A Rotation Invariant Shape Representation based on Wavelet Transform

Seiichi KON'YA and Kazuhiko KUSHIMA

NTT Information and Communication Systems Laboratories Yokosuka, Kanagawa, Japan

Abstract

This paper introduces an object shape representation which provides similarity search for databases with a large number of images. The proposed algorithm uses the following techniques: 1) the complex-valued wavelet transform to utilize the phase information, 2) the translation-invariant wavelet representation to normalize the shape orientation. Experimental results show that the proposed algorithm be able to normalize shape orientation more accurately than the normalized Fourier descriptors[1] and has better performance in similarity search than the planar curve descriptors[2].

1 Introduction

Image database systems are required to provide the similarity search mechanism based on multiple attributes such as color, texture, shape, etc., in order to meet various content-based retrieval requirements. To retrieve the object which is similar in shape from the database, objects' shapes are needed to be characterized by the orientation invariant features. Moreover, when the number of objects in database is very large (e.g., more than 100,000 objects), brute force search or template matching are very time consuming and not appropriate for similarity search, even if their time complexity are O(N). If the objects' features are represented by the vector space model, similarity search can be done efficiently using the multidimensional index such as R-tree or kd-tree.

We are developing ExSight, a prototype multimedia information retrieval system, which provides object-based image retrieval[3]. The system extracts objects from an input image automatically and characterizes them using features, such as color, shape, size, position, etc. The similarity search can be done by comparing between the stored objects and the reference object specified by an user. More specifically, it is performed by looking for the k-nearest neighbors of the reference, based on the similarity measure defined in the multidimensional feature vector space. Euclidean distance or Manhattan distance is used for the similarity measure. The system also has high performance data access methods based on multidimensional index, in order to fulfill the similarity search efficiently. This paper introduces the vector space representation of the object shape which is rotation invariant and less sensitive to deformation.

2 A Rotation Invariant Shape Representation

A rotation invariant shape representation is calculated as follows:

- 1. parameterize the contour to the discrete periodic signal,
- 2. apply the complex-valued wavelet transform to utilize the phase information of the contour,
- 3. apply shift normalization to the wavelet coefficients,
- 4. apply rotation normalization,
- 5. apply size normalization.

2.1 The Arc Length

If the object image is given, the object's contour can be detected using the contour tracking algorithm[7]. To apply dyadic wavelet transformation to the contour, it must be sampled to have 2^n points. We use the arc length in order to parameterize and sample the contour. We trace the contour counterclockwise and denote its coordinates (x'_i, y'_i) , $i = 1, \dots, M$. The parameterized contour is denoted by

$$x(t) = x'_k, t = \frac{l_k}{L}$$
$$y(t) = y'_k$$

where t is normalized arc length, l_k is arc length along the contour from (x'_1, y'_1) to (x'_k, y'_k) , and L is total arc length $(L = l_M)$. Then, we can sample 2^n points from the contour by $(x(\frac{j}{2^n}), y(\frac{j}{2^n})), j = 1, \dots, 2^n$. Figure 1 illustrates the arc length parameterization. The contour (center) is detected from the given image (left). The contour is sampled so that the length between points along the contour is constant.

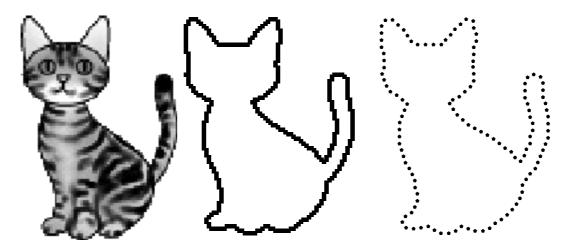


Figure 1: Image, contour and its parameterization.

2.2 The Wavelet Transform

The wavelet transform is the technique that decompose a given signal into the components in both time and frequency localization by the multiresolution analysis. The scaling function $\varphi(x)$ and the wavelet function $\psi(x)$ which constructs the multiresolution analysis must satisfy the two-scale relation:

$$\varphi(t) = 2\sum_{k=0}^{k} a_k \varphi(2t - k),$$

$$\psi(t) = 2\sum_{k=0}^{k} b_k \psi(2t - k).$$

If we have a representation of a signal at resolution level j + 1:

$$f(t) = \sum_{k=1}^{k} \lambda_{j+1,k} \varphi_{j+1,k}(t),$$

then we can also decompose it into its low pass and high pass components:

$$f(t) = \sum_{l}^{l} \lambda_{j,l} \varphi_{j,l}(t) + \sum_{l}^{l} \gamma_{j,l} \psi_{j,l}(t),$$

where

$$\varphi_{j,k}(t) \equiv 2^{\frac{j}{2}} \varphi(2^j t - k),$$

$$\psi_{i,k}(t) \equiv 2^{\frac{j}{2}} \psi(2^{j}t - k),$$

with $j,k\in\mathbf{Z}$, i.e., $\varphi_{j,k}(t)$ is a translated and dilated version of $\varphi(t)$. If the scaling function $\varphi(x)$ is orthonormal, the coefficients $\lambda_{j,l}$ and $\gamma_{j,l}$ can be calculated by the Fast Wavelet Transform:

$$\lambda_{j,l} = \sqrt{2} \sum_{k=0}^{k} a_{k-2l}^* \lambda_{j+1,k} ,$$

$$\gamma_{j,l} = \sqrt{2} \sum_{k=0}^{k} b_{k-2l}^* \gamma_{j+1,k} ,$$

where * denotes a complex conjugate.

The wavelets are qualified by the following properties: (i) compactness of the support, (ii) symmetry, (iii) accuracy of the approximation, (iv) orthonormality. It is well-known that only the complex-valued wavelets can achieve the above all properties[4]. By analogy to the Fourier descriptors, we used the complex Daubechies wavelets[5] to characterize the contour in the complex plane.

2.3 The Shift Normalization

The wavelet coefficients of two signals may be quite different, even if the two signals just differ by a time shift. Figure 2 shows the contour of cat's image (left) and its wavelet coefficients (right). The black and the grey polylines correspond to the wavelet coefficients whose contours start at the black and the grey points, respectively. It can be shown that shift of the starting point makes the coefficients quite different from the original.

The translation-invariant representation is achieved by computing the wavelet transform for all the circular time shifts of a signal and selecting the wavelet coefficients of the time shift which minimizes the cost function[6]. In this paper, we adopts the vector entropy, which is a measure of the concentration of the energy of the coefficients, as the cost function. The vector entropy μ_{ε} of the wavelet coefficients $c = \{c_k\}$ is given by

$$\mu_e(c) = -\sum_k |rac{c_k}{\|c\|}|^2 \log_2 |rac{c_k}{\|c\|}|^2,$$

where

$$c_k = u_k + iv_k, \ |c_k| = \sqrt{u_k^2 + v_k^2} \ ext{and} \ ||c|| = \sqrt{\sum_{k=1}^k (u_k^2 + v_k^2)}.$$

If the shift which minimizes the cost function is not unique, we choose the one that maximizes S, the area of triangle made by c_1, c_2, c_3 :

$$S = Im(c_1^*c_2 + c_2^*c_3 + c_3^*c_1),$$

where $Im(\cdot)$ denotes imaginary part.

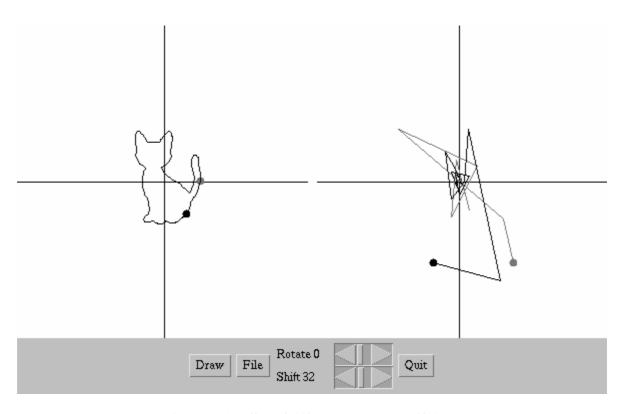


Figure 2: The effect of shift on the wavelet coefficients.

2.4 The Rotation Normalization

By the linearity of the wavelet transform, the rotation normalization of the contour can be performed using wavelet coefficients. Figure 3 shows the rotated contour (left) and its wavelet coefficients (right). The original contour and coefficients are drawn with grey color. If the contour is rotated 90 degrees (the black contour), the corresponding coefficients is also rotated 90 degrees (the black polyline).

We chose the rotation normalization method achieved by maximizing the summation of square of the coefficients' projection to the real line, i.e.,

$$P(\theta) = \sum_{j=1}^{n} \{ Re(c_j \cdot e^{i\theta}) \}^2,$$

where $Re(\cdot)$ denotes real part and $e^{\pmb{i}\theta}$ means rotation of θ radian. The θ which maximizes $P(\theta)$ holds the equation, $\frac{\partial P}{\partial \theta} = 0$. Then

$$\frac{\partial P}{\partial \theta} = -2(\cos^2 \theta - \sin^2 \theta) \sum u_j v_j - 2\sin\theta \cos\theta \sum (u_j^2 - v_j^2) = 0.$$

We obtain

$$\frac{\partial P}{\partial \theta} = 0 \implies \frac{\sum (u_j^2 - v_j^2)}{\sum u_j v_j} = \frac{\sin^2 \theta - \cos^2 \theta}{\sin \theta \cos \theta} = \frac{-2\cos 2\theta}{\sin 2\theta}$$

Hence.

$$\theta = \frac{1}{2} tan^{-1} \frac{-2\sum u_j v_j}{\sum (u_j^2 - v_j^2)}$$

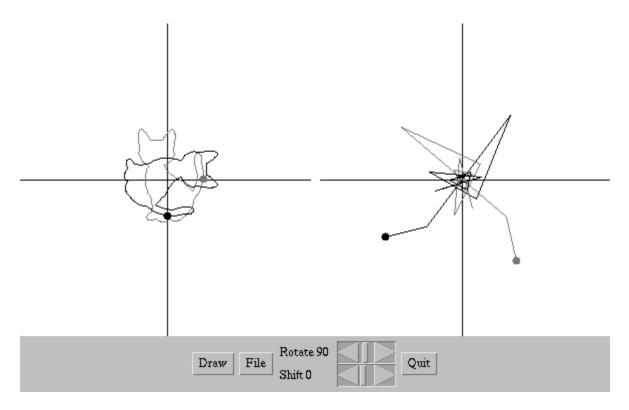


Figure 3: The effect of rotation on the wavelet coefficients.

The period of $tan\theta$ is π so that $P(\theta)$ has 4 extrema — 2 maxima and 2 minima. We select the θ that $Re(c_1 \cdot e^{i\theta}) > 0$.

2.5 The Size Normalization

The size normalization can be done simply by the wavelet coefficients are scaled to an unit vector:

$$c_k = \frac{1}{||c||} c_k.$$

3 Experiments

This section reports the numerical experiments of the proposed algorithm. The proposed method was tested on digital clip arts 1 . By applying the contour tracking algorithm[7] to the image, the contour of the objects in the image was extracted from the background of an uniform color. Excluding the same shape with different colors, 699 shapes were selected. The object contours were sampled to have 256 data points with the method in §2.1. And the shape features were approximated by 30 coefficients. (we chose 16 complex coefficients of lower resolution which carries the global approximation and excluded the lowest one which contains the position of shape. The complex coefficient has a real and a imaginary part. $(16-1) \times 2 = 30$.) Thus all objects were represented by the vector in 30 dimension space. The number of the coefficients affect on the discernibility of shapes and the space (memory) requirement and the search cost of the multidimensional index. If the number of coefficients decrease, the accuracy of similarity search degrades because the differences between shapes become small. If the number of coefficients increase, the performance of the multidimensional index degrades. We chose 30 from the experience. The similarity of two objects were measured by the Euclidean distance.

 $^{^1} http://home.impress.co.jp/books/dml/ips/book/4779/4779.htm \ (impress, 1997, ISBN 4-8443-4779-9)$

3.1 Accuracy of the Shift Normalization

To evaluate the shape orientation normalization, we applied the proposed method to the contours of 15, 30, 45, 60 and 75 degrees' rotated images and compared the starting points with the original contour. (90 degrees' rotation only switches x and y coordinates.) The contour shape of the rotated image is a little different from the original due to the quantization distortion. We also evaluated the effect of simple deformation on the proposed method using the images changed in aspect ratio, i.e., enlargement on horizontal or vertical. We compared the proposed algorithm with the Normalized Fourier Descriptors[1]. We measured the shift of the starting point from the original and classified it as an error if its shift is greater than 16 points. The experimental results are summarized in Table 1. These results show that the proposed method is more accurate than the NFD.

Table 1: The error rate of the starting point normalization.

change in aspect ratio	none	10 %	20 %
the NFD method[1]	17 %	25 %	31 %
the proposed method	2.6 %	9.1 %	17 %

3.2 Accuracy of Similarity Search

Next, we evaluated the accuracy of similarity search of the proposed method. The conventional measures, recall and precision, were used for our evaluation.

```
\operatorname{recall}(x) = \frac{\operatorname{number of objects found and relevant to } x}{\operatorname{the total number of objects relevant to } x} \operatorname{precision}(x) = \frac{\operatorname{number of objects found and relevant to } x}{\operatorname{the total number of objects found}}
```

To illustrate the effectiveness of the proposed method, we compared ours with the planar curve descriptors[2], which represents x and y coordinate components separately using the real-valued wavelets. In this experiments, the spline function was used as its wavelet base. Since the planar curve descriptors cannot normalize the starting points of contours, we added the shift normalization step described in $\S 2.3$ and modified the planar curve descriptors.

Figure 4,5 and 6 show the average performance of similarity search. Figure 4 is the case where the database has 699×6 (0, 15, 30, 45, 60, 75 degrees' rotated shapes) objects (case 1). In this case, we measured the precisions and recalls for retrieving 6 relevant shapes from 4194 images and averaged the performance over 699 kind of images. The proposed method has better performance than the planar curve descriptors. The complex-valued wavelet transform is seemed to be superior to the real-valued wavelets for the contour representation.

Figure 5 is the case where the database has $699 \times 6 \times 3$ (original size, 10% enlarge on horizontal, and 10% enlarge on vertical) objects (case 2). In this case, 18 relevant shapes were retrieved from 12582 images. And, figure 6 is the case where the database has $699 \times 6 \times 5$ (original size, 10% and 20% enlarge on horizontal, and 10% and 20% enlarge on vertical) objects (case 3). The performance becomes worse as the change in aspect ratio increases. However, the proposed method always performs better than the planar curve descriptors because of the complex-valued wavelets.

4 Conclusion

In this paper, we proposed the object shape representation which provides similarity search for databases with a large number of images. This method used the following techniques: 1) the complex-valued wavelet transform to utilize the phase information, 2) the translation-invariant wavelet representation to normalize the shape orientation.

Experimental results showed the proposed method can normalize the shape orientation even if the aspect ratio changes and is more accurate than the normalized Fourier descriptor[1]. And the performance measurements on similarity search showed the proposed method is superior to the conventional algorithm[2].

The proposed method is very suitable for the multidimensional indexing technique which provides the efficient similarity search, i.e., top k ranking of objects similar to the reference without a brute force computation.

A Rotation Invariant Shape Representation based on Wavelet Transform

References

- [1] Wallace T. P, Wintz P. A.: An Efficient Three-Dimensional Aircraft Recognition Algorithm using Normalized Fourier Descriptors. Comput. Graphics Image Processing 1980; vol. 13: 90-126
- [2] Chuang G. C.-H, Kuo C.-C. J.: Wavelet Descriptor of Planar Curves: Theory and Applications. IEEE Trans. Image Processing 1996; vol. 5, No. 1: 56-70
- [3] Yamamuro M, Kushima K, Kimoto H, et al: ExSight Multimedia Information Retrieval System. Proc. of 20th Annual Pacific Telecommunications Conference PTC'98 1998: 734-739
- [4] Daubechies I. Ten Lectures on Wavelets. SIAM, CBMS Series, 1992
- [5] Lina J.-M.: Image Processing with Complex Daubechies Wavelets. Journal of Math. Imaging and Vision 1997; vol. 7: 211-223
- [6] Liang J, Parks T. W.: A translation-Invariant Wavelet Representation Algorithm with Applications. IEEE Trans. Signal Processing 1996;vol. 44, No. 2: 225-232
- [7] Klette R, Zamperoni P. Handbook of Image Processing Operators. John Wiley & Sons Ltd., 1996

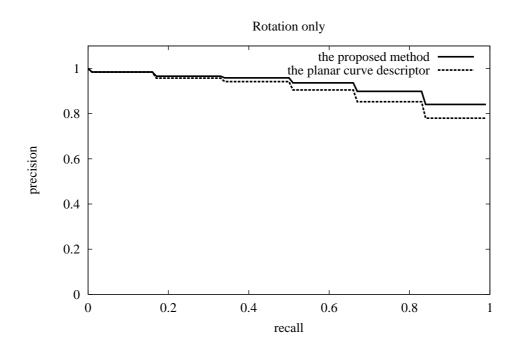


Figure 4: The recall precision curves of similarity search (case 1).

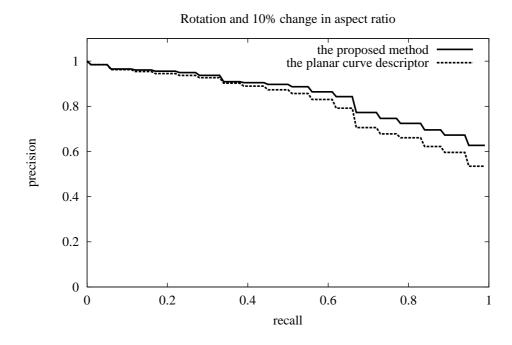


Figure 5: The recall precision curves of similarity search (case 2).

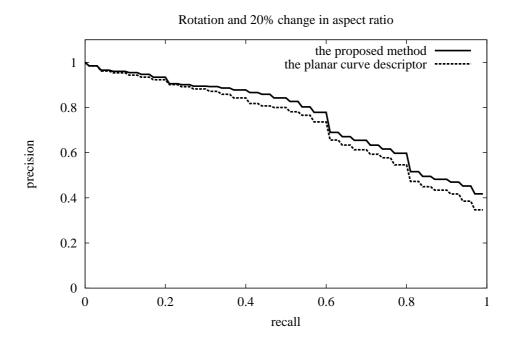


Figure 6: The recall precision curves of similarity search (case 3).