Research on Fourier Descriptor Image Retrieval Technology Based on Minimum Inertia Axis

Zhijie Zhao[1,2], Ze Gao[1,2](✉), Huadong Sun[1,2], and Xuesong Jin[1,2]

[1] School of Computer and Information Engineering, Harbin University of Commerce, Harbin 150028, China
1041451358@qq.com
[2] Heilongjiang Provincial Key Laboratory of Electronic Commerce and Information Processing, Harbin University of Commerce, Harbin 150028, China

Abstract. In image retrieval, the shape feature is one of the key features of image content description. At present, most of the widely used Fourier descriptors in shape descriptors are invariant in translation, rotation and scale expansion. But Fourier descriptors are susceptible to the location of the starting point. In this paper, an improved image retrieval method based on Fourier descriptors is proposed. First, the image is preprocessed and the edge of the image is extracted. Secondly, the starting point of the contour is determined by the minimum inertia axis. Then Fourier transform is used to get eigenvectors. Finally, the correlation coefficient is used to calculate the similarity. Experiments show that the Fourier Descriptor Image Retrieval Method based on the minimum inertia axis is more efficient than other methods in Swedish Leaf database.

Keywords: Image retrieval · Shape feature · Fourier descriptor · Minimum inertia axis

1 Introduction

Now, with the growing Internet technology, tens of thousands of product images are pouring into people's sight every day. Along with the influence of e-commerce on people's life and shopping habits, simple text information retrieval can no longer meet people's requirements for product retrieval. Content Based Image Retrieval [1, 2] mainly uses visual features such as color, shape, texture, spatial relationship, etc., and relies on image processing technology, recognition technology, artificial intelligence and computer vision technology to realize image retrieval [3]. In the image of the product, the shape is often associated with the target, which is in line with the visual habits of people to recognize the object, and has certain semantics. Therefore, the shape is widely used as a discriminating element in the field of content-based image retrieval. In many applications, shape captures most of the perceptual information of an object on an image, while colors and textures can often be omitted without affecting retrieval performance. But the shape may be affected by factors such as deformation, scaling,

M. Jia et al. (Eds.): WiSATS 2019, LNICST 281, pp. 241–249, 2019.
https://doi.org/10.1007/978-3-030-19156-6_23

orientation changes, noise and partial hiding. Therefore, accurate description of the shape is still a challenging technical issue.

For image retrieval systems based on shape features, the description of the shape of the image is very important. Shape features can be divided into two main categories: contour-based features and region-based features. The contour-based description method only uses the boundary information and loses the internal content of the shape, so the versatility is not high. The region-based description method utilizes the internal pixel information of the target shape, which can be applied to general occasions. However, all the region description methods currently extract the spatial features of the shape, so it is sensitive to subtle changes in noise and shape, and is resistant to interference. The ability is relatively poor.

In many existing shape feature descriptors, the Fourier descriptor has several desirable features, such as low computational complexity, sharpness and coarse to fine description, which makes it a popular descriptor. A new method for extracting Fourier descriptors that preserve the phase of Fourier coefficients is proposed by Sokic [4]. Introduce specific points, called pseudo-mirror points, and use them as shape orientation references. It helps extract the phase-preserving Fourier descriptor and is constant under translation, scaling, rotation and starting point changes. Performance and computational complexity measurements indicate that the proposed method is superior to other phase-based Fourier descriptors. El-ghazal et al. [5] proposed a new curvature-based Fourier descriptor (CBFD) shape retrieval. The proposed descriptor has an unconventional view of the curvature scale spatial representation of the shape profile because it is considered a two-dimensional binary image (hence the curvature scale image or CSI). The invariant descriptor is derived from the two-dimensional Fourier transform of the curvature scale image. This approach allows the descriptor to capture the detailed dynamics of the shape curvature and improve the efficiency of the shape matching process.

2 Image Shape Feature Extraction Based on Minimum Inertia Axis

2.1 Fourier Descriptor

First, the product image is binarized, and the binarized product image $p(x, y)$ is characterized by only two pixels of black and white. Black is the target pixel, denoted as $p(x, y) = 1$, white is the background pixel, denoted as $p(x, y) = 0$. Second extract the coordinates of the shape boundary from the image. In order to apply the shape description method, the equal length sampling is performed by the difference method, and the contour of the shape is resampled with a fixed number of points.

For subsequent analysis, it will be assumed that the shape profile is given by N boundary points $A(k) = (x_k, y_k),\ k = 1, 2, \cdots, N-1$. If set $x(k) = x_k,\ y(k) = y_k$, and to represent them in the plural form, then get the coordinate sequence of the boundary [6]:

$$s(k) = x(k) + jy(k)(k = 1, 2, \cdots, N-1) \tag{1}$$

for which the Discrete Fourier Transform may be computed as in the following equation [6]:

$$S(k) = \sum_{k=1}^{K-1} s(k)^{-j2\pi uk/K}(k = 1, 2, \cdots, K-1) \tag{2}$$

The Fourier coefficient $S(k)$ is used to derive the Fourier descriptor. In the case of translation, rotation, scaling and starting point changes, the coefficient $S(k)$ must be constant. It can be easily seen from the formula that the invariance under rotation and starting point variation is achieved only by the magnitude of the Fourier coefficient, while the invariance under translation is achieved by ignoring the DC component (coefficient $S(0)$).

Since the Fourier descriptor is related to the scale, direction, and position of the starting point of the image shape, in order to identify shapes with rotation, translation, and scale invariance, the Fourier descriptor needs to be normalized. According to the Fourier transform property, the position of the starting point of the object shape boundary is shifted by α length, the object is magnified by β times, the rotation angle φ and the translation displacement (x_0, y_0), and the new shape Fourier transform coefficient $S'(k)$ [7]:

$$\begin{aligned} S'(k) &= f[(x' + y')\beta e^{j\varphi} + (x_0 + iy_0)] \\ &= \beta e^{j\varphi} f(x' + y')\beta e^{j\varphi} + f(x_0 + iy_0) = \beta e^{j\varphi} e^{-j\frac{2\pi}{K}k\alpha} S(k) + f(x_0 + iy_0) \end{aligned} \tag{3}$$

$$\Rightarrow \begin{cases} S'(0) = \beta e^{-j\varphi} S(0) + f(x_0 + iy_0), k = 0 \\ S'(k) = \beta e^{-j\varphi} e^{-j\frac{2\pi}{K}k\alpha} S(k), k = 1, 2, \cdots, N-1 \end{cases} \tag{4}$$

$$\frac{\|S'(k)\|}{\|S'(1)\|} = \frac{\beta \left\| e^{-j\varphi} e^{-j\frac{2\pi}{K}k\alpha} S(k) \right\|}{\beta \left\| e^{-j\varphi} e^{-j\frac{2\pi}{K}\alpha} S(1) \right\|} = \frac{\|S(k)\|}{\|S(1)\|} \tag{5}$$

Where: $k = 1, 2, \cdots, N-1$, $\|\cdot\|$ indicates modulo. From the derivation formulas of Eqs. (3) and (4), it can be obtained that when the image shape is rotated φ and the starting position is transformed by α, the Fourier transform changes its phase $e^{j\varphi} e^{-j\frac{2\pi}{K}k\alpha}$; The object changes its amplitude β when it is scaled; when the object is translated, it only changes its $S(0)$ component $f(x_0 + iy_0)$. According to the ratio of the modulus of Eq. (5), the traditional normalized Fourier descriptor $Z(k)$ [8] is:

$$z(k) = \frac{\|S(k)\|}{\|S(1)\|}, k = 1, 2, \cdots, N-1 \tag{6}$$

The traditional normalized Fourier descriptor $z(k)$ eliminates the change of mode and phase through the ratio of the moduli, so that the Fourier descriptor has the rotation, translation and scale invariance of the shape. However, $z(k)$ is only identified using the modulus of the Fourier transform coefficient $S(k)$ as a feature, resulting in loss of image phase information. In general, phase information is very important for accurately identifying the shape of an object. All methods that are normalized by touch will result in an accuracy that is too low in image retrieval. In order to use the information contained in the phase of the Fourier coefficients, their invariance must be obtained at the starting point and direction change. If the starting point of the shape contour is determined, it can be easily implemented.

In the following sections, a novel method for determining the starting position of a shape will be introduced, which achieves the desired phase invariance.

2.2 Fourier Descriptor Based on Minimum Inertia Axis

Intra-class similarity and class similarity are actually improved by reducing the number of possible starting points. If compare one shape to all possible angles of another shape, even if they have no semantic relevance in those directions, it produces the best possible match between the shapes.

This paper presents a method for finding points with specific geometric and shape discriminating meanings. Set the starting point to the farthest point of the minimum inertia axis of the center of gravity and the intersection of the contours, and then fine tune around the intersection. Make sure the starting point of the contour C_1. The center of gravity formula is as follows [9]:

$$x_c = \frac{\sum p_i x_i}{\sum p_i}, y_c = \frac{\sum p_i y_i}{\sum p_i} \tag{7}$$

Where (x_i, y_i) is the coordinates of the pixel and p_i is the pixel value of the point.

In rigid body kinematics, it is pointed out that the axisymmetric rigid body with uniform mass distribution has the smallest moment of inertia when rotating around its axis of symmetry. The axisymmetric object presents two-dimensional shape information in the digital image. In the gray image, the target is composed of a pixel point set $\mathrm{p} = \{((\mathrm{x_i}, \mathrm{y_i})|\mathrm{i} \in [(\mathrm{x_i}, \mathrm{y_i})]\}$, where N is the target The number of pixels included, each of which can be considered equivalent to each of the quality micro-elements in the rigid body M. Then, the moment of inertia of the rigid body M for a straight line can be obtained by integrating the moment of inertia of each mass micro-element. Then the minimum inertia axis formula is as follows [10]:

$$I = \sum_{i=1}^{N} d_i^2 \tag{8}$$

Where $d_i = \frac{|y-(x-x_c)\cdot\tan\theta - y_c|}{\sqrt{1+\tan\theta^2}}$; (x_c, y_c) is the barycentric coordinates of the image; θ is the equal division of all pixel points of the image contour N parts, then the degree of gravity of the angle between the center of gravity and each boundary point and the main axis. After determining the minimum inertia axis, this paper determines the starting point as the farthest point between the minimum inertia axis and the contour intersection.

In the new normalization of the Fourier descriptor, the main direction φ of the normalized shape is the angle θ between the minimum inertia axis and the main coordinate axis. Since the main direction of the normalized shape is the horizontal direction, the main direction of the new shape is that the angle between the minimum inertia axis and the main direction is θ, and according to the derivation of Eq. (9), the boundary can be estimated by the main direction θ of the shape. The phase of the starting translational arc length α affects $e^{-j\frac{2\pi}{K}\alpha}$, there by eliminating the phase effect of the boundary starting point. Define a new normalized Fourier descriptor $z'(k)$, as shown in Eq. (10):

$$\frac{S'(1)}{\|S'(1)\|} = \frac{\beta e^{j\varphi} e^{-j\frac{2\pi}{K}k\alpha} S(1)}{\beta \left\| e^{j\varphi} e^{-j\frac{2\pi}{K}\alpha} S(1) \right\|} = e^{j\varphi} e^{-j\frac{2\pi}{K}\alpha} \tag{9}$$

$$z'(k) = \frac{s'(k) \cdot e^{j\theta}}{\|S(2)\|}, k = 1, 2, \cdots, N-1 \tag{10}$$

Since the details of the image correspond to the high-frequency part and the contour corresponds to the low-frequency part of the image, the high-frequency part of the outline is removed. Since the first point of the Fourier descriptor contains its high-frequency information, it is normalized. Use $\|S(2)\|$ when it is used.

2.3 Similarity Measure Algorithm and Evaluation Criteria

Similarity Measure Algorithm. In the content-based image retrieval, it is necessary to calculate the similarity matching between the image to be retrieved and the image in the database. Therefore, it is undoubtedly very important to define an appropriate visual feature similarity measurement method for image retrieval. Influence. The extracted visual features can be expressed in the form of vectors. In fact, the commonly used similarity measure methods are vector space models, that is, visual features are regarded as points in vector space, by calculating between two points. The proximity is a measure of the similarity between image features. This paper uses the Euclidean distance method.

The Euclidean Distance represents the distance between two points in the Euclidean space. The Euclidean distance between the two vectors f_{Q_i} and $f_{DB_{\mathrm{ji}}}$ is calculated as follows [7]:

$$\delta(f_{Q_i}, f_{DB_{\mathrm{ji}}}) = \sqrt{\sum_{i=1}^{Lg} (f_{DB_{\mathrm{ji}}} - f_{Q_i})^2} \tag{11}$$

Which $f_{DB_{\mathrm{ji}}}$ represents the i feature of the j image in database |DB|, Lg represents the number of images in the database.

Evaluation Criteria. In the field of image retrieval, the average retrieval precision and the average retrieval rate are generally used to evaluate the performance of the algorithm. For the image I_q to be retrieved, the precision and recall are defined as follows [11]:

$$\Pr ecision : P(I_q) = \frac{Number\ of\ related\ images\ retrieved}{Total\ number\ of\ images\ retrieved} \tag{12}$$

$$\Pr ecision : P(I_q) = \frac{Number\ of\ related\ images\ retrieved}{The\ total\ number\ of\ related\ images\ in\ the\ database} \tag{13}$$

For the image I_q to be retrieved, the average precision calculation formula (6) is shown in Eq. (14) [12]:

$$P(I_q, n) = \frac{1}{n}\sum_{i=1}^{|DB|} \left|Q(\varphi(I_i), \varphi(I_q))\right| \tag{14}$$

Where n represents the number of related pictures retrieved, |DB| represents the size of the database, and $\varphi(x)$ represents the type of x.

For the image to be retrieved I_q, the average recall calculation formula (6) is shown in Eq. (15) [12]:

$$P(I_q, n) = \frac{1}{N}\sum_{i=1}^{|DB|} \left|Q(\varphi(I_i), \varphi(I_q))\right| \tag{15}$$

Where N represents the total amount of images associated with the image I_q to be retrieved.

It can be seen from the above formulas (14) and (15) that each image in the image database is used as the image to be retrieved, and the calculated precision is averaged to obtain an average precision. Similarly, the image is obtained. Each image in the database is used as the image to be retrieved, and the calculated recall rate is averaged to obtain the average recall rate. The paper uses the average retrieval precision and the average retrieval rate as the evaluation criteria for retrieval performance.

3 Experimental Results and Analysis

3.1 Experimental Environment and Data

(1) Experimental environment: Matlab R2016b, Windows 10 operating system.
(2) Experimental data (Fig. 1):

Data preprocessing: Since the product images are all color images, the image is first converted into a binary image and the binary image is subjected to threshold processing; secondly, the coordinates of the edge points of the image contour are extracted, and the difference between the coordinate points is obtained with equal spacing. Contour coordinate point; then find the image centroid under the obtained coordinate

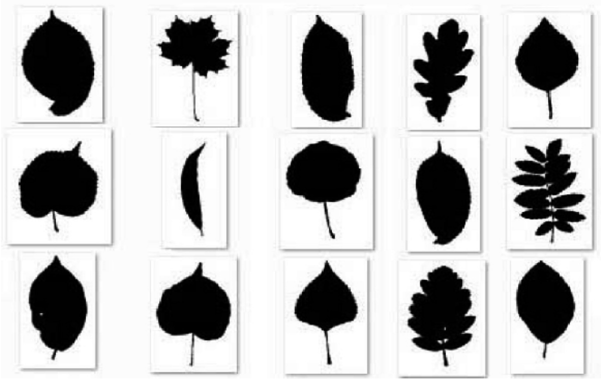
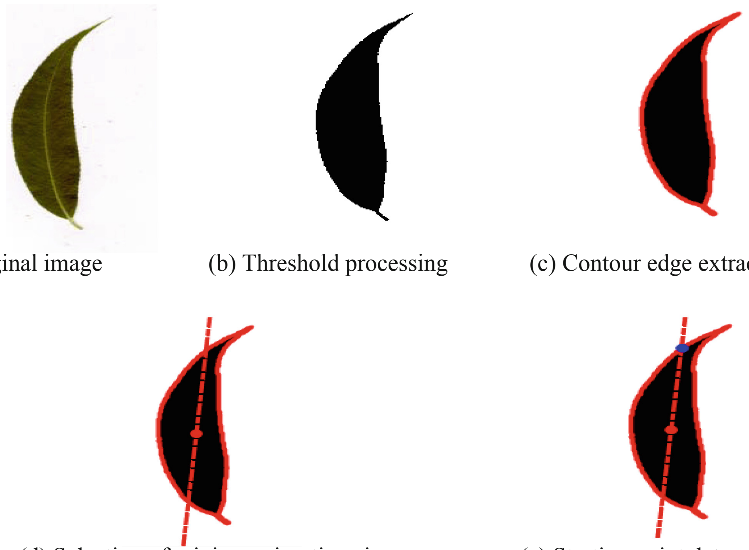


Fig. 1. An image of each plant in the Swedish Leaf database. Source: Swedish Leaf database as an experimental sample.

point to determine the minimum inertia axis of the image, and find the intersection of the minimum inertia axis and the image contour to determine the starting point of the Fourier transform next, and determine the image coordinate point after the starting point The Fourier transform is performed to obtain the feature vector. Finally, the image retrieval is realized by calculating the distance between the feature vector of the image to be retrieved and the feature vector of the image in the image database.

3.2 Image Shape Feature Extraction

As can be seen from the process of image feature extraction in Fig. 2, (a) is the original image of the leaves in the Swedish Leaf database, the size is 256×154, (b) is the



(a) Original image (b) Threshold processing (c) Contour edge extraction

(d) Selection of minimum inertia axis (e) Starting point determination

Fig. 2. Image feature extraction.

image threshold processing, using the automatic threshold processing method for the leaves The image is subjected to threshold processing; (c) is based on the contour edge extraction method. In the image retrieval based on Fourier descriptor, the required Fourier points are equidistant, so the equidistant interpolation method can guarantee The accuracy of the Fourier descriptor in the image retrieval process; (d) is the minimum inertia axis of the selected image; and (e) is the determination of the starting point of the image.

3.3 Feature Matching Results and Analysis

In the retrieval experiment, the number of boundary sample points for each shape contour is 256. A total of 50 low frequency Fourier coefficients are used to determine the similarity between the two shapes. Each shape in the database is treated as a query object. For each query, record the accuracy of the recall at each level. The FMAL accuracy retrieved is obtained by averaging all query results at each recall level. In theory, as the recall rate increases, so does the accuracy.

We compare the improved method with the retrieval of several widely used or recently introduced feature functions on two databases. These feature functions include CD, CC and FPD. 50 Fourier coefficients are reserved for each function to calculate the similarity between the two shapes.

Figure 3 shows the average accurate recall graph test on the Swedish leaf database, where each curve represents a method and the improved method is represented by MIAFD.

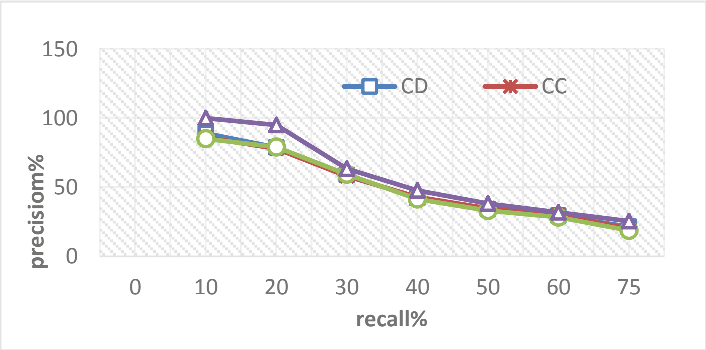


Fig. 3. Precision-Recall curve using five methods on the Swedish leaf database.

The experimental results show that the accuracy of the Fourier descriptor based on the minimum inertia axis is significantly improved on the Swedish leaf database. This shows that the method of determining the starting point position of the contour through the minimum inertia axis is effective for image retrieval in this paper, which not only ensures that the Fourier descriptor is not affected by the starting position, but also improves the accuracy of the image.

4 Conclusion

The shape description of the Fourier description is performed by determining the position of the starting point. Determining the starting point position by the minimum inertia axis is valuable for determining shape feature based retrieval. Where the starting point and rotation invariance are required, the minimum inertia axis determining starting point can be used to improve the shape description method. Combined with scale normalization and Fourier coefficient-based retention principles, they can be used for contour normalization to apply other more complex shape description techniques. Experiments performed on commonly used Swedish leaf databases have shown that this method is effective compared to other Fourier descriptors.

Acknowledgement. This work is supported by the Harbin Science and Technology Bureau outstanding subject leader fund project (2017RAXXJ055), and Nature Science Foundation of Heilongjiang Province (F2018020), and the Basic Scientific Research Operating Expense Project of Provincial Institutions of Higher Education in Heilongjiang (17XN003).

References

1. Datta, R., Joshi, D., Li, J., et al.: Image retrieval: Ideas, influences, and trends of the new age. ACM Comput. Surv. **40**(2), 1–60 (2008)
2. Madugunki, M., Bormane, D.S., Bhadoria, S., et al.: Comparison of different CBIR techniques. In: International Conference on Electronics Computer Technology, pp. 372–375 (2011)
3. Zhao, C., Ren, Y.: Research on shape-based image retrieval. Mod. Electron. Technol. **13**, 159–162 (2008)
4. Sokic, E., Konjicija, S.: Phase preserving Fourier descriptor for shape-based image retrieval. Signal Process. Image Commun. **40**(C), 82–96 (2016)
5. Bartolini, I., Ciaccia, P., Patella, M.: WARP: accurate retrieval of shapes using phase of fourier descriptors and time warping distance. IEEE Trans. Pattern Anal. Mach. Intell. **27**(1), 142–147 (2005)
6. Jain, A.K.: Fundamentals of Digital Image Processing, pp. 370–371. Prentice-Hall Press, Upper Saddle River (1989)
7. Wang, T., Liu, W., Sun, J., Zhang, H.: Using fourier descriptors to recognize object's shape. Comput. Res. Dev. **39**(12), 1714–1719 (2002)
8. Zhang, Y.: Image Processing and Analysis, 2nd edn. Tsinghua University Press, Beijing (1999)
9. Liang, N., Guo, L., Yu, Y.: A symmetry axis detected method based on the minimal value of moment of inertia. Microprocessor **6**, 62–64 (2009)
10. Shanmugavadivu, P., Sumathy, P., Vadivel, A.: FOSIR: fuzzy-object-shape for image retrieval applications. Neurocomputing **7**, 719–735 (2016)
11. Hong, T., Sun, H., Jin, X., et al.: Study on image retrieval algorithm based on local right-angle mean patterns. J. Harbin Univ. Commer. **7**(2), 172–174 (2017)