How to Overcome Perceptual Aliasing in ASIFT?

Nicolas Noury, Frédéric Sur, Marie-Odile Berger

How to Overcome Perceptual Aliasing in ASIFT?

Nicolas Noury, Frédéric Sur, Marie-Odile Berger

Magrit Project-Team, UHP / INPL / INRIA, Nancy, France

Abstract. SIFT is one of the most popular algorithms to extract points of interest from images. It is a scale+rotation invariant method. As a consequence, if one compares points of interest between two images subject to a large viewpoint change, then only a few, if any, common points will be retrieved. This may lead subsequent algorithms to failure, especially when considering structure and motion or object recognition problems. Reaching at least affine invariance is crucial for reliable point correspondences. Successful approaches have been recently proposed by several authors to strengthen scale+rotation invariance into affine invariance, using viewpoint simulation (*e.g.* the ASIFT algorithm). However, almost all resulting algorithms fail in presence of repeated patterns, which are common in man-made environments, because of the so-called perceptual aliasing. Focusing on ASIFT, we show how to overcome the perceptual aliasing problem. To the best of our knowledge, the resulting algorithm performs better than any existing generic point matching procedure.

1 Introduction and Related Works

One of the first steps in many computer vision applications is to find correspondences between points of interest from several images. Applications are *e.g.* photography stitching [1], object recognition [2], structure from motion [3], robot localization and mapping [4], etc. Points of interest belong to "objects" viewed from different camera positions. Thus, their definition ought to be insensitive to the aspect of the underlying object. Besides, it is desirable to attach vectors to these points which describe a surrounding patch of image, in order to find correspondences more easily. Ideally, these vectors should not change across the views. In the pinhole camera model, 3D objects are transformed via projective mappings. However, the underlying object is generally unknown. With the additional assumption that points of interest lie on planar structures, points and descriptors should be invariant to homographies. Since affine mappings are first-order approximations of homographies, this weaker invariance is often considered sufficient.

In his groundbreaking work [2], D. Lowe explains how to extract scale+rotation invariant keypoints, the so-called SIFT features. Some authors have tried to reach affine invariance (see *e.g.* MSER [5], Harris / Hessian Affine [6] and the survey [7], or [8] for semi-local descriptors). Although these latter methods have been proved to enable matching with a stronger viewpoint change, all of them

are prone to fail at a certain point. A more successful approach has been recently proposed by several authors (*e.g.* [9–11]), in which viewpoint simulation is used to increase scale+rotation to affine invariance. These papers demonstrate that this dramatically improves the number of matches between two views compared to MSER or Harris/Hessian Affine, even with a strong viewpoint change.

Let us explain viewpoint simulation, and especially Morel and Yu's ASIFT [10] which we aim at improving. In ASIFT, affine invariance of image descriptors is attained by remarking from Singular Value Decomposition that any affine mapping A (with positive determinant) can be decomposed as

$$A = \lambda R_\psi \begin{pmatrix} t & 0 \\ 0 & 1 \end{pmatrix} R_\phi \tag{1}$$

where $\lambda > 0$, R_ψ and R_ϕ are rotation matrices, $\phi \in [0, 180^\text{o})$, $t \geqslant 1$.

Since SIFT is scale+rotation invariant, a collection of affine invariant (ASIFT) descriptors of an image I is obtained by extracting SIFT features from the simulated images $I_{t,\phi}$ with

$$I_{t,\phi} = \begin{pmatrix} t & 0 \\ 0 & 1 \end{pmatrix} R_\phi(I). \tag{2}$$

Indeed, the location of the SIFT keypoints is (nearly) covariant with any scale and rotation change λR_ψ applied to $I_{t,\phi}$, and the associated descriptor does (almost) not change. From [10], it is sufficient to discretize t and ϕ as: $t \in \{1, \sqrt{2}, 2, 2\sqrt{2}, 4\}$ and $\phi = \{0, b/t, \ldots, kb/t\}$ with $b = 72^\text{o}$ and $k = \lfloor t/b \cdot 180^\text{o} \rfloor$.

The next step is to match ASIFT features between two images I and I'. A two-scale approach is proposed in [10]. First, the $I_{t,\phi}$ and $I'_{t',\phi'}$ are generated from downsampled images (factor 3), then SIFT features extracted from each pair $(I_{t,\phi}, I'_{t',\phi'})$ are matched via the standard algorithm from [2], namely that nearest neighbours are selected provided the ratio of the Euclidean distance between the nearest and the second nearest is below some threshold (0.6 in ASIFT). The deformations corresponding to the M pairs (M typically set to 5) that yield the largest number of matches are used on the full-resolution I and I', giving new SIFT features that are matched by the same above-mentioned criterion. The obtained correspondences are then projected back to I and I', provided already-placed correspondences are at a distance larger than $\sqrt{3}$. This strategy is used to limit the computational burden and also prevents redundancy between SIFT features from different deformations. A subsequent step consists in eliminating spurious correspondences with RANSAC by imposing epipolar constraints.

Lepetit and Fua [9] use the same decomposition as in eq. (1). Since their points of interest are invariant neither to scale nor to rotation, they have to discretize or randomly sample the whole set of parameters λ, t, ψ, ϕ.

Let us also mention Molton *et al.* work [11]. Small planar image patches are rectified through homographies in a monocular SLAM application. In this framework, an on-the-fly estimation of the camera motion and of the 3D normal of the patch is available. Thus there is no need to generate every possible rectification, making it effective in a real-time application. This provides a richer description of the 3D scene than with standard point features.

Aim and organization of the article. As noted in [10], ASIFT fails when confronted to repeated patterns. In this work we propose to secure ASIFT against it. Section 2 explains why repeated patterns are important and call for a special treatment in *every* image matching applications. Section 3 describes the proposed algorithm. We also improve the selection of the relevant simulated images and the back-projection step, while enabling non nearest neighbour matches. Experiments are presented in Section 4. The proposed algorithm has also an increased robustness to large viewpoint changes.

2 Perceptual Aliasing and Point Matching

Perceptual aliasing is a term coined by Whitehead and Ballard in 1991 [12]. It designates a situation where "*a state in the world, depending upon the configuration of the sensory-motor subsystem, may map to several internal states; [and] conversely, a single internal state may represent multiple world states*". In computer vision applications and especially point of interest matching, invariant features make it possible to overcome the first part of the perceptual aliasing. A viewpoint invariant feature such as ASIFT is indeed supposed to give a unique representation of the underlying 3D point, whatever the camera pose. However, repeated patterns are also uniquely represented although they do not correspond to the same 3D point. This makes almost all point matching algorithms fail when confronted to repeated patterns, except when explicitly taking them into account in an *ad hoc* application (*e.g.* [13]). Some authors even get rid of them at an early stage (*e.g.* in [14], patterns occurring more than five times are *a priori* discarded). The problem is of primary importance since repeated patterns are common in man-made environments. Just think of two views of a building: correctly matching the windows is simply impossible when considering only invariant descriptors. Additional geometric information is needed.

The problem is all the more relevant as in most applications, matching (or sometimes tracking) points of interest usually consists in two independent steps: 1) point of interest matching by keeping the "best" correspondence with respect to the distance between the associated descriptors, then 2) correspondence pruning by keeping those that are consistent with a viewpoint change. A popular choice for step 1) is nearest neighbour matching, which yet gives false correspondences, partly because of perceptual aliasing. The nearest neighbour has indeed no reason to be a correct match in case of repeated patterns. Step 2) is often a RANSAC scheme, which keeps only the correspondences consistent with the epipolar geometry (fundamental or essential matrix) or with a global homography (for planarly distributed points or for a camera rotating around its optical center). Since ASIFT uses this two-step scheme to match simulated images, it is not able to retrieve from perceptual aliasing. If the images mostly show repeated patterns, ASIFT even simply fails as in Section 4, Figures 5 and 6.

We have recently proposed [15] a new one-step method to replace both abovementioned steps 1) and 2). It is a general algorithm to match SIFT features between two views, and it is proved to be robust to repeated patterns. The

present contribution is to incorporate it into the ASIFT algorithm. Let us briefly describe the method (which is a generalization of [16]). Considering N_1 points of interest x_i with the associated SIFT descriptor D_i from image I, and N_2 points of interest x'_j with descriptor D'_j from image I', one aims at building a set of correspondences $(x_i, x'_j)_{(i,j)\in S}$ where S is a subset of $[1 \dots N_1] \times [1 \dots N_2]$, which is the "most consistent set with respect to a homography" among all possible sets of correspondences. Let us note that the model in [15] also copes with general epipolar geometry; we will see in Section 3 why we focus on homographies. The consistency of S is measured in [15] as a *Number of False Alarms* (NFA) derived from an *a contrario* model (see the books [17, 18] and references therein):

$$\mathrm{NFA}(S,H) = (\min\{N_1,N_2\}-4)\,k!\binom{N_1}{k}\binom{N_2}{k}\binom{k}{4} f_D(\delta_D)^k f_G(\delta_G)^{k-4} \qquad (3)$$

where:

– the homography H from I to I' is estimated from four pairs from S,

– k is the cardinality of S,

– $\delta_D = \max_{(i,j)\in S} \mathsf{dist}(D_i, D'_j)$ where dist is a metric over SIFT descriptors,

– f_D is the cumulative distribution function of δ_D and is empirically estimated from I and I', yielding an adaptive measure of resemblance,

– $\delta_G = \max_{(i,j)\in S} \max\{d(x'_j, Hx_i), d(x_i, H^{-1}x'_j)\}$ where d is the Euclidean distance between two points,

– f_G is the cumulative distribution function of δ_G.

Several possibilities for dist are investigated in [15]. We choose here to use the CEMD-SUM metric introduced in [19], based on an adaptation of the Earth's Mover Distance for SIFT descriptors. In particular, it is proved to behave better with respect to the quantization effects than the standard Euclidean distance.

For the sake of brevity, we elaborate here neither on the statistical model giving f_G and f_D nor on the definition of the NFA and kindly refer the reader to [15]. Let us simply say that $f_D(\delta_D)^k f_G(\delta_G)^{k-4}$ is the probability that *all* points in S are mapped to one another through H (with precision δ_G), while *simultaneously* the associated descriptors are similar enough (with precision δ_D), *assuming that points are independent*. If this probability is very low, then the independence assumption is rejected following the standard hypothesis testing framework. There must be a better explanation than independence, and each pair of points probably corresponds to the same 3D point. The advantage of this framework is that it automatically balances the resemblance between descriptors and the geometric constraint. Considering a group with a very low probability, all of its descriptors are close to one another (photometric constraint) and each of its points projects close to the corresponding point in the other image via H (geometric constraint, which is not covered at all by nearest neighbour matching). Mixing both constraints makes it possible to correctly associate repeated patterns. Additionally, it is permitted to match non-nearest neighbours, provided they satisfy the geometry. As we will see in Section 4, this provides a number of correspondences that are never considered in standard SIFT matching.

Now, instead of measuring the probability $f_D(\delta_D)^k f_G(\delta_G)^{k-4}$ (of a false positive in hypothesis testing) which naturally decreases as k grows, the NFA is introduced in the *a contrario* literature. One can prove (see [15, 17–19] for further information) that a group such that NFA $\leq \varepsilon$ is expected to appear less than ε times under independence hypothesis (hence the term *Number of False Alarms*). Thus, comparing groups of different sizes via the NFA is sound. As noted in [15], small groups can win over large ones if they are very accurate (that is, descriptors are very similar and points are nearly perfectly related by a homography).

Since the combinatorial complexity of the problem is very large, a heuristic-driven search based on random sampling is given in [15], in order to reach the group S with the (hopefully) lowest NFA. Let us also mention that this method does not need application-specific parameters.

Remark that Hsiao *et al.* [20] have very recently proposed to use ASIFT for 3D object recognition, improving pose estimation when facing strong viewpoint changes, thanks to the numerous point correspondences. As in [15] for 2D/2D matching, they solve correspondences between 3D points and 2D points by simultaneously taking account of photometric resemblance and pose consistency. Their algorithm is thus robust to repeated patterns.

3 Improving ASIFT

We explain here how we modify the ASIFT algorithm by incorporating the NFA criterion (yielding the Improved ASIFT algorithm, I-ASIFT in the sequel). The basic idea is to replace the nearest neighbour matching between generated images with the matching algorithm of [15], *i.e.* seek the group of correspondences consistent with a homography, with the lowest NFA. The back-projection of the matching features to the original images is also improved. The algorithm for both Improved ASIFT and Standard ASIFT is explained in Figure 1, where the proposed modifications are highlighted. A running-example is provided on Figure 2. Figure 3 compares with standard SIFT matching and ASIFT.

Let us discuss the modifications. First, we replace in step 3 the nearest neighbour criterion by the above-mentioned method. The reason to use homography constraint is that when simulating affine transformations, one expects that some of them will correctly approximate homographies related to planar parts of the scene (possibly related to *virtual* planes, in the sense that points may be distributed on a plane which has no physical meaning). Then, each group of correspondences between simulated images should correspond to points lying over a planar structure, and consequently be associated via a homography. In standard ASIFT, the number of groups (*i.e.* of considered pairs of generated images) is limited *a priori* to five. In our framework, it would lead to select correspondences from a fixed number of planar pieces. On the contrary, we keep groups with log(NFA) below -50. This amounts generally to keeping between 5 (fully planar scene) and $\simeq 70$ (multi-planar scene) groups. There is no need to keep a larger number of groups since groups with the largest NFA would be made of

Data: two images I and I'.

1. For both images, generate the new collection of images $I_{t,\phi}$ and $I'_{t',\phi'}$ (eq. (2)):
I-ASIFT - use $t, t' \in \{1, \sqrt{2}, 2\}$ and ϕ, ϕ' as in ASIFT (the range of t is the same as in [20], sufficient if the viewpoint change is not too extreme)
ASIFT - first low resolution, then full resolution simulation only for a limited number of $(t, \phi), (t'\phi')$, as explained in Section 2.

2. Extract the SIFT features from the generated images.

3. Match the SIFT features between the pairs of generated images:
I-ASIFT - for each pair $(I_{t,\phi}, I'_{t',\phi'})$ extract the group of point correspondences with the lowest NFA (eq. (3), see discussion).
ASIFT - for each pair from the limited set of step 1, match each feature from $I_{t,\phi}$ to its nearest neighbour in $I'_{t',\phi'}$, provided the ratio between the distances to the nearest and to the second nearest neighbour is below 0.6

4. Back-project the matched SIFT keypoints from the $I_{t,\phi}$'s and $I'_{t',\phi'}$'s to I and I':
I-ASIFT - keep groups with $\log(\mathbf{NFA}) < -50$, then sort them increasingly along their NFA. Starting from the first group, back-project a pair of matching features only if each feature do not fall in the vicinity of any already-placed feature. The vicinity is defined as the back-projection in I (resp. I') of the circle around the feature extracted from the simulated images, with radius equal to the SIFT scale (minimum $\simeq 2$ pixels).
ASIFT - back-project the matching features only if there is no already-placed feature at a distance less than $\sqrt{3}$ pixels.

5. Discard possible false correspondences:
I-ASIFT - use *a contrario* RANSAC [16] to check consistency with epipolar geometry or to homography, depending on the case of interest.
ASIFT - use a contrario *RANSAC to check consistency with epipolar geometry only (not mentioned in [10], but mandatory in the implementation from [21]).*

Output: a set of corresponding points of interest between I and I'.

Fig. 1. Improved ASIFT (**I-ASIFT**) and Standard ASIFT (*ASIFT*).

redundant points or would be made of a few inconsistent points filtered by the final RANSAC.

To improve the back-projection of step 4, we propose to use the NFA as a goodness-of-fit criterion. As remarked by Hsiao *et al.* [20], viewpoint simulation methods give anyway a large number of correspondences, some of them being concentrated in the same small area. The NFA criterion balances the size of a group and its accuracy as explained earlier. It seems to us that favouring groups with the lowest NFA is sounder than systematically favouring large groups. In addition, when back-projecting points we thoroughly select correspondences from their scale in order to prevent accumulations in small areas (note that our criterion is stricter than the one in ASIFT). Getting correspondences uniformly and

densely distributed across the 3D scene is important for structure and motion applications (as in [20]).

Let us remark that repeated patterns bring specific problems that RANSAC cannot manage. As remarked in [15], if the repeated patterns are distributed along the epipolar lines, then it is simply impossible to disambiguate them from two views (as in Figure 6, ACM+F). Theoretically, I-ASIFT could also suffer from it. However, it would require that: 1) one of the group consists in a bunch of shifted patterns consistent with a homography (as in group 51 on figure 2), 2) this group is large enough and has a very low NFA (otherwise most points are redundant with already-placed points), and 3) points are along the associated epipolar lines (otherwise they are discarded by the final RANSAC). Thus I-ASIFT is more robust to this phenomenon.

4 Experiments

We compare the proposed I-ASIFT, noted I-ASIFT+F (resp. I-ASIFT+H) when the final RANSAC is based on fundamental matrix (resp. homography), with:

– standard SIFT matching (that is, nearest neighbour + distance ratio condition, between 0.6 and 0.8 to give the best possible results), followed by the RANSAC from [16]. We note this algorithm NNR+F if RANSAC imposes epipolar geometry (fundamental matrix), or NNR+H for homography constraint;

– the *a contrario* matching algorithm from [15], which permits SIFT matching with repeated patterns, noted ACM+F or ACM+H;

– ASIFT, whose implementation is kindly provided by Morel and Yu [21].

We use Vedaldi and Fulkerson's code for SIFT [22]. The reader is kindly asked to zoom in the pdf file.

Figure 4 is an assessment on a pair of images with a very strong viewpoint change. NNR and ACM simply fail here, Harris/Hessian Affine and MSER give less than 2 matches (see [10]). Viewpoint simulation is thus needed. I-ASIFT provides more correspondences than ASIFT, which are distributed in a dense fashion while ASIFT accumulates them in small areas. This is mainly caused by the distance ratio threshold set to 0.6 in ASIFT, which discards too many correspondences in some generated image pairs. However, using a higher value leads to a larger rate of outliers, especially when considering large perspective deformations. The more sophisticated matching in I-ASIFT automatically adapts the resemblance metric between descriptors from each pair of simulated images.

Figure 5 shows an experiment with almost only repeated patterns which can still be disambiguated after a careful examination. In this case, ASIFT fails. More precisely, it gives some good correspondences, but they are buried in a large amount of false matches, and cannot be retrieved with the final RANSAC. NNR does not give any correspondence. Since the viewpoint change is not too strong, ACM+H still finds 21 correspondences, all correct. I-ASIFT+H finds 44 correspondences, all correct. One can see that the NFA criterion (eq. (3)) permits us to match features which are not nearest neighbours (30% in I-ASIFT+H). Of course, such correspondences are never considered in standard SIFT matching.

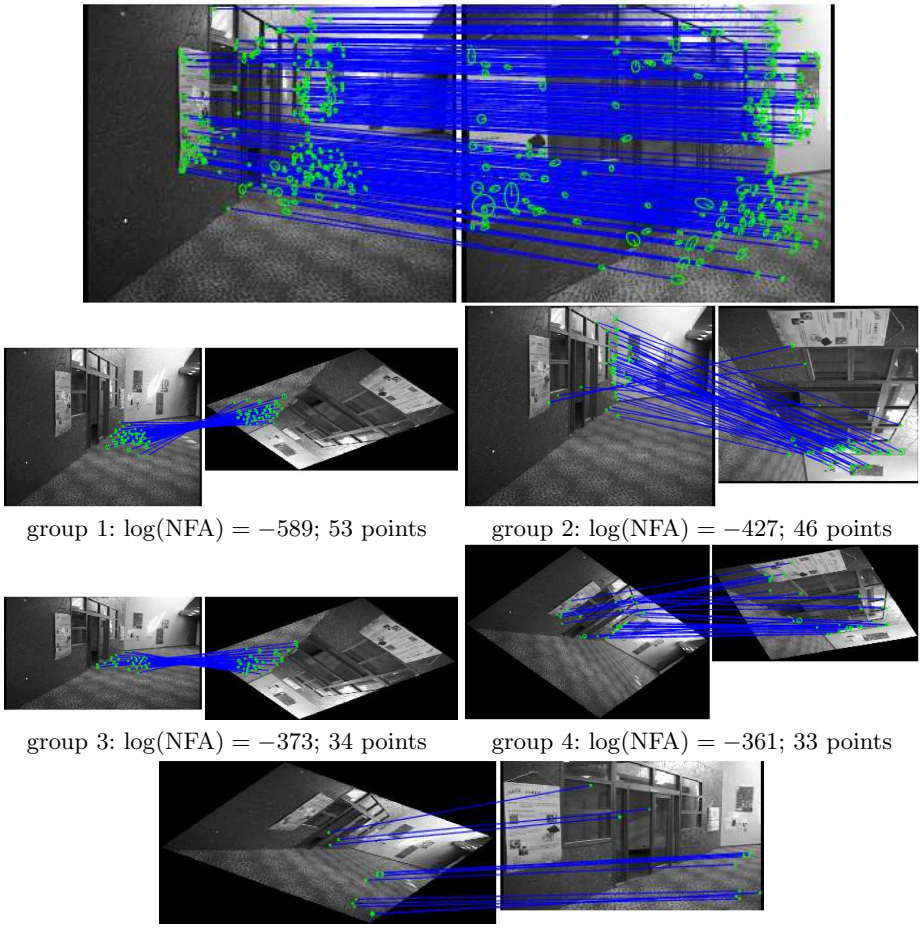

Fig. 2. *Running example.* **Top**: 210 correspondences found with I-ASIFT. Each green ellipse is the backprojection in the original images of the circle with a radius equal to the SIFT scale in the simulated image. **Below (groups 1 to 4)**: correspondences from the four pairs $(I_{t,\phi}, I'_{t',\phi'})$ corresponding to the groups with the lowest NFA. One can see that these groups actually correspond to points over a quite small piece of plane. In this experiment, 65 such groups are kept (with $\log(\text{NFA}) < -50$). The last 10 groups yield only 14 correspondences. As a comparison, the four groups shown here yield 116 correspondences. With our scheme, points from group 3 (resp. 4) redundant with those of group 1 (resp. 2) are not back-projected to I and I'. Note that points on the wall or on the carpet are scattered among several groups. Indeed, the strong induced homographies need to be approximated with several affine mappings. In **group 51**, the matching algorithm is trapped by perceptual aliasing: descriptors are alike but the correspondences are consistent with a homography "by chance". 10 points from this group are back-projected, but all of them are discarded by the final RANSAC imposing consistency to epipolar geometry.

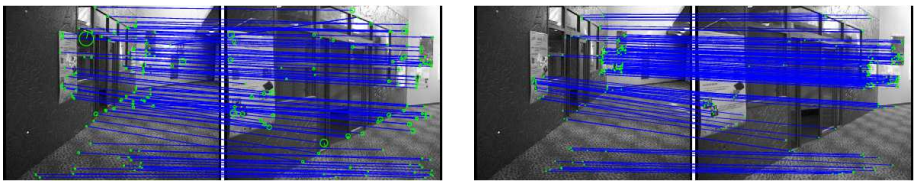

Fig. 3. *Running example.* **Left**: SIFT matching (nearest neighbour + ratio set to 0.8), cleaned by the same RANSAC as in ASIFT [16]. **Right**: ASIFT. 97 matches are found for SIFT, 153 for ASIFT. For a fair comparison, ASIFT was run with the same resolution as I-ASIFT (no downsampling). Some points on the carpet are not correctly matched. A bunch of wrong correspondences can indeed be seen on the foreground, because of repeated patterns falling by chance near the associated epipolar lines.

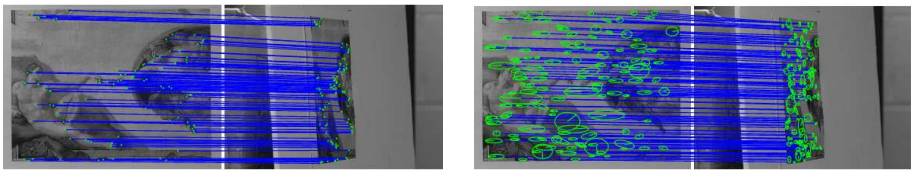

Fig. 4. *The Creation of Adam* (from [10, 21]). **Left**: ASIFT. 100 matches. **Right**: I-ASIFT+H. 124 matches are retrieved with t in the range $\{1, \sqrt{2}, 2, 2\sqrt{2}, 4\}$ as in ASIFT. 49 groups are kept. The range $\{1, \sqrt{2}, 2\}$ (which we use for all other experiments of this article with I-ASIFT) still gives 29 matches (19 groups), not shown here.

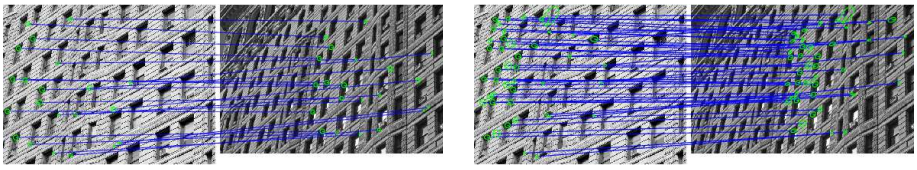

Fig. 5. *Flatiron Building.* **Left**: ACM+H finds 21 matches, 13 of which are nearest neighbours, 5 are second nearest neighbours, the 3 remaining matches are between 3rd and 7th nearest neighbours. **Right**: I-ASIFT+H finds 44 matches, 29 of which are nearest neighbours, 7 are 2nd nearest, the 8 remaining matches are between 3rd and 6th nearest neighbours. 15 groups are kept.

Figure 6 is another experiment with repeated patterns. To the best of our knowledge, I-ASIFT is the only generic point matching algorithm able to retrieve a large number of correct correspondences over the three visible sides of the cube. This is a highly desirable feature for structure and motion applications. One can also see that the ACM+H method (used in step 3 of I-ASIFT) is able to cope with repeated patterns. Let us remark that in this experiment we get *one* of the consistent solutions. However, we cannot eliminate the hypothesis that the cube has been rotated by 90°. In some cases, a certain amount of perceptual aliasing cannot be reduced from the information contained in images.

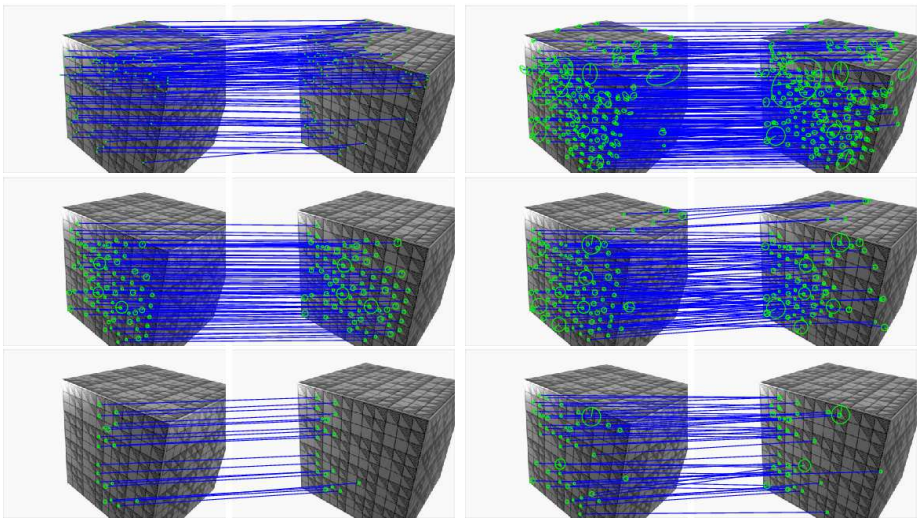

Fig. 6. *Synthetic Cube.* **Top left**: ASIFT. 101 correspondences, almost half of them are not correct. Many matching patterns are actually shifted. **Top right**: I-ASIFT+F. 192 correspondences. A careful examination proves that almost all are correct. Only 102 among them are nearest neighbours, the others match between 2nd and 8th nearest neighbours. 49 groups are kept. **Middle left**: ACM+H. 83 matches (only 40% of them are nearest neighbours), patterns of the "dominant" plane are correctly retrieved (homography constraint). **Middle right**: ACM+F. 102 matches (55% are nearest neighbours). False correspondences can be seen. This is simply unavoidable with two-view geometry, since in this experiment many wrongly associated repeated patterns (correct for the photometric constraint) lie along the corresponding epipolar lines (thus correct for the geometric constraint). **Bottom left**: NNR+H. 19 matches, corresponding to shifted patterns. **Bottom right**: NNR+F. 42 matches, many of them are not correct.

Figure 7 shows that I-ASIFT is also more robust to strong viewpoint changes than ASIFT. It is due to the proposed strategy consisting in back-projecting features from simulated images, which automatically selects a large number of groups of correspondences consistent with a local homography, contrary to

ASIFT where most correspondences actually come from the same pair of simulated images. NNR+F and ACM+F do not give any set of correspondences.

Fig. 7. *Leuven Castle*: two distant images from M. Pollefeys' sequence. **Left**: ASIFT. 94 matches, only among points from the same façade. Note that repeated windows yield false correspondences with the fourth window in the second image (which is not present in the first one.) **Right**: I-ASIFT+F. 118 matches (24% are not nearest neighbours), distributed over the whole building. Except for two, all of them are correct. The fundamental matrix is then estimated over the retrieved set of correspondences. The epipolar lines (in yellow in the right images) corresponding to some handpicked points (from the left images) prove that I-ASIFT permits to reliably estimate the camera motion. The points associated to the handpicked ones are indeed less than 1 pixel away from the corresponding epipolar line. In contrast, the camera motion cannot be retrieved from ASIFT. As a comparison, MSER gives 5 matches, and Harris/Hessian Affine 20-30 matches mainly between wrongly associated repeated patterns. (code from Mikolajczyk et al.'s `www.featurespace.org`)

5 Conclusion

The main contribution of this article is to change the matching paradigm of ASIFT (namely nearest neighbour matching) to a more sophisticated one which aggregates sets of correspondences consistent with a local homography. It is not limited to nearest neighbour and yields dramatic results when confronted to repeated patterns. The resulting algorithm is also more robust than ASIFT, MSER, or Harris/Hessian Affine to large viewpoint changes, showing promising capacities for Structure From Motion applications.

References

1. Brown, M., Lowe, D.: Automatic panoramic image stitching using invariant features. International Journal of Computer Vision **74** (2007) 59–73
2. Lowe, D.: Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision **60** (2004) 91–110
3. Gordon, I., Lowe, D.: Scene modelling, recognition and tracking with invariant image features. In: Proc. International Symposium on Mixed and Augmented Reality (ISMAR). (2004) 110–119

4. Se, S., Lowe, D., Little, J.: Vision-based global localization and mapping for mobile robots. IEEE Transactions on Robotics **21** (2005) 364–375
5. Matas, J., Chum, O., Urban, M., Pajdla, T.: Robust wide-baseline stereo from maximally stable extremal regions. Image and Vision Computing **22** (2004) 761–767
6. Mikolajczyk, K., Schmid, C.: Scale & affine invariant interest point detectors. International Journal of Computer Vision **60** (2004) 63–86
7. Mikolajczyk, K., Tuytelaars, T., Schmid, C., Zisserman, A., Matas, J., Schaffalitzky, F., Kadir, T., Gool, L.V.: A comparison of affine region detectors. International Journal of Computer Vision **65** (2006) 43–72
8. Musé, P., Sur, F., Cao, F., Gousseau, Y., Morel, J.M.: An a contrario decision method for shape element recognition. International Journal of Computer Vision **69** (2006) 295–315
9. Lepetit, V., Fua, P.: Keypoint recognition using randomized trees. IEEE Transactions on Pattern Analysis and Machine Intelligence **28** (2006) 1465–1479
10. Morel, J.M., Yu, G.: ASIFT: A new framework for fully affine invariant image comparison. SIAM Journal on Imaging Sciences **2** (2009) 438–469
11. Molton, N.D., Davison, A.J., Reid, I.D.: Locally planar patch features for real-time structure from motion. In: Proc. British Machine Vision Conference (BMVC). (2004)
12. Whitehead, S., Ballard, D.: Learning to perceive and act by trial and error. Machine Learning **7** (1991) 45–83
13. Schaffalitzky, F., Zisserman, A.: Planar grouping for automatic detection of vanishing lines and points. Image and Vision Computing **18** (2000) 647–658
14. Schaffalitzky, F., Zisserman, A.: Automated location matching in movies. Computer Vision and Image Understanding **92** (2003) 236–264
15. Noury, N., Sur, F., Berger, M.O.: Determining point correspondences between two views under geometric constraint and photometric consistency. Research Report 7246, INRIA (2010)
16. Moisan, L., Stival, B.: A probabilistic criterion to detect rigid point matches between two images and estimate the fundamental matrix. International Journal of Computer Vision **57** (2004) 201–218
17. Cao, F., Lisani, J., Morel, J.M., Musé, P., Sur, F.: A theory of shape identification. Number 1948 in Lecture Notes in Mathematics. Springer (2008)
18. Desolneux, A., Moisan, L., Morel, J.M.: From Gestalt theory to image analysis: a probabilistic approach. Interdisciplinary applied mathematics. Springer (2008)
19. Rabin, J., Delon, J., Gousseau, Y.: A statistical approach to the matching of local features. SIAM Journal on Imaging Sciences **2** (2009) 931–958
20. Hsiao, E., Collet, A., Hebert, M.: Making specific features less discriminative to improve point-based 3D object recognition. In: Proc. Conference on Computer Vision and Pattern Recognition (CVPR). (2010)
21. Morel, J.M., Yu, G.: ASIFT. IPOL Workshop (2009) `http://www.ipol.im/pub/algo/my_affine_sift`. Consulted 6.30.2010.
22. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms (2008) `http://www.vlfeat.org/`. Consulted 6.30.2010.