Image retrieval based on RST-invariant features

Zhe-Ming Lu†,††, Dan-Ni Li††, and Hans Burkhardt†

†University of Freiburg, Freiburg, Germany
††Harbin Institute of Technology Shenzhen Graduate School, Shenzhen, China

Summary
In the application of content-based image retrieval, the ideal characteristics should be invariance to geometrical transformations. That is, once the image undergoes geometrical transformations, we expect the features extracted from the image are invariant. Thus, in this paper, rotation, scaling and translation (RST) invariant features for image retrieval are investigated, and a new method is proposed to extract these features. This method performs log-polar transformations on images in order to convert the scaling and rotation transformations to translation transformations, and then utilizes the translation and rotation invariance property of the Burkhardt's features to extract RST-invariant features. Moreover we take the structural information into account and combine it with the histogram descriptor. By combining these techniques ingeniously, we can retrieve both the RST transformed images and the similar images of the query image. The retrieval performance of the proposed method is illustrated in experiments and its advantages are shown by comparing with other methods.
Key words:
Image retrieval, invariant features, RST invariant features, log-polar transformations.

1. Introduction

In view of the overwhelming accumulation of the digital databases, the development of retrieval systems which allow efficient browsing, searching and retrieving of digital images is urgently needed. CBIR (contend-based image retrieval) has become a hotspot of digital image processing techniques since the early 1990's. Many research groups in leading universities and companies are actively working in this area and a fairly large number of prototypes and commercial products are already available [1]. Generally, the existing retrieval methods utilize the features of an image to describe and retrieve similar images. However, among the valid features, many are lack of invariance especially in the case of geometrical transformations applied on image. This will result in the mismatch of the retrieval process when the image's orientation, position or scale is altered. Thus it is necessary to find image features with invariance to geometrical transformations and the content-based image retrieval methods which can retrieve both specific and generic objects become more and more important. The difficulty of this goal is to determine the identity of an object under arbitrary viewing conditions in the presence of cluttered real-world scenes or occlusions [2]. The histogram descriptor is proved to be robust to the changes of object's orientation, scale, partial occlusion or changes in the viewing direction. However all structural information is lost in the histogram. To solve this problem, the combination of DWT (Discrete Wavelet Transform) or DFT (Discrete Fourier Transform) with the feature extraction method is proposed. But these methods need time-consuming computations, and can not meet specific application requirements.

In this paper we propose a new method that doesn't need to transform the image to the frequency domain but utilizes the property of the log-polar transformation and Burkhardt's invariant features [2] to extract RST-invariant features. The texture structural information is also taken into account, because it represents the structural information of the original images in some sense. This method is proved to have good performance in retrieving both RST transformed images and other similar ones of the query image in a database with more than 1000 images. The rest parts of this paper are organized as follows. In Section 2, the principle of RST-invariant features is introduced. And then the detailed feature extraction method is described in Section 3. In Section 4, the simulation results of our method are shown in P-R curves. Section 5 concludes the whole paper.

2. The RST-Invariant Features

2.1 Translation and Rotation Invariant Features

The method mentioned in this paper draws inspiration from the invariant features proposed by Burkhardt [2]. The main idea is to utilize the property of the invariant features remaining unchanged when the data is transformed with a group of geometrical operations. In [2], a kind of global feature invariant to rotations and translations are presented. Given a gray-scale image

$$\mathbf{M} = \{\mathbf{M}(\mathbf{x}), \mathbf{x} = (x_0, x_1), 0 \le x_0 < N_0, 0 \le x_1 < N_1\} \qquad (1)$$

Manuscript revised January 2006.

and an element $g \in G$ of the group of image translations and rotations, the transformation can be expressed as:

$$(g\mathbf{M})(\mathbf{x}) = \mathbf{M}(\mathbf{x}') \text{, with}$$

$$\mathbf{x}' = \left\{ \begin{pmatrix} \cos\varphi & \sin\varphi \\ -\sin\varphi & \cos\varphi \end{pmatrix} \mathbf{x} + \begin{pmatrix} t_0 \\ t_1 \end{pmatrix} \right\} \bmod \begin{pmatrix} N_0 \\ N_1 \end{pmatrix} \tag{2}$$

Based on above definition, an invariant transformation T must satisfy

$$T(g\mathbf{M}) = T(\mathbf{M}), \quad \forall g \in G \tag{3}$$

For a given gray-scale image $\mathbf{M}$ and an arbitrary complex-valued function $f(\mathbf{M})$, it is possible to construct such an invariant transformation T by the following Haar integral:

$$\begin{aligned} T[f](\mathbf{M}) &:= \frac{1}{|G|} \int_G f(g\mathbf{M}) dg \\ &= \frac{1}{2\pi N_0 N_1} \int_{t_0=0}^{N_0} \int_{t_1=0}^{N_1} \int_{\varphi=0}^{2\pi} f(g(t_0, t_1, \varphi)\mathbf{M}) d\varphi dt_1 dt_0 \end{aligned} \tag{4}$$

The calculation strategy can be illustrated by Fig. 1. If we select a simple function

$$f(\mathbf{M}) = \sqrt{\mathbf{M}(0,1) \cdot \mathbf{M}(2,0)} \tag{5}$$

then

$$\begin{aligned} T[f](\mathbf{M}) &= \frac{1}{2\pi N_0 N_1} \int_{t_0=0}^{N_0} \int_{t_1=0}^{N_1} \int_{\varphi=0}^{2\pi} \sqrt{\mathbf{M}(\sin\varphi + t_0, \cos\varphi + t_1)} \\ &\cdot \sqrt{\mathbf{M}(2\sin(\varphi + \pi/2) + t_0, 2\cos(\varphi + \pi/2) + t_1)} d\varphi dt_1 dt_0 \end{aligned} \tag{6}$$

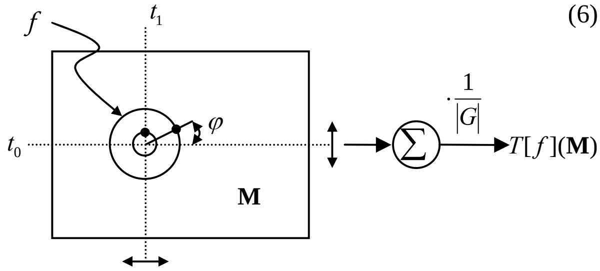

Fig. 1 Calculation strategy for invariant integration in the case of Euclidean motion.

For discrete images [2,3], because we choose integers for (t_0, t_1) and we use K steps for φ, we can obtain the following formula:

$$T[f](\mathbf{M}) \approx \frac{1}{KN_0 N_1} \sum_{t_0=0}^{N_0-1} \sum_{t_1=0}^{N_1-1} \sum_{k=0}^{K-1} f(g(t_0, t_1, \varphi = \frac{2\pi k}{K})\mathbf{M}) \tag{7}$$

However, because the above equation is of linear complexity in the number of pixels covered by the kernel function of finite support, we need a reduction in the calculation complexity. On the other hand, due to the above global averaging, non-local errors like greater occlusion or changing background can cause problems, thus we need another description method to preserve the local information. Considering the above two facts, we can apply the Monte-Carlo method together with the histogram description method as follows:

Step 1: Generate a set of n random 3-D vectors $\mathbf{P} = \{\boldsymbol{p}_0, \boldsymbol{p}_1, ..., \boldsymbol{p}_{n-1}\}$ that are equally distributed in $\{(t_0, t_1, \varphi) \mid 0 \le t_0 < A_0, 0 \le t_1 < A_1, 0 \le \varphi < 2\pi\}$.

Step 2: Compute

$$\mathbf{S} = \{f(g(t_0, t_1, \varphi)\mathbf{M}) \mid (t_0, t_1, \varphi) \in \mathbf{P}\} = \{s_0, s_1, ..., s_{n-1}\} \tag{8}$$

where $s_i, i = 0,1,..., n-1$ are scalars. Here inter-grid positions are handled applying bilinear interpolation.

Step 3: Compute the histogram for the array $\mathbf{S}$.

2.2 Scaling and Rotation Invariant Features

The method that combines the invariant features given above with the histogram descriptor has been successfully applied to the task of texture classification and texture defect detection [4]. It also has fast query performance in real image retrieval system after the reduction of the calculation complexity by using estimation of the features instead of deterministic calculation [5]. This method is proved to be good at constructing rotation and translation (RT) invariant features. However, it is sensitive to image scaling transformations [6]. In order to solve this problem, a method is proposed in this paper to improve Burkhardt's RT-invariant features. This method can construct rotation, translation and scaling (RST) invariant features of the image.

In this paper, we utilize the log-polar transformation to convert the scaling and rotation transformations to translation transformations. We can rewrite the rectangle coordinate (x_0, x_1) in the polar form as $x_0 = \rho\cos\theta,\ x_1 = \rho\sin\theta$. Then the image $\mathbf{M}(x_0, x_1)$ can be described as $\mathbf{M}(\rho,\theta)$. If an image is rotated by ϕ degrees, then it can be described by $\mathbf{M}(\rho,\theta+\phi)$, that is, the rotation transformation is converted into a translation transformation. On the other hand, assume that $\lambda = \ln\rho$, $\mathbf{M}(\rho,\theta)$ can then be written as $\mathbf{M}(e^{\lambda},\theta)$, that is $\mathbf{M}(e^{\lambda},\theta) = \mathbf{M}_1(\lambda,\theta)$. Thus the image is described as $\mathbf{M}_1(\lambda,\theta)$ in the log-polar coordinate system. If the scale of an image is changed, e.g., from ρ to $a\rho$ $(a > 0)$, the transformed image can then be described by $\mathbf{M}_1(\lambda + \ln a,\theta)$, thus the scaling transformation can be also converted into a translation transformation. After the log-polar transformation (LPM), we can then apply Equation (8) and the histogram descriptor to the log-polar transformed images to extract invariant features named RST-Invariant features in this paper. The two examples of log-polar transformation results are shown in Fig.2 and Fig.3.

Fig. 2 Log-polar transformation converts the rotation transformation of the original image (from up-left to bottom-left) into a translation transformation (from up-right to bottom-right).

3. RST-Invariant Feature Extraction

The main idea of our method used in the feature extraction process is to perform the log-polar transformation on the image first, and then use Equation (8) and the histogram descriptor to calculate the features. However, unlike the original Burkhardt's calculation method, we use a totally different method to calculate RST-invariant features. Here we only choose the H component based on the HSV color space to construct features to reduce the computation complexity. In fact, we can consider more components to improve the retrieval results if necessary. The entire process is showed in Fig. 4, which can be divided into four steps:

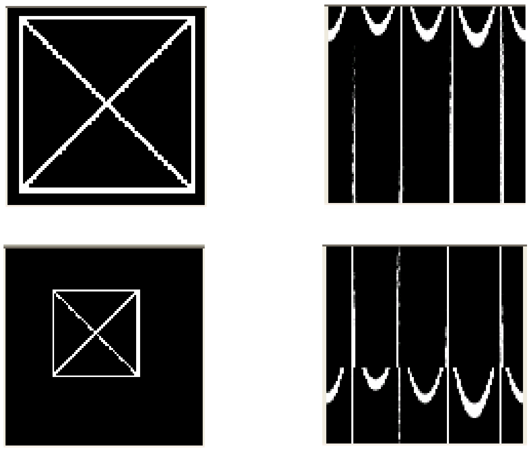

Fig. 3 Log-polar transformation converts the scaling transformation of the original images (from *up-left to bottom-left*) into a translation transformation (*from up-right to bottom-right*).

Step1: Select a part of the image pixels to calculate local translation invariant features, because it will be time-consuming if all the image pixels are taken into account to construct invariant features. Thus we only select K_1 $(0 < K_1 < N_0 \times N_1)$ pixels randomly from the image without losing the globality. For each selected point we use the following equation to calculate each feature value.

$$S_1(x,y) = \sum_{i=0}^{L-1} \mathbf{M}(x,y) \times g(x,y,r,\psi = 2\pi\frac{i}{L}) \quad (9)$$

where $S_1(x,y)$ is the feature value for the selected point (*x,y*), *r* is the radius of the kernel function that is confined by $r < x < N_0 - r$ and $r < y < N_1 - r$, and $g(x,y,r,\psi = 2\pi\frac{i}{L})$ is the pixel value of the point that is on the circle with radius r and angle $2\pi\frac{i}{L}$ around the coordinate (*x,y*). Here, we use the bilinear interpolation to get the pixel value of the point whose coordinates are not integer. Here all the pixel values are transformed to be in the interval [0,1] before feature calculation.

Step2: Perform the log-polar transformation on the original image. Change the rectangular coordinate (x, y) into the polar coordinate (ρ,θ), where $\rho \in (0, R)$, $R = \min(N_0/2, N_1/2)$, and $\theta \in (0,2\pi)$. Then we transform a part of Image $\mathbf{M}(\rho,\theta)$ into Image $\mathbf{M}(m,n)$ in the log-polar coordinates with bilinear interpolation. Set $\rho_{\min} = R/5, \rho_{\max} = R/2$, $\Delta\rho = (\ln\rho_{\max} - \ln\rho_{\min})/256$ and n=1, 2, ⋯, 256. Set $\Delta\theta = 2\times\pi/256$, m=1, 2, ⋯, 256, and $x_0 = N_0/2, y_0 = N_1/2$ [8]. The transformation can then be shown as follow:

$$\begin{cases} x = \rho\cos(\theta) = e^{n\times\Delta\rho+\rho_{\min}}\cos(m\times\Delta\theta) + x_0 \\ y = \rho\sin(\theta) = e^{n\times\Delta\rho+\rho_{\min}}\sin(m\times\Delta\theta) + y_0 \end{cases} \tag{10}$$

Here, the size of the transformed image $\mathbf{M}(m,n)$ is 256×256. So this process also normalizes different image sizes into the same size 256×256.

Step3: Select K_2 pixels randomly from the transformed image $\mathbf{M}(m,n)$ to calculate local rotation and scaling invariant features $S_2(x,y)$ for the original image. The specific process is the same as Step1.

Step4: Calculate the accumulation histogram for the extracted RST-invariant features [7]. Here we combine the features calculated in Step1 and Step3 with the histogram descriptor. Here, we calculate a $2L$-bin histogram where the first L bins are for the K_1 translation-invariant feature points and the remained L bins are for the K_2 scaling and rotation-invariant features, where L is the number of angle steps we set to calculate the local features in Step1 and Step3. Thus, we can obtain the following combined features.

$$F(J) = \begin{cases} hist(S_1(x,y)) & (0 \le J \le L-1) \\ hist(S_2(x,y)) & (L \le J \le 2L-1) \end{cases} \tag{11}$$

In this paper, we use the following equation to calculate the distance between two images p and q during the retrieving process.

$$D_j(p,q) = \sqrt{\sum_{i=0}^{2L-1}[F_p(i) - F_q(i)]^2} \tag{12}$$

4. Experimental Results

The proposed method has been implemented on Visual C++ 6.0 platform. Since most common evaluation measures used in Image Retrieval (IR) are precision and recall as shown in Eqs. (13) and (14) [9], which are usually presented as a precision vs. recall graph (PR-graph). Thus, to give a just and impersonal evaluation of the retrieval efficiency of our method [10], the PR graph is used in this paper. The method is tested on a certain database with 1000 images [11]. It consists of ten categories including people, beach, building, bus, dinosaur, elephant, flower, horse, mountain and food, each containing 100 images. The images are JPEG formatted color images of different sizes. From each category, five images are selected randomly and twelve geometry transformations are applied on each selected image, e.g. changing the scale of the image and then putting them back into the database. The purpose is to test the retrieval performance for transformed images. For the 50 images which are selected to be the query images we take down their precision and recall values when the user retrieves 1, 2, ⋯, 1600 images. For each number of retrieved images, we calculate the average precision and recall value over all test query images. And thus we can get a PR graph with 1600 points.

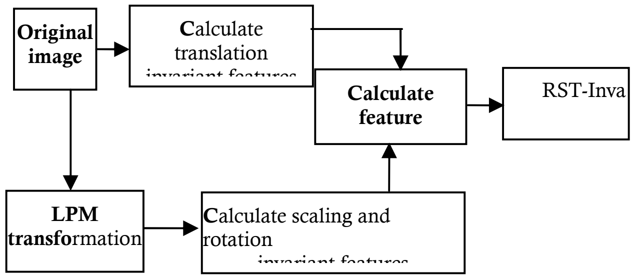

Fig. 4 RST Invariant feature extraction process.

In the simulation, we use L=20, K_1=K_2=3000 and test two cases to compare the retrieval performances among three kinds of features, i.e., the color histogram features, Burkhardt's RT-Invariant features and our RST-Invariant features. The first test is to show the performance of retrieving the RST-transformed images from the same image, and the comparison results are shown in Fig. 5. The second test is to show the performance of retrieving images similar to the query image, and the results are shown in Fig. 6.

5. Conclusions

This paper presents a method to extract RST-invariant features by performing LPM transformation on the image. Experimental results indicate that the retrieval based on our feature extraction method performs well on a database with 1600 JPEG color images of different sizes. Compared with the other two methods, it outperforms in retrieving both the RST transformed images and the similar images of the query image. Moreover, its computational complexity is much less than the methods that use the invariant property of DFT or DWT descriptor. Our future work will focus on extracting invariant features which are robust to illumination changes and other geometric operations.

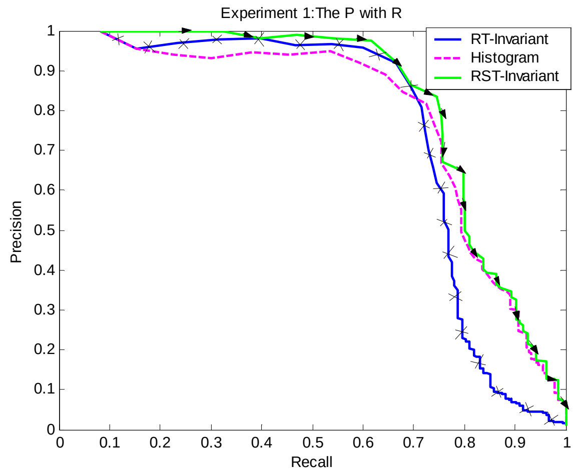

Fig. 5 Performance Comparisons of retrieving transformed images from the same image.

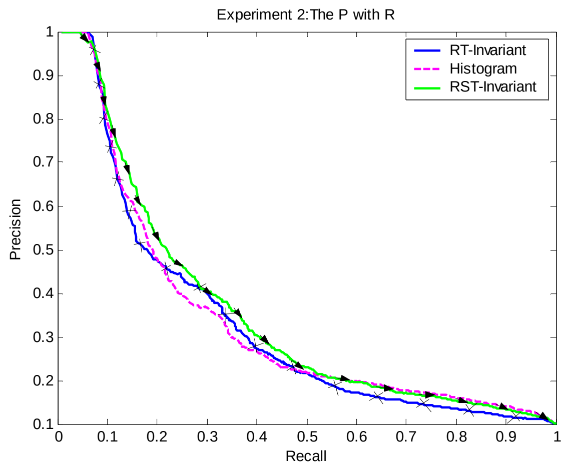

Fig. 6 Performance comparisons of retrieving images similar to the query image.

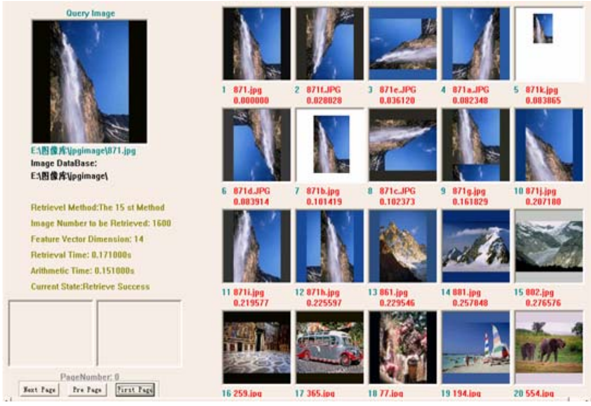

Fig. 7 Retrieval results from a query example.

Acknowledgments

This work was supported by Alexander von Humboldt Foundation Fellowship (Germany), ID: CHN 1115969 STP.

References

[1] Arnold W.M. Smeulders, Marcel Worring, Amarnath Gupta and Ramesh Jain:Content- Based Image Retrieval at the End of the Early Years, IEEE Transactions on pattern analysis and machine intelligence (2000) 1349-1380
[2] Sven Siggelkow and Hans Burkhardt: Fast invariant feature extraction for image retrieval, State-of-the-Art in Content-Based Image and Video Retrieval (2001) 43-68
[3] Hanns Schulz-Mirbach: Invariant features for gray scale images. In G.Sagerrer, S.Posch, and F.Kummert, editors, Mustererkennung, DAGM 1995 (1995) 1-14
[4] Marc Schael: Invariant texture classification using group averaging with relational kernel functions, Submitted to CVPR (2001) 11-13
[5] Hanns Schulz-Mirbach: Invariant features for gray scale images. In G.Sagerrer, S.Posch, and F.Kummert, editors, Mustererkennung, DAGM 1995 (1995)1-14
[6] S. Siggelkow: Improvement of histogram-based image retrieval and classification, IAPR International Conference on Pattern Recognition (ICPR), Quebec City, Canada (2002) 367-370
[7] H.Schulz-Mirbach, H.Burkhardt and S.Siggelkow: Using invariant features for content based data retrieval, Workshop on Nonlinear Methods in Model-Based Image Interpretation, Lausanne, Switzerland (1996) 1-5
[8] Jurgen Wolf, Wolfram Burgard and Hans Burkhardt: Robust vision-based localization for mobile robots using an image retrieval system based on invariant features, Proceedings of the 2002 IEEE international conference on robotics & automation (2002) 359-365
[9] Li Li, Ming-min Zhang, PAN Zhi-geng: Image watermarking algorithm resilient to geometrical transform. Journal of Zhejiang University (Engineering science) (2004)141-144

[10] Su-Zhi Li, Zhe-Ming Lu and Hai-jun Jin: Image retrieval method based on fast equal-average K nearest neighbor search and query-feature-vector-recomposition feedback, ISTM (2005) 6370-6373
[11] LI, J: Photography image database. http://www.stat.psu.edu/~jiali/index.download. html
[12] Yu-Jin Zhang: Content-based Visual Information Retrieval, Science Press. (2003) 47-50

Zhe-Ming Lu received the B.S., M.S. and Ph. D. degrees in Electrical Engineering from Harbin Institute of Technology in 1995, 1997 and 2001, respectively. He was the Alexander von Humboldt Research Fellow in University of Freiburg in Germany, from Oct., 2004 to Jan. 2006. He has published more than 110 papers and four books. He has been program committee members in several international conferences. He is now the Professor and Director of the Visual Information Analysis and Processing Research Center, Harbin Institute of Technology Shenzhen Graduate School. His research interests are image processing, pattern recognition, information hiding and visual information retrieval.

Dan-Ni Li received the B.S. degree in Automatic Test and Control from Harbin Institute of Technology in 1997. She is now the graduate student in the Visual Information Analysis and Processing Research Center, Harbin Institute of Technology Shenzhen Graduate School. Her research interests are image processing and visual information retrieval.

Hans Burkhardt obtained his Dipl.-Ing. Degree in electrical engineering in 1969, Dr.-Ing. Degree in 1974, and the Venia Legendi in 1979 from the University of Karlsruhe, Germany. From 1969 he was Research Assistant and in 1975 he became Lecturer at the University of Karlsruhe. In 1981 he became Professor for Control and Signal Theory at the University of Karlsruhe. During 1985-1996 he was full Professor at the Technical University of Hamburg. Since 1997 he is full Professor at the Computer Science Department of the University of Freiburg; director of an Institute for Pattern Recognition and Image Processing and currently Deputy Dean of the Faculty for Applied Sciences. He has published over 150 papers and given more than 200 lectures. He is a consultant for several national and international institutions e.g. the German Science Foundation (DFG), the European Commission and different international organizations and journals. In 1998 he was chair of the European Conference on Computer Vision (ECCV).