



MSLD: A robust descriptor for line matching

Zhiheng Wang^{*,1}, Fuchao Wu, Zhanyi Hu

National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences, 100080 Beijing, China

ARTICLE INFO

Article history:

Received 27 March 2008

Received in revised form 26 August 2008

Accepted 27 August 2008

Keywords:

Line matching

Line descriptor

MSLD descriptor

ABSTRACT

Line matching plays an important role in many applications, such as image registration, 3D reconstruction, object recognition and video understanding. However, compared with other features (such as point, region matching), it has made little progress in recent years.

In this paper, we investigate the problem of matching line segments automatically only from their neighborhood appearance, without resorting to any other constraints or priori knowledge. A novel line descriptor called mean-standard deviation line descriptor (MSLD) descriptor is proposed for this purpose, which is constructed by the following three steps: (1) For each pixel on the line segment, its pixel support region (PSR) is defined and then the PSR is divided into non-overlapped sub-regions. (2) Line gradient description matrix (GDM) is formed by characterizing each sub-region into a vector. (3) MSLD is built by computing the mean and standard deviation of GDM column vectors. Extensive experiments on real images show that MSLD descriptor is highly distinctive for line matching under rotation, illumination change, image blur, viewpoint change, noise, JPEG compression and partial occlusion.

In addition, the concept of MSLD descriptor can also be extended to creating curve descriptor (mean-standard deviation curve descriptor, MSCD), and promising MSCD-based results for both curve and region matching are also demonstrated in this work.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Feature matching has drawn a lot of attention in the last few years. Great progress has been made and various approaches have been proposed for wide baseline point [1,2] and region [3] matching. Most of such approaches characterize local regions into feature descriptors and perhaps the most famous one is SIFT descriptor [1]. Line matching also plays an important role and is irreplaceable in many scenes [4–7]. A typical example is man-made scenes, which mainly consist of line segments, and line matching often becomes an unavoidable step for their 3D reconstruction. Unfortunately, compared to point and region matching, line matching is rarely reported in the literature and is still a challenging task due to various reasons [8]: inaccuracy of line endpoint locations, no strong disambiguating geometric constraint available, lacking of rich textures in line local neighborhood and so on.

Only a few methods are reported in the literature. In Ref. [9] the trifocal tensor is used for line matching by finding lines

satisfying geometrical constraint in three views. Schmid and Zisserman [8,10] takes the epipolar constraint of line endpoints for short baseline matching, and one parameter family of plane homographies for wide baseline matching. However, both the trifocal tensor method [9] and the epipolar methods [8,10] demand known geometrical relations between images in advance. Lourakis et al. [11] present an approach using the “2 lines + 2 points” projective invariant for line matching in images of planar surfaces, and hence their method is limited to planar scenes. Herbert [12] proposes a method for automatic line matching in color images, where an initial set of line segment correspondences are generated using color histogram, then a topological filter is used to iteratively increase possible matches. The main drawback of this method is its heavy reliance on color rather than purely on texture. While color provides a very strong cue for discrimination, it may fail in the case where color feature is not distinctive, such as in gray images or remote sensing images. Although grouping matching strategy [13] has the advantage that more geometric information is available for removing ambiguities, and is able to cope with more significant camera motion, it often has high computational complexity and is sensitive to line topological connections or inaccuracy of endpoints.

As said in the above, most existing methods in the literature either require some prior knowledge [8–10] or are limited to some specific scenes [11,12] or are of high computational complexity [13], a method capable of automatically matching lines in general scene is

* Corresponding author.

E-mail address: zhwang@nlpr.ia.ac.cn (Z. Wang).

¹ Supported by the National Natural Science Foundation of China (60575019) and the National High-Tech Research and Development Program of China (2006AA01Z116 and 2007AA01Z341).

needed. This paper present a novel descriptor called mean-standard deviation line descriptor (MSLD) for automatic line matching without resorting to any prior knowledge. Compared with previous approaches, MSLD has two major appealing characteristics: one is that it is purely image content-based and can work without any other possible constraints, and the other is that MSLD is applicable to general scenes while some state-of-art methods are only scene-specific. Experiments show that MSLD descriptor is highly distinctive and robust against image rotation, illumination change, image blur, viewpoint change, noise, JPEG compression and partial occlusion.

The remainder of this paper is organized as follows. Section 2 introduces the definition of pixel support region (PSR) and the strategy of partitioning PSR into sub-regions, which is the key step to create line descriptor. In Section 3, each sub-region is characterized by a feature vector using a SIFT-like strategy. Section 4 elaborates the construction of MSLD descriptor. In Section 5, descriptor dimension and matching criteria are investigated, and Section 6 is the experiments. In Section 7, MSLD is extended for curve and region matching. Section 8 lists some concluding remarks.

2. Pixel support region (PSR)

Similar to creating point descriptor, how to select and partition line local neighborhood is the first step to construct our line descriptor. In this paper we propose a novel scheme to summarize local neighborhood of different-length lines into uniform description vectors.

Given a line segment L , as shown in Fig. 1, firstly two directions are introduced before defining the PSR: the average gradient direction \mathbf{d}_\perp of pixels on the line and its anticlockwise orthogonal direction \mathbf{d}_L . For each pixel on the line L , a rectangular region centered at it and aligned with the directions \mathbf{d}_\perp , \mathbf{d}_L is defined as the PSR. The PSRs of the pixels on the line along the direction \mathbf{d}_L are denoted as G_1, G_2, \dots, G_N (assuming L consists of N pixels). In order to give a more distinctive description for the PSR, each PSR is divided into M non-overlapped sub-regions with the same size along the direction \mathbf{d}_\perp : $G_i = G_{i1} \cup G_{i2} \cup \dots \cup G_{iM}$. It is noted that using \mathbf{d}_\perp is necessary for the final descriptor to be rotation invariant, otherwise, some ambiguity may occur when deciding the order of $G_{i1}, G_{i2}, \dots, G_{iM}$.

3. Sub-region representation

In this section, each sub-region will be characterized by a description vector using a SIFT-like strategy.

At first, it is noted that the gradient vector is not rotation invariant: suppose two images are related by $h(x') = f(\mathbf{R} \cdot x)$, the gradient vectors $\nabla \mathbf{f}$, $\nabla \mathbf{h}$ of a pair of corresponding points in the two images

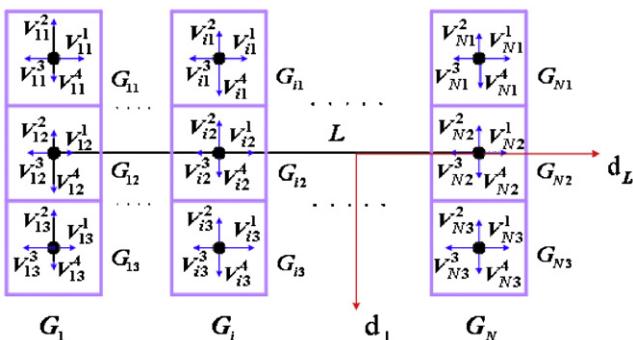


Fig. 1. Schematic figure of MSLD construction. In this figure, each PSR is divided into three sub-regions for illustration purpose, whereas nine sub-regions are adopted in our work.

must satisfy the relation $\nabla \mathbf{h} = \mathbf{R} \cdot \nabla \mathbf{f}$. To achieve rotation invariance, each sample gradient is rotated aligned with the directions \mathbf{d}_\perp and \mathbf{d}_L : $\nabla_L \mathbf{f} = (\nabla \mathbf{f} \cdot \mathbf{d}_\perp, \nabla \mathbf{f} \cdot \mathbf{d}_L)^T \triangleq (f_{d_\perp}, f_{d_L})^T$. This step is similar to that of SIFT, where the gradient orientations are aligned relative to the key-point orientation.

Motivated by SIFT, a Gaussian weighting function with a scale σ , equal to half of the PSR side along the direction \mathbf{d}_\perp , is used to assign a weight to each sample in the PSR: for a sample which has a distance d from the line L , its weight can be expressed as $w = 1/(\sqrt{2\pi}\sigma) e^{-d^2/2\sigma^2}$. The purpose of such a weighting is to give less importance on the gradients of those samples that are far from the line, which are most likely affected by mis-registration errors. Another reason is to reduce the descriptor's sensitivity to small change in the position of each PSR.

As one sample shifts smoothly from being within one sub-region to another, the descriptor may change abruptly and thus boundary effect arises. In order to reduce this effect, for a sample whose gradient is $\nabla \mathbf{f}$, it will contribute not only to its sub-region G_{ij} , but also to its nearest neighbor sub-region $G_{i(j-1)}$ (or $G_{i(j+1)}$) along the direction \mathbf{d}_\perp .

Denote the distances from it to the central lines (parallel to \mathbf{d}_L) of the two sub-regions are d_1, d_2 , then the contributions to the two sub-regions are $\nabla \mathbf{f} \cdot w_1, \nabla \mathbf{f} \cdot w_2$, respectively, where $w_1 = d_2/(d_1 + d_2)$, $w_2 = d_1/(d_1 + d_2)$. This step of linear interpolation is only adopted along the direction \mathbf{d}_\perp but not along the direction \mathbf{d}_L , because sub-regions are overlapped each other along the direction \mathbf{d}_L and thus boundary effect is negligible.

Denote the gradients distributed in a sub-region G_{ij} as $\{\tilde{f}_{d_\perp}, \tilde{f}_{d_L}\}^T$, then a 4D feature vector is formed by accumulating these gradients along the directions $\mathbf{d}_\perp, \mathbf{d}_L$ and their opposite directions, respectively (as shown in Fig. 1):

$$\mathbf{V}_{ij} = (V_{ij}^1, V_{ij}^2, V_{ij}^3, V_{ij}^4)^T \in \mathbb{R}^4 \quad (1)$$

where

$$V_{ij}^1 = \sum_{\tilde{f}_{d_\perp} > 0} \tilde{f}_{d_\perp}, \quad V_{ij}^2 = \sum_{\tilde{f}_{d_\perp} < 0} -\tilde{f}_{d_\perp}, \quad V_{ij}^3 = \sum_{\tilde{f}_{d_L} > 0} \tilde{f}_{d_L}, \quad V_{ij}^4 = \sum_{\tilde{f}_{d_L} < 0} -\tilde{f}_{d_L}$$

It can be proved that \mathbf{V}_{ij} constructed in such way is invariant to image rotation, and it is used as the description vector of the sub-region G_{ij} .

4. MSLD descriptor

By stacking the description vectors of all the sub-regions associated with a line segment, a $4M \times N$ matrix called line gradient description matrix (GDM) is formed as

$$\mathbf{GDM}(L) = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} & \cdots & \mathbf{V}_{1N} \\ \mathbf{V}_{21} & \mathbf{V}_{22} & \cdots & \mathbf{V}_{2N} \\ \cdots & \cdots & \cdots & \cdots \\ \mathbf{V}_{M1} & \mathbf{V}_{M2} & \cdots & \mathbf{V}_{MN} \end{pmatrix} \triangleq (\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_N) \quad (\mathbf{V}_i \in \mathbb{R}^{4M}) \quad (2)$$

GDM contains the most structural information in the line neighborhood region. However, it cannot be directly used as a line descriptor because its size still varies with line length. To make the descriptor independent of the line length, statistical entities from GDM column vectors are explored here. We have tested several popularly used statistic measures: mean, standard deviation, Fourier coefficients and moments. Among all the candidates and their combinations, we have found that the combination of mean and standard deviation can provide satisfying matching result, though the first four Fourier coefficients or moments can give slightly better performance. Considering descriptor's dimensional problem, the mean and standard deviation are adopted to construct our line descriptor

in this work, and it is called MSLD. MSLD is constructed using the followings steps.

First, we compute the mean vector and the standard deviation vector of GDM column vectors: $\mathbf{M}(\mathbf{GDM}(L)) = \text{Mean}\{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_N\}$, $\mathbf{S}(\mathbf{GDM}(L)) = \text{Std}\{\mathbf{V}_1, \mathbf{V}_2, \dots, \mathbf{V}_N\}$.

Second, in order to make the descriptor invariant to linear changes of illumination, the mean vector and the standard deviation vector are normalized to unit norm, respectively. Then, by concatenating the normalized mean vector with the normalized standard deviation vector into a single vector, we obtain a line description vector with a dimension of $8M$:

$$\mathbf{MSLD}(L) = \begin{pmatrix} \mathbf{M}(\mathbf{GDM}(L)) \\ \|\mathbf{M}(\mathbf{GDM}(L))\| \\ \mathbf{S}(\mathbf{GDM}(L)) \\ \|\mathbf{S}(\mathbf{GDM}(L))\| \end{pmatrix} \in R^{8M}$$

It is noted that the mean vector and the standard deviation vector should be normalized separately instead of as a whole. Consider a line with smooth gradient distribution in its neighborhood, the values of the mean vector's elements may be much larger than that of the standard deviation vector's, however, the most structural information is contained in the standard deviation vector. In this case, a global normalization will diminish the distinctiveness of the descriptor, while the separate normalization can reduce the influence of the mean vector and increase the influence of the standard deviation vector relatively, and consequently can enhance the discrimination power of descriptor. On the other hand, if a line's neighborhood is of rich texture, the values of the standard deviation vector's elements may be far larger than that of the mean vector's, and the separate normalization will make descriptor more stable.

In Ref. [1], in order to reduce the influence of non-linear illumination, Lowe thresholds the values in the unit feature vector such that each one is no larger than 0.2, which means that matching the magnitudes for large gradients is no longer as important, and that the distribution of orientations has greater influence. This strategy is also adopted here for the same reason. A value of 0.4 is found empirically to be a proper value and used in our work. After restraining the maximum value of each dimension, the line description vector is re-normalized to unit norm as a whole and the MSLD is finally achieved.

5. Descriptor dimension and matching criteria

Two parameters should be determined before constructing MSLD descriptor: the sub-region number M and the sub-region size. With the sub-region number and size increasing, MSLD becomes more discriminative while also more sensitive to shape distortions and partial occlusion. Fig. 2(a) gives some statistical matching results in which both two parameters vary. The graph is generated using seven image pairs. Obviously the correct ratio (CR) is much lower at start but increases rapidly before the sub-region size arrives at 3×3 , and then it becomes stable. As for the sub-region number, descriptor using 9 or 11 sub-regions performs distinctly better than those with less sub-regions. In addition, descriptor with 9 sub-regions provides quite similar result to that with 11 sub-regions, which implies that descriptor distinctiveness also becomes stable. Consequently, through this paper we use nine sub-regions each with a size of 5×5 , resulting in a 72-dimensional MSLD descriptor.

Having constructed line descriptors, two issues, matching metric and matching criteria, should also be addressed. In this paper, Euclidean distance between descriptors is used as the matching metric. Fig. 2(b) shows the probability density functions (PDF) for correct and incorrect matches in terms of Euclidean distance between descriptors, which are generated using seven image pairs. Clearly, correct and false matches are easy to recognize by their Euclidean distance of descriptors. Best-winning is the simplest criterion for matching

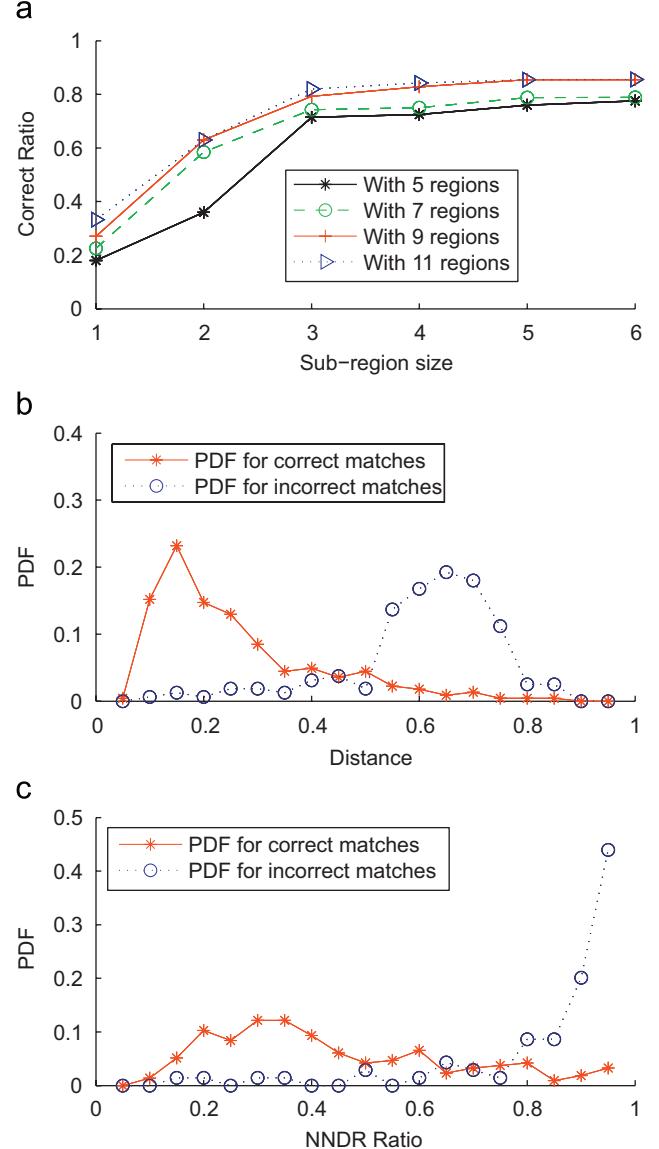


Fig. 2. Parameters selection. (a) Line matching results when the sub-region number and the sub-region size vary. (b) The PDF of distance between descriptors. (c) The PDF of NNDR ratio.

problem, however, we find it usually matches a single line segment in the first image to multi-candidates in the second image. Next we discuss other two effective matching criterions.

LRC (left/right checking): LRC is one of the most useful criterions for feature matching. According to the testing results provided by Egnal and Wildes [14], among common criterions for matching, LRC can give the best matching result with the highest CR in case of occlusion. Considering line matching is more sensitive to partial occlusion compared to point matching, we will test LRC for MSLD-based line matching in our experiments: for a line in the first image and its best match in the second image, the line pair will be accepted as a correct match (CM) only when the first line is also the best match of the second one. However, when a line segment in first image has no corresponding line segment in the second image, LRC usually classifies the nearest but false one as its CM. An extreme case is that only one line segment to match in the first image, and LRC will always find a match whether it is correct or not. To resolve this problem, a simple global threshold of 0.55 is used to improve the matching precision.

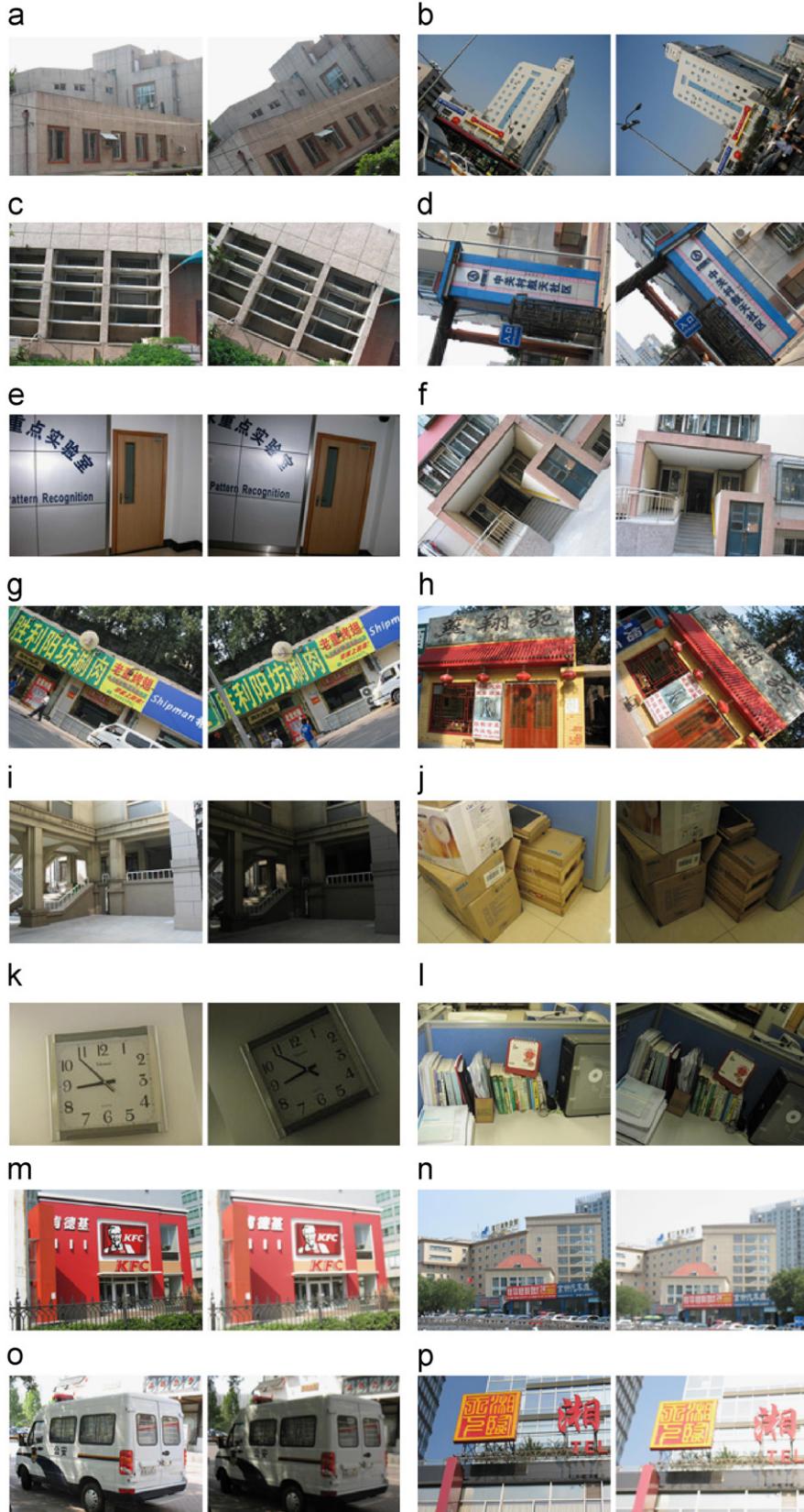


Fig. 3. Image set. (a)–(d) Rotation. (e)–(h) Viewpoint change. (i)–(l) Illumination change. (m)–(p) Image blur.

NNDR (nearest/next ratio): NNDR is another effective measure for feature matching, which is adopted by SIFT and has become one of the most popular criterions for point matching. Lowe [1] points

out that this measure performs well because CM need to have the closest neighbor significantly closer than the closest incorrect match to achieve reliable matching. For false matches, there will likely be

Table 1
Line matching results on images in Fig. 3

Images	a	b	c	d	CM	CR
LRC	33(6)	26(0)	101(5)	27(5)	171	91.4%
NNDR	34(7)	26(1)	103(5)	31(8)	173	89.2%
Images	e	f	g	h	CM	CR
LRC	21(2)	36(4)	41(2)	66(3)	153	93.4%
NNDR	19(3)	26(1)	41(5)	51(3)	125	91.6%
Images	i	j	k	l	CM	CR
LRC	45(3)	43(1)	16(1)	42(1)	140	95.9%
NNDR	43(3)	42(2)	13(0)	42(0)	135	96.4%
Images	m	n	o	p	CM	CR
LRC	47(1)	12(2)	22(4)	41(2)	113	91.9%
NNDR	48(2)	16(2)	22(3)	41(2)	118	94.5%

For each image pair, the first number represents the total matches and the number in bracket denotes the incorrect matches.

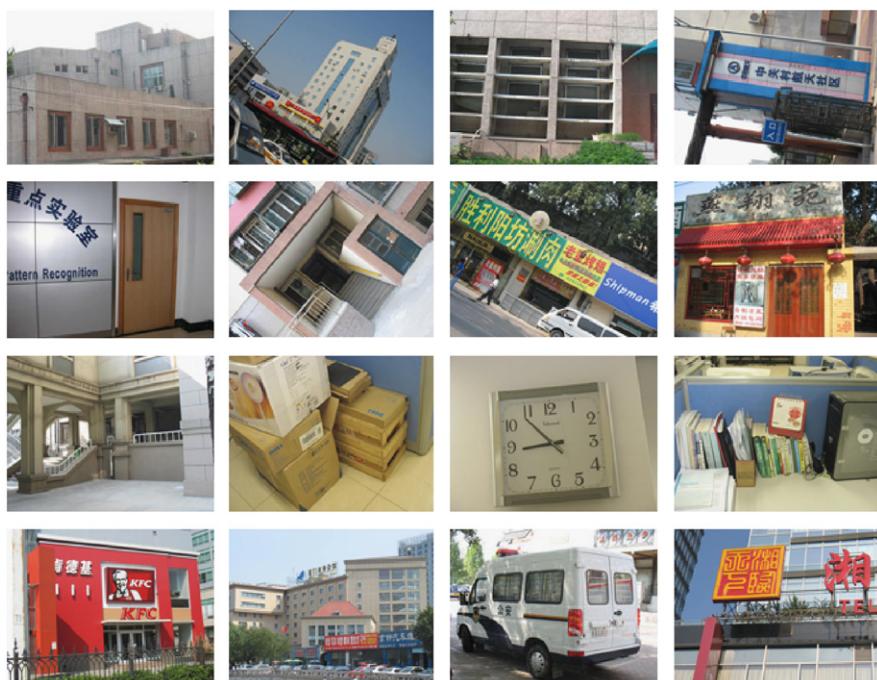


Fig. 4. Images for testing noise and JPEG compression.

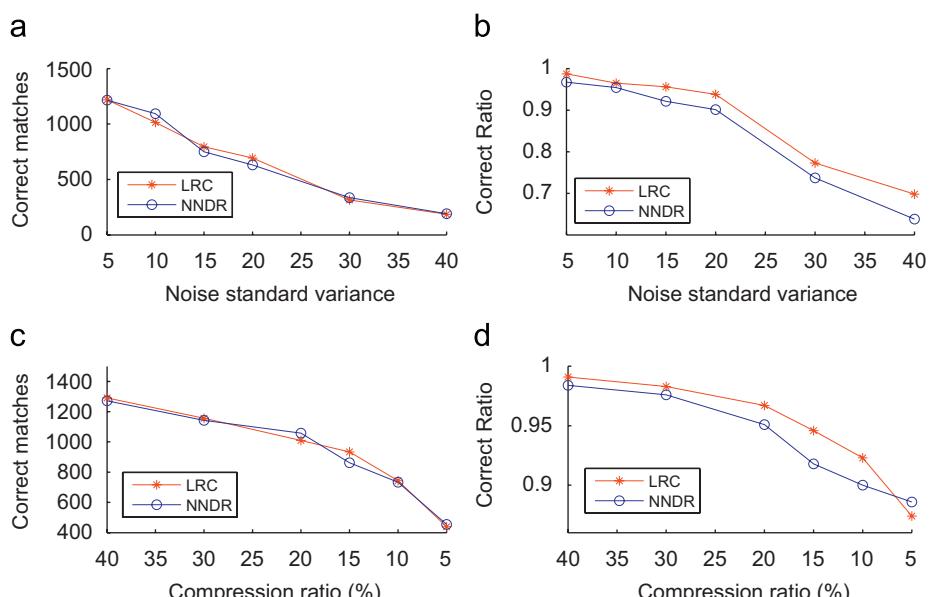


Fig. 5. Matching results under noise and JPEG compression. (a) CM–noise curve. (b) CR–noise curve. (c) CM–compression ratio curve. (d) CR–compression ratio curve.

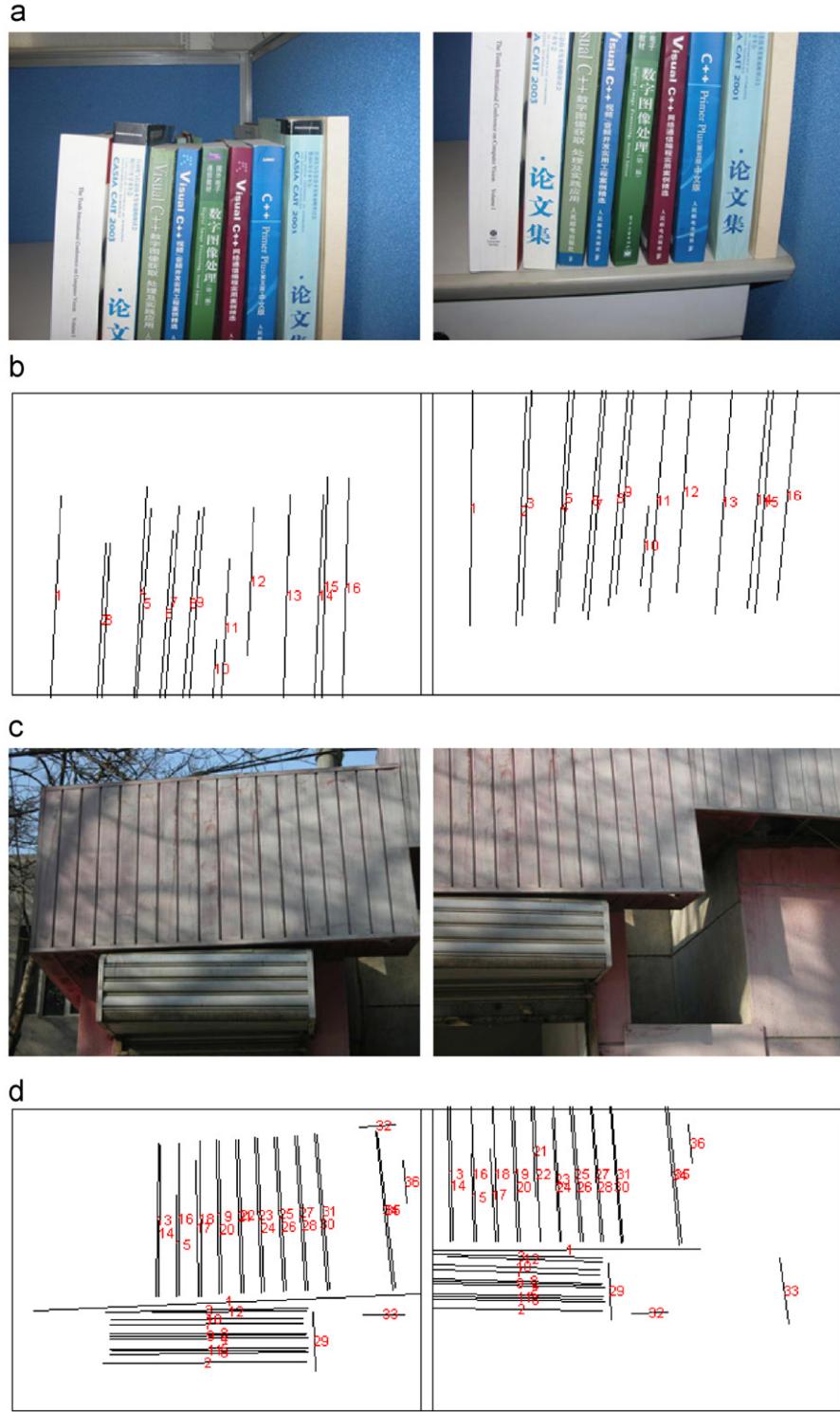


Fig. 6. Test MSLD robust to occlusion. (a) Extracted lines: 37, 32. (b) Matched lines: 16, incorrect: 0. (c) Extracted lines: 75, 55. (d) Matched lines: 36, incorrect: 2.

a number of other false matches within similar distances due to the high dimensionality of the feature space. Fig. 2(c) shows the PDF for correct and incorrect matches in term of NNDR ratio, which are also generated using seven image pairs. We can see that 0.6 seems to be a proper value for using NNDR criterion to recognize CM from false ones, and 0.8 will attain more matches but also with more false matches. From experiments, we have found that a global threshold of 0.55 can effectively remove most false matches that has

an NNDR between 0.6 and 0.8. Therefore, compared to only using NNDR 0.6, using NNDR 0.8 plus global threshold 0.55 at the same time can attain more matches while keeping a equivalent matching precision.

It is mentioned that, compared to using LRC or NNDR only, using global threshold can significantly improve the matching efficiency, since there is no need to verify LRC or NNDR when a candidate match has a distance higher than the global threshold. Thus, we have two

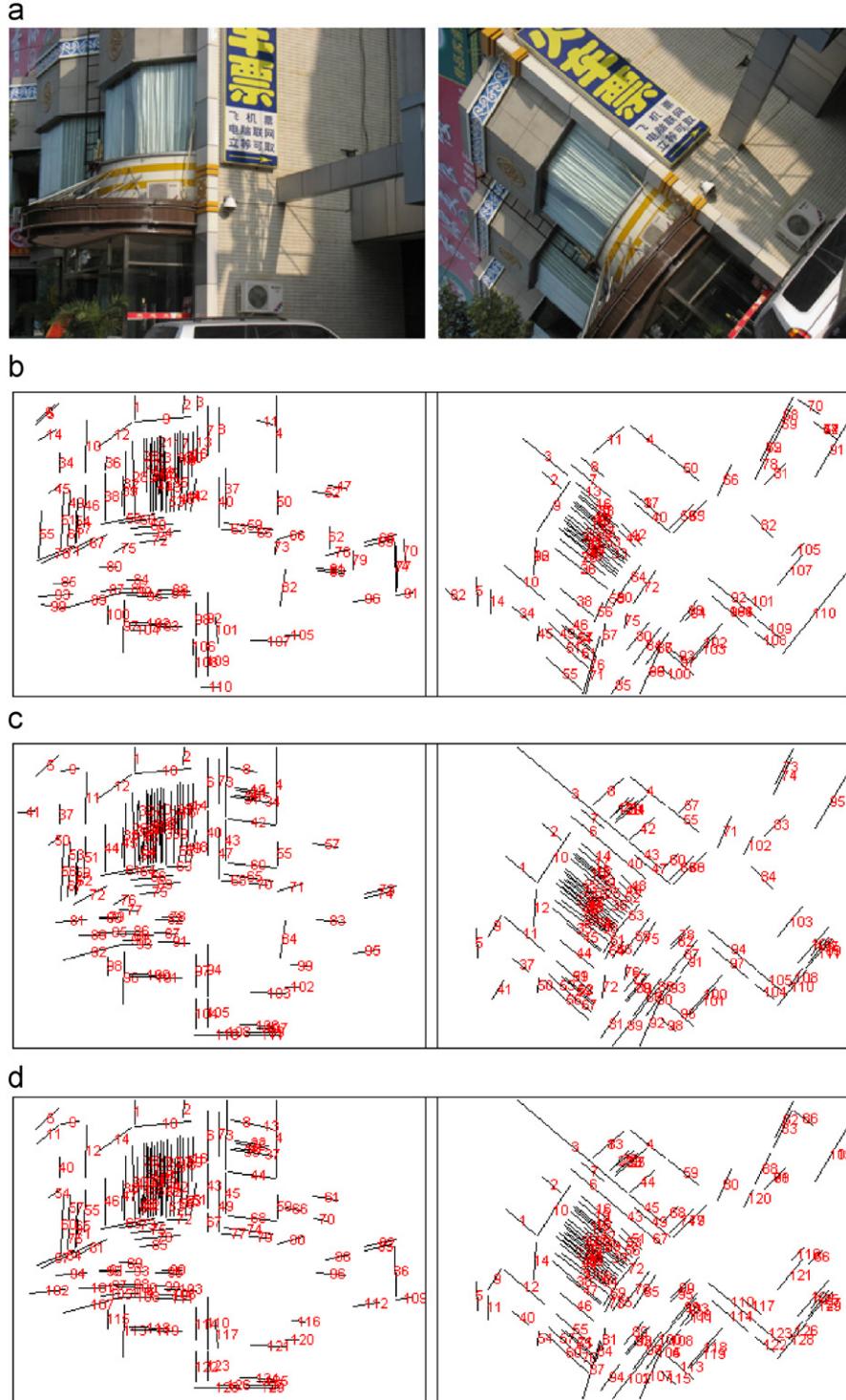


Fig. 7. Line matching using Schmid's method [10] and MSLD: Example 1. (a) Extracted lines: 193, 197. (b) Schmid's method [10] (cross-correlation + epipolar geometry). Matched lines: 110, incorrect: 15. (c) MSLD. Matched lines: 116, incorrect: 8. (d) MSLD + epipolar geometry. Matched lines: 129, incorrect: 3.

groups of criterions: the first one is called LRC-based criterion, using LRC and global threshold, and the second group is called NNDR-based criterion, using NNDR and global threshold.

6. Experiments

In this section, we will test the performance of MSLD descriptor on real images. Two criterions are used to evaluate its performance:

one is the CM, and the other is the CR, which is defined as the ratio of CM to the total matches. Each match is regarded as a CM or an incorrect one manually. In the experiments, the dimension of the used MSLD descriptor is 72. The NNDR ratio and the global threshold is set at 0.8 and 0.55, respectively.

Line segments are extracted using the method proposed in Ref. [12]: firstly Canny edge detector is used to extract edges, and then edges are split at points with high curvature. Finally, segments

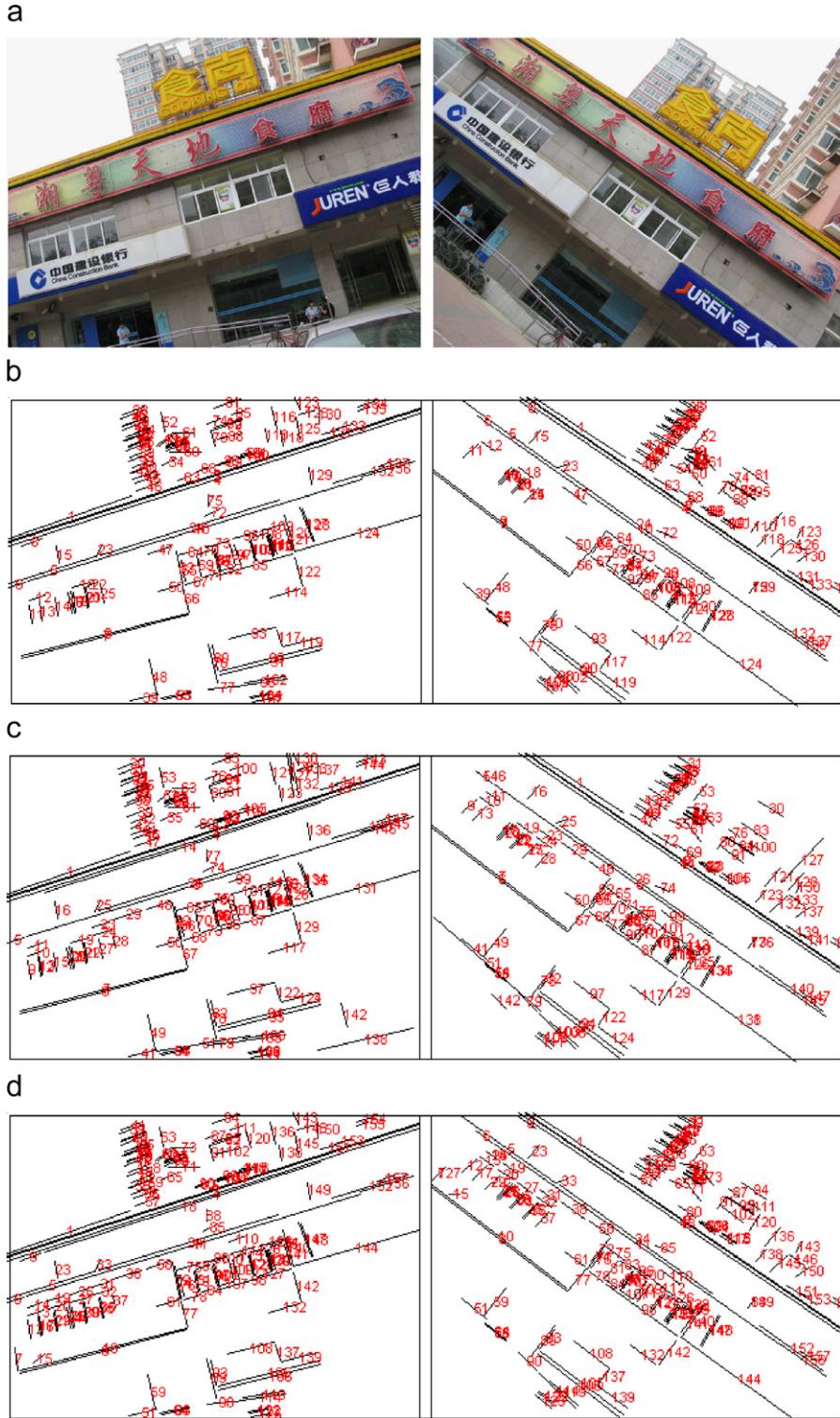


Fig. 8. Line matching using Schmid's method [10] and MSLD: Example 2. (a) Extracted lines: 387, 334. (b) Schmid's method [10] (cross-correlation + epipolar geometry). Matched lines: 137, incorrect: 6. (c) MSLD. Matched lines: 147, incorrect: 8. (d) MSLD + epipolar geometry. Matched lines: 159, incorrect: 5.

whose length is less than 20 pixels are discarded and lines are fitted to the split edges using the least squares method.

6.1. Test robustness to image transformation

Dataset: As is shown in Fig. 3, our dataset consists of 16 pairs of real images. (a)–(d) are rotation pairs obtained by rotating the camera around its optical axis for an arbitrarily selected degree within

the range of [0,90]. Image pairs (e)–(h) in viewpoint change are captured by varying the camera position. Illumination changes in image pairs (i)–(l) are introduced by varying the camera aperture, while blurred image pairs (m)–(p) are obtained by varying the camera focus.

Results: Table 1 gives the matching results on image pairs in Fig. 3 using LRC-based criterion and NNDR-based criterion, respectively.

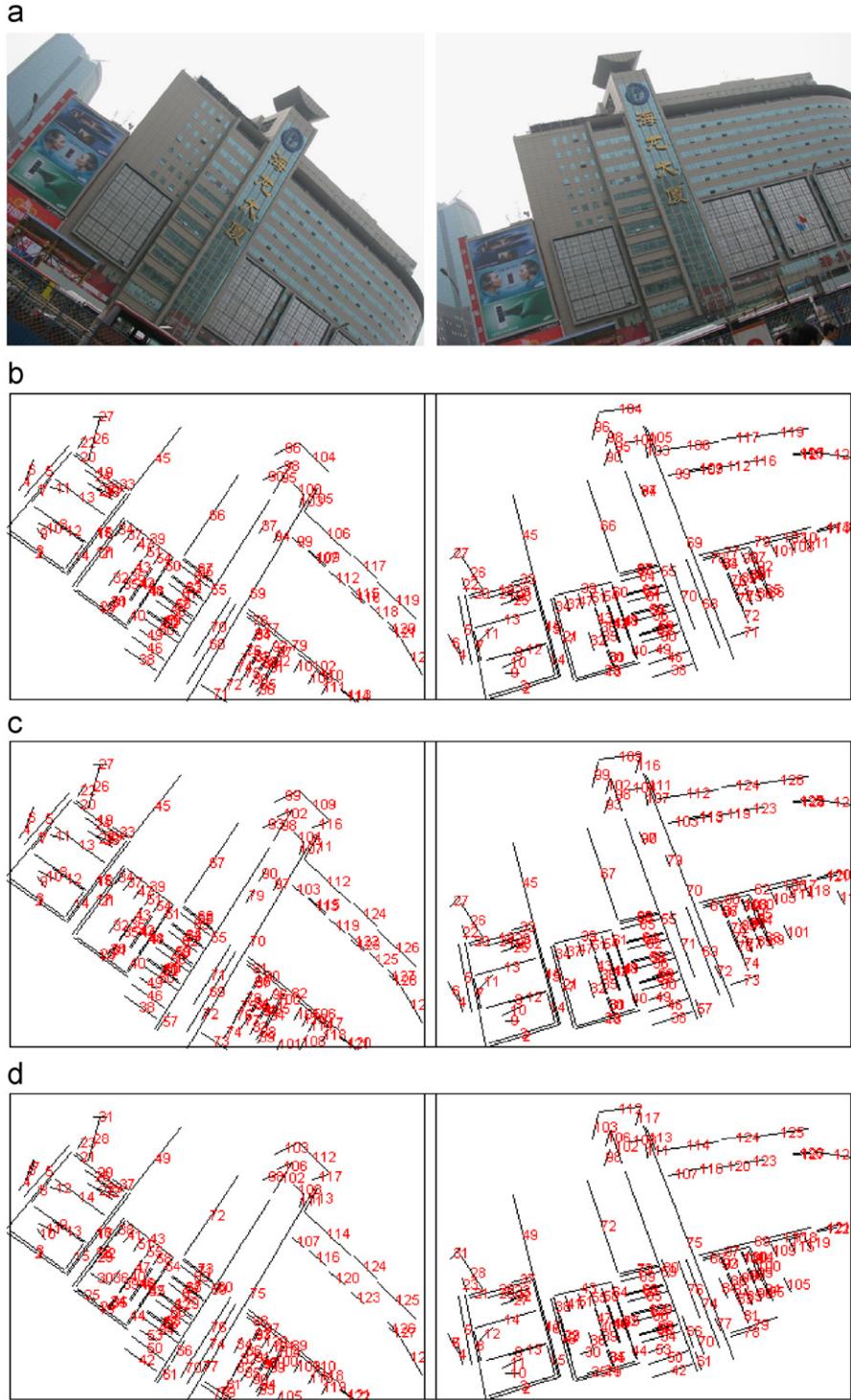


Fig. 9. Line matching using Schmid's method [10] and MSLD: Example 3. (a) Extracted lines: 296, 287. (b) Schmid's method [10] (cross-correlation + epipolar geometry). Matched lines: 122, incorrect: 5. (c) MSLD. Matched lines: 129, incorrect: 6. (d) MSLD + epipolar geometry. Matched lines: 130, incorrect: 1.

From the table, we can see that:

- Across image rotation, LRC-based criterion leads to a slightly higher CR of 91.4% while NNDR-based criterion has a CR of 89.2%. They provide similar performance in CM (171, 173).
- As for viewpoint change, LRC-based criterion clearly outperforms NNDR-based criterion either in the CM (153, 125) or in terms of the CR (93.4%, 91.6%).

- Under illumination change, in terms of CM, two criterions provide comparable performance (140, 135). While in terms of the CR, NNDR-based criterion has slightly better results (96.4%) than LRC-based criterion (95.9%).
- For image blur, NNDR-based criterion provides better results than LRC-based criterion either in the CM (118, 113) or in terms of the CR (94.5%, 91.9%).

Table 2

Matching results using MSLD and Schmid's method on images

Rotation	Images	a	b	c	d	CM	CR
	MSLD + epipolar	39(5)	32(0)	121(4)	35(5)	213	93.8%
	Schmid's method	32(6)	29(1)	102(7)	27(3)	173	91.5%
Viewpoint	Images	e	f	g	h	CM	CR
	MSLD + epipolar	25(2)	40(4)	51(4)	72(2)	176	93.6%
	Schmid's method	25(3)	32(2)	44(6)	61(1)	150	92.5%
Illumination	Images	i	j	k	l	CM	CR
	MSLD + epipolar	50(2)	47(2)	16(1)	45(1)	152	96.2%
	Schmid's method	49(2)	42(1)	15(0)	43(1)	145	97.3%
Blur	Images	m	n	o	p	CM	CR
	MSLD + epipolar	55(1)	16(2)	30(4)	52(4)	142	92.8%
	Schmid's method	51(2)	17(2)	29(4)	48(3)	134	92.4%
Noise	Images	q	r	s	t	CM	CR
	MSLD + epipolar	27(3)	19(4)	55(6)	18(3)	103	86.5%
	Schmid's method	28(2)	20(4)	52(6)	21(3)	106	87.6%
JPEG	Images	u	v	w	x	CM	CR
	MSLD + epipolar	30(3)	24(3)	47(5)	15(3)	116	87.9%
	Schmid's method	28(3)	25(3)	48(5)	16(4)	117	87.2%

For each image pair, the first number represents the total matches and the number in bracket denotes the incorrect matches.

The above matching results show that MSLD can perform well on real images across rotation, viewpoint change, illumination change and blur.

6.2. Test robustness to noise and compression

Fig. 4 shows the image set used to evaluate the performance of MSLD in addition to Gaussian noise and JPEG compression. Noisy and compressed image pairs are created by taking one original image as the first image and image in addition of Gaussian noise or compressed as the second one. The addition noise is Gaussian noise with zero mean and standard deviation of 5, 10, 15, 20, 30 and 40, and JPEG compressed images are generated by compressing the original images into 40%, 30%, 20%, 15%, 10% and 5%, respectively. Therefore, both noise image set and compressed image set include 96 image pairs.

Fig. 5 shows line matching results on noise image set and JPEG compressed image set. Graphs are generated using results on 96 image pairs. We can see from the figures that:

- The number of the CM declines almost linearly with the increase in the noise level and it decreases slowly with the increase in the compression degree.
- As for the CR, MSLD performs well when noise standard deviation is lower than 20, but its performance declines rapidly when the noise level is higher than 20. MSLD provides good results even when image is compressed into 5%.
- Two criterions lead to rather comparable results in terms of the total CM, but in terms of the CR, LRC-based criterion outperforms NNDR-based criterion.

This experiment shows that MSLD is robust to noise and JPEG compression.

6.3. Test robustness to occlusion

Considering that the line matching is more sensitive to partial occlusion than point matching, we test the performance of MSLD descriptor in the case of partial occlusion in this subsection. Fig. 6 provides two examples, in which there are 30% occlusion on average between the corresponding line pairs of two images because of camera viewpoint change. For the first pair, 37 and 32

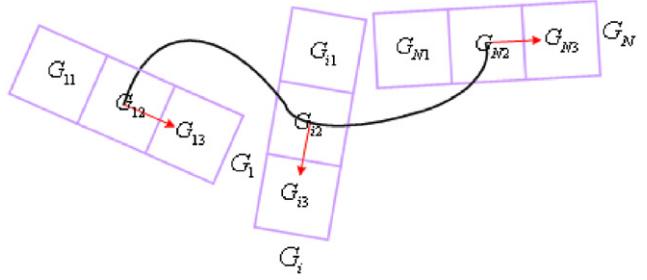


Fig. 10. Schematic figure of computing MSCD descriptor.

line segments are extracted in the two images, respectively, while 16 lines are matched and all of them are CM (b). As for the other group, 76 and 55 line segments are extracted and only two matches are incorrect among the obtained 36 matches (d). The experiment demonstrates the great robustness of our MSLD descriptor to partial occlusion.

6.4. MSLD versus Schmid's method

In this section, we compare the performance of MSLD with Schmid's method [10]: in this method firstly point correspondences are determined using the epipolar geometry on the lines in two images, and then similarity of lines is computed by averaging the cross-correlation scores of neighborhoods for all the corresponding points.

Figs. 7–9 give four examples. In each example, (a) includes images captured under two different viewpoints, (b) is the matching result using Schmid's method, (c) is the MSLD-based matching result (NNDR criterion is used) and (d) shows the result using MSLD and epipolar geometry at the same time. In Fig. 7, Schmid's method attains 110 matched lines and 15 of them are incorrect. In contrast, MSLD-based method provides a better result (116 lines are matched and 8 matches are incorrect). It is noted that MSLD-based method uses only image content information, while Schmid's method uses both image content information and the epipolar geometry. Compared to the result using only MSLD descriptor, more CM and less false matches can be achieved by using MSLD descriptor plus epipolar geometry (129 lines are matched and 3 matches are incorrect).

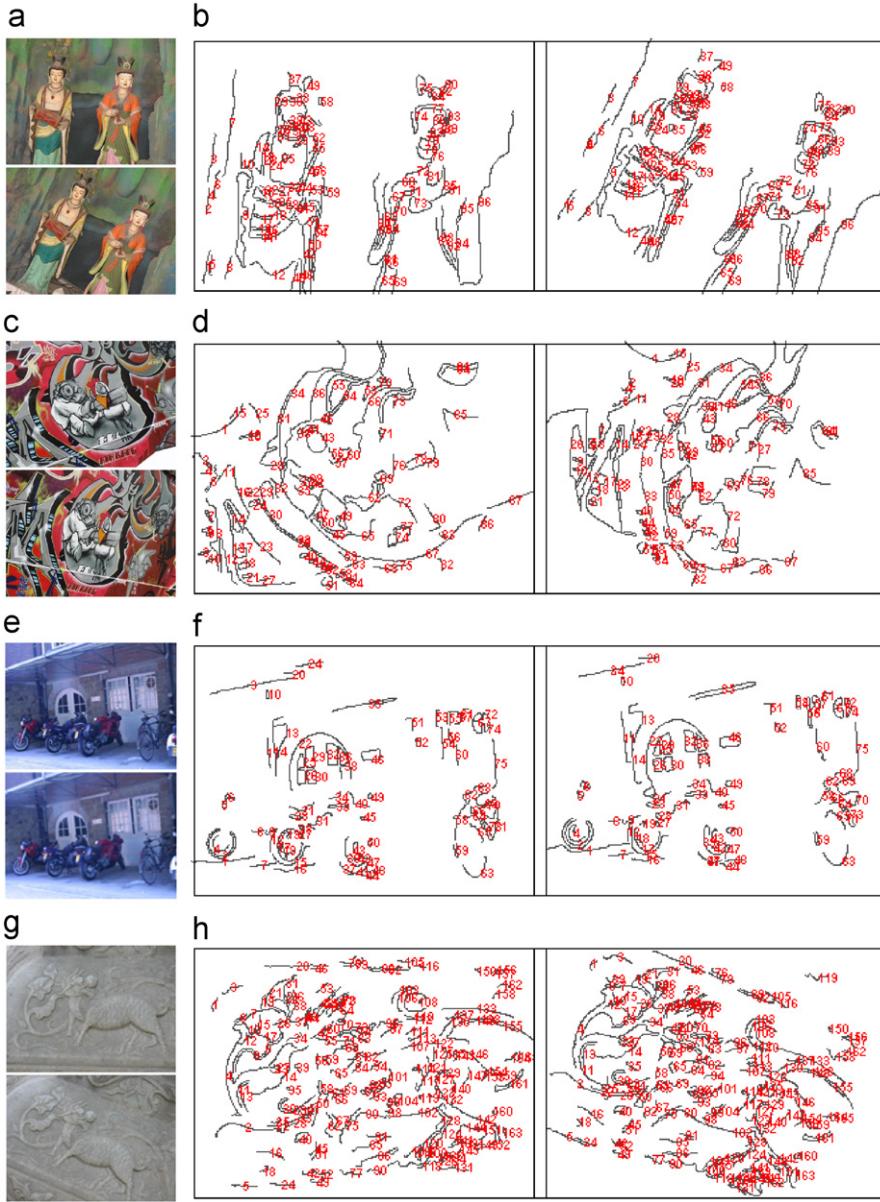


Fig. 11. Curve matching using MSCLD (image size is 640 × 480). (a), (b) Extracted curves: 136, 131, matched curves: 96, incorrect: 1. (c), (d) Extracted curves: 313, 354, matched curves: 87, incorrect: 5. (e), (f) Extracted curves: 210, 137, matched curves: 75, incorrect: 2. (g), (h) Extracted curves: 429, 423, matched curves: 165, incorrect: 3.

In examples of Figs. 8 and 9, MSLD and Schmid's method share almost equivalent performance, while using MSLD and epipolar geometry synchronously attains obviously better result.

Table 2 provides the matching results under various transformations. (a)–(o) are image pairs in Fig. 3. The noise pairs (q)–(t) and JPEG compressed image pairs (u)–(x) are created using the first four images in Fig. 4, where the added Gaussian noise level is 30 and the compression ratio is set at 5%. From the table, it can be concluded that MSLD with geometry restraint outperforms distinctly Schmid's method in both rotation and viewpoint change. As for illumination change and blur, they attain almost the same CR, while MSLD is superior to Schmid's method in the total CM. As for noise and JPEG compression, they share equivalent performance.

7. Curve and region matching

The concept used to construct line descriptor proposed in this paper is also extended to creating curve descriptor, which is called

mean–standard deviation curve descriptor (MSCLD). Fig. 10 illustrates how to construct MSCLD for a curve. A little difference exists between constructing MSLD and MSCLD: all the PSRs share the same direction when constructing MSLD for a line, while each PSR is determined by the pixel gradient direction when constructing MSCLD for a curve. Fig. 11 gives several examples using our MSCLD for curve matching. Curves are extracted using Canny edge detector. NNDR and global threshold, which are set at 0.8 and 0.5, respectively, are used as matching criterions here. There exists some rotation in image pair (a), while two images in pair (c) are captured under different viewpoint and thus leads to large affine distortion. Image blur is introduced by varying the camera focus in (e) and the last pair (g) includes two fresco images. The results indicate that our MSCLD descriptor can achieve good curve matching results on real images.

Since the boundary of a region is a closed curve or a set of closed curves and MSCLD descriptor can also be constructed on a set of curves, our curve descriptor is applicable to region matching using

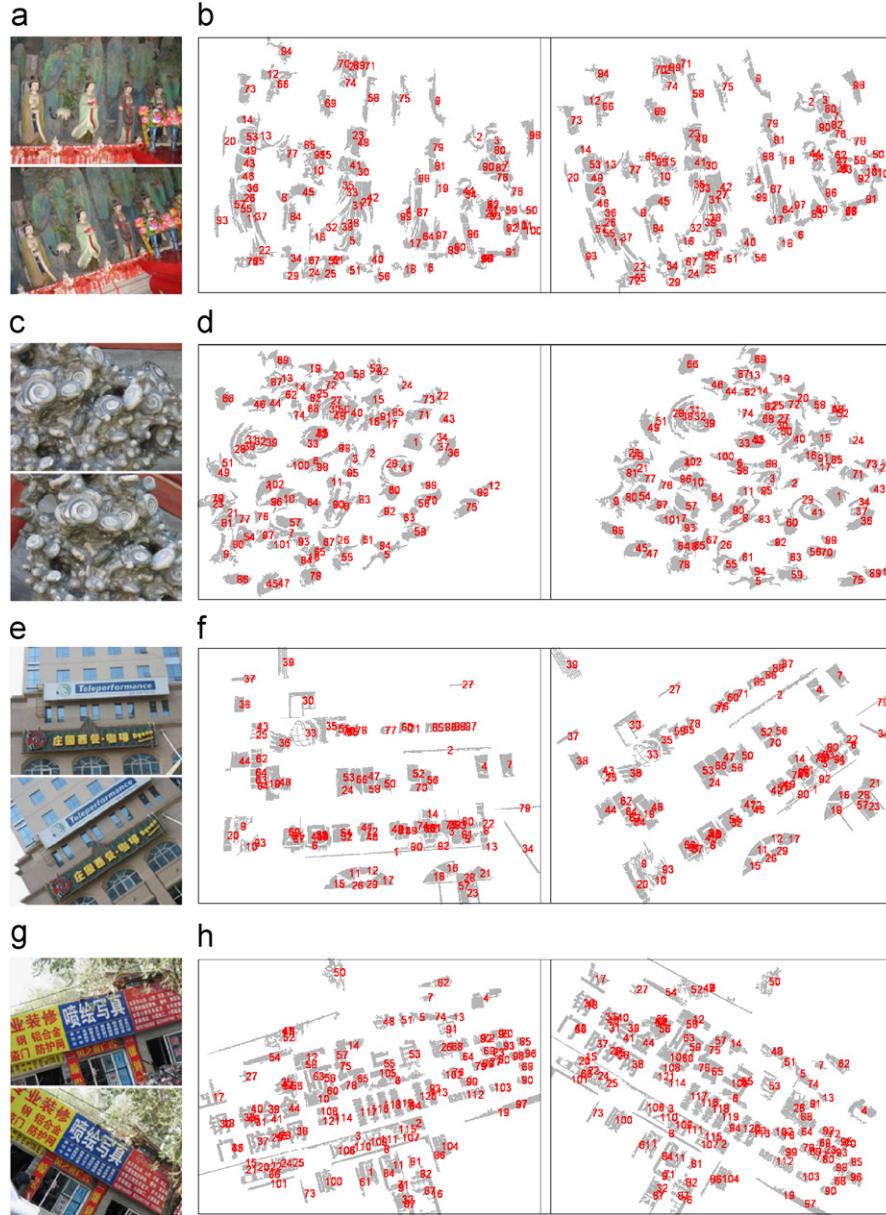


Fig. 12. Region matching using our descriptor (image size is 640×480). (a), (b) Extracted regions: 243, 233, matched regions: 101, incorrect: 1. (c), (d) Extracted regions: 258, 237, matched regions: 102, incorrect: 0. (e), (f) Extracted regions: 202, 189, matched regions: 93, incorrect: 4. (g), (h) Extracted regions: 332, 366, matched regions: 121, incorrect: 2.

MSCD of the boundary curve/curves of a region. Compared to the popularly used descriptors SIFT and shape context, both of which need an ellipse-fitting step while matching irregularly shaped regions, our MSCD-based method can avoid this step and is more effective. Fig. 12 provides several examples of our MSCD-based region matching. The feature regions are extracted using the MSER detector [15]. Experiments show that MSCD-based method performs surprisingly well for matching irregularly shaped regions.

8. Conclusion

This paper reports a new and robust descriptor MSLD for automatic wide baseline matching, which is also extensible to curve and region matching. MSLD can be used for automatic matching without resorting to any prior knowledge about scene or camera position, or can work for better performance with other conditions if other

prior knowledge is available. Experiments show that MSLD is both effective and robust to various image transformations. MSLD can be useful in 3D-reconstruction and object recognition. Fig. 13 shows a simple example of 3D reconstruction from line matching. The type of scene in example is often difficult to reconstruct from point matching, while MSLD is competent for such a scene. In the example, line segments are matched by using MSLD descriptor plus epipolar constraint.

We have also tried the way of directly partitioning line neighborhood into strips parallel to the line, and then constructing the descriptor by computing mean and standard deviation of the rotated sample gradients in each strips. However, such an undertaking is found less effective and robust than MSLD developed in this paper.

Up to now, the work described in this paper allows only to construct descriptor on a single scale, and thus our descriptor is not

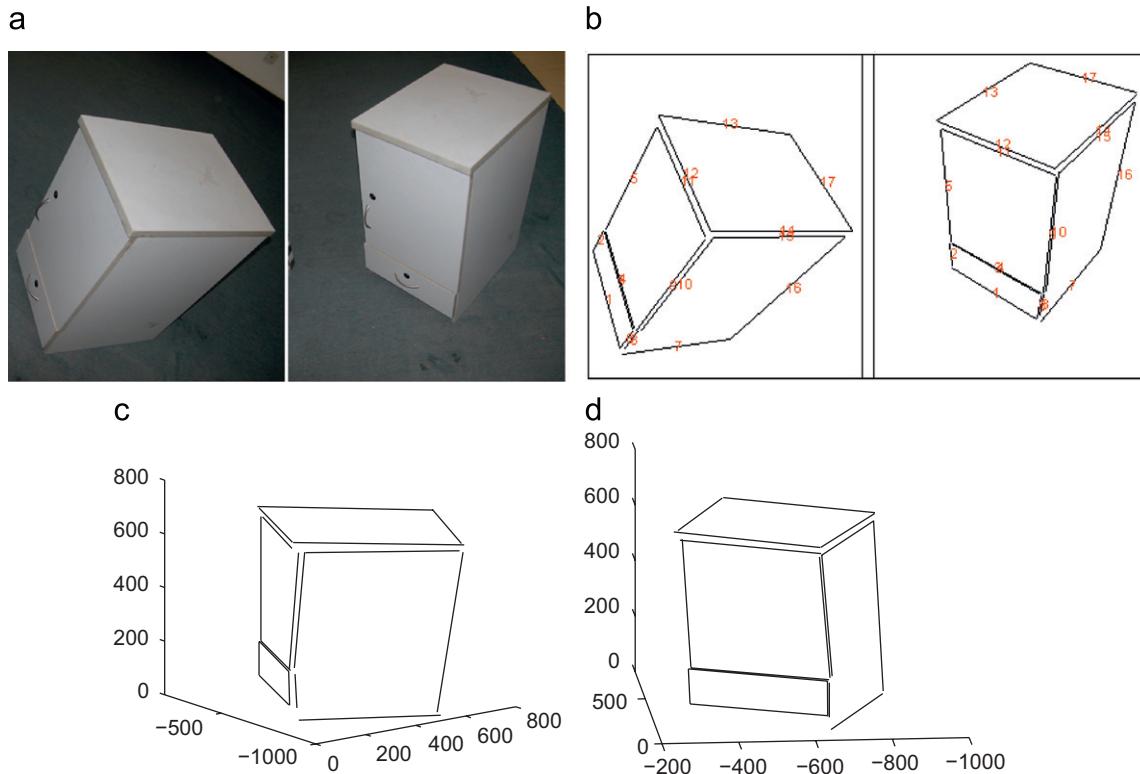


Fig. 13. An example of 3D-reconstruction from line matching. (a) Images. (b) Matched lines. (c) 3D lines from view 1. (d) 3D lines from view 2.

scale-invariant. Our future work will focus on multi-scale analyses and make the descriptor scale-invariant.

References

- [1] D.G. Lowe, Distinctive image features from scale-invariant key-points, *Int. J. Comput. Vision* 60 (2) (2004) 91–110.
- [2] K. Mikolajczyk, C. Schmid, A performance evaluation of local descriptors, *IEEE Trans. Pattern Anal. Mach. Intell.* 27 (10) (2005) 1615–1630.
- [3] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, A comparison of affine region detectors, *Int. J. Comput. Vision* 65 (1) (2006) 43–72.
- [4] A. Tang, T. Ng, Y. Hung, C. Leung, Projective reconstruction from line-correspondences in multiple un-calibrated images, *Pattern Recognition* 39 (5) (2006) 889–896.
- [5] A. Bartoli, P. Sturm, Multiple-view structure and motion from line correspondences, in: *IEEE International Conference on Computer Vision*, 2003.
- [6] F.H. Shi, J.H. Wang, J. Zhang, Y.C. Liu, Motion segmentation of multiple translating objects from line correspondences, *Pattern Recognition* 38 (10) (2005) 1775–1778.
- [7] O.A. Aider, P. Hoppenot, E. Colle, A model-based method for indoor mobile robot localization using monocular vision and straight-line correspondences, *Robotics Autonomous Syst.* 52 (2005) 229–246.
- [8] C. Schmid, A. Zisserman, Automatic line matching across views, in: *IEEE International Conference on Computer Vision and Pattern Recognition*, 1997.
- [9] R. Hartley, A linear method for reconstruction from lines and points, in: *IEEE International Conference on Computer Vision*, 1995.
- [10] C. Schmid, A. Zisserman, The geometry and matching of lines and curves over multiple views, *Int. J. Comput. Vision* 40 (3) (2000) 199–233.
- [11] M.J.A. Lourakis, S.T. Halkidis, S.C. Orphanoudakis, Matching disparate views of planar surfaces using projective invariants, *Image Vision Comput.* 18 (9) (2000) 673–683.
- [12] B. Herbert, F. Vittorio, V.G. Luc, Wide-baseline stereo matching with line segments, in: *IEEE International Conference on Computer Vision and Pattern Recognition*, 2005.
- [13] Y. Deng, X.Y. Lin, A fast line segment based dense stereo algorithm using tree dynamic programming, in: *European Conference on Computer Vision*, 2006.
- [14] G. Egnal, R.P. Wildes, Detecting binocular half-occlusions: empirical comparisons of four approaches, in: *IEEE International Conference on Computer Vision and Pattern Recognition*, 2000.
- [15] J. Matas, O. Chum, M. Urban, T. Pajdla, Robust wide baseline stereo from maximally stable extremal regions, in: *Proceedings of the British Machine Vision Conference*, 2000.

About the Author—ZHIHENG WANG is a Ph.D. candidate at the Institute of Automation, Chinese Academy of Sciences. He received his B.S. degree from Beijing Institute of Technology in 2004. His research interest covers feature extraction and image correspondence.

About the Author—FUCHAO WU is a Professor at the Institute of Automation, Chinese Academy of Sciences. His research interest covers computer vision, which include camera calibration, 3D reconstruction, active vision and image-based modeling and rendering.

About the Author—ZHANYI HU is a Professor at the Institute of Automation, Chinese Academy of Sciences. His research interests include computer vision, geometric primitive extraction, vision guided robot navigation and pattern recognition.