MVA2013 IAPR International Conference on Machine Vision Applications, May 20-23, 2013, Kyoto, JAPAN

Photometric Based 3D Registration

Narut Soontranon[1,2]
[1]MATIS, IGN
73 Avenue de Paris 94160
Saint Mandé, France
firstname.name@ign.fr

Marc Pierrot-Deseilligny
ENSG, IGN
73 Avenue de Paris 94160
Saint Mandé, France
firstname.name@ensg.eu

El Mustapha Mouaddib
[2]MIS, UPJV
33 Rue Saint Leu 80039
Amiens, France
mouaddib@u-picardi.fr

Abstract

Suppose we have two sets of 3D points with their associated 2D images from two distant positions (they can be taken by LIDAR, stereo cameras or multi-stereoscopic acquisition). Our work aims to improve the performance of matching between these 2D images in order to register the two sets of 3D points. We propose two approaches which use the 3D information in order to transform the original images and to apply the SIFT detector on the transformed images. The first one is consisted of two algorithms, to segment the planar regions and to transform the initial images into ortho-rectified images. The second one is based on conformal mapping. These approaches have been validated and compared with a method which apply SIFT on initial images. The comparison shows that proposed methods increase the number of homologous points and that the points are distributed all around common part of the images.

1 Introduction

Computation of homologous points from a pair of corresponding images (two different viewpoints) is one of the most studied problems in computer vision and photogrammetry. For example, it is the first mandatory step in all the photo-based 3D modeling pipeline [1, 11, 18, 19, 20, 21] that has emerged in the past few years.

Although the problem has received several satisfying solutions [4, 6] when the images are taken from closed viewpoints, it has been currently observed that, due to geometric deformation, the number of matches is decreased when the angles between the images are increased. Some recent work (e.g. [8]) tends to solve this problem using a combinatorial approach. Their results are generally satisfying for the number of matches but they still have drawbacks in terms of computation time and outliers that need to be filtered.

Nowadays, it becomes more and more current that for an image, we do not have only the photometric information but also some 3D information. Most current examples are given by modern LIDAR acquisition, stereo cameras or multi-stereoscopic acquisition where viewpoints are disconnected (see figure 1).

In this paper, we focus on the following problem: when, for two viewpoints, there exists both depth and image information, how can 3D information be used to improve the performance of matching between these images?

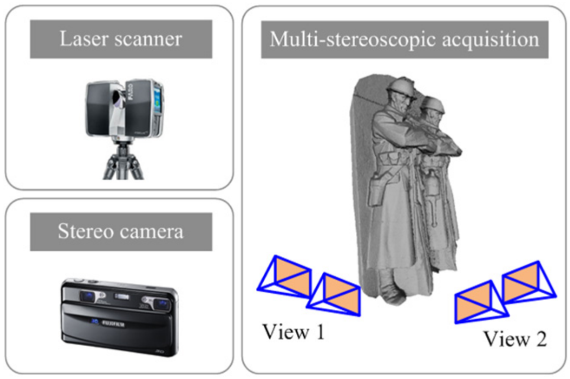


Figure 1. Acquisition devices used to acquire "3D+image": laser scanner, stereo camera and multi-stereoscopic acquisition.

2 Related work

Three types of problems and approaches can be distinguished in the literature for computation of homologous points from different views:

- Image to Image registration: For this case, there are only two images of the same scene, the research focuses on how radiometry can be used to give some invariance to geometric deformation. Due to the high dissemination of digital camera, the research has given the most attention by the community since several years ago.
- 3D to 3D registration: The matching is based exclusively on geometric primitives. This topic has become more popular since the emergence of laser scanner at the beginning of the year 2000.
- Mixed approach: Each viewpoint contains both radiometric and geometric information. The approach is generally to privilege the radiometric information, which is more discriminative, using the geometric information to correct the distortion due to difference in viewpoints.

In image to image registration, the process is generally composed of three steps: 1) Extraction of salient points like corner [16, 17] or maximum Laplacian [4, 5], extraction of salient regions is also a common alternative [6]. 2) Computation of local descriptors in the neighborhood of these points, the gradient based SIFT descriptor being the most currently used in stereo-vision problems. 3) Matching between the descriptors to detect the homologous points, where an algorithm [3] can efficiently match large dataset of descriptors.

To limit the problem of distortions due to wide baselines, much attention has been given to the invariance of description [7, 8].

In 3D to 3D registration, one of the most popular methods is the Iterative Closest Point method [14]. It is basic and efficient in general, but it requires some initialization, which can be solved using identification of common primitives like plane and some RANSAC strategy [15]. However, it requires that the scene contains at least three significant planes, which is not always satisfied.

For mixed approach, recent work [9, 10] extracts putative planes from the depth information and uses these planes to correct the perspective distortion. Another study [12], working on face recognition, applies conformal mapping to transform an initial image and uses global correlation between the developed images to improve the recognition performance.

3 Proposed approach

3.1 General invariance strategy

Suppose we have two images acquired with wide baselines and associated depth information (see figure 2). Due to state-of-the-art research on image matching, we now have very efficient tools to extract points and compute their descriptors that are invariant to translation, rotation and scaling. For each point of an image, if we know the relevant 3D surface, then we have all necessary information to compute locally an image that corresponds to a virtual acquisition done from a camera which is orthogonal to the surface. It will be called *"ortho-rectified image"*. Whatever maybe the surfaces and the geometry of acquisitions, for a given homologous pair, the two ortho-rectified images are identical up to a planar similitude (translation, rotation and scaling).

Our general strategy uses the depth information in the matching process by following steps:

1) Compute images that are locally ortho-rectified, and memorize the mapping from initial into ortho-rectified image.

2) Compute the matching between the ortho-rectified images.

3) Import the point matches into the initial geometry, using the inverse mapping.

The idea of this process is that the ortho-rectified image being identical up to a similitude and SIFT approach being invariant to scaling and rotation. We obtain the matching process that is completely invariant to geometric deformation.

In the following sections, we discuss in more details how this can be done. The first method is usable for planar scenes (3.2) and the second method is usable for smooth surface scenes (3.3).

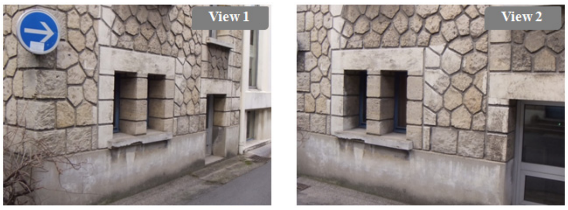
3.2 Planar scene

The first approach is applicable in specific scenes, like urban scenes, where 3D structure of the scene is essentially made from planes. The implemented process can be described as following:

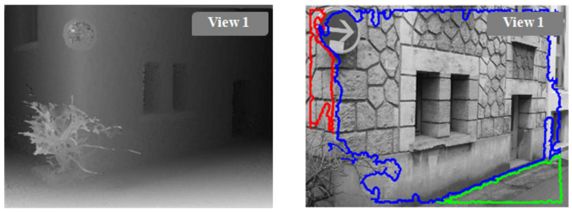
- Extract plane regions from the depth images. This is done by a standard method. We randomly generate square windows. Each window is a seed to compute a planar region by a region growing method. At the end, we iteratively select the region having the maximal area of point not already affected to an existing region.

- For each planar region, compute the ortho-rectified image corresponding to a virtual camera having its orthogonal axis to the plane (figure 3b). Then the homography matrix is computed between original and ortho-rectified images, which can be used to resample the original image.

- Perform the matching, using SIFT algorithm, between the ortho-rectified images and convert the point matches into the initial geometry (figure 3a and 3b).



a) Two corresponding images: "Façade" view 1 and 2.



b) Disparity map and segmenation regions of view 1.

Figure 2. An example of planar scene: a) two views, b) disparity map and segmentation results.

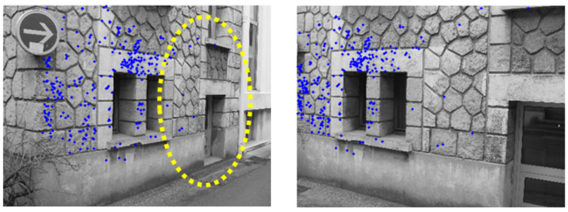
Figure 3 and table 1 present a comparison of the results obtained from standard SIFT and the ortho-rectified images. Table 1 shows that the number of good matches obtained from the ortho-rectified is better than the standard SIFT. However, the number of matches is not only an issue. Another advantage is that the point matches are distributed on the common part of the image (more homogeneous than the standard SIFT as shown in figure 3b), which is the point matches to use for registration in 3D adjustment.

Table 1. Comparison of #matches between original and ortho-rectified images (façade).

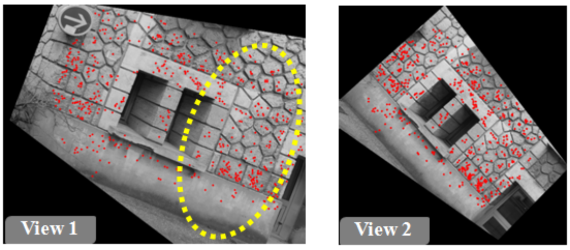
Façade	Standard SIFT	Ortho-rectified
#keypoints 1	25,629	12,190
#keypoints 2	22,678	4,942
#matches	338	600

3.3 Smooth surface scene

The previous approach gives satisfying results on 3D scenes made from planar surfaces, like urban or interior scenes. However, it is impossible to use it with more



a) SIFT match on original images.



b) SIFT match on ortho-rectified (blue region of figure 2b).

Figure 3. Comparison of point matches between a) original and b) ortho-rectified images (improved matches in dashed line).

complicated surfaces like human faces or sculptures. To generalize this approach with any smooth surface, without multiplying the number of rectification planes, we need to map the 3D surface on a unique plane with a single one to one mapping. To preserve the invariance between two viewpoints, the Jacobian of this mapping must be a similitude in every point (i.e. conserves the angle). The theory of conformal mapping is perfectly suited for this goal [13]. Let's recall some classical points of conformal mapping:

- They have been used for a long time by cartographer to map the sphere on the plane while preserving the angles (well known conformal projections are Lambert, Mercator etc.)
- From a formal point of view, the Riemann's uniformization theorem tells us that any 2D regular manifold can be mapped by a conformal mapping to some part of the plane.
- From a computational point of view, for a 3D surface given by a mesh, there exist efficient linear algorithms [13] to compute the conformal mapping that send the 3D mesh in the plane.
- From a practical point of view, the least square conformal mapping algorithm can be found in several libraries. In our work, we have used a version of the CGAL library [2].

The pipeline, we implemented for this case, is similar to the planar case:

- Compute a smooth 3D mesh (the smoothing is done by image low frequency filtering on the depth map).
- Use the least square conformal map of CGAL library to compute a planar triangulation which is the conformal development of the 3D mesh.
- Use the conformal map to resample the image, so named *"conformal image"*. Figure 5 shows the conformal images corresponding to the input data in figure 4.
- Use SIFT to match on the conformal images.
- Use the inverse conformal mapping to transform point matches into initial geometry.

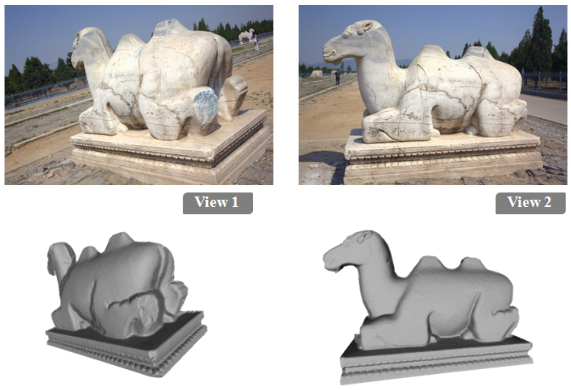


Figure 4. An example of input data to our process: two images and their associated 3D models.

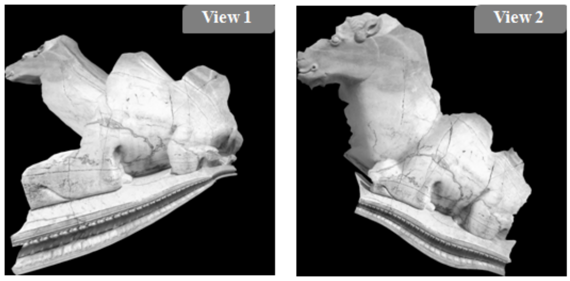


Figure 5. Conformal images of figure 4.

Figure 6 presents a realistic example for using our method on a camel sculpture. It can be seen that with conformal mapping, detected points are distributed on a much larger area.

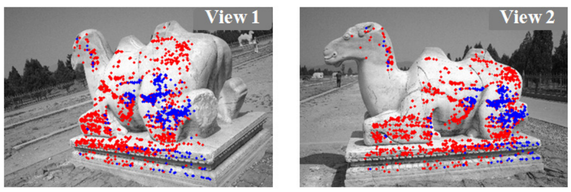


Figure 6. Results: blue-the standard SIFT matching, red-SIFT matching on conformal images (and converted into initial geometry).

Table 2 presents some quantitative results that show the number of correct matches is significantly better with conformal images than with initial images. Note that, as with the planar case, the important improvement is not quantitative but qualitative in the spatial repartition of the point matches as shown in figure 6.

Figure 7 is more artificial. It presents the results of our method on a sphere, which is the archetype of

Table 2. Comparison of #matches between original and conformal images (camel).

Camel	Standard SIFT	Conformal
#keypoints 1	29,947	25,712
#keypoints 2	28,328	27,647
#matches	360	1,624

smooth surface; two images are very wide baselines (90 degrees). It can be seen that with standard SIFT (blue: in dashed line), the points are all located on a very small part of the images. The part where can be found the matches, in fact, there is a little distortion between the images. Conversely, it can be observed that with conformal matching (red), the point are distributed all around common part of the images. Table 3 shows a quantitative comparison of match points between SIFT original and conformal images.

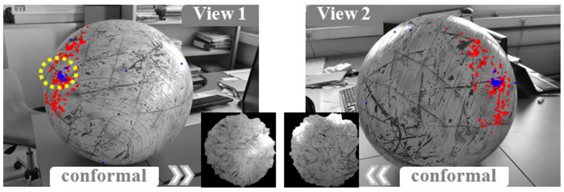


Figure 7. Same as figure 6, with very wide baseline images (90 degrees).

Table 3. Comparison of #matches between original and conformal images (sphere).

Sphere	Standard SIFT	Conformal
#keypoints 1	58,664	42,825
#keypoints 2	68,765	52,926
#matches	71	509

4 Conclusions

The aim of this work is to develop an algorithm and software for efficiently computing homologous points from a pair of images with associated depth information. In this paper, we have presented two approaches for solving specific sub-problems: one for the case of planar scenes and one for the case of smooth surface scenes. Both quantitative and qualitative evaluations of the methods show improvement in the matching process. Using the depth information to correct image distortions, it is possible to obtain matches homogeneously located on the common part of the images.

However, the presented work is only the first step that, to our opinion, validates our approach from a theoretical point of view. Further developments are required in order to obtain a fully automatic process for matching of radiometric and depth image. The main questions, we intend to address in the near future, are:

- Automate the segmentation of depth image between plane area, smooth area and "noisy" area (like trees), this is necessary to obtain a fully automatic process.
- Use more sophisticated development than least square conformal mapping, for example non linear isometric constraint to limit possible degeneracy with unsmooth surface.

In a more long-term development, we intend to compute directly the "3D isotropic SIFT on the 3D surface". This means that instead of computing an anamorphous image, we will compute directly on the initial image the result of convolutions by Gaussians that are isotropic on the 3D surface.

References

[1] M. Pierrot-Deseilligny and I. Cléry : "APERO, An Open Source Bundle Adjustment Software for Automatic Calibration and Orientation of A Set of Images," *ISPRS Commission V*, 2011. (`www.micmac.ign.fr`)

[2] CGAL, Computational Geometry Algorithms Library: `http://www.cgal.org`

[3] ANN, A Library for Approximate Nearest Neighbor Searching: `http://www.cs.umd.edu/~mount/ANN/`

[4] D. G. Lowe: "Distinctive Image Features from Scale-Invariant Keypoints," *IJCV*, pp.91–110, 2004.

[5] H. Bay et al.: "SURF: Speeded Up Robust Features," *CVIU*, pp.346–359, 2008.

[6] J. Matas et al.: "Robust Wide Baseline Stereo from Maximally Stable Extremal Regions," *BMVC*, pp.384–393, 2002.

[7] K. Mikolajczyk et al.: "A Comparison of Affine Region Detectors," *IJCV*, 2005.

[8] J. M. Morel and G. Yu: "ASIFT: A New Framework for Fully Affine Invariant Image Comparison," *SIAM, on Imaging Sciences*, pp.438–469, 2009.

[9] C. Wu et al.: "3D Model Matching with Viewpoint-Invariant Patches (VIP)," *IEEE CVPR*, 2008.

[10] K. Köser and R. Koch: "Perspectively Invariant Normal Features," *ICCV Workshop on 3D Representation for Recognition*, 2007.

[11] M. Pollefeys et al.: "Detailed Real-Time Urban 3D Reconstruction From Video," *IJCV*, 2008.

[12] S. Wang et al.: "Conformal Geometry and Its Applications on 3D Shape Matching, Recognition, and Stitching," *PAMI*, pp.1209–1220, 2007.

[13] B. Lévy et al.: "Least Squares Conformal Maps for Automatic Texture Atlas Generation," *ACM Transactions on Graphics*, pp.362–371, 2002.

[14] P. J. Besl and N. D. McKay: "A Method for Registration of 3-D Shapes," *PAMI*, pp.239–256, 1992.

[15] M. A. Fischler and R. C. Bolles: "Random Sample Consensus: A Paradigm for Model Fitting with Applications to Image Analysis and Automated Cartography," *Comm. of the ACM*, pp.381–395, 1981.

[16] W. Förstner and E. Gülchl: "A Fast Operator for Detection and Precise Location of Distinct Points, Corners and Centres of Circular Features," *ISPRS Conf. on Fast Processing of Photogrammetric Data*, pp.281–305, 1987.

[17] C. Harris and M. Stephens: "A Combined Corner and Edge Detector," *in* 4^{th} *Alvey Vision Conf.*, pp.147–151, 1988.

[18] CMVS, Clustering views for Multi-View Stereo: `http://www.di.ens.fr/cmvs/`

[19] N. Snavely et al.: "Photo Tourism: Exploring Image Collections in 3D," *ACM Transactions on Graphics (Proceedings of SIGGRAPH 2006)*, 2006.

[20] VisualSFM, A Visual Structure from Motion System: `http://homes.cs.washington.edu/~ccwu/vsfm/`

[21] Acute3D Software: `http://www.acute3d.com`