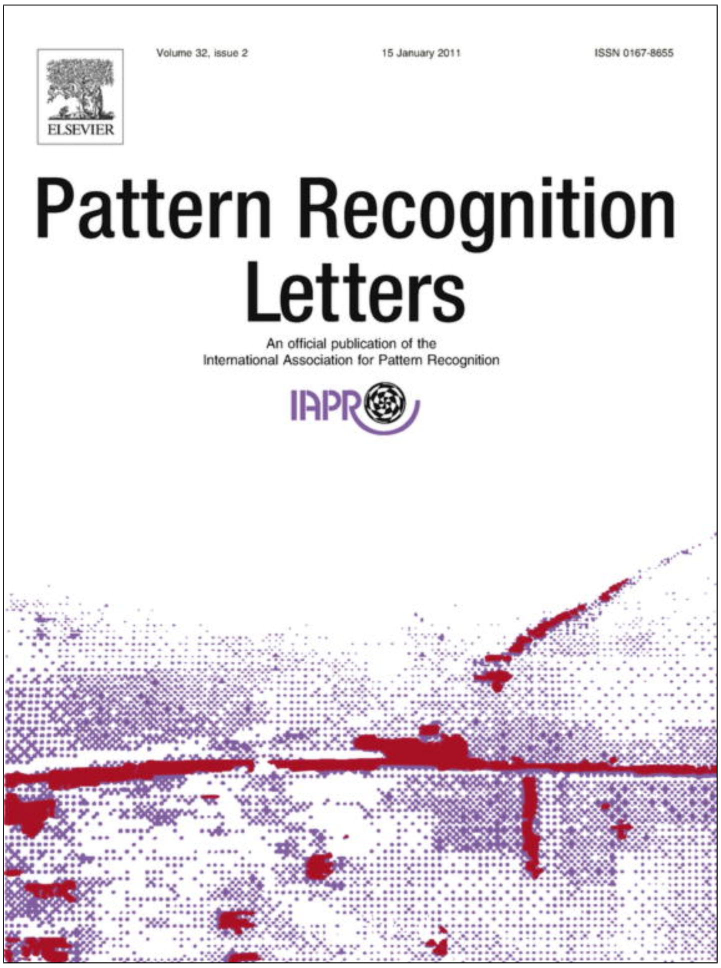

Volume 32, issue 2
15 January 2011
ISSN 0167-8655
ELSEVIER
Pattern Recognition Letters
An official publication of the
International Association for Pattern Recognition
IAPR

Pattern Recognition Letters 32 (2011) 1851–1859

Contents lists available at SciVerse ScienceDirect

Pattern Recognition Letters

journal homepage: www.elsevier.com/locate/patrec

Towards reliable matching of images containing repetitive patterns

Bin Fan *, Fuchao Wu, Zhanyi Hu

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, 100190 Beijing, China

ARTICLE INFO

Article history:
Received 14 September 2010
Available online 16 August 2011
Communicated by S. Todorovic

Keywords:
Image matching
Repetitive patterns
Local descriptors
Point matching
Geometric consistency

ABSTRACT

This paper aims to solve the problem of matching images containing repetitive patterns. Although repetitive patterns widely exist in real world images, these images are difficult to be matched due to local ambiguities even if the viewpoint changes are not very large. It is still an open and challenging problem. To solve the problem, this paper proposes to match pairs of interest points and then obtain point correspondences from the matched pairs of interest points based on the low distortion constraint, which is meant that the distortions of point groups should be small across images. By matching pairs of interest points, local ambiguities induced by repetitive patterns can be reduced to some extent since information in a much larger region is used. Meanwhile, owing to our newly defined compatibility measure between one correspondence and a set of point correspondences, the obtained point correspondences are very reliable. Experiments have demonstrated the effectiveness of our method and its superiority to the existing methods.

1. Introduction

In many computer vision tasks, such as 3D reconstruction (Agarwal et al., 2009; Furukawa and Ponce, 2010), structure from motion (Snavely et al., 2006), object recognition (Lowe, 2004), texture classification (Zhang et al., 2007) and image stitch (Brown and Lowe, 2003; Szeliski, 2006), one of the core steps is to establish reliable correspondences of points between two images. The most popular matching methods may be those based on local descriptors (Lowe, 2004; Mikolajczyk and Schmid, 2005; Tang et al., 2009; Tola et al., 2010; Gupta and Mittal, 2008; Heikkila et al., 2009; Brown et al., 2010). By local descriptors, point matching becomes robust to scale changes, rotation changes, illumination changes as well as viewpoint changes to some extent.

However, these methods usually fail when the matching images containing repetitive patterns that confuse matches, cf Fig. 1(a). Although such repetitive patterns widely exist in images of man-made objects, such as buildings, these images are hard to be matched with the current techniques due to local ambiguities, even if the viewpoint does not change too much. This paper is primarily focused on matching of this kind of images, i.e. images containing repetitive patterns. Different from the traditional matching framework, which is to match individual points by directly comparing their local descriptors, this paper presents a novel method based on matching pairs of interest points and defines a new compatibility measure in order to obtain reliable corresponding points from the tentative matched pairs of interest points between two images. The newly defined compatibility measure is based on the assumption that the spatial arrangement of point groups does not change too much across images. Pairs of points are generated from interest points and then they are matched by comparing their descriptors. Since the main challenge in matching images with repetitive patterns is the local ambiguities induced by repetitive patterns, describing pairs of interest points can utilize information in a much larger region, thus the local ambiguities may be reduced to some extent compared with previous work (Lowe, 2004; Mikolajczyk and Schmid, 2005; Tang et al., 2009; Tola et al., 2010; Gupta and Mittal, 2008; Heikkila et al., 2009). Meanwhile, owing to our newly defined compatibility measure between one correspondence and a set of point correspondences, consistent point correspondences are established reliably from the tentative matched point-pairs. As can be seen from Fig. 1(b), our proposed method can match images containing repetitive structures successfully. The effectiveness of our proposed method and its superiority to the state-of-the-art methods have been tested on images containing repetitive patterns (structures and textures) of both planar and 3D scenes (indoor and outdoor).

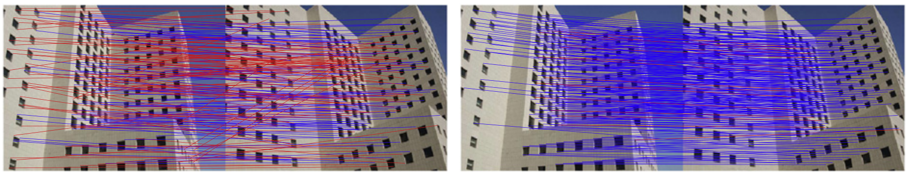
The remainder of this paper is organized as follows. Related work is discussed in Section 2. Then how to match pairs of interest points is elaborated in Section 3, followed by the algorithm to obtain reliable point correspondences from the tentative matched pairs of interest points in Section 4. Experiments are reported in Section 5, and Section 6 concludes this paper.

2. Related work

Repetitive patterns are widely existed in natural and man-made scenes. Detection of such repeated elements/texels in images has

* Corresponding author. Tel.: +86 10 62542946; fax: +86 10 62551993.
E-mail address: bfan@nlpr.ia.ac.cn (B. Fan).
URL: http://vision.ia.ac.cn/Students/bfan/index.htm (B. Fan).

0167-8655/$ - see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.patrec.2011.07.029



(a) (b)

Fig. 1. Comparison of matching results for images in Fig. 5. Blue lines indicate correct matches while red lines indicate incorrect ones. (a) Matching results by local descriptor based method (DAISY Tola et al., 2010), 113/259 correct matches (43.6%). (b) Matching results by our proposed method, 344/360 correct matches (95.6%). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

been widely studied in the computer vision and graphics community (Liu et al., 2004; Pauly et al., 2008; Cheng et al., 2010; Ahuja and Todorovic, 2007; Todorovic and Ahuja, 2009). Leung and Malik (1996) proposed a greedy algorithm to extract repetitive elements by constructing a graph with individual elements as nodes and their affine transforms as edges. Hays and Leordeanu (2006) proposed to discover texture regularity by formulating it as a higher-order correspondence problem which is solved by a spectral method. Ahuja and Todorovic (2007) represented the image by a segmentation tree and extracted texels in 2.1D textures by learning substructures in the segmentation tree. Cheng et al. (2010) proposed a framework to find approximately repeated elements for image editing with a litter user interaction.

Due to the ambiguities induced by repeated elements, matching images with repetitive patterns is not an easy task. The traditional local descriptor-based matching methods, which match interest points by comparing their local descriptors, can not achieve satisfactory results when dealing with such images. Mortensen et al. (2005) incorporated global context into SIFT descriptor to match interest points in images of repetitive patterns. The global context is computed from curvilinear shape information in a much larger neighborhood of interest point to disambiguate the confusion induced by repetitive patterns. Another attempt about matching such images is to explore the inter-relationship among correspondences since the geometric relationship of correct correspondences should be preserved across images is a reasonable assumption (Leordeanu and Hebert, 2005; Berg et al., 2005; Duchenne et al., 2009; Choi and Kweon, 2009). Tell and Carlsson (2002) incorporated topological constraints in their matching framework and reported the improved results. Leordeanu and Hebert (2005) used a spectral technique to solve the correspondence problem using pairwise constraints between correspondences by constructing a graph with potential correspondences as nodes and the pairwise agreements between potential correspondences as edges. Duchenne et al. (2009) used tensors to generalize the idea of spectral method, thus higher-order constraints (such as relationship between tuples) can be incorporated in the correspondence problem. Berg et al. (2005) used low distortion correspondences, which are meant that the geometric relationships (in their paper are distance and angle) between correct correspondences should not changed too much across images, for shape matching and object recognition.

The idea of matching images with interest point groups is not a new one in the computer vision community. Brown and Lowe (2002) used invariant features computed from interest point groups for image matching. The matched point groups are refined by a Hough transform (Ballard, 1981) followed by RANSAC (Fischler and Bolles, 1981) to compute the fundamental matrix as well as to find a final set of point correspondences that are consistent with epipolar geometry. Tell and Carlsson (2000) used the Fourier coefficients of the intensity profile between two points in a pair to describe a pair of interest points and matched interest point-pairs by comparing their Fourier coefficients. The final set of point correspondences is obtained from the tentative matched point-pairs using a voting strategy.

Our work is basically different from the previous ones. Firstly, although we use pairs of interest points for image matching, an improved version of DAISY (Tola et al., 2010; Winder et al., 2009) is used for describing point-pairs (Section 3.2). Secondly, the voting strategy used in (Tell and Carlsson, 2000) for obtaining point correspondences from the tentative matched point-pairs does not take into account the inter-relationship between correct correspondences, while our method generates point correspondences from the matched point-pairs under the low distortion constraint between point correspondences (Section 4). Note that here the low distortion constraint is used in a different way from that in (Berg et al., 2005). In our work, a measure of compatibility between one correspondence and a set of correspondences is defined based on the low distortion constraint so as to reliably obtain point correspondences from the matched point-pairs. While in (Berg et al., 2005), the low distortion constraint is used in formulating objective function which results in their optimization problem be an integer quadratic programming problem that is computational expensive.

3. Matching pairs of interest points

The proposed method starts from matching pairs of interest points, and the final set of point correspondences is then obtained from the tentative matched pairs of interest points on the basis of geometrical consistency. In order to get a number of tentative matched point-pairs for an image pair, we need to generate, describe and match pairs of interest points successively.

3.1. Generating pairs of interest points

Given a set of interest points in an image, intuitively, any two of them can form one pair. However, such a way to form pairs of points is redundant and will produce too many pairs since the number of pairs is quadratic in the number of points. If the number of points is large, then the computational burden of matching will be high. Therefore, the number of pairs should be limited. For one point, there is no need to choose all the other points to form pairs with it. As we will describe in the next subsection, the direction decided by the two points in a pair is used to construct descriptor for this pair in order to be rotation invariant. Therefore, the two points formed a pair should not too far from each other since the rotation invariance holds locally in real images. In the other aspect, if the two points of a pair are near, they do not contain more information than using only one of them, thus may reducing the distinctive ability of pairs of interest points. As a result, for each interest point X_i, pairs are formed by X_i and the other interest points $X_j \in \mathrm{SubR}(X_i)$, in which $\mathrm{SubR}(X) = \{Y : \sigma_1 \leqslant \|X - Y\|_2 < \sigma_2\}$. Setting σ_1 and σ_2 is not trivial and they are typically set to 50 pixels and 100 pixels respectively in our experiments. Therefore, for a given set of interest points $\{X_i, i = 1, 2, \ldots, n\}$, pairs of interest points are formed as: $\{(X_i, X_j) : X_j \in \mathrm{SubR}(X_i),\ i,\ j = 1, 2, \ldots, n\}$in which n is the number of interest points.

3.2. Constructing descriptors for pairs of interest points

Local descriptors have been developed rapidly in the past decade (Mikolajczyk and Schmid, 2005; Tang et al., 2009; Gupta and Mittal, 2008; Heikkila et al., 2009). They are designed to be distinctive while robust to many image transformations including illumination changes, rotation changes, scale changes and viewpoint changes. In this paper, we use local descriptors to describe pairs of interest points. Two descriptors of the two points in a pair are concatenated together to form the descriptor of this pair. Recently, DAISY (Tola et al., 2010) is proposed and reported with better performance than SIFT as well as other local descriptors (Winder et al., 2009; Brown et al., 2010). Besides, it can be computed very efficiently and the source code is available on the Internet.[1] Therefore, we choose DAISY to describe pairs of interest points.

Similar to SIFT, DAISY is obtained by accumulating distributions of gradient orientation in the local neighborhood of an interest point. As shown in Fig. 2, the crosses represent sample points in the neighborhood of the interest point. For each sample point, a vector is obtained by convolving its neighboring points' gradient orientation maps with a Gaussian kernel. Then all the vectors of sample points are concatenated together as the DAISY descriptor of the interest point. The circles around the crosses are proportional to the standard deviation of their Gaussian kernels. In order to be rotation invariant, the DAISY descriptor is computed relative to a direction. The reader is referred to Tola et al. (2010) and Winder et al. (2009) for more details about DAISY.

Note that a direction is required to be assigned for each interest point as shown in Fig. 2, in order to make the descriptor be rotation invariant. Such a direction is usually assigned by a histogramming technique (Lowe, 2004). In brief, the orientation corresponding to the largest peak in the orientation histogram within a region around the interest point is taken as the direction of this point. Any other local peak that is within 80% of the highest peak is also used as direction for descriptor extraction. Different from one single point, a pair of points supplies a direction inherently, i.e. the direction from the first point to the second point. Therefore, in this paper the direction decided by the two points of a pair is used to extract descriptor of the pair as shown in Fig. 3. Thus the DAISY descriptors of the two points in a pair is not identical to those extracted by the original DAISY method due to different direction assignment. In the following, we would like to experimentally show that the direction assigned according to the two points of a pair is much more stable than the direction assigned by the histogramming technique. Therefore, the DAISY descriptors extracted by our method are much more robust than the original ones. We have obtained statistics of direction errors of corresponding interest point-pairs for images download from the Oxford dataset.[2] Specifically, for each image, we applied 100 random synthetic affine warps to it along with 1 unit of additive Gaussian noise. Then we detected DoG interest points and computed their directions in all images. Thus, according to the synthetic affine transform of each pair of reference and warped image, corresponding points can be obtained. Finally, we generated pairs of interest points from these corresponding points and histogrammed direction errors of the corresponding pairs of interest points. Let us denote a corresponding pair of interest points as $\{(X_1^r, \theta_1^r; X_1^w, \theta_1^w), (X_2^r, \theta_2^r; X_2^w, \theta_2^w)\}$ in which $(X_i^r, \theta_i^r; X_i^w, \theta_i^w)$, $i = 1, 2$ are two correspondences of interest points between the reference image and the warped image, and X_i^r, θ_i^r denote the position and direction of one interest point in the pair in the reference image, X_i^w, θ_i^w are the position and direction of its corresponding point in the warped image. Then the direction error of this corresponding pair of interest points can be calculated as:

$$\varepsilon_i = \theta_i^w - f(\theta_i^r : H), \quad i = 1, 2, \tag{1}$$

where H is the affine transform between the reference image and the warped image and $f(\theta : H)$ is a function that warps θ from the reference image to the warped image according to H. Fig. 4 shows the statistics of direction errors with our method and the traditional method respectively. It is obviously that using the traditional method to assign directions for each pair of interest point has large errors while our method is much more robust. When assigning directions according to the two points of a pair, the direction errors of most corresponding pairs are no more than 3 degrees.

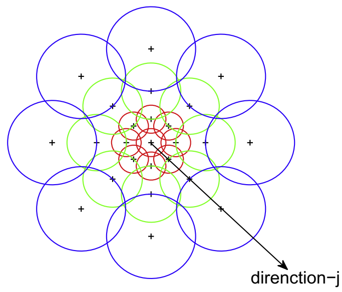


Fig. 2. The DAISY descriptor (Tola et al., 2010).

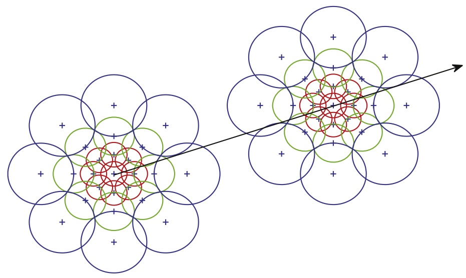


Fig. 3. Descriptor for a pair of points.

3.3. Matching pairs of interest points

Given two images, a set of interest point-pairs can be detected in each image, and descriptors can be constructed for those pairs. Similar to point matching based on local descriptors, interest point-pairs are matched by the nearest neighbor of the distances of their descriptors. The distance between descriptors of two pairs is computed by the Euclidean distance. Although with a threshold on the nearest neighbor distance one can reject many mismatches, more correct matches will also be rejected. As we will introduce in the next section, an algorithm is proposed in this paper in order to obtain a final set of point correspondences from all tentative matched point-pairs based on the low distortion constraint. Most of mismatches will be rejected by such an algorithm, hence we do not need to set a threshold for matching pairs and all the nearest neighbor matches are taken as the tentative matched pairs of interest points. The set of tentative matched point-pairs is denoted as $\mathcal{P} = \{(M_{i1}, M_{i2}, s_i), i = 1, 2, \ldots, N\}$ in which N is the number of matched point-pairs. In each matched pair of interest points, M_{i1}, M_{i2} are formed by the points in the matched pair and either is a tentative correspondence of points between two images. s_i is the

[1] http://cvlab.epfl.ch/~tola/daisy.html.
[2] http://www.robots.ox.ac.uk/~vgg/data/data-aff.html.

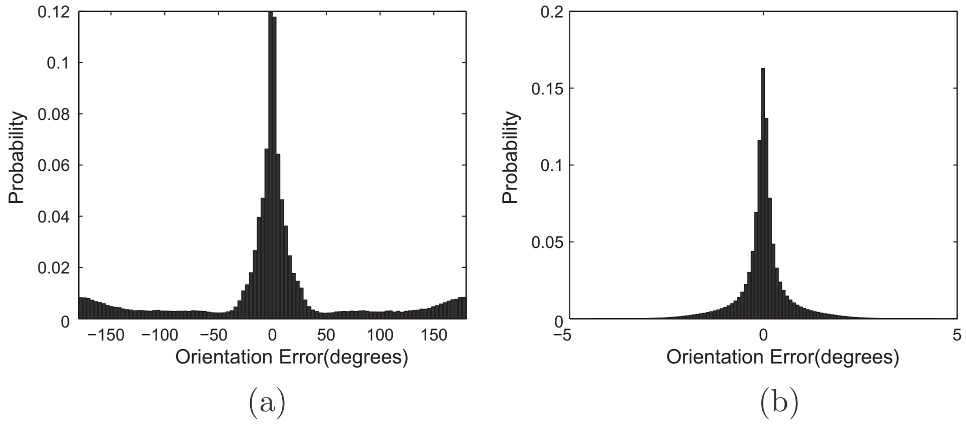


Fig. 4. Statistics of direction errors between corresponding interest point-pairs. (a) Directions are assigned by the traditional method, i.e. histogramming technique; (b) Directions are assigned by the proposed method in this paper.

Table 1
Numbers of detected interest points, candidate point correspondences and inlier point correspondences in the tested images of Fig. 6. See text for details.

(a)	(b)	(c)	(d)	(e)	(f)	(g)	(h)
Detected interest points							
(352,334)	(494,499)	(495,493)	(487,484)	(197,231)	(500,500)	(498,498)	(495,489)
Candidate point correspondences							
3197	5657	4103	3554	1192	4899	7345	5596
Inlier point correspondences							
270	302	347	312	174	348	323	331

nearest neighbor distance ratio (NNDR) (Mikolajczyk and Schmid, 2005) of this matched pair. The s_i of each matched pair can be considered as its confidence to be a correct one. The smaller the s_i is, the higher confidence of this matched pair to be a correct one. We choose the NNDR but not the nearest neighbor distance because the NNDR is more reliable and has the ability to discard ambiguous matches to some extent. Such a set of tentative matched pairs of interest points is used to obtain point correspondences in the subsequent step.

4. Obtaining point correspondences from the tentative matched pairs of interest points

Once the tentative matched point-pairs have been generated, we need to obtain a final set of point correspondences since our purpose is to establish point correspondences between two images. For each matched point-pair (M_{i1}, M_{i2}, s_i), it contains two tentative correspondences of points (i.e. M_{i1}, M_{i2}) formed by the points in the matched pair. However, these tentative point correspondences usually contain large amount of outliers, e.g. in our experiments about 90% outliers exist (see Table 1). Therefore, we propose a method exploring the inter-correspondence geometric relationship in order to establish point correspondences reliably. It takes the assumption that the spatial arrangement of point groups does not change too much across images.

4.1. The low distortion constraint

As shown in Fig. 5, $\left(X_i^1, X_i^2\right)$ and $\left(X_k^1, X_k^2\right)$ are two correspondences of points, the geometric relationship (such as distance) between X_i^1 and X_k^1 in the left image does not deviate too much from the relationship between X_i^2 and X_k^2 in the right image. Such an observation is called low distortion constraint. More specifically, given two point correspondences, $M_i = \left(X_i^1, X_i^2\right)$ and $M_k = \left(X_k^1, X_k^2\right)$, they are said to be compatible mutually if they are low distortion correspondences, i.e. they satisfy the following constraints:

$$\left| \left\| X_i^1 - X_k^1 \right\|_2 - \left\| X_i^2 - X_k^2 \right\|_2 \right| \leqslant t, \tag{2}$$

$$\left\| X_i^1 - X_k^1 \right\|_2 \leqslant t_n \quad \text{or} \quad \left\| X_i^2 - X_k^2 \right\|_2 \leqslant t_n, \tag{3}$$

where $\left| \left\| X_i^1 - X_k^1 \right\|_2 - \left\| X_i^2 - X_k^2 \right\|_2 \right|$ measures the deformation across images between two correspondences of points. The t controls the

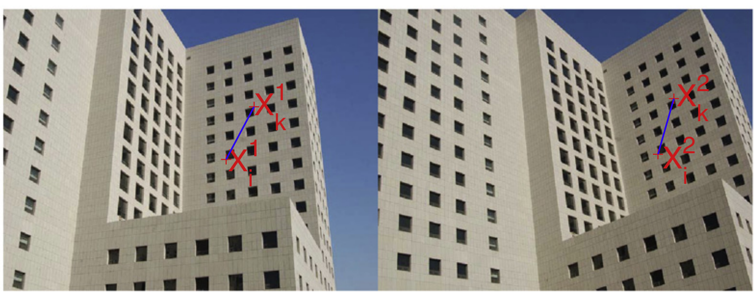


Fig. 5. Illustration of low distortion correspondences.

sensitivity on deformations. The larger t is, the more inter-image deformations can accommodate, also incorrect point correspondences may more likely to be considered as compatible. It is worth noting that (3) imposes the constraint of low distortion on neighboring correspondences (at least in one image), because the assumption of low distortion may not hold when point correspondences are far from each other. In all of our experiments, t is set to 15 pixels while t_n is set to 50 pixels. For convenience, we define an indicator function of compatibility as:

$$g(M_i, M_k) = \begin{cases} 1, & \text{if (2) (3) hold}, \\ 0, & \text{otherwise}. \end{cases} \tag{4}$$

4.2. Compatibility measure between one correspondence and a set of point correspondences

Having the definition of compatibility between two point correspondences, the compatibility between one correspondence and a set of point correspondences can be further defined, which leads to our proposed algorithm of obtaining a set of point correspondences from the tentative matched point-pairs. Given an initial set of point correspondences, reliable point correspondences are successively added to the set from the tentative matched pairs of interest points according to their compatibilities to the set. Suppose we have one corresponding point $M = (X^1, X^2)$ and a set of point correspondences, denoted as $\mathcal{C} = \{M_i, i = 1, 2, \ldots, m\}$, then define

$$f_c(M, \mathcal{C}) = \frac{\sum_{M_i \in \mathcal{C}} g(M_i, M)}{\sum_{M_i \in \mathcal{C}} f(M_i, M)}, \tag{5}$$

where

$$f(M_i, M) = \begin{cases} 1, & \text{if } \left\|X_i^1 - X^1\right\|_2 < t_n \text{ or } \left\|X_i^2 - X^2\right\|_2 < t_n, \\ 0, & \text{otherwise}, \end{cases} \tag{6}$$

indicates whether M_i and M are neighboring correspondences.

Actually, $f_c(M, \mathcal{C})$ is the ratio of the number of correspondences that are compatible with M to the number of M's neighboring correspondences. It can be considered as a qualitative measure of the compatibility between M and $\mathcal{C}$, and it reflects that how well the point correspondence M is compatible with the set $\mathcal{C}$.

4.3. The algorithm

Algorithm 1: ParseMatchedPair ($\mathcal{P}$)

Input:
  A set of tentative matched pairs
  $\mathcal{P} = \{(M_{i1}, M_{i2}, s_i), i = 1, 2, \ldots, N\}$.
Output:
  A set of point correspondences $\mathcal{C} = \{M_i, i = 1, 2 \ldots, m\}$.
1: Initialization: $\mathcal{U} = \emptyset, \mathcal{C} = \{M_{j1}, M_{j2}\}, \mathcal{P} = \mathcal{P} - (M_{j1}, M_{j2}, s_j)$.
2: **while** $\mathcal{P} \neq \emptyset$ **do**
3:   Select (M_{i1}, M_{i2}, s_i) such that $s_i \leqslant s_k, \ \forall (M_{k1}, M_{k2}, s_k) \in \mathcal{P}$.
4:   $\mathcal{P} = \mathcal{P} - (M_{i1}, M_{i2}, s_i)$.
5:   **if** $\sum_{M \in \mathcal{C}} f(M_{i1}, M) > 0$ and $\sum_{M \in \mathcal{C}} f(M_{i2}, M) > 0$ **then**
6:     **if** $f_c(M_{i1}, \mathcal{C}) \geqslant 0.85$ and $f_c(M_{i2}, \mathcal{C}) \geqslant 0.85$ **then**
7:       $\mathcal{C} = \{\mathcal{C}, M_{i1} : M_{i1} \notin \mathcal{C}\}, \mathcal{C} = \{\mathcal{C}, M_{i2} : M_{i2} \notin \mathcal{C}\}$
8:     **end if**
9:   **else**
10:     $\mathcal{U} = \{\mathcal{U}, (M_{i1}, M_{i2}, s_i)\}$
11:   **end if**
12: **end while**
13: $\mathcal{P} = \mathcal{U}, \ \mathcal{U} = \emptyset$
14: Repeat 2–13 until $\mathcal{P} = \emptyset$ or the size of $\mathcal{C}$ does not change any more.
15: **if** The size of $\mathcal{C}$ is very small **then**
16:   $\mathcal{C} = \emptyset$
17: **end if**
18: **if** $\mathcal{P} \neq \emptyset$ **then**
19:   $\mathcal{C} = \mathcal{C} \cup \text{ParseMatchedPair}(\mathcal{P})$
20: **end if**
21: return $\mathcal{C}$

Based on the definition of compatibility measure between one correspondence and a set of correspondences as in Eq. (5), an algorithm is proposed in order to reliably obtain point correspondences from the tentative matched pairs of interest points which is outlined in Algorithm 1. The input is a set of tentative matched pairs $\mathcal{P}$ while the output is a set of point correspondences $\mathcal{C}$. Firstly, the set of point correspondences $\mathcal{C}$ is initialized by one matched pair (such as the matched pair with the highest confidence, assumed to be the jth matched pair), that is $\mathcal{C} = \{M_{j1}, M_{j2}\}$. Once $\mathcal{C}$ has been initialized, selecting and removing the matched pair (M_{i1}, M_{i2}, s_i) from $\mathcal{P}$ with the highest confidence, i.e. the lowest NNDR, then checking its compatibility with $\mathcal{C}$. If both M_{i1} and M_{i2} are compatible with $\mathcal{C}$ in a high score, they are added into $\mathcal{C}$. Here we would note that since the compatibility measure (Eq. (5)) is defined on the basis of point correspondences belonging to $\mathcal{C}$ that are also located in the t_n-neighborhood of M, there may be no such correspondence at all. In this case, Eq. (5) is meaningless. Therefore, if either two tentative point correspondences (M_{i1} and M_{i2}) in the selected matched pair does not have any t_n-neighboring correspondence in $\mathcal{C}$, (M_{i1}, M_{i2}, s_i) is added to another auxiliary set $\mathcal{U}$. After all the matched pairs in $\mathcal{P}$ have been processed, the procedure is repeated with the matched pairs in $\mathcal{U}$ until convergence. It typically needs less than 5 iterations to converge according to our experiments. If non or very few point correspondences are added to $\mathcal{C}$ after initialization which is meant that the size of $\mathcal{C}$ is very small (e.g. less than a small proportion of the number of detected interest points), it is very possible that the $\mathcal{C}$ is initialized with incorrect correspondences. Therefore, we discard it in this case (line 15–17 in Algorithm 1). Meanwhile, since such a greedy strategy tends to generate point correspondences in the same cluster as the initialized correspondences, we continue the process to obtain point correspondences from the remaining matched pairs (line 18–20 in Algorithm 1). As the NNDR of a matched pair reflects its confidence to be a correct one, it is reasonable to initialize $\mathcal{C}$ with the matched pair whose NNDR is very low. To reduce the risk that the matched pair with the lowest NNDR is actually wrong, each of the first 5 lowest matched pairs is used to initialize $\mathcal{C}$ separately. The matching result with the largest number of obtained point correspondences is served as the final set of point correspondences.

5. Experimental results

We have conducted experiments to validate the effectiveness of our proposed method. As shown in Fig. 6, eight pairs of images are used for experimental evaluation. All of them contain either repetitive structures or repetitive textures. Half of them are planar scenes while the rest are general 3D scenes. In our experiments, we used the Harris interest point (Harris and Stephens, 1988) with the threshold set as 0.03. We have systematically compared our method with the following four state-of-the-art methods:

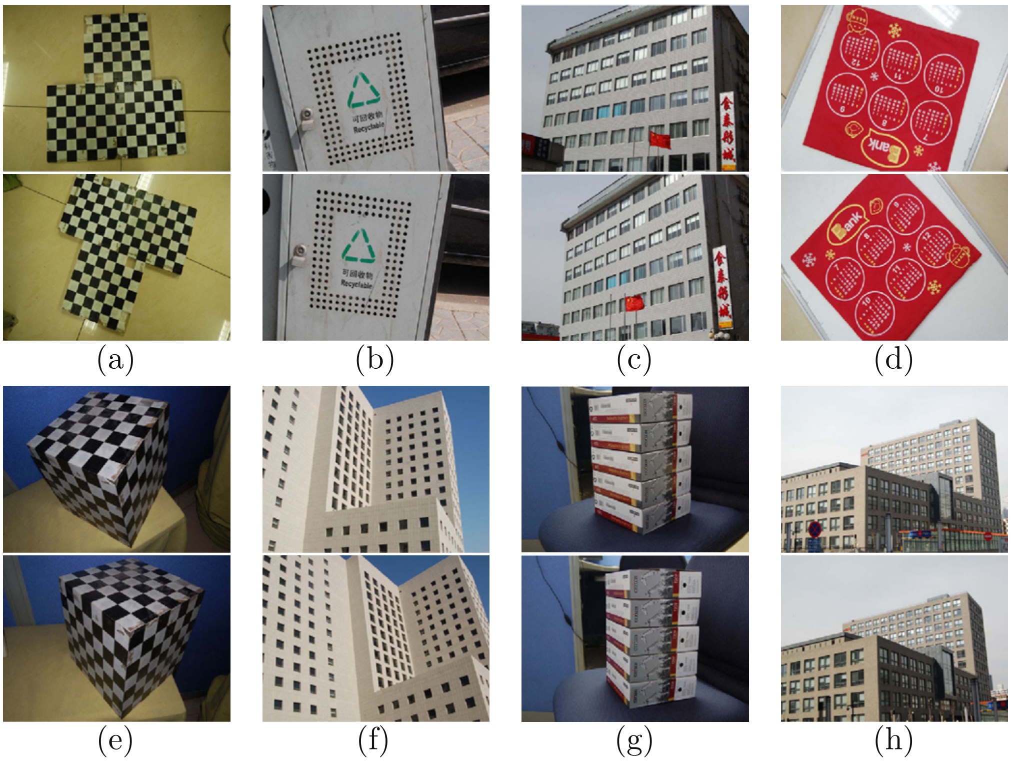
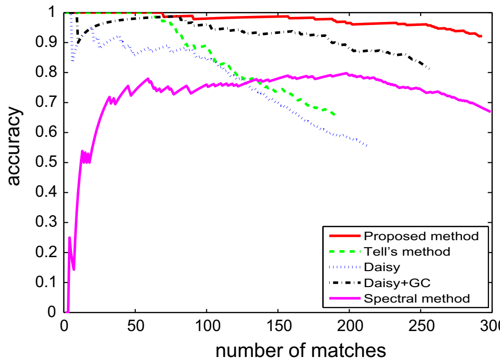
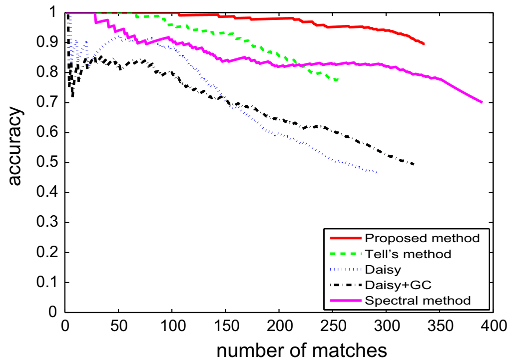


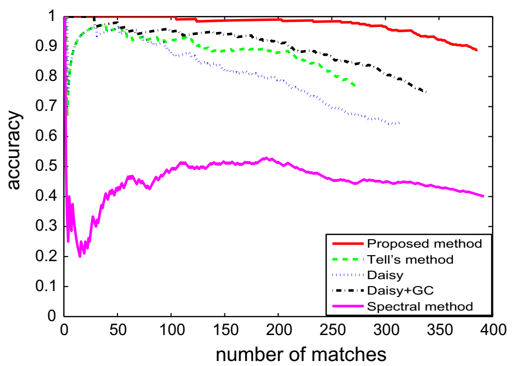
Fig. 6. Image pairs used for experimental evaluation.



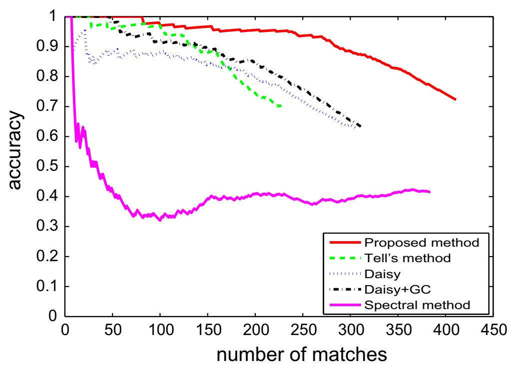
(a) Results for image pair in Fig.6(a).



(b) Results for image pair in Fig.6(b).



(c) Results for image pair in Fig.6(c).



(d) Results for image pair in Fig.6(d).

Fig. 7. Experimental results for images of planar scenes. The numbers of detected interest points in the tested images are listed in Table 1.

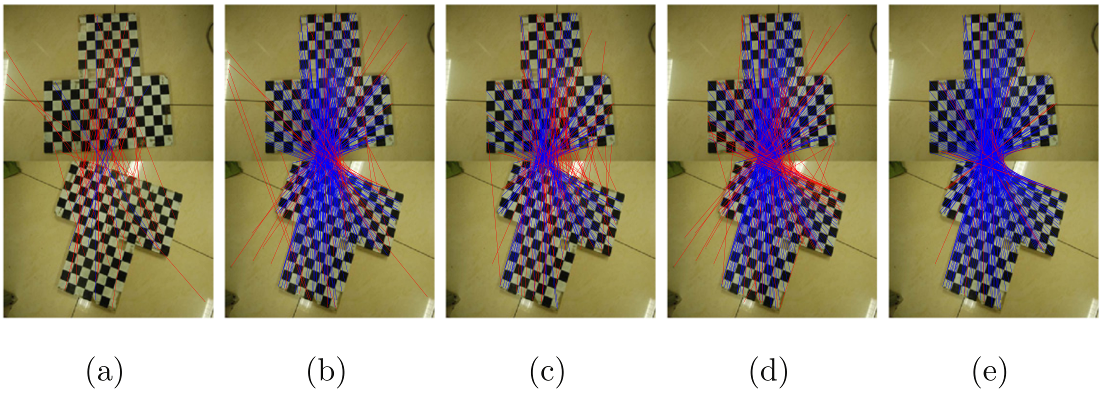


Fig. 8. Matching results for image pair in Fig. 6(a). The blue lines are correct matches, while the incorrect matches are marked by red lines. (a) 118/214 correct matches with Daisy method (55.1%); (b) 208/256 correct matches with Daisy + GC method (81.3%); (c) 125/192 correct matches with Tell's method (65.1%); (d) 200/299 correct matches with Spectral method (66.9%); (e) 270/293 correct matches with our method (92.2%). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Daisy: The point matching method based on DAISY descriptor for each individual interest point. The matching strategy used here is NNDR.

Tell's method (Tell and Carlsson, 2000): Another matching method that uses pairs of interest points. It uses the Fourier coefficients of the intensity profile between two points of a pair to describe the pair. Then the point correspondences are obtained from the tentative matched pairs of points by voting.

Daisy + GC: Similar to Mortensen et al. (2005), describing an interest point with both its local and global information. The local information is encoded in DAISY while the global information is encoded in GC (Global Context).[3] The matching strategy is NNDR, the same one as that used in the Daisy method.

Spectral method (Leordeanu and Hebert, 2005): In our experiments, the candidate point correspondences input to the Spectral method are those tentative point correspondences constituting the tentative matched pairs of interest points. Therefore, in our method and the Spectral method, the candidate point correspondences are same but their methods of generating final set of point correspondences are different. In the Spectral method, the final set of point correspondences is obtained by the spectral technique while our method utilizes our defined measure of compatibility.

In all of these methods, a one-to-one correspondence constraint has been imposed: if two or more points are matched to a single point in another image, we just keep the best one and discard the others. The matching results are sorted from the best to the worst. In our method and the Spectral method, the point correspondence that is earlier added to the set of point correspondences is considered to be better. In Tell's method, the point correspondence that has more votes is considered to be better while in Daisy and Daisy + GC, the better correspondence is the one with smaller NNDR score.

5.1. Results on planar scenes

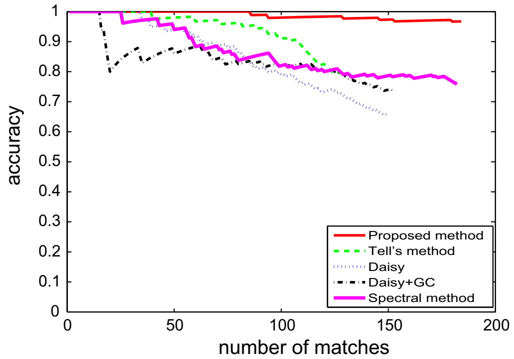
Figs. 6(a)–(d) are four image pairs of planar scenes containing repetitive patterns. For planar scenes, a correct point correspondence is related by a homography. However, since the coordinates of corresponding points are often corrupted by noise, they usually do not strictly satisfy the homography. In our experiments, if one point transformed by the homography is within 3 pixels of its corresponding point, then they are considered to be a correct match. Handcrafted point correspondences are used for estimating the ground-truth homographies in our experiments.

The matching results are plotted from Figs. 7(a)–(d) and the number of detected interest points for each tested image pair is listed in Table 1. It can be seen from these figures that our proposed method significantly outperforms the other four methods. Due to repetitive patterns, the traditional local descriptor based method can not achieve satisfactory results, as shown by plots of Daisy in figures. By incorporating global context to local descriptor or matching pairs of interest points, Daisy + GC, Tell's method and our method can achieve better results than the Daisy method. In fact, matching pairs of interest points also incorporates information in a much larger region than matching local descriptors of individual interest points. This is one of the reasons why our proposed method works well. It is clearly that our method significantly outperforms Tell's method although both of them are based on pairs of interest points. In Table 1, it shows the number of candidate point correspondences consist in the tentative matched pairs of points in each pair of tested images. The numbers of inliers among these correspondences are listed in the table too. Obviously, there are large amount of outliers exist in the candidate point correspondences. We can find that the Spectral method does not perform well and we think the reason is mainly because of many outliers exist. Owing to our newly defined measure of compatibility, the Algorithm 1 can successfully obtain reliable point correspondences in case of many outliers exist, which just reflects the effectiveness of our proposed method. Fig. 8 shows the matching results by the five tested methods for images in Fig. 6(a).

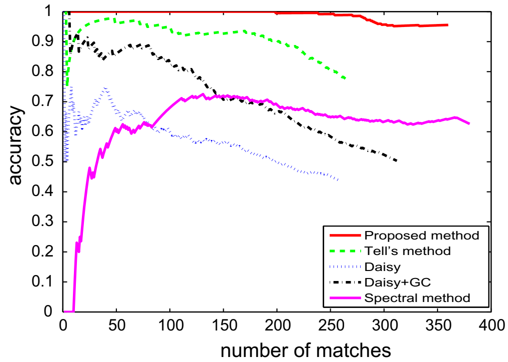
5.2. Results on 3D scenes

We have also tested our proposed method on images of general 3D scenes. Figs. 6(e)–(h) are four image pairs of 3D scenes that contain repetitive patterns. For 3D scenes, the epipolar geometry is used to check whether a pair of corresponding points is correct or not. Given a correspondence of points, if both of them lie in the corresponding epipolar lines, then they are regarded as a correct correspondence. Since the coordinates of corresponding points are usually corrupted by noise, they usually do not strictly satisfy the epipolar constraint. In our work, if the distance from a point to its corresponding epipolar line is within 3 pixels, it is considered to be on the epipolar line. The ground-truth fundamental matrices are estimated from handcrafted point correspondences for each tested pair. Since such an epipolar relationship is just a necessary condition for a correct correspondence, some false correspondences may pass through. However, since all the methods are evaluated with the same criterion, it is still a meaningful indicator.

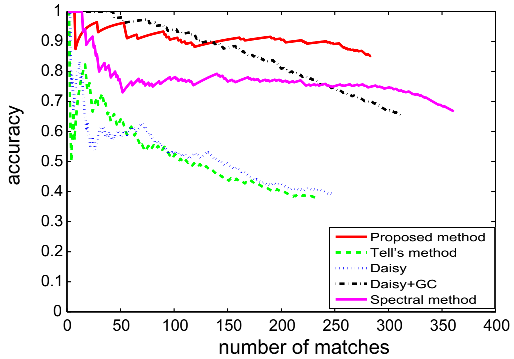
[3] In our experiments we have found that the Euclidean distance performs better than the χ^2 measure proposed in (Mortensen et al., 2005) when measuring the distance between two GC vectors. Therefore, in our experiments, the Euclidean distance is used.



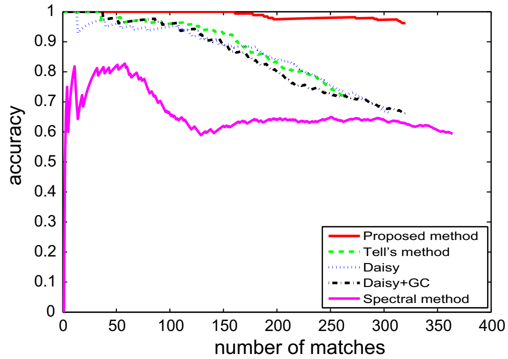
(a)Results for image pairin Fig.6(e).



(b)Results for image pairin Fig.6(f).

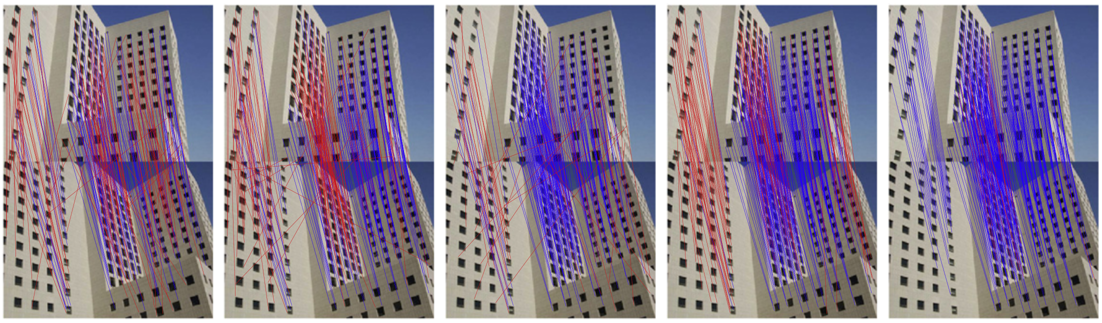


(c)Results for image pairin Fig.6(g).



(d)Results for image pairin Fig.6(h).

Fig. 9. Experimental results for images of 3D scenes. The numbers of detected interest points in the tested images are listed in Table 1.



(a) (b) (c) (d) (e)

Fig. 10. Matching results for image pair in Fig. 6(f). The blue lines are correct matches, while the incorrect matches are marked by red lines. (a) 113/259 correct matches with Daisy method (43.6%); (b) 157/312 correct matches with Daisy + GC method (50.3%); (c) 205/264 correct matches with Tell's method (77.7%); (d) 238/380 correct matches with Spectral method (62.6%); (e) 344/360 correct matches with our method (95.6%). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Fig. 9 shows the comparative results of all the tested methods and the number of detected interest points for each tested image pair is listed in Table 1. With the help of pairs of interest points and our defined compatibility measure, our proposed method outperforms the other tested methods, as indicated by these figures. Moreover, from Figs. 7, 9 and Table 1, we have found that almost all the correct point correspondences have been recalled by our method and the precision of the obtained correspondences is still very high (higher than 90% in most cases). Fig. 10 shows the concrete matching results for image pair in Fig. 6(f) by the five tested methods respectively. It can be clearly seen from Fig. 10 that our method not only has more correct correspondences but also with higher accuracy than the other four methods.

6. Conclusion

Although image matching methods based on local descriptors have been well studied, they are usually fail to match images containing repetitive patterns even if the viewpoint changes between these images are not very large, due to the local ambiguities induced by repetitive patterns. Since such repetitive patterns widely exist in images of man-made scenes or objects, it is meaningful to study the problem of matching this kind of images. A novel method for matching images containing repetitive patterns is presented in this paper. The proposed method starts from matching pairs of interest points and then finds reliable point correspondences from the tentative matched pairs of interest points under the low distortion constraint. By our newly defined compatibility measure between one correspondence and a set of point correspondences, we can establish correspondences of points between two images with very high precision. Experiments on images of both planar and 3D scenes as well as comparisons with the state-of-the-art methods have demonstrated the effectiveness of the proposed method.

Acknowledgement

This work is supported by the National Science Foundation of China (60835003, 61075038).

References

Agarwal, S., Snavely, N., Simon, I., Seitz, S.M., Szeliski, R., 2009. Building rome in a day. In: Proc. Internat. Conf. on Computer Vision (ICCV), pp. 72–79.

Ahuja, N., Todorovic, S., 2007. Extracting texels in 2.1D natural textures. In: Proc. Internat. Conf. on Computer Vision (ICCV), pp. 1–8.

Ballard, D., 1981. Generalizing the hough transform to detect arbitrary shapes. Pattern Recognition 13 (2), 111–122.

Berg, A.C., Berg, T.L., Malik, J., 2005. Shape matching and object recognition using low distortion correspondences. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 26–33.

Brown, M., Lowe, D., 2002. Invariant features from interest point groups. In: Proc. British Machine Vison Conf., pp. 656–665.

Brown, M., Lowe, D., 2003. Recognising panoramas. In: Proc. Internat. Conf. on Computer Vision (ICCV)., vol. 2. pp. 1218–1225.

Brown, M., Hua, G., Winder, S., 2010. Discriminative learning of local image descriptors. IEEE Trans. Pattern Anal. Mach. Intell., 99.

Cheng, M.-M., Zhang, F.-L., Mitra, N.J., Hu, S.-M., 2010. Repfinder: finding approximately repeated scene elements for image editing. ACM Trans. Graphics (SIGGRAPH 2010) 29 (3), 1–8.

Choi, O., Kweon, I.S., 2009. Robust feature point matching by preserving local geometric consistency. Comput. Vision Image Understand. 113, 726–742.

Duchenne, O., Bach, F., Kweon, I., Ponce, J., 2009. A tensor-based algorithm for high-order graph matching. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 1980–1987.

Fischler, M., Bolles, R., 1981. Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography. Comm. ACM 24 (6), 381–395.

Furukawa, Y., Ponce, J., 2010. Accurate, dense, and robust multiview stereopsis. IEEE Trans. Pattern Anal. Machine Intell. 32 (8), 1362–1376.

Gupta, R., Mittal, A., 2008. SMD: A locally stable monotonic change invariant feature descriptor. In: Proc. European Conf. of Computer Vision (ECCV), pp. 265–277.

Harris, C., Stephens, M., 1988. A combined corner and edge detector. In: Proc. Fourth Alvey Vision Conf., pp. 147–151.

Hays, J., Leordeanu, M., Efros, A.A., Liu, Y., 2006. Discovering texture regularity as a higher-order correspondence problem. In: Proc. European Conf. of Computer Vision (ECCV), pp. 522–535.

Heikkila, M., Pietikainen, M., Schmid, C., 2009. Description of interest regions with local binary patterns. Pattern Recognition 42 (3), 425–436.

Leordeanu, M., Hebert, M., 2005. A spectral technique for correspondence problems using pairwise constraints. In: Proc. Internat. Conf. on Computer Vision (ICCV), vol. 2, pp. 1482–1489.

Leung, T.K., Malik, J., 1996. Detecting, localizing and grouping repeated scene elements from an image. In: Proc. European Conf. of Computer Vision (ECCV), pp. 546–555.

Liu, Y., Collins, R., Tsin, Y., 2004. A computational model for periodic pattern perception based on frieze and wallpaper groups. IEEE Trans. Pattern Anal. Machine Intell. 26 (1), 354–371.

Lowe, D.G., 2004. Distinctive image features from scale-invariant keypoints. Internat. J. Comput. Vision 60 (2), 91–110.

Mikolajczyk, K., Schmid, C., 2005. A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Machine Intell. 27 (10), 1615–1630.

Mortensen, E., Deng, H., Shapiro, L., 2005. A SIFT descriptor with global context. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), vol. 1, pp. 184–190.

Pauly, M., Mitra, N.J., Wallner, J., Pottmann, H., Guibas, L., 2008. Discovering structural regularity in 3D geometry. ACM Transactions on Graphics (SIGGRAPH 2008) 27 (3), 1–11, No. 43.

Snavely, N., Seitz, S.M., Szeliski, R., 2006. Photo tourism: exploring photo collections in 3D. ACM Trans. Graphics (TOG) 25 (3), 835–846.

Szeliski, R., 2006. Image alignment and stitching: a tutorial. Found. Trends Comput. Graphics Vision 2 (1), 1–104.

Tang, F., Lim, S.H., Chang, N., Tao, H., 2009. A novel feature descriptor invariant to complex brightness change. In: Proc. IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), pp. 2631–2638.

Tell, D., Carlsson, S., 2000. Wide baseline point matching using affine invariants computed from intensity profiles. In: Proc. European Conf. of Computer Vision (ECCV), pp. 814–828.

Tell, D., Carlsson, S., 2002. Combining appearance and topology for wide baseline matching. In: Proc. European Conf. of Computer Vision (ECCV), pp. 68–81.

Todorovic, S., Ahuja, N., 2009. Texel-based texture segmentation. In: Proc. Internat. Conf. on Computer Vision (ICCV), pp. 841–848.

Tola, E., Lepetit, V., Fua, P., 2010. DAISY: an efficient dense descriptor applied to wide-baseline stereo. IEEE Trans. Pattern Anal. Machine. Intell. 32 (5), 815–830.

Winder, S., Hua, G., Brown, M., 2009. Picking the best DAISY. In: Proc. IEEE Conf. on Comput. Vision Pattern Recognition (CVPR), pp. 178–185.

Zhang, J., Marszalek, M., Lazebnik, S., Schmid, C., 2007. Local features and kernels for classification of texture and object categories: a comprehensive study. Internat. J. Comput. Vision 73 (2), 213–238.