Periodic Textures in Wide Baseline Stereo

Dmitry Chetverikov and Zsolt Jankó
Computer and Automation Research Institute and Eötvös Loránd University
1111 Budapest, Hungary
Email: csetverikov@sztaki.hu

***Abstract*— This study addresses the problem of finding correspondences for wide baseline stereo. Texture has traditionally been utilised as a single-image cue for 3D shape reconstruction (shape-from-texture); at the same time, its role in multiview scene reconstruction has been very limited. In stereo image matching, repetitive patterns are usually considered as disturbing factor since they tend to produce multiple peaks of correlation, which results in matching ambiguity. We argue that presence and proper analysis of distinct, compact periodic texture areas can facilitate wide baseline matching by providing periodic distinguished regions (PDRs) that efficiently constrain the search for correspondences. We demonstrate how PDRs can be used to find a few initial correspondences in a wide baseline stereo pair and to establish precise correspondences for building the epipolar geometry. Experimental results for various wide baseline stereo pairs are shown.**

I. Introduction

Finding correspondences is critical for fully automatic reconstruction of a 3D scene from two images taken from significantly different viewpoints and in varying illumination conditions. Such reconstruction is usually called wide baseline stereo. In this case, local image deformations cannot be approximated by translation and rotation; an affine model is more adequate. Photometric variations, shadows and self-occlusion make finding the correspondences even more difficult.

Regions selected for wide baseline matching should possess some distinguishing, highly invariant and stable property. In the recent paper [1], Matas et.al call such image elements *distinguished regions*, abbreviated as DRs. The authors introduce and apply a new type of non-texture DRs called maximally stable extremal regions. The number of such regions is usually large, while the values of the descriptors computed for them are quite unstable. A feature vector of high dimensionality is used, which provides input to a RANSAC-like epipolar geometry estimation procedure.

A similar scheme can be traced in most of the previous studies on wide baseline stereo. The paper [1] provides an up-to-date overview of previous work. In particular, regions for which invariant properties are defined are often based on Harris interest points [2], [3] or other elements such as groups of line segments [4] or curve features [5]. Local intensity extrema were previously considered in [6].

Affine invariance of region descriptors is achieved in different ways. A frequently used method was proposed by Baumberg [3], where rotation-invariant descriptors are computed after applying a transformation that diagonalises the covariance matrix of the intensity gradient of a region. Such transformation normalising region image up to a rotation is used in [1] and [7].

Texture has been traditionally utilised as a single-image cue for 3D shape reconstruction (shape-from-texture). At the same time, its role in multiview scene reconstruction has been limited. (By texture we mean statistically repetitive patterns such as those characterising surfaces of materials.) Most of the existing methods avoid using texture as a tool for matching, since repetitive patterns tend to correlate in multiple positions, which may lead to an ambiguity. Perhaps the first – and a rare – use of texture for wide baseline stereo is described in the paper of Schaffalitzky and Zisserman [7]. Their method works as follows.

The images are roughly segmented into regions (segments) of different texture. Affine normalisation of the segments up to rotation is then performed with a method similar to [3]. Putative inter-image correspondences between the texture segments are established using rotation-invariant texture features based on Zernike moments. The putative correspondences are verified by matching Harris interest points within the affine-normalised segments. Finally, more point matches are obtained for the verified segments, and the fundamental matrix is estimated.

In [7], the operation of the method is illustrated by building the epipolar geometry for the pair of the Valbonne images shown in figure 9. Most of the point matches come from the brick wall areas. Since the images are deliberately over-segmented to avoid grouping over depth discontinuities, the segmentation results in a large number (about 50) of unstable regions. 1600 interest points are computed in each image to verify the putative region correspondences. The method works in this particular case; however, given the homogeneity of textures, the idea of reducing ambiguity in region correspondences by matching points within textures does not seem to be sound. In fact, the point matching relies here on differences between the bricks, which might be too subtle to utilise.

By nature, texture is a global rather than a local property. It is unsuitable for establishing precise correspondences. We agree with Schaffalitzky and Zisserman [7] in that texture can serve as a constraint that limits the search for point correspondences. However, we believe that this constraint should be made more firm and strict – if necessary, at the price of less generality. Another point is that texture based constraints should support point matching *outside* regions of repetitive texture, since matching *inside* such regions will never be reliable. (In fact, as far as traditional homogeneous texture is concerned, the term 'texture matching' can only mean classification of two regions as same or different texture.)

In this paper, we limit the class of textures considered and show that the presence of compact periodic textures can facilitate wide baseline matching by providing the periodic distinguished regions (PDRs) that efficiently constrain the search for correspondences. Regularity features introduced earlier in [8] are used as affine-invariant descriptors to identify PDRs in a wide baseline stereo pair.

This study is the continuation of our previous research [9]. The novel *contribution of this paper* is as follows. We demonstrate that in certain cases three initial correspondences between PDRs can be used to obtain a 2D affine transformation that roughly aligns the two images. This allows us to apply a conventional matching technique to find a sufficient number of precise correspondences and build the epipolar geometry.

II. Outline of the method

In this section we informally discuss and illustrate the main steps of the proposed method. Then we present some geometric considerations that motivate and justify our method.

Consider a wide baseline stereo pair. Assume that a large *part* of the scene has the following property: relation between the two images of this part can be *roughly* approximated by a 2D affine transformation in the image plane. (Conditions when this is possible will be discussed later; weak perspective [10] is an example.) The transformation has 6 parameters, which means that 3 correspondences are required to estimate it. A key idea is that once an *approximate affine alignment* has been found and applied, the corresponding image features get much closer and a conventional matching algorithm can be used.

Figures 1a and 1b show two images from the RADIUS database [11]. The viewing geometry can loosely be treated as weak perspective, but the corresponding points are very far away. Assume that one can somehow identify three pairs of approximately corresponding points. Then one can roughly align the two images in the hope that many more corresponding features will become matchable.

Which elements of the images can be used to induce the rough alignment? The pair contains a few compact periodic texture regions which are suitable for this purpose. Finding and identifying in both images the regions marked in 1a, we establish three rough correspondences between the centroids of the regions. Then we estimate the 2D affine transformation that approximately aligns image 1b to image 1a. The result of the alignment is shown in figure 1c.

Due to approximate affine model and imprecision of the initial correspondences, the alignment is quite rough. Still, it brings the corresponding features into proximity, which is sufficient for successful application of a conventional feature matching algorithm. We use the well-known KLT feature tracker [12] that incorporates an efficient matching procedure tolerating a moderate affine distortion. The algorithm provides a significant number of good matches between the two approximately aligned images. Applying the inverse affine transformation, we obtain correspondences for the original pair. A standard robust algorithm [10] is then used to estimate the epipolar geometry. Sample epipolar lines are shown in the top row of figure 2. The bottom row demonstrates the rectified images which are suitable for dense matching.

(a) (b) (c)

Fig. 1. (a–b) A wide baseline RADIUS pair. The regions indicated in (a) are used to find the initial correspondences and estimate the affine alignment. (c) Rough affine alignment of (b) to (a).

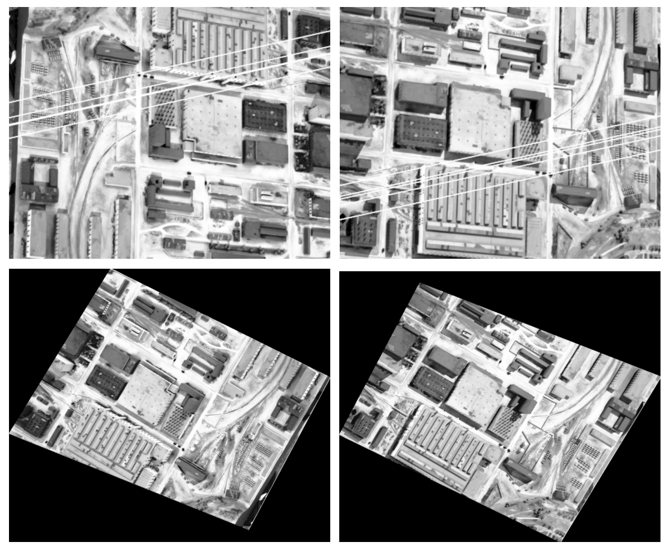

Fig. 2. Top row: Sample epipolar lines. Bottom row: Rectified images.

The idea of rough alignment is based on the assumption that the images of a large part of the scene can be approximately related by a 2D affine transformation in the image plane (2D affine homography). Precise analysis of the necessary and sufficient geometric conditions is beyond the scope of this paper. Presence of a large planar region is helpful but not sufficient. Hartley and Zisserman [10] show how the (projective) 2D homography induced by a plane can facilitate the estimation of the fundamental matrix when correspondences for two more additional off-the-plane points are known. The method is based on measuring the plane induced parallax of the two points.

Our method uses only three initial (region) correspondences and does not assume the presence of a plane. Instead, it implicitly assumes that the projections of the 3D points lying in a vicinity of the triangle formed by the three initial points will be brought into closer proximity by an affine alignment. This assumption is certainly reasonable when the depth variation of the scene is much smaller than the distance to the observer, that is, when the weak perspective camera model [10] is applicable. When the depth variation is relatively large, the parallax of the 3D points located far from the triangle will be large; such points will not be matched. However, other points may still match. Our experimental results demonstrate that the method can be successfully applied under viewing conditions more general than weak perspective.

III. The proposed algorithm

The three initial correspondences can in principle be obtained in different ways depending on the typical features of the images being considered. Our approach is completely automatic, but it needs the presence in the stereo pair of at least three compact periodic textures, referred to as periodic distinguished regions. The following framework algorithm summarises the proposed method.

Algorithm 1: Wide Baseline Stereo from PDRs

1) Detect periodic distinguished regions in both images and at varying resolution levels.
2) Describe PDRs by their affine-invariant regularity features and periodicity vectors.
3) Identify PDRs in the two images and establish tentative region correspondences.
4) Test each triple of region correspondences by computing rough affine alignment and applying a conventional matching algorithm to the roughly aligned pair.
5) Select alignment that produces the highest matching score and apply inverse transformation to obtain point matches for original pair.
6) Estimate fundamental matrix and rectify the images.
7) Refine point matches by local affine matching in rectified images.
8) Re-estimate fundamental matrix from refined point matches.

Steps 1–3 of the algorithm are described in our previous paper [9]. Step 1 involves computation of our regularity features [8] in a sliding window, which is called *regularity filtering*. Figure 3 exemplifies the regularity filtering of the images shown in the left column of figure 4. The maximal regularity feature [8] is displayed intensity-coded, with light regions having higher regularity. The two results were obtained for two different resolutions of an image and with different parameters of the filter.

Fig. 3. Examples of regularity filtering. The detected PDRs are shown in the left column of figure 4.

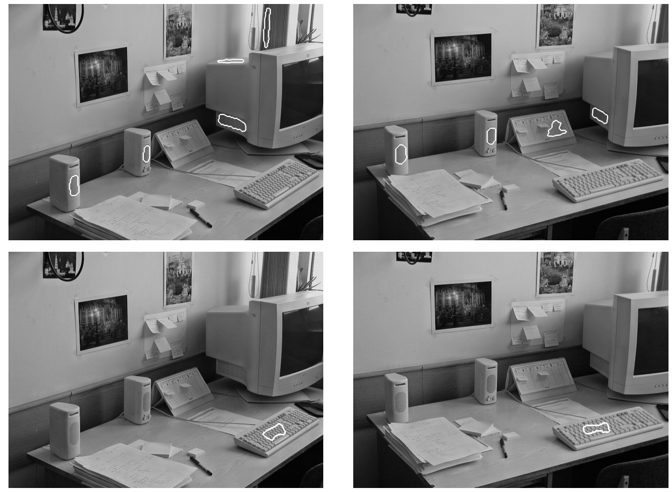

Fig. 4. The JANKO pair with PDRs detected at two resolution levels.

PDRs are detected in both images of a wide baseline stereo pair and at different resolution levels. In figure 4, fine periodic structures are detected at the initial resolution; the coarser structure of the keyboard appears at a lower resolution. Relatively large, dominant regions are normally detected. Their number is small. (Typically, less than 10.)

Search for correspondences involves two different but related operations: PDR matching and feature point matching. First, tentative PDR correspondences are established (step 3) as discussed in [9]. (Note that matching of PDRs from different resolution levels is prohibited.) Then each triple of tentative correspondences is tested and the best one is selected. Testing a triple includes computing the rough affine alignment for this triple, then applying the modified KLT Tracker [12] to the roughly aligned pair. To speed up the procedure, the KLT matching is done at a reduced image resolution. The best alignment is the one that produces the highest matching score, that is, the largest number of successfully matched feature points. Finally, the inverse affine transformation is applied to obtain point matches for the original pair.

The process of PDR matching is illustrated in figure 5 where the tentative correspondences obtained at both resolution levels are shown by lines connecting the centroids of the regions. Note that the same region may appear at different levels. Correspondences forming the best matching triangle are indicated by small circles marking the centroids. (Because of smaller image resolution, these circles are better visible in similar figures 11 and 12.)

Given three pairs of corresponding points, a rough affine alignment can be determined by solving a simple linear system. A modified version of the KLT Tracker [12] is applied to the roughly aligned pair. The original code has been modified to discard wrong matches (outliers) by reversing the order of the images and checking if the reverse match is sufficiently close to the original one. Since an alignment is either invalid or valid but still approximate, many of the original matches are discarded in all cases. When an alignment is invalid, only a

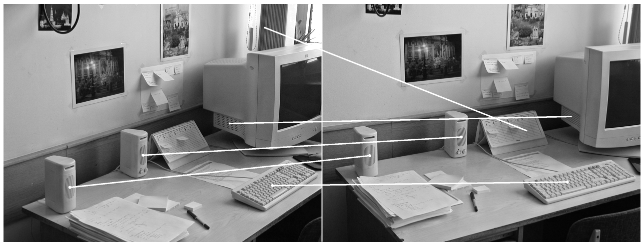

Fig. 5. PDR matching in the JANKO pair.

few spurious matches remain. When it is valid, the number of matches increases drastically, indicating that the tested triangle is a valid candidate. The valid candidate with the highest matching score is selected.

Figure 6 shows the matches obtained for the best candidate triangle marked in the two images. (See figure 4 for the PDRs detected in the pair.) Three valid candidates were available; still another one, formed by the two speakers and the ventilation grid of the monitor, fails to produce a usable alignment, since the three PDR centroids are almost collinear. Note that the successfully matched points lie close to the plane of the best triangle. The regions that are definitely off the plane, such as the pictures on the wall, were not matched because of the large parallax.

Fig. 6. Points matched in the JANKO pair.

When a sufficient number of point matches has been obtained, the fundamental matrix is estimated using a robust method such as RANSAC. A selection of efficient methods are available [10]. An example of the epipolar geometry built for the JANKO pair is demonstrated in figure 9. More results will be shown in section IV.

Based on the epipolar geometry, the images are rectified using a standard technique [10]. Then the point matches are *refined* in the rectified images in order to make the epipolar geometry more stable and precise. The task is twofold: to remove the outliers that may still be present, and to refine the locations of the matches. This is achieved by local affine pattern matching constrained by the epipolar geometry. An alternative solution is presented in [1]. Finally, the fundamental matrix is re-estimated from the refined point matches.

IV. WIDE BASELINE STEREO TESTS

In this section we show sample wide baseline stereo results obtained by the proposed method. All of the results were obtained automatically; some of the critical parameters, such as the number of the resolution levels and the regularity threshold, were manually selected for each particular stereo pair. The main goal of the tests was to demonstrate the feasibility of the method. No serious effort was made to optimise the algorithms, for example, for the precision of the epipolar geometry.

Figures 7 and 8 show sample epipolar lines and the rectified images, respectively, for the already mentioned JANKO pair. The PDRs concentrate in the central part of the image. Note that the keyboard is clearly a periodic *3D texture*; the speaker PDRs are also 3D textures, although the third dimension is smaller. The epipolar geometry obtained is quite precise; although it is seemingly close to affine, a projective method was used to estimate the fundamental matrix.

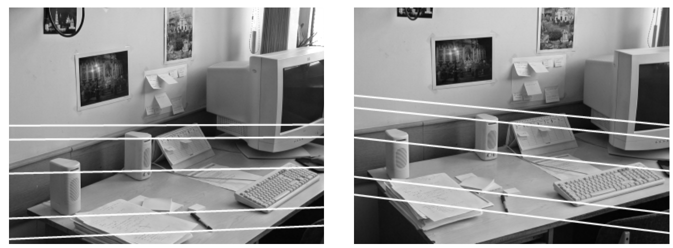

Fig. 7. Sample epipolar lines in the JANKO pair.

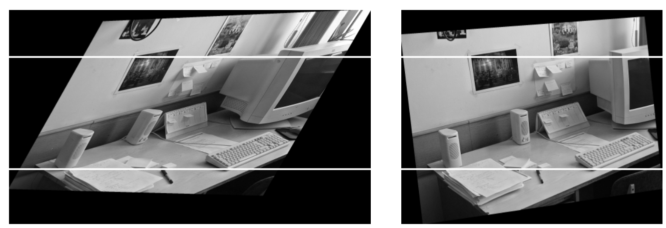

Fig. 8. Rectified JANKO pair with sample corresponding rows marked.

The well-known VALBONNE pair (figure 9) was also successfully processed, although its viewing geometry at first glance does not seem suitable for the proposed approach. In addition, most of the matched points concentrate in the almost planar region of the facade. Nevertheless, three PDRs, including the two pieces of the tower front wall, were successfully detected and identified. In this case, refining the point matches is essential, as it results in a much more stable solution. A single resolution level was used. Note that two of the PDRs are essentially the same brick wall texture. To find the proper correspondences, one has to test two alternative variants.

Finally, the CORNER pair shown in figure 10 was a challenging data because the periodic regions are not as distinct and compact as in the previous cases. Here, more triangles had to be tested before an acceptable alignment was found. Two resolution levels were used.

Processing time of the method depends on the size and content of the images. For the largest JANKO images (1600×1200), the regularity filtering at the initial resolution takes about 30 sec on a 2.4 GHz PC running Linux. Testing all

Fig. 9. Sample epipolar lines in the VALBONNE pair.

Fig. 10. Sample epipolar lines in the CORNER pair.

possible alignments may take up to 60 sec, but usually it is much faster.

V. SUMMARY AND CONCLUSIONS

It is obvious that the PDR method can only be applied in the presence of distinct periodic structures, and that PDRs provide constraints rather than precise correspondences. The success of the method depends on whether the assumptions of partial affine alignment are met; if there are few feature points in the roughly aligned area, the method might fail. A related potential drawback is that the matched 3D points tend to concentrate near the plane defined by the best matching triangle. If the area of the triangle is small, or if any of its heights is small, the alignment might be of no use.

If for some reasons the tentative PDR correspondences are poor and no valid alignment is found, it may be necessary to extend the testing beyond the tentative PDR correspondences. For example, it may happen that the magnifications (levels of detail) of two images are significantly different. Then PDRs should be identified at *different scales*, which needs scale-invariant texture classification. As periodicity is, in general, scale-dependent, the scale-invariance cannot be guaranteed. However, it is still highly probable that the PDRs will be detected. In such cases one should not rely on the tentative correspondences, and all reasonable candidates must be tested.

This is computationally feasible since the number of PDRs is small. This is illustrated in figures 11 and 12 that show the tentative correspondences considered in these two pairs. (As before, the small circles indicate the best matches.) Additional constrains may be imposed to speed up the procedure by discarding small or narrow triangles, as well as the correspondences that reverse the order of the vertices.

Fig. 11. PDR matching in the VALBONNE pair. The borders of the detected PDRs are shown.

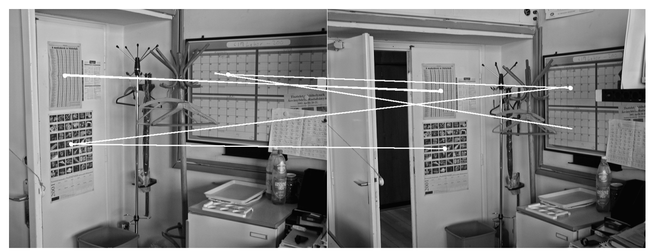

Fig. 12. PDR matching in the CORNER pair. The borders of the PDRs are not shown for clarity.

An advantage of the proposed approach is that it is based on patterns of perceptual value. Matching feature points within textures, or matching a large number of very small, unstable regions are avoided when only possible. The PDRs are stable dominant regions that can be relied upon. Although the regularity features of a periodic 3D texture region are not in general affine invariant, the region is still easy to detect and test against a small number of regions in the other image.

To summarise the contribution of this study, we hope that the results presented clearly demonstrate that the periodic distinguished regions can be used as an efficient aid to wide baseline stereo. Further analysis is needed to clarify the assumptions of the method and to evaluate its performance and precision.

ACKNOWLEDGMENTS

This work was supported by the Hungarian Scientific Research Fund (OTKA) grants T038355 and M28078.

References

[1] J. Matas, O. Chum, M. Urban, and T. Pajdla, "Robust Wide baseline Stereo from Maximally Stable Extremal Regions," in *Proc. of the British Machine Vision Conference*, P. Rosin and D. Marshall, Eds., vol. 1, 2002, pp. 384–393.

[2] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *PAMI*, vol. 19, no. 5, pp. 530–535, May 1997.

[3] A. Baumberg, "Reliable feature matching across widely separated views," in *Proc. Conf. on Computer Vision Pattern Recognition*, 2000, pp. I:774–781.

[4] P. Pritchett and A. Zisserman, "Wide baseline stereo matching," in *Proc. Int. Conf. on Computer Vision*, 1998, pp. 754–760.

[5] T. Tuytelaars and L. V. Gool, "Content-based image retrieval based on local affinely invariant regions," in *Proc* 3^{rd} *Int. Conf. on Visual Information Systems*, 1999, pp. 493–500.

[6] ——, "Wide baseline stereo based on local, affinely invariant regions," in *Proc. British Machine Vision Conf.*, 2000, pp. 412–422.

[7] F. Schaffalitzky and A. Zisserman, "Viewpoint invariant texture matching and wide baseline stereo," in *Proc. Int. Conf. on Computer Vision*, 2001.

[8] D. Chetverikov, "Pattern regularity as a visual key," *Image and Vision Computing*, vol. 18, pp. 975–985, 2000.

[9] D. Chetverikov and J. Matas, "Periodic Textures as Distinguished Regions for Wide-Baseline Stereo Correspondence," in *Proc.* 2^{nd} *International workshop on texture analysis and synthesis Texture'2002*, Coppenhagen, 2002, pp. 25–30, available at `http://www.cee.hw.ac.uk/texture2002/`.

[10] R. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*. Cambridge, UK: Cambridge University Press, 2000.

[11] *RADIUS Model Board Imagery Database*, University of Washington, Seattle, WA.

[12] S. Birchfield, *KLT: An Implementation of the Kanade-Lucas-Tomasi Feature Tracker*, Stanford University, `http://robotics.stanford.edu/~birch/klt/`.