georgia, March 2005

51. Closed DARPA workshop at Office of Naval Research (US Navy), Washington, D.C., March 2003

52. *"Représentation morphologique d'images numériques et applications"*, seminary Algorithmique et Programmation, CIRM, Luminy, France, May 2001

53. *"Morphological representation of images and application to registration"*, lecture at School on Mathematical Problems in Image Processing, International Center for Theoretical Physics, Trieste, Italy, 2000

Service to the scientific community

Editorial activities

– Associate editor for the journal Image Processing On Line

Journal reviewer

– IEEE Transactions on Pattern Analysis and Machine Intelligence
– SIAM Multiscale Modeling and Simulation
– IEEE Transactions on Image Processing
– Journal of Mathematical Imaging and Vision
– Elsevier Image and Vision Computing
– Image Processing On Line
– Computer Vision and Image Understanding
– Journal of Graphics Tools
– Journal of Selected Topics in Signal Processing
– IET Computer Vision
– GRETSI Traitement du Signal

Program committee at conferences

- IEEE International Conference on Computer Vision (2013)
- IEEE Conference on Computer Vision and Pattern Recognition (2013)
- IEEE Asian Conference on Computer Vision (2012)
- IEEE European Conference on Computer Vision (2012) workshop on "Unsolved Problems in Optical Flow and Stereo Estimation"
- IEEE International Conference on Computer Vision (2011)
- IEEE European Conference on Computer Vision (2010)

PhD advisor

- Zhongwei Tang (2007-2011): co-direction with Jean-Michel Morel (50%)
- Pierre Moulon (2010-2013): co-direction with Renaud Marlet (50%)
- Victoria Rudakova (2010-2013)
- Zhe Liu (2011-): co-direction with Renaud Marlet (50%)
- Pauline Tan (2013-): co-direction with Antonin Chambolle (50%)
- Yohann Salaun (2013-): co-direction with Renaud Marlet (50%)
- Bruno Conejo (2013-): co-direction avec Jean-Philippe Avouac (50%)

PhD committees

- Eric Bughin "Vers une vectorisation automatique, précise et validée en stéréoscopie satellitaire en milieu urbain", ENS Cachan, France, October 26, 2011
- Gui-Song Xia "Some geometric methods for the analysis of images and textures", Telecom ParisTech, France, March 18, 2011
- Souleymane Kadri-Harouna "Ondelettes pour la prise en compte de conditions aux limites en turbulence incompressible", INPG, Grenoble, France, September 13, 2010
- Narut Soontranon "Appariement entre images de point de vue éloignés par utilisation de carte de profondeur", Université de Picardie Jules Verne, France, October 21, 2013
- Yongchao Xu, "Tree-based shape spaces for applications in image processing and computer vision", Université Paris Est, France, December 2013

Teaching

- "Algorithmics and Programmation", Oct. 2009 - Mar. 2013, coordinator and lecturer for a course on computer science for all 130 undergraduate engineering students at École des Ponts ParisTech (4×70h)
- "Algorithmics and Programmation", Oct. 2008 - Mar. 2009, lecturer for undergraduate engineering students at École des Ponts ParisTech (70h)
- "Modelling, Programmation and Simulation", Oct. 2008 - Feb. 2013, lectures on advanced C++ programming for undergraduate engineering students at École des Ponts ParisTech (5×12h)
- "Vision and 3D reconstruction", Oct. 2009 - Nov. 2012, lectures on computer vision for MSc Mathématiques, Vision, Apprentissage, Paris (4×15h)

- Oct. 1999, Tutorials of image processing for postgraduate students, Mathematisches Forschungsinstitüt, Oberwolfach, Germany.
- Sept. 1997 - June 1999: Tutorials on optimization, differential calculus and calculus of variations, to undergraduate engineering students at the Ecole Nationale Supérieure des Mines de Paris.
- Sept. 1996 - June 1997: Tutorials of mathematical analysis (University Paris 9) and linear algebra (University Paris 8) to junior undergraduate students.
- Sept. 1992 - May 1993: Oral training in mathematics in Mathématiques Spéciales M, lycée Louis-le-Grand, Paris.