AMIFT: AFFINE-MIRROR INVARIANT FEATURE TRANSFORM

Noureddine Mohtaram, Amina Radgui, Guillaume Caron, El Mustapha Mouaddib

AMIFT: AFFINE-MIRROR INVARIANT FEATURE TRANSFORM

Noureddine Mohtaram [(a,b)], *Amina Radgui* [(c)], *Guillaume Caron* [(b)], *El Mustapha Mouaddib* [(b)]

[(a)] LRIT, Rabat IT Center, Faculty of Sciences, Mohammed V University, B.P. 1014, Rabat, Morocco
[(b)] University of Picardie Jules Verne, MIS Laboratory 33 rue Saint Leu, 80039 Amiens Cedex 1, France
[(c)] INPT, Madinat AL Irfane, Rabat, Morocco

ABSTRACT

In this paper, we propose a descriptor for image matching under multiple mirror reflections. Indeed, existing adaptations of SIFT for the mirror transformation are not successful when object and mirrors orientations are not constrained. Hence, we propose to combine MIFT and Affine-SIFT descriptors as the Affine Mirror Invariant Feature Transform (AMIFT). The experimental results and given evaluation show that our proposed descriptor outperforms MIFT and ASIFT on both synthetic and real images datasets.

***Index Terms*—** SIFT, Affine transformation, local features, reflection invariant feature.

1. INTRODUCTION

A Planar Catadioptric System (PCS) is an imaging system made of refracting lenses (dioptrics) and reflecting mirrors (catoptrics) [1]. Using planar mirrors, one can change the view direction, expand the field of view of a pinhole camera, and, with two mirrors, acquire multiple views of the same object in a single image (Fig. 1). Considered in this paper, the latter captures, in a single PCS image, a direct view of the scene and its single and multiple reflections on mirrors, providing a surrounding image acquisition. Enough information is then available for the 3D reconstruction of the entire object, despite the part in contact with the ground, obviously.

3D reconstruction of objects, observed by PCS made with two planar mirrors, is well known considering the object envelop [2] but, to our knowledge, not considering the object texture as in common 3D reconstruction pipelines. The latter, in a few words, would provide more details of the object surface than the former, but need to match every view of the same object (or scene). Such matching is the key issue of the pipeline. It is done from detected image features, and computation leads to their corresponding 3D point coordinates.

Matching detected features is done thanks to feature descriptors as SIFT, the most popular method. SIFT [3] computes a descriptor for each feature point that is invariant to scale and image plane rotation transformations. Extensions of SIFT consider more degrees of freedom of the transformations that a feature point descriptor is invariant

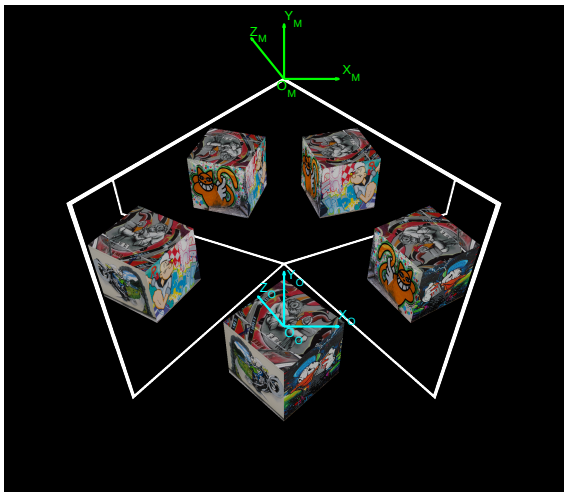


Fig. 1: A captured image of the PCS.

to: mirror transformation (MIFT [4], ISIFT[5], MI-SIFT[6], Max-SIFT[7], FIND[8], MBR-SIFT[9]) or limited-range affine transformation (ASIFT [10]). Once candidate matches are computed, a random sample consensus (RANSAC) [11], exploiting epipolar geometry [12], is performed to keep correct matches (inliers) only, for a robust 3D reconstruction. However, the performance of these descriptors-based matching decreases a lot when encountering both affine and mirror transformations, which is the case when considering PCS systems. To overcome the issues encountered by SIFT methods-based matching in a single image acquired by a PCS, we propose to:

- Combine ASIFT and MIFT together as AMIFT, for Affine-Mirror Invariant Feature Transform.
- Consider the modeling of the epipolar geometry for a PCS in a single image in order to extend RANSAC-based robust point matching to a PCS image, as Reflective-RANSAC.

AMIFT and Reflective-RANSAC are the contributions of the paper. Both are described in Section 2. Section 3 presents simulation and actual scene results and their evaluation with

respect to state-of-the-art methods. Section 4 concludes the paper.

2. PROPOSED METHOD

This section presents our Affine-Mirror Invariant Feature Transform used for point matching further improved by Reflective-RANSAC. The latter is based on Reflective Fundamental matrix computed in the PCS. The latter system consists of a perspective camera and two planar mirrors and is equivalent to a system of five cameras pointed to the same object surrounding it. Hence, this system provides four pairs of views in a single image. Each pair is composed of an image of the object and its reflection through one of the mirrors.

2.1. Epipolar geometry

The epipolar constraint in PCS with two mirrors was investigated in [2], [13], [14] and [15]. Forbes et al [2] determined the epipoles from the silhouette outlines of the object and its reflection. Gluckman et al [15] shew that the degrees of freedom of the fundamental matrix in mirror reflection case is reduced from 7 to 6. In [13] and [14], authors gave some properties of multi-views camera-mirrors systems and studied a large variety of their geometric constraints. In this section, we show how it is possible to obtain the epipolar geometry by estimating the fundamental matrix from the correspondence between some points, without additional information on the intrinsic or extrinsic parameters of the used camera.
For each pair of views and since the two cameras are symmetric, their epipolar lines and epipoles are identical. This allows to reduce the number of degrees of freedom of the fundamental matrix from 7 (in standard two view geometry) to 2 in our case [13], [14]. So, from minimum two couples of non-coplanar matches, we are able to find an estimation of the antisymmetric matrix F given by :

$$F = \begin{bmatrix} 0 & -e_3 & e_2 \\ e_3 & 0 & -e_1 \\ -e_2 & e_1 & 0 \end{bmatrix} \quad (1)$$

where $\mathbf{e} = \begin{bmatrix} e_1, e_2, e_3 \end{bmatrix}^T$ is the epipole between two symmetric cameras.
By using this simple method, the two epipoles ($\{e_{V1}, e_{V2}\}$ see Fig . 2) from the considered four pairs of views in one image can be determined. The fundamental matrix will be used in this work to eliminate the false matching points.

2.2. Affine Mirror reflection invariant feature transform

In this section, we propose a new approach, named AMIFT which combines affine and mirror reflection transformation for points matching within the PCS. In addition, AMIFT preserves all the advantages such as rotation and scale invariances exhibited by SIFT. The image given by the PCS includes 4 pairs of views. Each pair, selected manually, is composed of an image of the object and its reflection on a mirror called the sub-images. The AMIFT approach is based on four steps: (1) affine transformation of sub-images, (2) detection of keypoints in each transformed sub-image and creation of ASIFT descriptor vectors, (3) flipping and matching these vectors and finally (4) eliminating the false matching points based on Reflective-RANSAC constrained by the reflective fundamental matrix (Equation. 1).

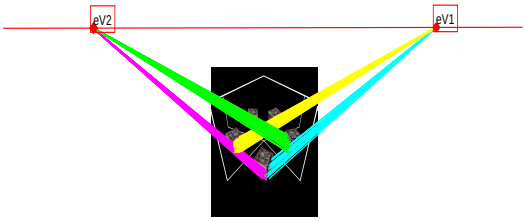


Fig. 2: Epipoles and epipolar lines in PCS

The affine transformation simulates some possible affine distortions caused by the change of the camera optical axis orientation from a frontal position. The distortion depends on two parameters: the longitude angle ϕ and the latitude angle θ. The sub-images undergo ϕ rotations followed by tilts with parameters t given by:

$$|t| = |\frac{1}{\cos\theta}| \quad (2)$$

The affine transformation f, defined in Equation. 3, allows to determine the sub-image $u'(x, y)$ obtained using bilinear interpolated transform of the initial sub-image $u(x, y)$:

$$f : \begin{pmatrix} x \\ y \end{pmatrix} \mapsto \begin{bmatrix} a & b \\ c & d \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \quad (3)$$

where $\begin{bmatrix} a & b \\ c & d \end{bmatrix} = \lambda \begin{bmatrix} \cos\psi & -\sin\psi \\ \sin\psi & \cos\psi \end{bmatrix} \begin{bmatrix} t & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\phi & -\sin\phi \\ \sin\phi & \cos\phi \end{bmatrix}$

λ is the scale parameter, and ψ is the rotation angle of camera around optical axis.

In order to match the key-points computed in two transformed sub-images, we used the same approach as the original ASIFT [10]. In the phase of descriptor establishing, as SIFT, this descriptors are made of 128 values using 8 orientation bins to describe each keypoint. The modification of the MIFT algorithm [4] is used in this process to flip the descriptor vectors. Then, we need to reorganize the sequence of bins in anticlockwise direction as shown in Fig. 3 to ensure the symmetric transformation of the features due to the use of the mirror. The descriptors resulting from the application of AMIFT can be matched using any matching algorithm that would work for SIFT since their structure is preserved. The

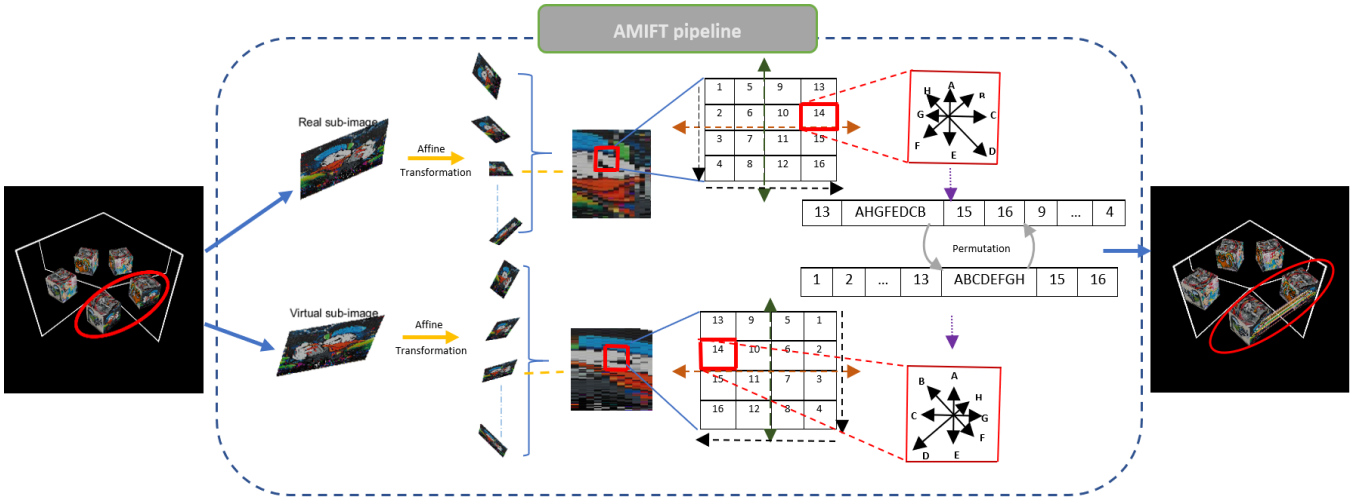


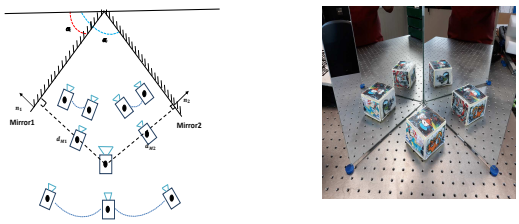
Fig. 3: AMIFT pipeline

basic idea for the matching stage is to find the minimum distance between each feature descriptor from each transformed sub-image and all other features from all their transformed reflection. The final step discards the false matches using the constraint of epipolar geometry as Reflective-RANSAC.

3. EXPERIMENTAL RESULTS

3.1. Dataset construction

First, we consider a dataset of mirror images using two static planar mirrors, fixed using angle $\alpha_l = 35°, \alpha_r = 125°$, and various camera poses as shown in Fig. 4(a). A representative sample of this database are illustrated in Fig. 4(b). As can be seen, varying one-shot views enables acquiring various views of the object of interest. Each image was divided into five sub-images (direct view sub-image and four virtual sub-images). Grouping the real view with its reflections enables to obtain four correspondences groups. Therefore, a number of 40 views was used to evaluate the relevance of the proposed approach.



(a) (b)

Fig. 4: Experiment setup: (a) different camera positions within PCS. (b) multi-view image captured by the PCS.

3.2. AMIFT evaluation

We compare AMIFT to SIFT, MIFT and ASIFT. According to results depicted in Fig. 5, which is one of 40 images in the database, the total number of correspondences is obtained by taking sums of each one of the four correspondence groups in the synthetic image. AMIFT generates the highest number of correct correspondences under high viewpoint mirror changes (612 points for AMIFT, 301 points for MIFT, 138 points for ASIFT and 65 points for SIFT). MIFT generates many incorrect correspondences in the case where the principal axis of the camera is not perfectly perpendicular to the planar mirror. Only a few number of correct correspondences is recovered using ASIFT.

Recall-Precision analysis: To deeply evaluate our approach, we create synthetic images using the Pov-Ray Tool in order to generate the ground truth. Thereafter, to evaluate the performance of each method, the Precision and Recall values of each image from the acquired dataset are computed [16]. Recall-Precision curve analysis was used to compare AMIFT behavior, on synthetic data, to SIFT, MIFT and ASIFT. Recall represents the ratio of correct matches M_C to the total number of correspondences M_T between two sub-images:

$$Recall = \frac{M_C}{M_T}. \quad (4)$$

Precision describes the ratio of correctly matched descriptors to the total positive matches P_T determined by the ground truth.

$$Precision = \frac{M_C}{P_T}. \quad (5)$$

Precision and Recall for our 40 PCS images dataset are reported in Fig. 6. As can be seen, curves obtained, with and without Reflective-RANSAC, demonstrate that our approach depicts high performances comparing to the other ap-

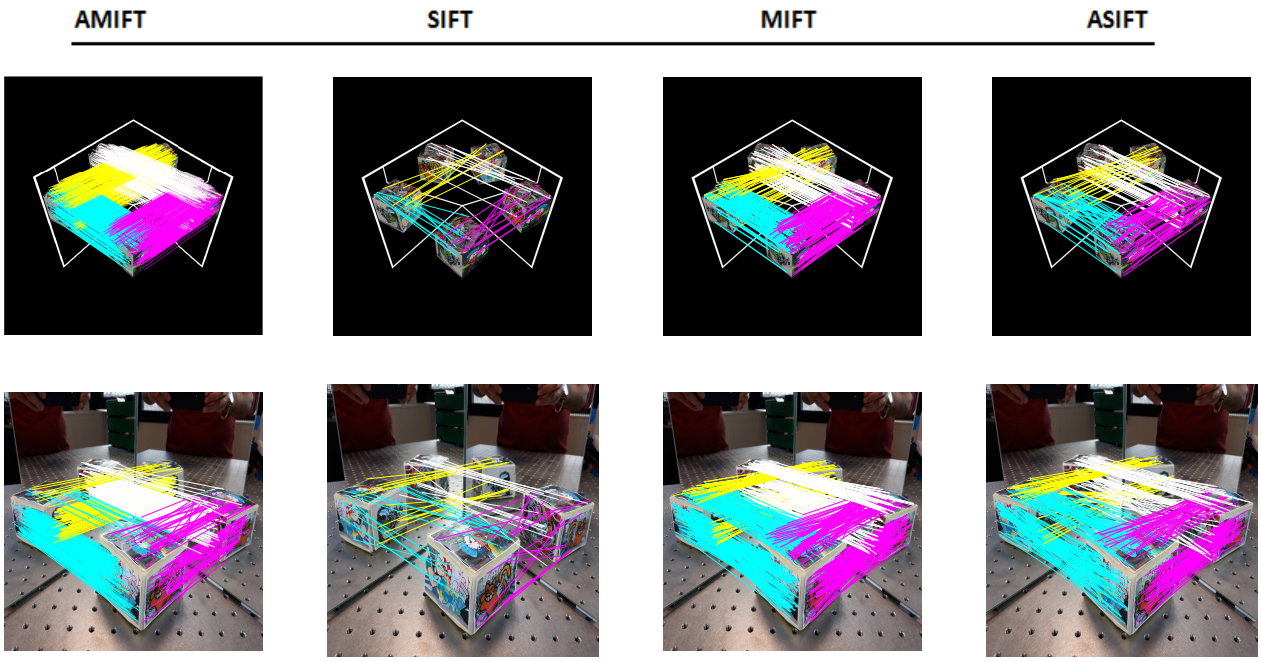
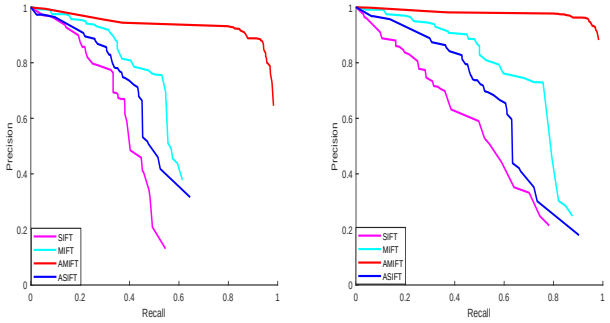


Fig. 5: Recovered matching correspondences obtained on the synthetic and real sets using the different methods.

proaches. This proves the ability of AMIFT to deal efficiently with changes including mirror reflection and important transformations with respect to mirrors.



(a) PR Curve (b) PR Curve after outliers elimination step

Fig. 6: PR curves obtained from synthetic dataset

4. CONCLUSION

In this paper, we proposed a new feature descriptor that is robust to multiple planar mirror reflections: AMIFT. We also derived geometric constraints for corresponding features in direct-reflection or reflection-reflection pairs of views. Based on the geometric constraints, we built the Reflective-RANSAC to enhance even more the robustness of AMIFT. Results demonstrate the clearly higher robustness of AMIFT, over SIFT, MIFT and ASIFT, on both synthetic and real datasets. Future works will be dedicated to evaluate AMIFT on some higher scale PCS, for outdoor scenes where large objects back-faces are physically inaccessible by direct camera views.

5. REFERENCES

[1] Simon Baker and Shree K Nayar, "A theory of catadioptric image formation," in *Computer Vision, 1998. Sixth International Conference on*. IEEE, 1998, pp. 35–42.

[2] Keith Forbes, Fred Nicolls, Gerhard De Jager, and Anthon Voigt, "Shape-from-silhouette with two mirrors and an uncalibrated camera," in *European Conference on Computer Vision*. Springer, 2006, pp. 165–178.

[3] David G Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[4] Xiaojie Guo and Xiaochun Cao, "Mift: A framework for feature descriptors to be mirror reflection invariant," *Image and Vision Computing*, vol. 30, no. 8, pp. 546–556, 2012.

[5] Jiaji Song, Xiaobo Lu, Wei Wang, and Cong Chen, "Isift: Improving the performance of sift for mirror images," in *Computer and Communications (ICCC), 2016 2nd IEEE International Conference on*. IEEE, 2016, pp. 742–745.

[6] Rui Ma, Jian Chen, and Zhong Su, "Mi-sift: mirror and inversion invariant generalization for sift descriptor," in *Proceedings of the ACM International Conference on Image and Video Retrieval*. ACM, 2010, pp. 228–235.

[7] Lingxi Xie, Qi Tian, and Bo Zhang, "Max-sift: Flipping invariant descriptors for web logo search," in *Image Processing (ICIP), 2014 IEEE International Conference on*. IEEE, 2014, pp. 5716–5720.

[8] Xiaojie Guo and Xiaochun Cao, "Find: A neat flip invariant descriptor," in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 515–518.

[9] Mingzhe Su, Yan Ma, Xiangfen Zhang, Yan Wang, and Yuping Zhang, "Mbr-sift: A mirror reflected invariant feature descriptor using a binary representation for image matching," *PloS one*, vol. 12, no. 5, pp. e0178090, 2017.

[10] Jean-Michel Morel and Guoshen Yu, "Asift: A new framework for fully affine invariant image comparison," *SIAM journal on imaging sciences*, vol. 2, no. 2, pp. 438–469, 2009.

[11] Martin A Fischler and Robert C Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," in *Readings in computer vision*, pp. 726–740. Elsevier, 1987.

[12] Richard Hartley and Andrew Zisserman, *Multiple view geometry in computer vision*, Cambridge university press, 2003.

[13] Gian Luca Mariottini, Stefano Scheggi, Fabio Morbidi, and Domenico Prattichizzo, "Planar mirrors for image-based robot localization and 3-d reconstruction," *Mechatronics*, vol. 22, no. 4, pp. 398–409, 2012.

[14] Shinji Fujiyama, Fumihiko Sakaue, and Jun Sato, "Multiple view geometries for mirrors and cameras," in *Pattern Recognition (ICPR), 2010 20th International Conference on*. IEEE, 2010, pp. 45–48.

[15] Joshua Gluckman and Shree K Nayar, "Catadioptric stereo using planar mirrors," *International Journal of Computer Vision*, vol. 44, no. 1, pp. 65–79, 2001.

[16] Krystian Mikolajczyk and Cordelia Schmid, "A performance evaluation of local descriptors," *IEEE transactions on pattern analysis and machine intelligence*, vol. 27, no. 10, pp. 1615–1630, 2005.