Epipolar Geometry Estimation for Urban Scenes with Repetitive Structures

Maria Kushnir and Ilan Shimshoni

Department of Information Systems
University of Haifa, Israel
mkushn01@campus.haifa.ac.il ishimshoni@mis.haifa.ac.il

Abstract. Algorithms for the estimation of epipolar geometry from a pair of images have been very successful in recent years, being able to deal with wide baseline images. The algorithms succeed even when the percentage of correct matches from the initial set of matches is very low. In this paper the problem of scenes with repeated structures is addressed, concentrating on the common case of building facades. In these cases a large number of repeated features is found and can not be matched initially, causing state-of-the-art algorithms to fail. Our algorithm therefore clusters similar features in each of the two images and matches clusters of features. From these cluster pairs, a set of hypothesized homographies of the building facade are generated and ranked mainly according the support of matches of non-repeating features. Then in a separate step the epipole is recovered yielding the fundamental matrix. The algorithm then decides whether the fundamental matrix has been recovered reliably enough and if not returns only the homography. The algorithm has been tested successfully on a large number of pairs of images of buildings from the benchmark ZuBuD database for which several state-of-the-art algorithms nearly always fail.

1 Introduction

Repeated structures are commonly seen in many types of scenes. They are especially prevalent in man made scenes such as buildings as can be seen for example in Fig. 1. For reasons which will be explained shortly, algorithms for epipolar geometry estimation from two images tend to fail on such scenes. The goal of this paper is to present an algorithm to deal with these cases. In this paper we will concentrate on building facades which are one of the most common cases of repeated structures.

In recent years there has been significant progress in developing algorithms for epipolar geometry estimation for wide baseline image pairs. Generally speaking, the algorithm is given as input two images. On both images a feature detection algorithm is run yielding a set of features and their associated descriptors (e.g., SIFT [1]). The two feature sets are then matched yielding a set of pairs of similar features from the two images. On this set of putative matches a robust algorithm from the RANSAC [2] family is run resulting in a model which in some

cases is the fundamental matrix or an homography in others. The matches are also classified as inliers or outliers. On this general framework many advances have been made. The LO-RANSAC [3] algorithm performs local optimization on candidate solutions suggested by RANSAC, reducing the number of iterations. Other algorithms suggest methods to guide the selection of subsets selected by the RANSAC process [4–7]. Finally, methods were suggested to reduce the number of putative matches selected at each iteration resulting in a much faster algorithm which can deal with a much higher percentage of outliers [8–10].

(a) (b)

Fig. 1. Possible cases of images with repeating structures. (a) A building with repeating elements appearing on the same vertical and horizontal lines. (b) A building with repeating elements appearing periodically in a grid structure.

As a result of all these advances wide baseline stereo image registration systems are successful in many hard cases with very low inlier match percentage. However, for scenes with repeated structures they often fail. The reason for this is that repeated structures yield similar sets of local features for which humans and automated systems fail to match correctly. In most cases the algorithm is able to recognize that there are several very similar matches to such a feature and it therefore discards the feature altogether. As a result, when the overlap between the two scenes contains mainly repeated structures, the alignment algorithms tend to fail.

1.1 Related work on repeated elements

In this work we will be dealing with image registration but repeated elements have been extensively studied in different contexts such as detection and grouping of similar elements [11–13], classification and identification [14], matching [15, 16], geo-tagging and location recognition [17–19] as well as structure from motion methods [20].

From these works we would like to elaborate on several papers. In [20] the problem addressed was of recovering the structure from a large number of images (SfM) when the scene contains multiple instances of the same object. The

challenge which is addressed in that paper is to eliminate the incorrect fundamental matrices from the set of fundamental matrices recovered from matching all image pairs. This is done using geometric and image-based cues.

Perhaps the papers most closely related to our work, dealing with image registration and repeated patterns are [16, 17, 19, 21–24], as they present different approaches for matching images of building facades, without analyzing or modeling of the entire structure as in [11–13].

In [24] a guided RANSAC algorithm is presented. A large number of putative matches is generated by matching all possible similar points but giving repeated features low probabilities. Thus, they are not used in the model generation step but only in the verification step. When the number of correct unique correspondences is small the running time of the algorithm can be long.

In [16] it is assumed that the objects investigated are comprised of planar quadrilaterals bounded by straight lines. For each hypothesized match between a pair of quadrilaterals, the homography between images is calculated. The score of the homography is given by counting the number of corresponding Harris corners within the region. It should be noted that there is no descriptor extraction for the detected Harris points, and that this method results with a projective homography, that matches two building facades without any estimation of the epipolar geometry. [19] recovers the position of a mobile robot by matching building facades. The algorithm exploits the fact that the views were obtained from similar heights, and thus matches are restricted to a narrow margin surrounding a 1D scan line. Similarly, in [21] invalid correspondences are eliminated based on geometric constraints generated from approximate knowledge of internal and external camera calibration parameters. [22, 23] deal with scenes with multiple objects using an a-contrario approach. They concentrate on the post-processing step in which the algorithm has to decide which of the matches belong to the current solution. Finally, [17] extracts calibrated images from an existing database and matches it to an input image. The transformation supported by the maximal number of matches is returned. It is therefore possible that a shifted solution will be returned by the algorithm.

1.2 Our approach

In this paper we suggest an algorithm to deal with the case of repeated structures placed on planar or close to planar surfaces. The main application of such an algorithm is for images containing mainly building facades. Without such an algorithm, systems with an image registration component will fail from time to time unexpectedly when the overlap between the images are mainly building facades.

The algorithm exploits three important characteristics of the scenes that we are dealing with. First, a large number of repeated structures lie on a planar surface in an ordered fashion and second that the local feature descriptors detected in the image can be clustered and the clusters between the two images can be matched without determining initially how the individual members of a matched

cluster are matched. Using the repeated features usually several possible solutions are generated with similar support. We therefore extract the small number of regular unique matches in order to select the correct solution, in contrast to the methods described above.

The algorithm divides the task of epipolar geometry estimation into two steps. It first recovers the homography associated with the building's facade and then recovers the epipole. Finally it decides whether the fundamental matrix is reliable and if not returns only the homography.

The paper continues as follows. In the next section we will present our general approach. In Section 3 we present experimental results run on challenging image pairs from the ZuBuD database of images of buildings from Zurich [25] for which general purpose state of the art algorithms usually fail. We compare our method to a SIFT [1] matching step followed by a standard RANSAC [2] and to two state-of-the-art wide baseline registration algorithms BEEM [10] and BLOGS [7]. Conclusions and plans for future work are discussed in Section 4.

2 The algorithm

Scenes with repeated structures are very common. In this paper we will concentrate on the special case where most of the repeated elements lie on planar surfaces or close to planar surfaces such as building facades.

In our algorithm we consider two cases. In the first case we only assume that the repeated objects are partially organized horizontally or vertically (Fig. 1(a)). In the second case we assume that there exists a grid of repeated objects (Fig. 1(b)). We will first describe the algorithm which deals with the non-periodic case and then in Section 2.4 the modifications required to deal with grids of repeated objects will be presented.

When two images containing a planar surface with repeated objects are given, the first step of the algorithm (described in Section 2.1) is to find for each image an homography which will transform the image into a fronto parallel view. This step is performed for two reasons. First, eliminating the projective distortion makes the descriptors recovered from the repeated features more similar and thus easier to cluster. Second, when given two fronto-parallel images of a planar surface, the transformation between them is much simpler. All that has to be recovered, is the 2D translation and the scale factor.

On each of the rectified images SIFT features are extracted and features with similar descriptors are clustered. We then match pairs of clusters from the two images. There are of course features which do not cluster and will be called non-repeating features.

In Section 2.2 we generate a set of hypothesized transformations of the plane appearing in both rectified images. This is done by matching minimal subsets of features from a cluster generated from the first image to a subset of features from its corresponding cluster from the second image. The hypothesized transformations are ranked by the number of matched features that satisfy $\mathbf{x}' = H\mathbf{x}$.

In Section 2.3 we exploit the fact that the fundamental matrix F can be factored into $F = [\mathbf{e}']_\times H$. Therefore it can be computed by estimating the epipole $\mathbf{e}'$ for a given homography. Once F has been found the algorithm decides whether there is enough evidence to support it, and if not it returns only H.

2.1 Image rectification

In our algorithm, we use the Canny edge detector, to detect edges and from them extract line segments in the image. Then, we apply RANSAC [2] twice to find the vertical vanishing point $\mathbf{V}_{\mathbf{p}_v}$ and the horizontal vanishing point $\mathbf{V}_{\mathbf{p}_h}$, although as will be shown later, our method can handle a swap in those directions.

Under the standard assumptions of square pixels, zero skew, and that the principal point is at the image center, the internal calibration matrix K and the rotation matrix R can be recovered [26, Chapter 8]. Consequently, the original image is rectified by: $H = KRK^{-1}$, resulting in a fronto-parallel view. An example of the results of this procedure can be seen in Fig. 2.

(a) (b)

Fig. 2. Image rectification. (a) The original image with detected line segments, that are consistent with the two vanishing points. (b) Fronto-parallel rectified image.

From the rectified images we extract SIFT features and descriptors (using the implementation provided by [27]). This step is performed on the rectified images, since in the case of repeated features, descriptors are more similar due to elimination of the projective distortion. In general, each SIFT key-point can be assigned with an orientation, based on the local image gradient direction, which is the key step in achieving invariance to rotation. In our case, we use upright SIFT key-points, for which the key-point orientation is set to be vertical. The single fixed orientation for all features is a natural choice, given that the rotation is compensated through the rectification. Moreover, it prevents features such as for example window corners of different orientations to be considered as the same feature.

We then cluster the SIFT key-points within a single image. Repeating points are identified and clustered if their appearances are similar, i.e., their normalized cross correlation is larger than a threshold (in our experiments 0.9). For each

cluster, we select the medoid of the repeating points' descriptors as the cluster descriptor. The result of this step is not perfect. Not all clusters represent real repeating objects and not all repeating objects are represented by a feature in a cluster of a repeating feature. Still as can be seen in the supplementary material, the number of recovered clusters can be used to differentiate between images with or without repeating structures.

2.2 Planar homography estimation

We start the image registration process by searching for a specific transformation H, induced by the rectified plane with repeating elements on it, that maps one rectified image to another. For that purpose we assume that repeating key-points, appear on the same vertical and horizontal lines, without specific requirements about distances or periodicity. We build a list of candidate transformations.

As a first step, we match key-point clusters from the rectification stage. We check all possible cluster pairs from two images and compute the Euclidean distance between the cluster descriptor vectors of each pair. Similar to Lowe's approach, we define the best match as the one with minimal distance. We determine the probability that a cluster match is correct, by taking the ratio of distance from the closest neighbor to the distance of the second closest. Small clusters (smaller than 5 points), or those that do not have any good match (distance ratio larger than 0.8) are discarded.

When searching for all possible homographies, we use the rectified images extracted previously. Due to that, both transformed images are fronto-parallel. As a result, instead of searching for eight degrees of freedom of a general projective transformation H, we are left with only three, namely the two coordinates of a relative translation t_x,t_y and the relative scale s

$$H = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \implies \begin{bmatrix} s & 0 & t_x \\ 0 & s & t_y \\ 0 & 0 & 1 \end{bmatrix}. \tag{1}$$

To detect H candidates, we check all the feature points from two images. For each point of the first image $\mathbf{x}_c$ we try to find its approximate vertical $\mathbf{x}_v$ and horizontal $\mathbf{x}_h$ nearest neighbors within the same cluster if they exist. Such a point triplet will be denoted $\mathcal{T}$. We perform an identical procedure on the second image yielding point triplets, each denoted $\mathcal{T}'$. We then match pairs of point triplets from the two rectified images, belonging to matched clusters. Every such matched triplet $\mathcal{MT} = \{\mathcal{T}, \mathcal{T}'\}$ is used to compute a transformation H. In general two feature points from each image would be sufficient for transformation estimation, but relying on triplets gives rise to less candidates to handle and much more accurate results. Exploiting the scale constraint, we eliminate transformations that do not satisfy it. When the triplet strategy fails, we resort to using pairs of feature points from each image. In general the scale ratio is not always accurately estimated. In this case there are two scales s_x s_y instead of one s. We have implemented both versions dealing with two/one scale and they both always succeeded.

For each candidate transformation, other feature point pairs from different clusters which satisfy the transformation relation are accumulated and are considered point matches which support the transformation. Based on them, we improve the accuracy of H using LO-RANSAC [3] as follows. We iteratively calculate an homography based on randomly selecting half of supporting point matches, and compute H using a non-linear method which minimizes Sampson's approximation to the geometric re-projection error. As a result we obtain, an accurate homography, which relies on many points, instead of the single localized triplet and is immune to the inaccuracies of the rectification procedure and the proximity between the point triplet. The result of this step is a list of candidate homographies.

(a) (b)

Fig. 3. Typical results, when building a list of all possible homographies. (a) Blue stars: An arbitrary point with its vertical and horizontal nearest neighbors within the same cluster. Red points: Additional points from the same cluster, that support the same H. (b) Blue stars: correct point match and its nearest neighbors. Green squares: An alternative point match.

To illustrate the process we present in Fig. 3 a typical result. In both images two point triplets are marked by blue stars. These two point triplets can be used to compute the correct H. The red points in Fig. 3(a) indicate additional points that support that H. The green squares in Fig. 3(b) represent an alternative point triplet from which an additional H candidate is computed.

Homography ranking: Once the set of homographies has been generated the next task is to rank them. In order to better deal with this issue, we assume that not only repeating elements appear on the plane, but also several unique key-points, which we plan to exploit to break the symmetry. We therefore match the SIFT key-points from the two rectified images, by the standard technique, proposed by Lowe.

We rank each homography from the list, by the number of key-point matches that are consistent with it. If the homography H is the correct one, it should return not only repeating key-points from several clusters, but also corresponding locations of unique SIFT key-points. Therefore, we sort homographies based on

the following score:

$$S_H = N_{rep} + \alpha N_{non-rep}, \tag{2}$$

where α is a weight constant and N_{rep} and $N_{non-rep}$ are the numbers of supporting correspondences from repeating and non-repeating key-points respectively. In our experiments we set $\alpha = 100$ (the algorithm works for $\alpha > 10$), to emphasize that we mostly rely on the small number of matches of unique key-points, to rank the homographies. When sorted, we iteratively check the homography with the maximal S_H until no improvement is reached.

One of the advantages of our method is that we can empirically tune this weight constant α, by changing the preference of one type of key-points over the other. In [16, 17, 19], this would be impossible, since only the number of matches is counted. As a result, for an image pair with partial occlusion of the repeating elements, the homography H having the maximal overlap would be chosen, as there are naturally more repeating key-points in the images. In our method on the other side a few highly weighted non-repeating key-points would be sufficient to detect the correct H, regardless of occlusion or a partially non-overlapping scene.

In addition, there always is a possibility that when rectifying one of the images, horizontal and vertical directions were swapped. This is especially common when the original images were taken with a roll angle of approximately 90°, as shown in Fig. 2(b). We therefore keep the first rectified image unchanged and check for three possible alignments of the second image: the one obtained from its rectification and the two rotations by ±90°. We also change all the key-point locations and descriptors respectively. For each one of the three alignments, we rank the homographies as described above.

2.3 Image registration

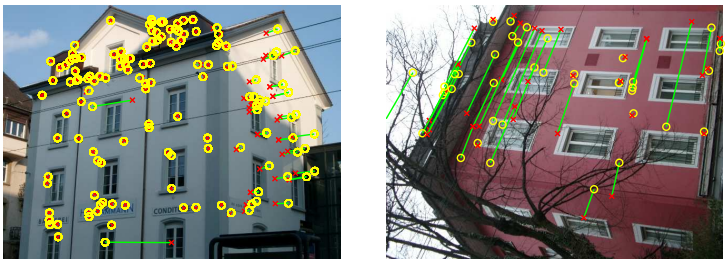
After the homographies have been ranked, for each one of them a RANSAC process will be run to estimate the epipole $\mathbf{e}'$. Combining it with the homography H yields F. When looking for the correct fundamental matrix F, we assume that the repeating elements are bounded to the underlying plane, and therefore they are not considered in this step. Matching correspondences of the non-repeating key-points however, can appear on, as well as off the plane. Thus, we select non-repeating key-points, that can contribute to the estimation of F. Those point pairs $\{\mathbf{x}_i, \mathbf{x}'_i\}$ must satisfy:

$$\|H\mathbf{x}_i - \mathbf{x}'_i\| = \|\mathbf{x}''_i - \mathbf{x}'_i\| \propto |\rho_i| > d_{proj}, \tag{3}$$

where $\mathbf{x}''_i = H\mathbf{x}_i$, ρ_i is the projective depth, relative to the underlying plane, and d_{proj} is a constant distance threshold. In our experiments d_{proj} was set to five pixels.

In Fig. 4(a) we show non-repeating key-points $\mathbf{x}'_i$ marked by yellow circles and $\mathbf{x}''_i$ by red crosses. The green lines are proportional to the projective depth ρ_i of the pairs. For most of the in-plane key-points, we can see the yellow circles merge with the red crosses, which indicates zero projective depth. There are only

two mistaken matched pairs, that are associated with a visible green line, despite being on the plane. In addition, we can observe for the off-plane points, that the further the point is from the plane, the longer is the green line associated with it.



(a) (b)

Fig. 4. (a) Matching correspondences of the non-repeating key-points. Original image with non-repeating key-points $\mathbf{x}'_i$ marked by yellow circles and $\mathbf{x}''_i$ by red crosses. (b) Wrong matches of non-repeating key-points, that result from repetitive elements. Original image with non-repeating key-points $\mathbf{x}'_i$ marked by yellow circles and $\mathbf{x}''_i$ by red crosses. Green lines are proportional to the projective depth ρ_i.

Another problem demonstrated in Fig. 4(b) is of putative matches, which are due to incorrect matches between repetitive features that were not detected as such during the clustering phase. In general RANSAC is able to deal with outliers. However, when the feature pairs lie on a horizontal or vertical line on the facade as can be seen in the figure, these incorrect matches will vote together for an incorrect epipole, the horizontal or vertical vanishing point which in many cases will produce an incorrect solution. We therefore remove these putative matches from consideration. These removed matches satisfy $(H\mathbf{x}-\mathbf{x}')\times\mathbf{V_p}\approx 0$.

All the remaining matches, termed candidate F supporters, are used in the RANSAC step to recover the epipole $\mathbf{e}'$. The candidate H and the recovered $\mathbf{e}'$ will then be combined to yield the fundamental matrix $F=[\mathbf{e}']_\times H$. In this step the putative matches come from two sources: matched features extracted from the rectified images which mainly come from the parallel planes consisting of the building's facade and matched features extracted from the original images. These matches usually come from off-plane 3D points, since they become too distorted in the rectification process to be matched using the rectified images.

Once the RANSAC step has been completed all the matches that support the fundamental matrix F (including the ones that support the homography) are given to a final RANSAC step which recovers the homography H accurately.

The question that remains is whether the algorithm should return F or that there is not enough evidence to support a fundamental matrix (when for example the overlap between the two images is close to planar) and only H should be returned. We answer this question by counting the number of matches that support F and do not support H. If there are more than a certain number of

supporters (10 in our experiments) F is returned by the algorithm and if not, only H is returned.

2.4 Grid of repeating structures

In this section we switch to a more demanding case than discussed earlier, the case of periodic repeating elements. We describe here the additional steps and required changes to the full algorithm flow, previously presented.

During the first step of image rectification and key-points extraction, additional information can be evaluated. Assuming periodicity, we estimate optimal horizontal and vertical repetition intervals separately for each rectified image. We define the difference between every pair of intersections, of all the detected line segments with one of the axes, as a possible horizontal or vertical repetition interval. In other words, if two intersection points x_1 and x_2 support an interval I then $x_1 - x_2 = kI$ for some integer k or $\mod(x_1, I) = \mod(x_2, I)$.

Therefore, for every possible interval I, we build a histogram h_n at a resolution of one pixel, of $\mod(x_i, I)$. Thus, the number of supporting pairs of lines for an interval will be

$$N_I = \sum_{n=0}^{\lceil I-1 \rceil} h_n(I)(h_n(I) - 1)/2. \tag{4}$$

If an interval I is a good candidate, we expect to have sharp peaks in the histogram, coming from the unification of repeating lines' intersection. An example of such a histogram for a good vs. bad candidate can be seen in Fig. 5. However, for intervals $\left\{\frac{I}{2}, \frac{I}{3}, \frac{I}{4} \ldots\right\}$, it is expected to have even sharper histograms.

Thus, the score for interval is set to $S_I = N_I I$ to induce a preference for I and not for its fractions. The algorithm builds a list of several (three in our implementation) candidates for I with the maximal scores. The value of $S_I(I_{max})$ can be used to detect images with a grid of repeating structures, as can be seen in the submitted supplementary material.

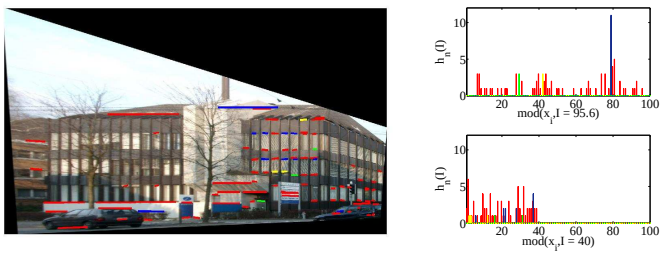


Fig. 5. Estimation of optimal I_y. Rectified image with all the detected horizontal line segments and histograms for good ($I_y = 95.6$) and bad ($I_y = 40$) candidates of I_y.

It is during the next step, when building a list of all possible transformations between the two rectified images, we exploit the list of optimal horizontal and

vertical repetition intervals extracted previously. We compute the relative scale s from Eq. 1, by:

$$s = I_{x_2}/I_{x_1} \qquad s = I_{y_2}/I_{y_1}, \tag{5}$$

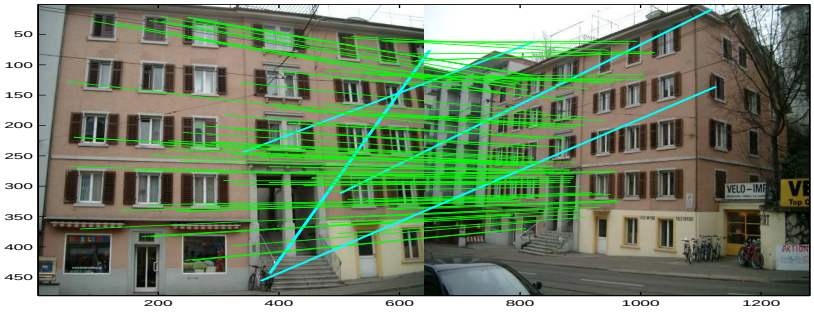
where I_{x_i} and I_{y_i} are the horizontal and vertical repetition intervals in the rectified image respectively. Eq. 5 is used to select sets of consistent interval values. Thus, when estimating the transformation between the two rectified images we are left with only two out of eight degrees of freedom of a general projective transformation H: the two coordinates of the relative translation t_x and t_y.

In order to detect H candidates, following a similar strategy as in the non-periodic case, it is required to check all the points from two images belonging to matched clusters and compute the translation between each pair. Every such point pair could yield a transformation H. As a result there would be many more candidates in this case than in the non-periodic case. Therefore, we change our approach slightly. As we have mentioned earlier, we assume that if the H is correct, it should be supported not only by repeating key-points, but also by corresponding locations of unique SIFT key-points. Thus, it is possible to build H candidate from each non-repeating key-point correspondence. Identically to the non-periodic case, for each candidate transformation, all feature point pairs from different clusters which satisfy the transformation relation are found and are considered point matches which support the homography. Based on them we compute H using LO-RANSAC [3]. The output of this step is a list of candidate homographies. The following steps of homography ranking and image registration are identical to those in the non-periodic case.

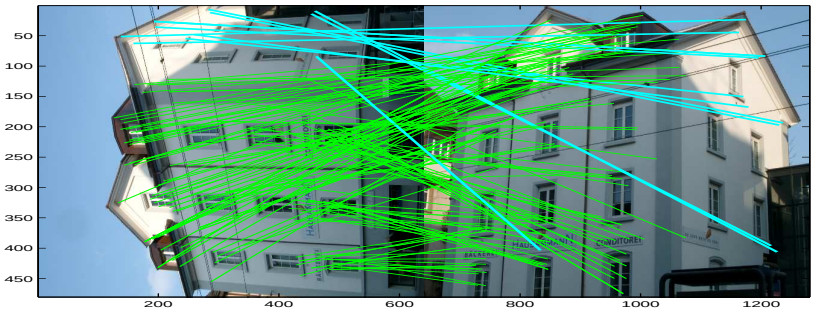
3 Experimental results

We will now present experimental results of our implementation of the algorithm. We ran experiments with the same settings on all the results included in this work. We used the publicly available ZubuD database [25] to test our method. The database contains 1005 color images of 201 buildings (5 images per building) of scenes in Zurich, taken from different viewpoints and illumination conditions.

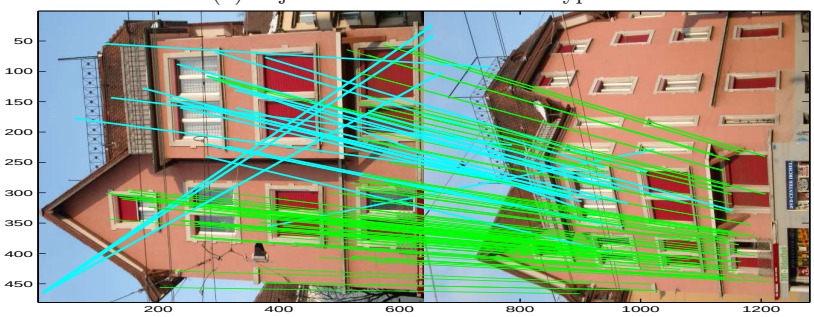
As we were interested in the additional value that our method can contribute, we compared it to the state-of-the-art wide baseline registration algorithms BLOGS [7] and BEEM [10], which can estimate the epipolar geometry in many difficult cases. We first automatically selected all the image pairs, that at least one of them failed to find a correct fundamental matrix for. We successfully ran our algorithm on 20 such image pairs of different buildings. Due to the low number of unique feature points in those images BLOGS succeeded only for 4 image pairs, whereas BEEM succeeded for 3. We also verified that a SIFT matching step followed by a standard RANSAC implementation failed on all image pairs. The results of running our algorithm on all of them, as well a comparison to other registration algorithms are included in the supplementary material submitted with this paper. They include the image pairs and a table presenting for each run numerical results of the various steps of the algorithm.



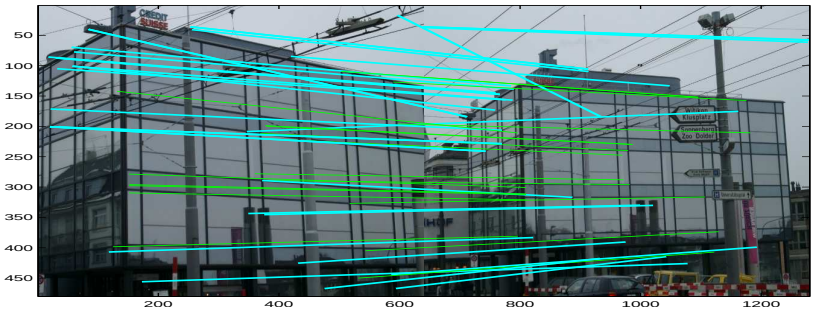
(a) object0010. Our method: H type.



(b) object0033. Our method: F type.



(c) object0066. Our method: F type.



(d) object0131. Our method: P-F type.

Fig. 6. Experimental results of our method on object0010, object0033, object0066 and object0131 from the ZuBuD database.

In Fig. 6 we present four representative results. For each image pair we present non-repeating key-points that are inliers of a fundamental matrix F. The key-points, that are also inliers of the homography H are connected with green lines and, those that were considered as F supporters are connected with the cyan lines.

In Fig. 6(a) we can see an example of a fully planar case, since there is only one building facade in the left image. As a result, there is an infinite number of fundamental matrices that could be chosen, one of which is shown here. In that case, as discussed earlier, the confidence in F is low due to a small number of its supporters (cyan lines) and we report only the recovered H with its inliers (green lines).

Fig. 6(b), on the contrary presents both images having two facades. As a consequence, we obtain a large number of key-points at different depths and report a fundamental matrix F along with its inliers. We can clearly see a color differentiation between on-plane and off-plane key-points. Key-points located on the plane are connected by green lines, whereas off-plane matches are in cyan.

A building with two parallel planes on the same facade is presented in Fig. 6(c). In this situation, the correct H, maps only one of the planes, whereas the key-points from the other have different depths and are colored in cyan. In this example the correct H maps key-points from the inner plane. The matches from the other plane are used to estimate the fundamental matrix correctly.

Finally, in Fig. 6(d) we demonstrate a periodic case. Here there are enough matches on the second facade to estimate the fundamental matrix correctly.

4 Conclusions and future work

In this paper we presented a wide baseline registration algorithm for scenes of building facades with repeating structures. The algorithm was implemented and tested successfully on a large number of image pairs for which general state-of-the-art algorithms usually fail.

Future research will be dedicated to developing an algorithm that can deal also with scenes of non-planar man-made objects and natural scenes with repeating objects.

Acknowledgment. This research was supported by the VULCAN consortium of the ministry of industry and commerce.

References

1. Lowe, D.: Distinctive image features from scale-invariant keypoints. IJCV **60** (2004) 91–110
2. Fischler, M., Bolles, R.: Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Communications of the ACM **24** (1981) 381–395
3. Chum, O., Matas, J., Kittler, J.: Locally optimized random sample consensus. In: German Pattern Recognition Symposium. (2003) 236–243

4. Tordoff, B., Murray, D.: Guided sampling and consensus for motion estimation. In: ECCV. (2002) 82–98
5. Chum, O., Matas, J.: Matching with PROSAC progressive sample consensus. In: CVPR. (2005) 220–226
6. Goshen, L., Shimshoni, I.: Guided sampling via weak motion models and outlier sample generation for epipolar geometry estimation. IJCV **80** (2008) 275–288
7. Brahmachari, A., Sarkar, S.: BLOGS: Balanced local and global search for non-degenerate two view epipolar geometry. In: ICCV. (2009) 1685–1692
8. Schaffalitzky, F., Zisserman, A.: Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?". In: ECCV. (2002) I: 414–431
9. Chum, O., Matas, J., Obdrzalek, S.: Enhancing RANSAC by generalized model optimization. In: ACCV. (2004) II: 812–817
10. Goshen, L., Shimshoni, I.: Balanced exploration and exploitation model search for efficient epipolar geometry estimation. PAMI **30** (2008) 1230–1242
11. Wu, C., Frahm, J., Pollefeys, M.: Detecting large repetitive structures with salient boundaries. In: ECCV. (2010) 142–155
12. Wenzel, S., Drauschke, M., Forstner, W.: Detection of repeated structures in facade images. PRAI **18** (2008) 406–411
13. Jiang, N., Tan, P., Cheong, L.: Multi-view repetitive structure detection. In: ICCV. (2011)
14. Liu, Y., Collins, R., Tsin, Y.: A computational model for periodic pattern perception based on frieze and wallpaper groups. PAMI **26** (2004) 354–371
15. Hays, J., Leordeanu, M., Efros, A., Liu, Y.: Discovering texture regularity as a higher-order correspondence problem. In: ECCV. (2006) 522–535
16. Lee, J., Yow, K., Chia, A.S.: Robust matching of building facades under large viewpoint changes. In: ICCV. (2009) 1258 – 1264
17. Baatz, G., Koser, K., Chen, D., Grzeszczuk, R., Pollefeys, M.: Handling urban location recognition as a 2D homothetic problem. In: ECCV. (2010) 266–279
18. Schindler, G., Krishnamurthy, P., Lublinerman, R., Liu, Y., Dellaert, F.: Detecting and matching repeated patterns for automatic geo-tagging in urban environments. In: CVPR. (2008) 1–7
19. Robertson, D., Cipolla, R.: An image-based system for urban navigation. In: BMVC. (2004) 819–828
20. Roberts, R., Sinha, S., Szeliski, R., Steedly, D.: Structure from Motion for Scenes with Large Duplicate Structures. In: CVPR. (2011) 3137–3144
21. Serradell, E., Ozuysal, M., Lepetit, V., Fua, P., Moreno-Noguer, F.: Combining geometric and appearance priors for robust homography estimation. In: ECCV. Volume 6313. (2010) 58–72
22. Rabin, J., Delon, J., Gousseau, Y., Moisan, L.: MAC-RANSAC: a robust algorithm for the recognition of multiple objects. In: 3DPVT. (2010)
23. Sur, F., Noury, N., Berger, M.O.: Image point correspondences and repeated patterns. Technical Report RR-7693, INRIA (2011)
24. Zhang, W., Kosecka, J.: Generalized RANSAC framework for relaxed correspondence problems. In: 3DPVT. (2006) 854–860
25. Shao, H., Svoboda, T., Gool, L.: ZuBuD Zurich Buildings Database for Image Based Recognition. In: Technical Report 260, CVL, ETH Zurich. (2003)
26. Hartley, R., Zisserman, A.: Mltiple View Geometry in Computer Vision. Second edn. Cambridge University Press (2004)
27. Vedaldi, A., Fulkerson, B.: VLFeat: An open and portable library of computer vision algorithms (2008)