Reconstruction spatio-temporelle de la ville de Reims
Appariement robuste d'anciennes cartes postales

Lara YOUNES, Barbara ROMANIUK, Éric BITTAR

Laboratoire CReSTIC-SIC
Département Informatique, IUT de RCC, rue des crayères BP 1035, 51687 Reims Cedex 2
Lara.Younes@univ-reims.fr,Barbara.Romaniuk@univ-reims.fr
Eric.Bittar@univ-reims.fr

Résumé – De nombreuses approches de reconstruction 3D des milieux urbains exploitent aujourd'hui la richesse qu'offrent les Systèmes d'Information Géographique (SIG). C'est dans ce contexte que nous proposons de localiser, reconstruire et visualiser en 3D la ville de Reims à partir de données anciennes telles que des cartes postales et des cadastres. Ces données imparfaites tant au niveau spatial que temporel témoignent de l'évolution de la ville sur la première moitié du XX^{ieme} siècle. Dans cet article nous nous focalisons sur une étape clé de notre projet qui est l'appariement d'images. Nous proposons ainsi une comparaison des performances de plusieurs couples détecteur-déscripteur dans le cadre d'une mise en correspondance partielle d'images.

Abstract – Nowadays, many 3D urban spaces reconstruction approaches exploit performance and flexibility offered by the Geographic Information Systems (GIS). It is in this context that we propose to locate, reconstruct and visualize in 3D the city of Reims (France) from old data such as postcards and cadastral surveys. This data, both spatially and temporally imperfect, reflect the evolution of the city in the first half of XX^{th} century. In this paper we focus on a crucial step in our project that is image matching. We thus propose a performance comparison of several state-of-the-art detector-descriptor couples.

1 Introduction

La reconstruction 3D et la géolocalisation de bâtiments à partir d'images est un domaine de recherche en plein essor, qui débouche sur de nombreuses applications comme la planification urbaine, l'archéologie, le tourisme virtuel ou encore la restauration. Depuis peu, les Système d'Information Géographique (SIG) permettent de rassembler des informations sur l'espace urbain de natures variées pour enrichir ces modèles 3D. La reconstruction 3D à partir d'importantes bases de données d'images prises au sol a été abordée dans de nombreux travaux récents. Agarwal et co. [1] ont étudié la reconstruction 3D de Rome à partir d'images trouvées sur Flickr. La base de données qu'ils gèrent est de très forte cardinalité, redondante, dense et non structurée. Les auteurs proposent une approche classique dans le domaine de la vision par ordinateur appelée *Structure from Motion*, dont la première étape consiste à détecter des caractéri- stiques robustes, reproductibles et distinctes dans les images. Une fois détectées elles sont mises en correspondance en appariant les images deux à deux selon un critère de similarité.

Dans notre cas, nous avons pour objectif de reconnaître, d'extraire et de reconstruire en 3D les bâtiments géolocalisés dans la ville de Reims à partir de données anciennes. Nous souhaitons proposer un outil permettant aux citoyens de s'approprier l'histoire de leur ville en leur proposant une navigation spatio-temporelle dans la ville de Reims utilisant à cet effet des SIG interactifs. Reims ayant joué un rôle prédominant dans l'histoire de la France, de nombreux documents attestent de l'évolution de la ville. Nous nous intéressons ici tout particulièrement aux cartes postales anciennes couvrant la période du début au milieu du vingtième siècle que nous associons aux plans cadastraux. La ville de Reims ayant été fortement endommagée pendant les deux guerres, les cartes postales témoignent de l'évolution de l'espace urbain en terme de destruction, restauration et reconstruction. Par ailleurs, ces images de faible résolution contiennent du texte, des timbres et des tampons. Reconstruire Reims à partir de ces données s'apparente donc à un vrai défi et l'originalité de notre approche réside dans l'utilisation de ces données éparses et imparfaites.

L'appariement d'images est un préalable incontournable à ce projet. Il s'agit d'estimer la transformation géométrique homographique au minimum ou projective si ceci est possible permettant de faire correspondre au mieux les informations visuelles communes aux images. Cette estimation est obtenue par le biais de l'identification des propriétés photométriques ou géométriques particulières (contours, coins, *blobs*, $\cdots$) dans l'image. Ces propriétés sont traduites par des descripteurs locaux. La mise en correspondance entre éléments visuels communs aux deux images est réalisée en comparant les descripteurs à l'aide de métriques adaptées. La démarche se trouve décomposée en trois principales étapes : détection des points d'intérêts ; calcul des descripteurs associés ; mise en correspondance des points d'intérêt entre les deux images à l'aide

de leurs descripteurs et estimation de la transformation géométrique par élimination des fausses correspondances. Dans la littérature il n'est pas d' usage de traiter ces trois étapes simultanément. Compte tenu de la complexité de ce sujet nous estimons important de nous intéresser aux trois étapes.

2 Processus d'évaluation

De nombreux travaux comparent les performances des couples détecteur-descripteur. Une manière de procéder est d'appliquer des transformations géométriques (en particulier des homographies) et photométriques contrôlées. En particulier, Mikolajczyk et co. [7] ont réalisé une étude générale de descripteurs de régions invariantes à l'échelle et aux transformations affines. Moreels et Perona [9] ont étendu l'évaluation des couples de détecteur-descripteurs aux scènes 3D. Ils ont comparé plusieurs couples dans le cadre de la mise en correspondance des caractéristiques d'objets 3D selon plusieurs points de vue et conditions d'illumination. Ils ont par ailleurs proposé une nouvelle approche automatique permettant d'estimer la vérité terrain pour des scènes 3D. D'autres auteurs ont comparé différents couples détecteur-descripteur dans le contexte de la reprise photographique historique [3] ou du suivi temps-réel [4].

Dans nos travaux antérieurs [10] nous nous sommes intéressés à la mise en place d'un processus d'évaluation des couples détecteur-descripteur étudiés par [7] dans le cadre de notre application. Ce processus consiste à mettre en correspondance des couples d'images sous trois conditions liées à la connaissance d'une vérité terrain.

Nous souhaitons aujourd'hui revenir sur le choix des détecteurs, des descripteurs et l'appariement proprement dit. Nous avons décidé de tester trois détecteurs récents dont la nature des caractéristiques recherchées diffère. En ce qui concerne le descripteur, nous comparons le SIFT à sa variante récente DAISY. Finalement, nous estimons les transformations géométri- ques existantes entre deux images en utilisant l'algorithme RANSAC et sa variante ORSA.

2.1 Détecteurs de points caractéristiques

Les détections et représentations des caractéristiques doivent présenter un certain nombre d'invariances, en terme de point de vue, de changement des conditions d'éclairage, d'occlusion et de déformations locales. Le caractère local des détecteurs assure la robustesse aux occlusions et aux déformations locales. C'est l'organisation relative de la majorité de ces points qui caractérise un objet. Ils doivent donc être assez nombreux et discriminants. Les invariances au point de vue et à l'illumination sont assurées par les détecteurs et les descripteurs. Un point d'intérêt peut être assimilé à un motif dans l'image qui diffère de son voisinage de par ses propriétés photométriques (intensité, couleur, texture), sa forme locale particulière ou encore la stabilité de sa segmentation. Les caractéristiques extraites sont classiquement des points, des lignes ou des régions. Nous avons souhaité tester trois détecteurs : FAST qui repose sur la détection de coins dans l'image, SIFT de *blobs* et le MSER de régions stables.

Le détecteur FAST [11] , conçu pour les applications temps réel, permet de détecter des coins dans les images. Un point est un coin lorsqu'un nombre n suffisant de points connexes situés sur un cercle de rayon fixé diffèrent en intensité du point central étudié. Les paramètres initiaux de cet algorithme sont fixés à un rayon de 3 pixels pour un cercle discret de 16 pixels et n à 12.

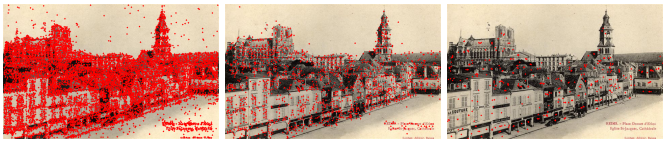

FIGURE 1 – Détection de caractéristiques avec trois détecteurs : FAST (gauche), SIFT (centre) et MSER (droite).

Le détecteur SIFT [5] permet de localiser des *blobs* dans l'image. C'est un détecteur invariant à l'échelle. Il construit une pyramide gaussienne, calcule une mesure d'intérêt spatiale normalisée (laplacien de gaussienne) pour chaque pixel et pour chaque niveau de la pyramide, puis détecte les extrema en x, y, σ dans cet espace. Les candidats de faible contraste sont éliminés, de même que les points positionnés sur des contours de faible courbure. Pour diminuer le coût algorithmique de cette approche, le laplacien de gaussienne est approximé par une différence de gaussiennes.

Le détecteur MSER [6] permet de localiser des régions stables dans l'image. Il repose sur l'idée que l'intensité varie rapidement aux bords des objets. Les régions considérées par ce détecteur sont les composantes connexes de l'image seuillée dont la surface varie peu en fonction de ce seuil. Leur détection est réalisée par une ligne de partage des eaux, au cours de laquelle on analyse la surface des bassins d'inondation. Ces régions sont invariantes aux transformations monotones d'intensité et aux transformations géométriques homographiques ou non-linéaires mais continues.

2.2 Descripteurs de points caractéristiques

Une description des points caractéristiques est nécessaire pour les discriminer dans le cadre de la mise en correspondance d'images. Dans la littérature de nombreuses méthodes de description ont été proposées et conduisent au calcul de vecteurs descripteurs pour chaque point. Le descripteur d'un point traduit la distribution des informations du voisinage centré sur le point caractéristique considéré. Ces descripteurs sont conçus de manière à être invariants aux petites déformations, aux erreurs de localisation, aux transformations rigides ou affines ainsi qu'aux changements d'illumination. Dans nos travaux antérieurs [10] nous avons comparé les performances de plusieurs descripteurs et avons conclu à la supériorité du descripteur SIFT. En conséquence, nous avons retenu dans cette étude le descripteur SIFT et ainsi que sa variante plus récente DAISY.

Le descripteur SIFT [5] décrit la distribution des gradients

dans le voisinage d'un point. L'échelle de détection du point est utilisée pour définir la taille du voisinage du point dans lequel le descripteur sera calculé. Elle pondère de manière adaptative l'influence du voisinage sur le descripteur en donnant moins d'importance aux pixels les plus lointains. L'invariance à la rotation est assurée par la rotation de la fenêtre de calcul carré du descripteur dans la direction de l'orientation dominante de gradient. Cette orientation est obtenue en analysant l'histogramme d'orientations des gradients. La fenêtre de calcul est subdivisée en sous-régions de taille 4×4. Un histogramme d'orientations des gradients (8 orientations à 45°) est ensuite calculé pour chacune de ces sous-régions. Les différents histogrammes sont stockés dans un vecteur descripteur de dimension $4 \times 4 \times 8 = 128$. Le vecteur descripteur est ensuite normalisé pour assurer une invariance aux changements d'illumination.

Le descripteur DAISY [12] est similaire au descripteur SIFT par son principe de calcul des histogrammes d'orientations. La différence essentielle réside dans la forme de la région dans laquelle est calculé le descripteur. Cette région est constituée de plusieurs cercles se chevauchant, centrée sur le point d'intérêt. Les rayons des cercles augmentent avec la distance au centre et la puissance du lissage Gaussien est proportionnelle aux rayons des cercles. Pour chaque région circulaire, un histogramme d'orientation est calculé et normalisé. Le vecteur descripteur d'un point d'intérêt est le résultat de la concaténation de tous les histogrammes d'orientations de la région, atteignant des dimensions importantes (544). Grâce à la forme de sa région de calcul, le descripteur DAISY est naturellement invariant aux rotations. Il présente par ailleurs l'intérêt de pouvoir être redéfini dans différentes directions sans qu'il soit nécessaire de calculer à nouveau les convolutions. Il atteint des performances supérieures aux autres au prix d'une grande dimensionnalité.

2.3 Appariement d'images

Le processus d'appariement consiste à comparer et faire correspondre entre eux les descripteurs extraits dans une image inconnue avec les descripteurs extraits dans une image cible. Cette comparaison repose sur le calcul d'une mesure de dissimilarité et un critère de sélection. Retenir une correspondance entre un nouveau descripteur avec un ou plusieurs descripteurs candidats revient donc à définir un seuil portant sur les distances de similarités. L'approche utilisée pour déterminer l'appariement entre deux descripteurs [5, 7] repose sur la définition empirique des seuils et exploite la notion de plus proche voisin spatial (distance euclidienne) pour établir les correspondances valides entre les descripteurs. Les distances de similarités aux deux plus proches voisins d'un descripeur sont calculées et comparées permettant ainsi de ne pas retenir les descripteurs qui ont de nombreux correspondants proches (ce qui rendrait la correspondance peu fiable). Parmis les correspondances ainsi retenues certaines sont de faux positifs. Afin de les éliminer une seconde phase est nécessaire. Elle consiste à estimer une transformation géométrique en détectant des groupes de correspondances cohérentes. Plusieurs familles de méthodes globales permettent d'éliminer les faux positifs parmi lesquels nous pouvons citer la transformée de Hough généralisée, l'appariement de graphes ou les approches basées sur l'algorithme RANSAC. Il est nécessaire d'étudier des groupes de 2 correspondances au minimum pour estimer une similitude, 3 correspondances pour une transformation affine, 4 correspondances pour une homographie et 7 correspondances dans le cas de la géométrie épipolaire. La complexité algorithmique de la mise en correspondance est étroitement liée à la transformation géométrique qui est considérée, les deux premières familles d'approches précitées se voient ainsi confrontées à leurs limites ce qui nous oriente naturellement vers le troisième type de méthodes.

L'algorithme RANSAC [2] permet de séparer efficacement les (*inliers*) des (*outliers*). Il échantillonne aléatoirement des sous-ensembles de correspondances pour prédire la transformation géométrique (homographie ou géométrie épipolaire). Les correspondances cohérentes avec ce modèle sont alors comptabilisées. L'ensemble des correspondances rassemblant le consensus maximal est retenu et forme les *inliers*. La stratégie adoptée dans l'algorithme RANSAC est rapide et robuste, mais nécessite le réglage de paramètres sensibles. Par ailleurs elle s'avère peu concluante lorsque le taux d'*inliers* est faible.

L'algorithme ORSA [8] permet d'apporter une solution efficace à ce problème. Nous avons donc retenu cette méthode. Elle permet de s'affranchir du réglage des paramètres. Elle repose sur le principe de détection *a contrario* qui consiste à détecter des groupements de descripteurs dont l'existence est très peu probable sous l'hypothèse que ces descripteurs sont indépendants les uns des autres.

3 Résultats

Dans cette partie nous comparons les cartes postales deux à deux en respectant deux contraintes fortes : les deux cartes doivent être différentes mais doivent impérativement se recouvrir partiellement.

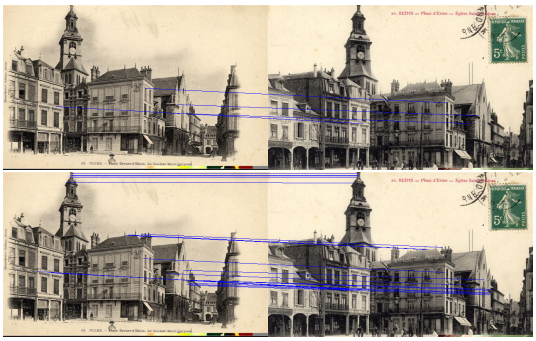

FIGURE 2 – Appariement utilisant la méthode ORSA. En haut le couple SIFT-DAISY, en bas le FAST-DAISY.

Les résultats obtenus sont illustrés par la figure 2. Le tableau 3 permet de constater que, indépendamment du descripteur, lorsque le détecteur MSER est utilisé, la phase d'appariement ne permet pas d'établir un nombre de correspondances

TABLE 1 – Evaluation des couples détecteur-descripteur sans filtrage de faux-positifs. Dans la colonne moyenne est reporté le nombre moyen de correspondances trouvées, dans les colonnes Min et Max leur nombre minimal et maximal.

Détecteur	Descripteur	Moyenne	Min	Max
MSER	SIFT	1.92	0	8
SIFT	SIFT	19.19	3	90
FAST	SIFT	5.65	0	20
MSER	DAISY	1.69	0	7
SIFT	DAISY	17.23	0	132
FAST	DAISY	20.5	0	189

suffisant. Par ailleurs le couple FAST-SIFT s'avère peu concluant.

Le tableau 2 permet de conclure à une nette supériorité du couple SIFT-SIFT qui s'avère être la méthode la plus robuste dans le cadre de notre application. Il est cependant intéressant de retenir les bonnes performances du couple FAST-DAISY lorsque le nombre initial de correspondances est suffisant pour atteindre l'étape du filtrage. Ces deux couples s'avèrent par ailleurs complémentaires. En effet, pour certaines combinaisons d'images le premier couple échoue alors que le second fournit de bons résultats.

TABLE 2 – Evaluation des couples détecteur-descripteur avec filtrage de faux-positifs par estimation de la transformation géométrique épipolaire. Dans la colonne Efficacité est reporté le pourcentage de couples d'images sur lesquels le nombre de correspondances initial a été suffisant pour estimer la matrice fondamentale (≥ 7), dans les colonnes Ransac et Orsa le nombre moyen de correspondances retenues après filtrage.

Détecteur	Descripteur	Efficacité	RANSAC	ORSA
SIFT	SIFT	73.08	19.89	14.42
FAST	SIFT	30.77	20.38	16.13
SIFT	DAISY	42.31	37.55	32.27
FAST	DAISY	46.15	42.58	37.58

L'appariement d'images est un point clé dans notre projet. Il intervient dans la construction des modèles 3D des bâtiments de la ville de Reims et permet d'interagir avec le SIG que nous souhaitons pouvoir enrichir par le biais d'une approche collaborative. Dans ce cas il peut permettre de géolocaliser les bâtiments présents sur une nouvelle carte postale grâce aux données présentes dans le SIG, de rendre plus dense le modèle 3D des bâtiments correspondant à une période donnée ou encore de conduire à la construction d'un nouveau modèle. L'appariement est effectué entre des cartes postales anciennes de faible résolution spatiale et de datation imprécise. Ceci peut avoir pour conséquence que nous ne disposons pas d'un nombre suffisant de correspondances pour pouvoir estimer une transformation géométrique épipolaire entre deux images. On peut dans ce cas là supposer que dans un milieu urbain dense un ensemble de façades alignées se trouve dans un même plan, ce qui permettrait de limiter l'estimation de la transformation géométrique à une homographique.

Références

[1] Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Brian Curless, Steven M. Seitz, and Richard Szeliski. Reconstructing Rome. *Computer*, 43 :40–47, 2010.

[2] Martin A. Fischler and Robert C. Bolles. Random sample consensus : a paradigm for model fitting with applications to image analysis and automated cartography. *Commun. ACM*, 24(6) :381–395, June 1981.

[3] Christopher Gat, Alexandra Branzan Albu, Daniel German, and Eric Higgs. A comparative evaluation of feature detectors on historic repeat photography. In *Proceedings of the 7th International Conference on Advances in Visual Computing - Volume Part II*, ISVC'11, pages 701–714, 2011.

[4] Steffen Gauglitz, Tobias Höllerer, and Matthew Turk. Evaluation of interest point detectors and feature descriptors for visual tracking. *International Journal of Computer Vision*, 94(3) :335–360, September 2011.

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

[6] J Matas, O Chum, U Martin, and T Pajdla. Robust wide baseline stereo from maximally stable extremal regions. In *British Machine Vision Conference*, volume 1, pages 384–393, London, 2002.

[7] Krystian Mikolajczyk and Cordelia Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 27(10) :1615–1630, 2005.

[8] Lionel Moisan and Bérenger Stival. A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. *Int. J. Comput. Vision*, 57(3) :201–218, May 2004.

[9] Pierre Moreels and Pietro Perona. Evaluation of features detectors and descriptors based on 3D objects. *International Journal of Computer Vision*, 73(3) :263–284, July 2007.

[10] Barbara Romaniuk, Lara Younes, and Eric Bittar. First steps toward spatio-temporal rheims reconstruction using old postcards. In *SITIS*, pages 374–380, 2012.

[11] Edward Rosten and Tom Drummond. Fusing points and lines for high performance tracking. In *IEEE International Conference on Computer Vision*, volume 2, pages 1508–1511, October 2005.

[12] Engin Tola, Vincent Lepetit, and Pascal Fua. Daisy : An efficient dense descriptor applied to wide-baseline stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(5) :815–830, 2010.