Rotation invariant Shape Contexts based on 2D Fourier transform and Eigenshapes for Radiological image retrieval

Alaidine Ben Ayed¹, Mustapha Kardouchi³ and Sid-Ahmed Selouani²

¹ Université de Moncton, Campus de Moncton. 18, avenue Antonine-Maillet, Moncton, NB, Canada E1A 3E9
alaidine.ben.ayed@umoncton.ca,

² Université de Moncton, Campus de Shippagan. 218, boulevard J-D Gauthier, Shippagan, NB, Canada E8S 1P6

Abstract. This paper presents a new descriptor based on Shape Contexts, Fourier transform and Eigenshapes for radiological medical image retrieval. First shape context histograms are computed. Then, 2D FFT are performed on each 2D histogram to achieve rotation invariance. Finally, histograms are projected onto a more representative and lower dimensionality feature space that highlights the most important variations between shapes by computing eigenshapes. Eigenshapes are the more representative features, they are the principal components for radiological images. The proposed approach is translation, scale and rotation invariant, furthermore, retrieving operation is robust and fast due to space dimensionality. Experimental results with classes of the medical IRMA database demonstrate that the proposed approach produces better performance than known Rotation Invariant Shape Contexts based on Feature-space Fourier transformation.

keywords: Image retrieval; Shape Contexts; Fourier transform; Eigenshapes; Radiological images.

1 Introduction

Medical image processing has been one of the most fundamental areas of research over the last years. The major goal is to close the gap between informatics and medicine to improve health care. The radiology is considered as the second producer of digital images. So, the growing amounts of radiological data underlines the need to create new efficient tools for medical image processing.

Among the fields of medical image processing is image retrieval. The goal is to retrieve the most similar images to a query image used for diagnostic or therapy. Each image in the database is identified through its corresponding signature. Choosing the best descriptor to extract those signatures is a challenge. In the case of radiological images we do not have textures. Also, all images are
in gray scale levels whose distributions are very close. So using textures or gray scale information is not sufficient to distinguish between two radiological images. The better way to define signatures is then shape study. There are two approaches: a global one (known as continuous approach) and a structural one (known as non continuous). For the global approach, objects are considered as one unit curve. The structural approach is usually based on points of interest to divide the contour of the object into segments known as primitives [1]. Among the most robust shape descriptors is Shape context [2]. In this paper we improve the rotation invariant shape contexts based on Feature-space Fourier Transformation [7] by projecting shape images onto a more representative and lower dimensionality feature space that highlights the most important variations between shapes. Eigenshapes are the more representative features, they are the principal components for radiological images.

This paper is organized as follows: Section 2 illustrates previous work on shape contexts. The proposed approach is described in Section 3. Experimental results are presented in Section 4. Finally, we conclude this paper in Section 5.

2 Previous work

Serge Belongie et al. proposed the Shape context feature descriptor used for shape matching and object recognition [2] [3] [4]. The main idea is to extract the contour of the object, then to pick \( n \) points. Selected points do not need to be key points such as maxima or minima of curvature. Computing shape context at a given point is performed by taking it as the center of a log polar coordinate system and to focus on the distribution of vectors originating from that point over relative positions. This distribution is a reach description of the shape localized at that point. The use of a log polar coordinate system makes descriptor more sensitive to differences in nearby points. In this paper we use 12 equally spaced bins and 5 equally spaced log-radius bins (Figure 1).

![Log polar grid with 60 bins used to compute shape contexts.](image)
Shape context for a given point \( p_i \) is an histogram providing the the distribution of the the remaining \( n - 1 \) points over \( l \) bins. Mathematically, it is described as follows:

\[
h_i(k) = \# \{q \neq p_i : (q - p_i) \in \text{bin}(k) | k = 1, 2, ..., l \}
\]

Su Yang et al. proposed a 2D FFT-RISC technique (Rotation invariant shape contexts via FFT) [7]. Computation of this feature is performed by using coordinates and tangents at every point to compute a set of modules and angles \( \{(r_{ij}, \alpha_{ij}) | i, j = 1, 2, ..., n\} \) for every point \( p_i \). This set is used to obtain 2D histograms defining shape contexts. Application of a 2D FFT on those 2D histograms provides rotation-invariance. In the next section, we propose to ameliorate FFT-RISC by mapping data in a lower dimensionality space that keeps only most relevant information. This makes retrieval procedure more robust. Also, it reduces execution time.

3 ROTATION INVARIANT SHAPE CONTEXTS BASED ON 2D FFT AND EIGENSHAPES

For a given image, we consider a reference point and we pick other \( n - 1 \) equidistant points. Every point is described via its shape context which is a 2D histogram. Each histogram is then reshaped onto 1D vector which is added as a new line to the matrix representing the signature of that image. The signature is so an \( M_{nl} \) matrix where \( n \) denotes the number of picked points and \( l \) denotes the number of bins. The next two subsections describe the training and retrieval procedures.

3.1 Training

In the training phase, we consider a set of \( m \) training images \( S = \{S_1, S_2, ..., S_m\} \). Each image is represented by an \( M_{nl} \) matrix which is converted onto a column vector \( \zeta_i \). \( \zeta_i \) is a vector of \( z \) components where \( z = n \times l \). Then the average shape vector \( \tau \) is computed as follows:

\[
\tau = \frac{1}{m} \sum_{i=1}^{m} \zeta_i
\]

Note that each \( \zeta_i \) should be normalized to get rid of redundant information. This is performed by subtracting the mean shape:

\[
\Theta_i = \zeta_i - \tau
\]

In the next step, we compute the covariance matrix defined as follows :

\[
C = \frac{1}{m} \sum_{n=1}^{m} \Theta_n \Theta_n^t = AA^t
\]
Where $A = [\Theta_1, ..., \Theta_m]$. Note that $C$ in (4) is a $z \times z$ matrix and $A$ is a $z \times m$ matrix.

Eigenshapes are the eigenvectors of the covariance matrix. They are obtained by performing a singular value decomposition of $A$:

$$A = U.S.V^T$$

(5)

Where dimensions of matrix $U$, $S$ and $V$ are respectively $z \times z$, $z \times m$ and $m \times m$. Also, $U$ and $V$ are orthonormal matrix ($UU^T = U^TU = Id_z$ and $VV^T = V^TV = Id_m$). In addition to that:

- Columns of $V$ are eigenvectors of $A^tA$.
- Columns of $U$ are eigenvectors of $AA^t$.
- Squares of singular values $s_k$ of $S$ are the eigenvalues $\lambda_k$ of $AA^t$ and $A^tA$.

Note that $m < z$. So eigenvalues $\lambda_k$ of $AA^t$ are equal to zero when $k > m$ and their associated eigenvectors are not necessary. So matrix $U$ and $S$ can be truncated. We consider so $A = USV^t$ where dimensions of $U$, $S$ and $V$ are respectively $z \times m$, $m \times m$ and $m \times m$. Then, eigenshape space is composed by the largest $K$ eigenvectors:

$$\Xi_K = [U_1, U_2, ..., U_K]$$

(6)

Each projected image in the Eigenshape space is represented as a linear combination of $K$ eigenshapes:

$$\Theta_{i}^{proj} = \sum_k C_{\Theta_i}(k)U_k$$

(7)

Where $C_{\Theta_i}(k) = U_k^t\Theta_i$ is a vector providing coordinates of the projected image in the Eigenshape space.

### 3.2 Retrieval

Now, given a query image $\zeta$, the goal is to retrieve the most similar image to it in the database. First of all, $\zeta$ is normalized: $\Theta_q = \zeta - \tau$. Then it is projected on the Eigenshape space.

$$\Theta_{q}^{proj} = \sum_k C_{\Theta_q}(k)U_k$$

(8)

The distance between the projected image and any other image is defined as follows:

$$d_i(\Theta_{q}^{proj}) = ||\Theta_{q}^{proj} - \Theta_{i}^{proj}||$$

(9)

the most similar image is that one corresponding to min($d_i$).
4 EXPERIMENTAL RESULTS

4.1 Image collection

To evaluate the performance of the proposed approach we use the radiological IRMA database. It contains images of several body parts. Figure 2 shows some samples of this database.

![IRMA samples](image)

Fig. 2. IRMA samples.

We randomly selected 1000 images belonging to four classes: Hands, Breasts, Chests and Heads. The number of images per class is the same. Figure 3 shows sample images of these classes. Images in figure 3 are used in the next sub-section as targets to evaluate the performance of the proposed approach. To measure the similarity between images, we use the Euclidean distance.

![Sample images](image)

Fig. 3. Four classes used for performance measurement.

4.2 EXPERIMENTS

To evaluate the proposed approach, we use recall and precision measurements. Precision is defined as the ratio between the number of correctly retrieved images and the total number of images retrieved, while recall is defined as the ratio between the number of correctly retrieved images by search and the total number of samples of the class to which belongs the query image. For each measure of recall precision, we consider the 10, 20, 40, 60, 80, 100, 150, 200 and 250 most similar images. We compare the proposed method (RISC-FFT-EIG) with the
FFT-RISC approach.

Figure 4 plot the recall precision curve for the Hand sample image (a) showing that the RISC-FFT-EIG outperforms significantly the FFT-RISC. Precision rate remains superior to 90 % even when we consider the best 250 retrieved images.

![Fig. 4. Recall versus Precision: Hand sample image.](image)

Recall and precision curves for the Breast sample image (b) are illustrated in figure 5 showing that the precision rate is equal to 100 % for the five first measurements. The proposed approach provides better recognition rates when recall is higher than 0.4

![Fig. 5. Recall versus Precision: Breast sample image.](image)
Figure 6 plot recall precision for the Chest image sample (c) showing that the precision rate is 100% for both of the tested approaches, then it is higher for the FFT-RISC. However the FFT-RISC-EIG outperforms when recall is higher than 0.5. This means that the proposed approach gives better recognition rates when the number of best retrieved images taken in consideration is high.

To further prove the performance of the proposed approach we compute average of the precision rate per class considering the best 200 images retrieved. Results are illustrated by Table 1 showing that the recognition rate with FFT-RISC is equal to 90.72% while it reaches 95.65% when we use the RISC-FFT-EIG method. This is due to elimination of noisy data. Another advantage of the proposed method is reduction of execution time which is related to space dimensionality.

Table 1. Average of the precision rate per class considering the best 200 images retrieved.

<table>
<thead>
<tr>
<th>Image</th>
<th>FFT-RISC</th>
<th>RISC-FFT-EIG</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hands</td>
<td>83.09</td>
<td>98.08</td>
</tr>
<tr>
<td>Breasts</td>
<td>85.65</td>
<td>93.47</td>
</tr>
<tr>
<td>Chests</td>
<td>98.49</td>
<td>97.01</td>
</tr>
<tr>
<td>Heads</td>
<td>95.67</td>
<td>94.06</td>
</tr>
<tr>
<td><strong>Average</strong></td>
<td><strong>90.72</strong></td>
<td><strong>95.65</strong></td>
</tr>
</tbody>
</table>

Figure 7 gives an example of image retrieval. Each sample in the top line is taken as a target. Then, 10 most similar images are extracted. The retrieved
images are listed in the descending order of similarity from top to down.

Fig. 7. Example of image retrieval

5 CONCLUSIONS

Shape contexts have been subject of several studies due to their simplicity and efficiency. They are translation and scale invariant. Performing a 2D FFT on each 2D shape context histogram provides rotation invariance.

We have shown that we obtain better precision rates when we project data in a lower dimensionality space that highlights the most important variations.
Furthermore, retrieving operation becomes faster. This is due to the lower dimensionality of the new space. Note that the major drawback of shape contexts and their extensions is the fact that they can not deal with images having many textures.

The proposed approach can be improved by using the Fuzzy logic concepts: the probability of belonging to a given bin is not absolute. So a contour point does not belong to a single bin, however, it also belongs to the surrounding bins with lower belonging probabilities.

Acknowledgment

This work was supported by the New Brunswick Innovation Foundation (NBIF) and the Natural Sciences and Engineering Research Council of Canada (NSERC). Authors would like to thank University of Ashen for the IRMA database.

References