Robust Matching Method for Scale and Rotation Invariant Local Descriptors and Its Application to Image Indexing

Kengo Terasawa, Takeshi Nagasaki, and Toshio Kawashima

School of Systems Information Science, Future University-Hakodate,
116–2 Kamedanakano-cho, Hakodate-shi, Hokkaido, 041–8655, Japan

Abstract. Interest point matching is widely used for image indexing. In this paper we introduce a new distance measure between two local descriptors instead of conventional Mahalanobis distance to improve matching accuracy. From experiments with synthetic images we show that the error distribution of local jet is gaussian but the distribution of the descriptors derived from local jet is not gaussian. Based on the observation, we design a new distance measure between two local descriptors and improve accuracy of point matching. We also reduce the number of candidate points and reduce the computational cost by taking into account the characteristic scale ratio. Experimental results confirm the validity of our method.

1 Introduction

Appearance-based image indexing can be roughly divided into two approaches: global method and local method. Local appearance describes the characteristics of small area around some specific points. This approach has advantage of robustness to partial occlusions and background changes. Furthermore, this technique is appropriate to multiple object search and pose recognition because it matches corresponding features in images and is consequently able to recover the structure of images.

Usually, this approach begins with the extraction of points where image information is concentrated: such points are called interest points. After the extraction of interest points, the characteristics of local area around each extracted interest point is described by a local descriptors vector. Finally, the interest points are matched according to the similarity of local descriptors. At every step of these processes, a lot of studies for improvement have been proposed and evaluated.

There are many way to extract interest points. Most commonly used methods are a corner detector based on Harris function [4] and a blob detector based on Laplacian. Both of them use the responses of Gaussian derivative filters for implementation and a problem is that Gaussian derivatives are dependent on image scale. Scale-space theory introduced by Lindeberg [6] normalize these

derivatives and generalizes interest points. Based on his scale normalized differentiation, many type of scale invariant interest point detectors are derived in the past few years [7–9].

Local feature of these interest points are described by a feature descriptor. Local jet [5] is often used to describe the characteristics of local feature. It is a set of the responses of Gaussian derivative filters which describes the neighborhood of a point. Unfortunately, it again poses a problem that Gaussian derivatives are dependent on image orientation and scale. To avoid the rotation dependency, rotation invariant vectors based on local jet components are devised [11, 13, 15]. Another approach to obtain rotation invariance is to normalize local jets to dominant direction [1, 9]. This normalization is implemented by applying steerable filter [3] to the gradient direction of that point. Scale invariance is realized by describing local descriptor of interest points at multiple-scales [2, 13] or by using characteristic scale to determine a radius of neighbor region [1, 9]. Alternative approaches also exist, such as: SIFT descriptors proposed by Lowe [7, 8] and complex filters proposed by Schaffalitzky and Zisserman [12]. Mikolajczyk and Schmid [10] reported that SIFT shows the best performance among of them, and the steerable filter follows it.

In interest point matching implementation, Mahalanobis distance is commonly used as a similarity measure of two local descriptors. This distance measure has an advantage of simplicity in computation, however, it also has some disadvantages. One such is that it uses a single covariance matrix to all the points — this hypothesis is inadequate for some cases. Another problem is that they hypothesize error distribution as Gaussian normal without verification.

In this paper, we investigate the error distribution of local descriptors by observing synthetic images with controlled displacements such as small translations, small stretches, and random noises. According to the observation, we design a new distance measure for the rotation invariant local descriptors. The distance definition improves the precision of point matching.

In addition, we refine indexing process by taking into account the characteristic scale ratio information. Experimental result shows effectiveness of our methods.

2 Local Jet and Its Scale Invariant Formulation

In this paper we use local jets both in interest point detection and in description of its local neighborhoods. The following is brief introduction to local jets and its normalization to scale. Local jets are the responses of Gaussian derivative filters, and are written as

$$L_{i_1 \ldots i_n}(\boldsymbol{x}, \sigma) = G_{i_1 \ldots i_n}(\boldsymbol{x}, \sigma) * I(\boldsymbol{x}), \tag{1}$$

where $I(\boldsymbol{x})$ is image intensity, $G(\boldsymbol{x}, \sigma)$ is Gaussian distribution function, and subscripts represent partial differentiation. These $L_{i_1 \ldots i_n}$ are dependent on image resolution and Gaussian parameter σ. They are, therefore, inconvenient to use for image indexing with different scale.

To be robust to scale change, normalized Gaussian derivatives and characteristic scale are often used [6, 9]. The normalized Gaussian derivatives of m-th order is written as

$$D_{i_1 \ldots i_m}(\boldsymbol{x}, \sigma) = \sigma^m L_{i_1 \ldots i_m}(\boldsymbol{x}, \sigma). \tag{2}$$

To prove that $D_{i_1 \ldots i_m}$ is normalized for scale, consider two images of different scale, I and I', which are connected with $I(\boldsymbol{x}) = I'(\boldsymbol{x}')$, where $\boldsymbol{x}' = t\boldsymbol{x}$. Applying Gaussian derivatives to this equation, we obtain

$$\sigma^m G_{i_1 \ldots i_m}(\boldsymbol{x}, \sigma) * I(\boldsymbol{x}) = t^m \sigma^m G_{i_1 \ldots i_m}(\boldsymbol{x}, t\sigma) * I'(\boldsymbol{x}'), \tag{3}$$

so that we have

$$D_{i_1 \ldots i_m}(\boldsymbol{x}, \sigma) = D'_{i_1 \ldots i_m}(\boldsymbol{x}', t\sigma). \tag{4}$$

This equation indicates that the values of D are independent of the scale of image, if σ is properly chosen. In practice, since the scale ratio between two images is unknown, it is not clear how to choose appropriate σ. It should satisfy $\sigma' = t\sigma$.

The use of characteristic scale solves this problem. Characteristic scale can be obtained by examining some kind of function of normalized Gaussian derivatives, e.g. squared gradient, Laplacian, Harris function, and so force. The function should be chosen depending on the purpose. The characteristic scale is defined as the scale σ where a function of normalized Gaussian derivatives takes a peak value as shown in Fig. 1. The characteristic scale is proportional to image scales, therefore we can make local jet invariant to scale by substituting characteristic scale for σ in (2). The bottom row of Fig. 1 represents the Harris function value plotted over σ at the points displayed center of circles in the images in the top row. Characteristic scale is represented as dashed line in the bottom row, and also represented as the radius of the circle in the top row. It indicates that the ratio between two σ at corresponding point of each image gives the scale ratio between two images.

3 Interest Point Detection

In this section, we describe how to extract scale invariant interest points. Several methods are proposed to extract interest points. Among of them, the method of Harris and Stephens [4] is reported to show good repeatability [14]. In the followings, Harris method extended to scale space is introduced.

The idea is to search for maxima in the 3D space of x, y and σ. Here again σ represents the Gaussian parameter. For implementation, all coordinates should be represented in discrete domain. x and y are the location of a pixel, and σ is the scale sampled at exponential intervals, $\sigma_n = k^n \sigma_0$. In this 3D grid space, we calculate the 2×2 matrix:

$$M = \exp\left(-\frac{x^2 + y^2}{2\widetilde{\sigma}^2}\right) \otimes \begin{bmatrix} D_x^2 & D_x D_y \\ D_x D_y & D_y^2 \end{bmatrix}, \tag{5}$$

where $\widetilde{\sigma}$ must be proportional to σ. In this study, we set $\widetilde{\sigma} = \sigma$.

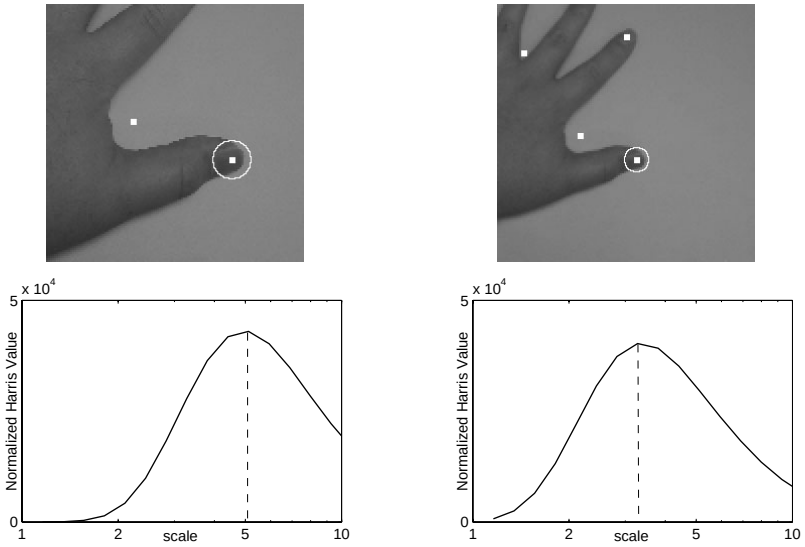


Fig. 1. Two images of different scale and the characteristic scale of corresponding point (the top of the thumb)

Next, we calculate Harris function R:

$$R = \det(M) - k \ \mathrm{trace}(M)^2, \tag{6}$$

where k is a constant. In this study, we set $k = 0.06$, which is commonly used value. This R is invariant to rotation and scale. Scale invariant interest point is defined as coordinates (x, y, σ) where the function gives local extrema of R. In practice, certain threshold value t is set to avoid the clutters. Summarizing above, point (x_n, y_m, σ_l) is extracted as scale invariant interest point if it satisfies

$$\begin{aligned} &R(x_n, y_m, \sigma_l) \geq R(x_{n+i}, y_{m+j}, \sigma_{l+k}), \quad \forall i, j, k \in \{-1, 0, 1\} \\ &R(x_n, y_m, \sigma_l) \geq t. \end{aligned} \tag{7}$$

In this paper, we set threshold value $t = 4000$. As a result, the average number of extracted point per image is about 50 to 200.

In this way, we can extract interest points that is invariant to image rotation and scales. Figure 2 is an example. This figure shows interest points of two images with different scale and rotation. Interest points are represented as white dot in the figure. Each point has 3D coordinates, x, y, σ-coordinates. If they are compared in each same row respectively, they do not correspond at all. However, they do correspond if taken as point group in 3D space.

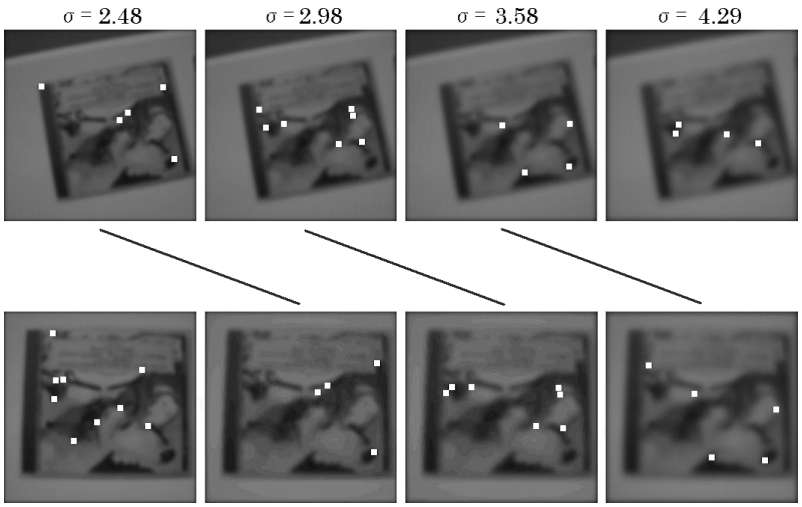


Fig. 2. Scale invariant interest points of two images with different scale and orientation. Interest points have 3D coordinates (x, y, σ)

4 Description of Local Neighborhoods

In this section, we consider how to describe the characteristics of the region around each interest point. It is desirable that this descriptor is robust to illumination change, camera position, camera noise, etc.

One idea is to use local jets introduced in Sect. 2. If we set characteristic scale to σ in (2), it obtains invariance to scale. There are several candidates for function to determine the characteristic scale as described before. In this study, we already calculated Harris function to extract interest points, so we use this again. In this case, characteristic scale is the same as σ-coordinate of interest point. This local jets are invariant to scale but not to rotation. To obtain the invariance to rotation, combine the components as:

$$\nu[0...8] = \begin{bmatrix} D \\ D_i D_i \\ D_i D_{ij} D_j \\ D_{ii} \\ D_{ij} D_{ji} \\ \varepsilon_{ij}(D_{jkl} D_i D_k D_l - D_{jkk} D_i D_l D_l) \\ D_{iij} D_j D_k D_k - D_{ijk} D_i D_j D_k \\ -\varepsilon_{ij} D_{jkl} D_i D_k D_l \\ D_{ijk} D_i D_j D_k \end{bmatrix}, \tag{8}$$

where $\varepsilon_{xy} = -\varepsilon_{yx} = 1$, $\varepsilon_{xx} = \varepsilon_{yy} = 0$. This vector $\boldsymbol{\nu}$ is invariant to rotation [11, 13].

Furthermore, to get robustness to illumination change, substitute $D_{i_1 \ldots i_m}$ to $D_{i_1 \ldots i_m}/D$. By this substitution (so we don't use $\nu[0]$), the vector becomes

invariant to linear illumination change. We use this 8-dimensional vector $\boldsymbol{\nu}$ as the interest point descriptor.

In this time, only components up to third order are used. Higher order derivatives may provide more accurate descriptor and may improve the precision of point matching, however, higher order term is likely to make descriptors more sensitive to noise, and have disadvantage that increase computational cost. For this reason, we use terms up to third order derivatives.

5 Interest Point Matching

5.1 Test for Normality of Error Distribution

Owing to the local descriptor derived in the last section, interest point matching based on their local features is now available. Matching candidates are determined by measuring similarity between descriptors, i.e. the point whose descriptor gives the smallest distance from the descriptor of query point in the feature space is considered as a matching candidate. To implement this process, the distance measure between two descriptors must be defined. In existing method, Maharanobis distance

$$d_M^2(\boldsymbol{\nu}_i, \boldsymbol{\nu}_j) = (\boldsymbol{\nu}_i - \boldsymbol{\nu}_j)^T \Lambda^{-1} (\boldsymbol{\nu}_i - \boldsymbol{\nu}_j) \tag{9}$$

(where Λ^{-1} is inverse of covariance matrix) is used frequently [13, 9]. This method has advantage that it can be calculated easily, but it also has disadvantage or incompleteness. One incompleteness is caused by the fact that it ideally requires covariance matrix for each and every cluster, where in this case cluster means the set of interest points that should be regarded to correspond. In the real situation, since there exist infinite variety of interest point, it is impossible to know all covariance matrices in advance. Ordinary implementation uses only one global covariance matrix as a substitute for such covariance matrices. Another incompleteness is that it is based on the assumption that the error of descriptors should follow normal distribution, but this assumption is verified neither theoretically nor experimentally.

To know the characteristics of error distribution, we executed an experiment as follows. We define the error as the difference between true value and observed value for each local descriptor of the interest point. Observed value is the summation of true value and observation noise.

We suppose that the noise comes mainly from imaging sensor and digitization process. To simulate these two noise artificially, we employ the function

$$I(\boldsymbol{x}) = \frac{1}{1 + e^{-(\boldsymbol{x}^T A \boldsymbol{x} - r^2)}}, \tag{10}$$

which gives ellipse image (Fig. 3). We modify this image by slight translation (less than one pixel), and slight scale change, and adding 10% white noise, i.e.

$$I'(\boldsymbol{x}) = I(a\boldsymbol{x} + \boldsymbol{u}) + w, \tag{11}$$

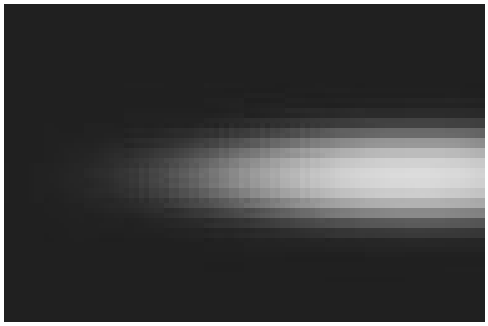


Fig. 3. Synthetic image used in our experiment

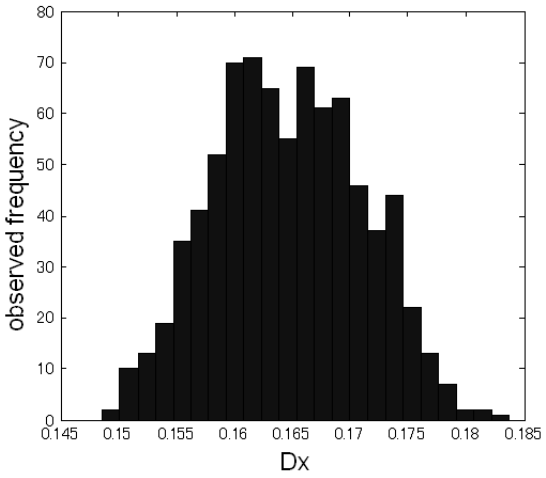


Fig. 4. Distribution of D_x

where a represents stretch ratio, $\boldsymbol{u}$ represents small translation, and w represents white noise. Numerous I' are made, and the statistics of local descriptors observed in whole these images are considered as the error distribution of local descriptors under noise. Figure 4 is the result of D_x. This distribution is symmetric, has a single peak, forms a bell curve, looks like a normal distribution. To verify the normality of distributions, we also apply the Kolmogorov-Smirnov test for normality. Table 1 represents the result. The result indicates that it is adequate to regard the distribution of original local jets have normality, but it is not adequate to regard the distribution of rotation invariant descriptors constructed by combination of local jets have normality.

The non-normality of rotation invariant descriptors can be understood by following discussion. The observation error of rotation invariant descriptors $\boldsymbol{e}$ is

Table 1. Result of Kolmogorov-Smirnov test for normality to local descriptors. Level of significance is set at 0.05

descriptor	:	result	descriptor	:	result
D_x	:	Not Rejected	$\nu[1]$	:	Not Rejected
D_y	:	Not Rejected	$\nu[2]$	:	Rejected
D_{xx}	:	Not Rejected	$\nu[3]$	:	Rejected
D_{xy}	:	Not Rejected	$\nu[4]$	:	Rejected
D_{yy}	:	Not Rejected	$\nu[5]$	:	Not Rejected
D_{xxx}	:	Not Rejected	$\nu[6]$	:	Not Rejected
D_{xxy}	:	Not Rejected	$\nu[7]$	:	Not Rejected
D_{xyy}	:	Not Rejected	$\nu[8]$	:	Rejected
D_{yyy}	:	Not Rejected			

expressed as:

$$\widetilde{\nu[i]} = \nu[i] + e[i], \tag{12}$$

where $\boldsymbol{\nu}$ represents true value, $\widetilde{\boldsymbol{\nu}}$ represents observed value, and $\nu[i]$ represents the i-th element of vector $\boldsymbol{\nu}$. On the other hand, the error of local jet ε_* is defined as

$$\widetilde{D_*} = D_* + \varepsilon_*. \tag{13}$$

As mentioned above, it is adequate to hypothesize that ε_* follows normal distribution. If we hypothesize that, the error distribution for the combination of some D_* is obtained by calculation. For example, the error of $\widetilde{D_x}\widetilde{D_x}$ is calculated as:

$$\begin{aligned} \widetilde{D_x}\widetilde{D_x} &= (D_x + \varepsilon_x)(D_x + \varepsilon_x) \\ &= D_x D_x + 2\varepsilon_x D_x + \varepsilon_x^2 \end{aligned} \tag{14}$$

This equation indicates that errors expand proportional to the values of D_x. It derives the non-normality of $\boldsymbol{\nu}$, corresponding to our observation.

5.2 The Arranged Definition of Distance Measure

Since we have found that the error of $\boldsymbol{\nu}$ does not follow normal distribution, we should rearrange distance measure according to this speculation.

Assuming that ε_* follows normal distribution $N(0, \sigma)$, the error of $\nu[1]$ is evaluated as

$$\begin{aligned} e[1] &= \widetilde{\nu[1]} - \nu[1] \\ &= (\widetilde{D_x}\widetilde{D_x} + \widetilde{D_y}\widetilde{D_y}) - (D_x D_x + D_y D_y) \\ &\approx 2\varepsilon_x D_x + 2\varepsilon_y D_y \\ &= 2D_x N(0, \sigma) + 2D_y N(0, \sigma). \end{aligned} \tag{15}$$

Using the additivity of variance of independent variable, the variance of this equation is evaluated as $4\sigma^2(D_x^2 + D_y^2)$. Therefore, if we divide this equation

by $\sqrt{4(D_x^2+D_y^2)}$, the variance of error is normalized to constant value, σ^2. (In practice, since the true value of D_x, D_y is unknown, the maximum likelihood estimate $\widetilde{D_x}, \widetilde{D_x}$ is used instead). Same calculation is made to the rest element of $\boldsymbol{\nu}$, and we obtain

$$
\begin{aligned}
e[2] \approx\ & 2\varepsilon_x(D_{xx}D_x + D_{xy}D_y) + 2\varepsilon_y(D_{yy}D_y + D_xD_{xy}) \\
& + \varepsilon_{xx}D_x^2 + \varepsilon_{yy}D_y^2 + 2\varepsilon_{xy}D_xD_y, \\
e[3] \approx\ & \varepsilon_{xx} + \varepsilon_{yy}, \\
e[4] \approx\ & 2\varepsilon_{xx}D_{xx} + 4\varepsilon_{xy}D_{xy} + 2\varepsilon_{yy}D_{yy}, \\
e[5] \approx\ & \varepsilon_{xxx}D_y^3 - 3\varepsilon_{xxy}D_xD_y^2 + 3\varepsilon_{xyy}D_x^2D_y - \varepsilon_{yyy}D_x^3 \\
& - 3\varepsilon_x(D_{xxy}D_y^2 - 2D_{xyy}D_xD_y + D_{yyy}D_x^2) \\
& + 3\varepsilon_y(D_{xxx}D_y^2 - 2D_{xxy}D_xD_y + D_{xyy}D_x^2), \\
e[6] \approx\ & \varepsilon_{xxx}D_xD_y^2 + \varepsilon_{xxy}(D_y^3 - 2D_x^2D_y) + \varepsilon_{xyy}(D_x^3 - 2D_xD_y^2) + \varepsilon_{yyy}D_x^2D_y \\
& + \varepsilon_x(D_{xxx}D_y^2 - 4D_xxyD_xD_y + 3D_{xyy}D_x^2 - 2D_{xyy}D_y^2 + 2D_{yyy}D_xD_y) \\
& + \varepsilon_y(2D_{xxx}D_xD_y + 3D_{xxy}D_y^2 - 2D_{xxy}D_x^2 - 4D_{xyy}D_xD_y + D_{yyy}D_x^2), \\
e[7] \approx\ & \varepsilon_{xxx}D_xD_y^2 + \varepsilon_{xxy}(2D_xD_y^2 - D_x^3) + \varepsilon_{xyy}(D_y^3 - 2D_x^2D_y) + \varepsilon_{yyy}D_xD_y^2 \\
& + \varepsilon_x(2D_{xxx}D_xD_y + 2D_{xxy}D_y^2 - 3D_{xxy}D_x^2 - 2D_{xyy}D_xD_y - D_{yyy}D_y^2) \\
& + \varepsilon_y(D_{xxx}D_x^2 + 2D_{xxy}D_xD_y + 3D_{xyy}D_y^2 - 2D_{xyy}D_xD_y - 2D_{yyy}D_xD_y), \\
e[8] \approx\ & \varepsilon_{xxx}D_x^3 + 3\varepsilon_{xxy}D_x^2D_y + 3\varepsilon_{xyy}D_xD_y^2 + \varepsilon_{yyy}D_y^3 \\
& + 3\varepsilon_x(D_{xxx}D_x^2 + 2D_{xxy}D_xD_y + D_{xyy}D_y^2) \\
& + 3\varepsilon_y(D_{xxy}D_x^2 + 2D_{xyy}D_xD_y + D_{yyy}D_y^2).
\end{aligned}
\tag{16}
$$

These equations lead us to normalization of error variance for each coordinate of $\boldsymbol{\nu}$.

Our new measure is summarized as follows. The distance measure between two local descriptor vector $d(\boldsymbol{\nu}_i, \boldsymbol{\nu}_j)$ is defined as

$$
d^2(\boldsymbol{\nu}_i, \boldsymbol{\nu}_j) = \sum_k \frac{(\nu_i[k] - \nu_j[k])^2}{\alpha_i[k] + \alpha_j[k]}, \tag{17}
$$

where α is

$$
\begin{aligned}
\alpha[1] &= 4(D_x^2 + D_y^2), \\
\alpha[2] &= 4(D_{xx}D_x + D_{xy}D_y)^2 + 4(D_{yy}D_y + D_xD_{xy})^2 + D_x^4 + D_y^4 + 4D_x^2D_y^2,
\end{aligned}
$$

and so force (omitted $\alpha[3..8]$ are easily derived from (16)). α_i is calculated by the derivatives correspond to $\boldsymbol{\nu}_i$, and α_j is calculated by the derivatives correspond to $\boldsymbol{\nu}_j$.

Table 2. Rank of corresponding point with noise

	4dim Mahalanobis	8dim Mahalanobis	4dim arranged	8dim arranged
worst	194556	180447	53913	28502
mean	22561	20744	3641	380

5.3 Experimental Verification of Arranged Distance Measure

We also made a experiment to verify the effect of our new definition of distance measure. First, we extract 1306 interest points from 21 test images, and then calculate distance for all combinations (1306×1305/2=852165 pairs), and sort it in increasing order. In following experiment, this sorted distance table is referred to evaluate how small the distance is. On the other hand, we calculate distance between each pair of corresponding point of synthetic images. In this time synthetic modification are made by

$$I'(\boldsymbol{x}) = I(aR\boldsymbol{x} + \boldsymbol{u}) + w, \tag{18}$$

where R represents rotation matrix. The descriptor $\boldsymbol{\nu}$ is ideally robust to such modifications. At last, we evaluate how small the distances of two $\boldsymbol{\nu}$ with different modification is, by use of ranked distance table obtained above. In the evaluation, we use the largest distance (worst) and average distance (mean). Same experiment are executed in four cases, that is, the dimensions of vector $\boldsymbol{\nu}$ is set either 8-dimension (it means up to third order derivatives are used) or 4-dimension (up to second order derivatives are used), and ordinary Mahalanobis distance and our arranged distance is applied respectively. In ordinary Mahalanobis distance case, Λ is covariance matrix of all 1306 descriptors.

Table 2 represents the result. In mean distance evaluation, it is found that if we use the Mahalanobis distance measure of 4-dimension, corresponding point stand at rank 22561 of 852165, it means that the correct matching stand inner of about 2.6%. On the other hand, if we use our arranged distance measure of 4-dimension, the correct match stand at rank 3641 of 852165, it means that stand inner of 0.43%. Furthermore, if we use arranged measure of 8-dimension, the result is improved moreover. These result represents the advantage of our method.

In addition, we have made same experiment for descriptors normalized to dominant direction by steerable filter. However, this type of descriptor needs highly precise dominant direction estimation. We found it difficult to obtain sufficient repeatability and distinctiveness if the dimension of descriptor is relatively low as this experiment.

6 Elimination of False Matching

Since the characteristic scale is proportional to image resolution, the proportion of characteristic scale of corresponding point in two image should be constant.

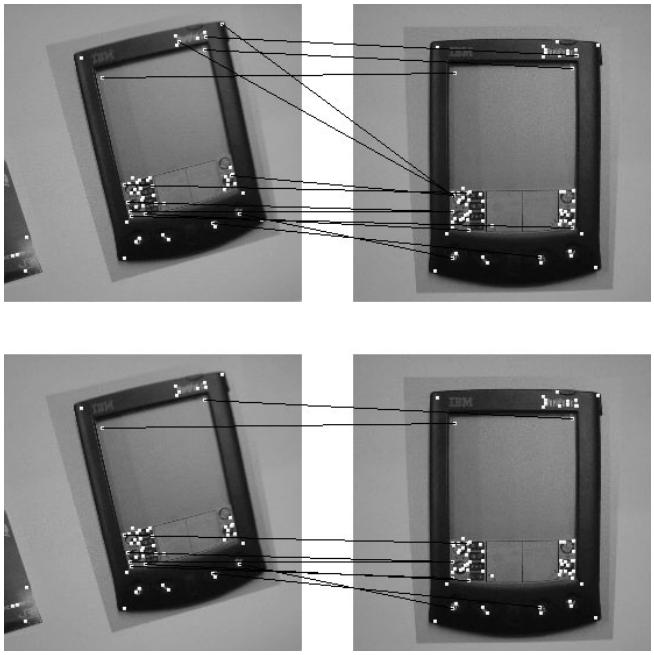


Fig. 5. Improvement of point-to-point matching. The top row is the result of simple nearest neighbor method, and the bottom row is the result after elimination of false matching

By using this properties, we can roughly guess the scale proportion between two images. It leads to elimination of false matching. In this section, according to this consideration, we propose a new modification to image indexing method.

At the first step, for each interest point of query image, calculate the distance for all interest points in the database. For each point in query, the nearest neighbor point (the point which gives minimum distance) in the each image of database is set as first candidate for corresponding point (top of Fig. 5).

The second step is the estimation of scale proportion between two images. This estimate is obtained by voting method based on the first candidates for corresponding point obtained in the last step. Table 3 is the example. As plotting candidate of corresponding point on the table, we may find the diagonal area where much votes concentrates if two image is corresponding (as top of Table 3). In the case of this example, we can find diagonal line one block upper from main diagonal line acquires much votes. This is the good estimate of scale proportion of two images. In practice, we accumulate the number per scale ratio (as bottom of Table 3), and the proportion with maximum vote is taken as estimate for scale proportion. In this example, the proportion 1:1.2 acquires the maximum vote, therefore we can estimate the scale proportion of two images as 1:1.2.

As we can estimate scale proportion according to this method, we can eliminate false positive matching candidates. It is, if the candidate combination is

Table 3. Match table of characteristic scale

Query Image	Database Image 1.44	1.72	2.07	2.48	2.98	3.58	4.29	5.15
1.44	3	2	1	2			1	
1.72			3	1			3	
2.07			1	4				
2.48					4			
2.98				1		4	1	
3.58							4	
4.29								2
5.15	1					1		

Diagonally accumulated count

···	1.44:1	1.2:1	1:1	1:1.2	1:1.44	1:1.72	···
	1	1	4	23	3	2	

off the line of Table 3, it is likely to false matching. It should be discarded and new candidate should be taken among the on-line combination. That is to say, we choose a first candidate of corresponding point as the nearest neighbor point in the whole points in the database at first, and after the estimation of scale between two images, the candidate point is re-chosen as the nearest neighbor point in the limited points which has estimated proportion of characteristic scale to the query point. By this operation, we can eliminate the false positive matches, and able to improve the accuracy of point matching (bottom of Fig. 5).

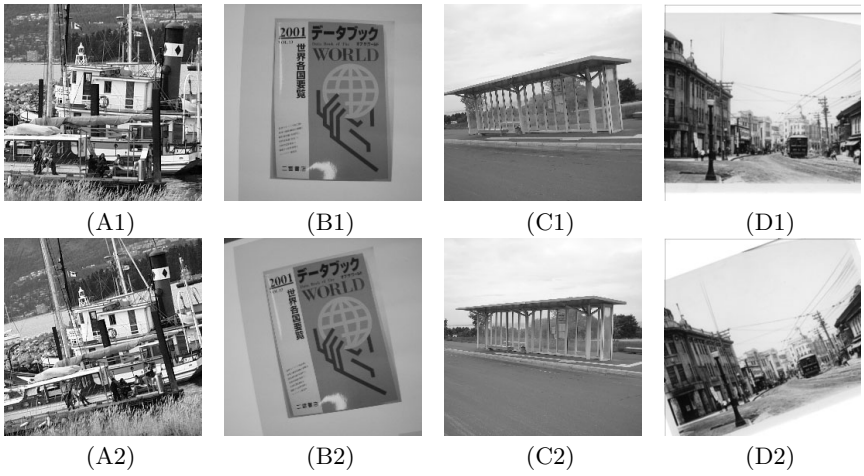
In literature, there exist other method to eliminate false match such as geometric coherence check or RANSAC. However test of such type has problem that the calculation cost explode as number of interest points increases. In contrast to it, the cost of elimination method proposed in this section is linear to the number of interest points. Furthermore, we may also apply geometric elimination method after proposed elimination method with far lower computational cost. The reduction of the number of candidate points is experimentally verified in the next section.

7 Experiment

In order to verify the effect of our method for indexing, we made an image retrieval experiment for an image database consists of 852 images. The database contains scanned picture postcards published in our city in the last hundred years. Parts of them are shown in Fig. 6.

In the retrieval test, we use 32 image pairs; each pair consists of images of same scene but with different scale and camera angle. For each pair, one image is used as the query image, and the other is included in the database. Among these pairs, 9 pairs are taken from publicly available database provided by Mikolajczyk's website [16], and 10 pairs are photoshots taken near our laboratory with

Fig. 6. Parts of images in our database



(A1) (B1) (C1) (D1)

(A2) (B2) (C2) (D2)

Fig. 7. Parts of images used in our experiment. A1 and A2 are provided by Mikolajczyk, B1,B2,C1 and C2 are photoshots taken near our laboratory, D1 and D2 are scanned postcards

different camera angle, and the rest 12 pairs are a scanned picture image and its transformed one. Applied transformations are 20 degree rotation and scale changes. Parts of them are shown in Fig. 7.

Our implementation goes as:

0. Extract interest points and calculate local descriptors for all images in the database. Images are 256-level grayscale, 256×256 pixels, and discrete σ is 10 step ($\sigma = 1.2^n,\ n = 1, 2, ..., 10$). The threshold value of Harris function for interest point extraction is set as $t = 4000$. Descriptors are 8-dimensional rotation invariant descriptors, where up to third order local jets are used.
1. In the same way, extract interest points and calculate local descriptors for query image.
2. Calculate distance for each pair of interest points between query image and database image. For each interest point of query image, the nearest neighbor point in each database image is chosen as first candidate of corresponding point.

3. According to first candidate, estimate image scale ratio by voting method. The scale ratio which acquire the maximum vote is chosen as estimate.
4. Re-choose the candidate point as its characteristic scale correspond to that of query point. In this step, only ± 1 step error is allowed.
5. If the distance between query point and final candidate point is under certain constant, determine them as the corresponding point (this constant is adjusted experimentally: in this case we choose 0.04 for arranged measure, and 0.20 for Mahalanobis measure). The number of determined corresponding point is regarded as matching score of the two images.
6. The image which marks highest matching score is chosen as indexed image. In this step, because characteristic scale found at minimum scale or maximum scale is not true characteristic scale, they are excluded for calculation.

In usual image retrieval system, some refinement such as geometric coherence check is processed after above process, but its computational cost explode as number of candidate matching increases. The reduction of the number of candidate points significantly save the computational cost of geometric coherence check. From this viewpoint, we display the result without geometric refinement. Table 4 is the result. In the table, the recognition rate and average matching score of correct image and incorrect image is represented.

By comparing these result, we can verify the advantage of our method. Our indexing method without elimination choose true image in 56.3% cases, it is higher that of Mahalanobis method, 21.9%. It shows the effect of our distance measure arrangement. Furthermore, by use of our false match elimination method, the recognition rate of our method grows above 80%. Remark that this refinement cost is proportional to the number of corresponding point, in contrast to cost of geometric refinement method explode as the number increases.

It is also observed that our method can eliminate false matching significantly. Although the matching score of correct image decline only 24%, the matching score of incorrect image decline 61%. It leads to save the computational cost of geometric coherence check, if applied.

8 Conclusion

In this paper, we proposed two ideas which improve image indexing based on local descriptors. The first idea is a new distance measure for local descriptors which improves the accuracy of point-to-point matching. The other idea reduces the number of matching candidates by voting characteristic scale ratio. The experimental result has proved the effectiveness of our method.

For further improvement of our method, a promising approach is to use higher dimensional local descriptors, or to add the constraints of semi-local coherence. In case of these arranged implementation, the advantage of our method still exist as the reduction of computational cost, because of its contribution to the reduction of candidate point.

Table 4. Results of image indexing. 'our method with elimination' is the result after all step(1–6) executed, and 'our method without elimination' is the result without step 3 and 4

Recognition Rate

Method	recognition rate
Mahalanobis distance based method	21.9%
our method without elimination	56.3%
our method with elimination	81.3%

Matching Scores

Method	average score for correct image (a)	average score for incorrect image (b)	(a)/(b)
Mahalanobis distance based method	34.2	26.9	1.27
our method without elimination	37.1	20.4	1.82
our method with elimination	28.3	8.0	3.53

References

1. O. Chomat, V.C. de Verdière, D. Hall, and J.L. Crowley, "Local scale selection for Gaussian based description techniques," *In ECCV, LNCS 1842*, vol. 1, pp. 117–133, 2000
2. Y. Dufournaud, C. Schmid, and R. Horaud, "Matching Images with Different Resolutions," *Proc. of the Conference on Computer Vision and Pattern Recognition, CVPR '00*, vol. 1, pp. 612–618, South Carolina, Jun.2000.
3. W.T. Freeman, E.H. Adelson, "The Design and Use of Steerable Filters," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 13, No. 9, pp. 891–906, Sep.1991
4. C. Harris and M. Stephens, "A Combined Corner and Edge Detector," *Alvey Vision Conf.*, pp. 147–151, 1988.
5. J.J. Koenderink and A.J. van Doorn, "Representation of Local Geometry in the Visual System," *Biol. Cybern.*, vol. 55, pp. 367–375, 1987.
6. T. Lindeberg, "Feature detection with automatic scale selection," *International Journal of Computer Vision*, vol. 30, No. 2, pp. 77–116, 1998
7. D.G. Lowe, "Object Recognition from Local Scale-Invariant Features," *Proc. of the International Conference on Computer Vision, ICCV '99*, vol. 2, pp. 1150–1157, Kerkyra, Corfu, Greece, Sep.1999.
8. D.G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, No. 2, pp. 91–110, Nov.2004
9. K. Mikolajczyk and C. Schmid, "Indexing based on scale invariant interest points," *Proc. 8th IEEE International Conference on Computer Vision, ICCV '01*, vol. 1, pp. 525–531, Vancouver, B.C., Canada, Jul.2001.
10. K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *Proc. of the Conference on Computer Vision and Pattern Recognition, CVPR '03*, vol. 2, pp. 257–563, Madison, Wisconsin, Jun.2003.
11. B.M. ter Haar Romeny, L.M.J. Florack, A.H. Salden, M.A. Viergever "Higher order differential structure of images," *Image and Vision Computing*, 12(6), pp. 317–325, 1994.

12. F. Schaffalitzky and A. Zisserman, "Multi-view matching for unordered image sets, or "How do I organize my holiday snaps?"," *In ECCV, LNCS 2350*, pp. 414–431, 2002.
13. C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, No. 5, pp. 530–534, May.1997
14. C. Schmid, R. Mohr, and C. Bauckhage, "Comparing and Evaluating Interest Points," *IEEE Proc. of the 6th International Conference on Computer Vision, ICCV '98*, pp. 230–235, Bombay, India, Jan.1998
15. N. Sebe, M.S. Lew, "Comparing salient point detectors," *IEEE International Conference on Multimedia and Expo*, Tokyo, Japan, Aug.2001.
16. http://www.inrialpes.fr/lear/people/Mikolajczyk/