ABSTRACT

In this paper, we proposed a new affine invariant region-based shape descriptor, the ICA Zernike Moment Shape Descriptor (ICAZMSD). Independent Component Analysis (ICA) is first used to turn the original shape into a canonical form, in which the effects of scaling and skewing are eliminated. Next, the properties of the Zernike transform is used to further eliminate the effects of any possible rotation and reflection of the canonical shapes, in extracting the Zernike moments as the affine invariant region-based descriptors. Using the proposed ICAMSD as shape feature, shape-based image retrieval experiments on a 4000 complex shape image database and a 5600 simple shape image database, show promising retrieval rates of 99.80% and 92.25%, respectively.

Index Terms— Affine Invariant, Shape Descriptor, Independent Component Analysis, Zernike moments

1. INTRODUCTION

Shape is an important low level feature in object recognition and image retrieval. The shape of an object from one viewpoint to another can be linked through an affine transformation, if it is viewed from a much larger distance than its size along the line of sight. Since images in database are often geometrically distorted by the change of viewpoints, it is often desirable to have affine invariant shape descriptors and algorithms for image retrieval. Currently, almost all of the affine invariant shape descriptors, such as [1, 2, 3], are contour-based, and few of them are region-based. However, contour-based shape descriptors, which utilize only the contour of the shape, have their limitations, i.e., they can describe only a simple shape with a single connected region as in Fig.1(a), but not a complex shape with holes in the object or consisting of several disjoint regions, as in Fig.1(b) and Fig.1(c), respectively. By contrast, region-based shape descriptors, which utilize all the pixels of the shape, have no such limitations. They can describe both simple and complex shapes.

To fill the need for affine invariant region-based shape descriptors, we propose the ICA Zernike Moment Shape Descriptor (ICAZMSD). In [3], Independent Component Analysis (ICA) was used in transforming object contour to a canonical form for shape matching. Because of the inherited order and sign ambiguities of ICA, and the mismatching of the starting point, exhaustive trial approaches must be used after the ICA transformation. Also, since the method is contour-based, it can only be applied to simple shape images, but not to complex shape images, which have far richer content. In our newly proposed ICAMSD, ICA is applied to all pixels constituting the shape, in order to transform the shape into its canonical form. Zernike moments are then extracted from the canonical shape as the affine invariant region-based shape descriptor. Such a region-based approach makes the ICAZMSD applicable to both simple and complex shape images. The invariant properties of the Zernike transform are utilized to tackle the problem of the ICA ambiguities, so that no exhaustive trial approaches are required.

2. AFFINE TRANSFORM

Affine transform can be decomposed into scale, skew, rotation and translation. Using a vector-matrix notation, a linear mixing model can be written as:

$$\tilde{x} = Ax + T,$$  \hspace{1cm} (1)

where \(x\) and \(\tilde{x}\) are the vectors that contain the coordinates of the original and the affine transformed shape images, respectively. The \(2 \times 2\) nonsingular matrix \(A\) can be decomposed as follow:

$$A = \begin{bmatrix} s_x & 0 \\ 0 & s_y \end{bmatrix} \begin{bmatrix} 1 & \alpha \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix},$$  \hspace{1cm} (2)
where \( s, \alpha \) and \( \theta \) represent scale, skew and rotation, respectively. \( s_x \) and \( s_y \) are the scaling factors in the \( x \) and the \( y \) directions, respectively. When \( s_x \) and \( s_y \) are of different signs, it corresponds to a reflection. The \( 2 \times 1 \) vector \( T \) represents the translation. When the effect of translation is eliminated by setting the origin of the coordinate system to the centroid of the shape, (1) reduces to:

\[
\bar{x} = Ax. \tag{3}
\]

3. ICA ZERNIKE MOMENT SHAPE DESCRIPTOR

3.1. Canonical shape by ICA

In order to compare images from a database where they may be affine related, we need to first transform the shape images into their canonical form. We do this by using Independent Component Analysis[4], which can demix mixed signals even though both the mixture matrix and the source signals are unknown. Using the pixel coordinates of the shape image as two “observed mixed signals", ICA could extract two “independent source signals” which are the pixel coordinates of the new shape image in the canonical form[3]. Different from [3] where ICA is applied only to the pixels of the shape contour, ICA is applied to all the pixels constituting the shape when extracting our proposed region-based shape descriptor. The \( x \) and \( y \) coordinate values of all the pixels constituting the shape, are used as two input signals of the ICA algorithm.

In Fig.2, the canonical shape images shown in the second column of the figure are transformed by ICA from the corresponding shape images shown in the first column. The shape images in (a1) and (a2) are affine related to each other, and the one in (a3) is affine related to neither (a1) nor (a2).

3.2. Zernike Moments Extraction from Canonical Shape

ICA eliminates the scaling and skewing effects of any affine transformations. However, the shapes in the canonical form are not yet ready for comparison. There are the coordinate and sign ambiguities of the canonical shapes, which are inherited from the order and sign ambiguities of ICA. As shown in Fig.2, (b1) is a reflected and 90-degree rotated version of (b2), although their corresponding shape images are affine related. In other cases, the coordinate and/or the signs may or may not be switched. To tackle the problem, we explore the reflection and rotation invariant properties of the Zernike moments[5] and extract the Zernike moments from the canonical shape images as theICAZMDS.

Zernike polynomials, \( \{V_{nm}\} \), are complex sets of orthogonal functions within a unit circle.

\[
V_{nm}(x, y) = V_{nm}(\rho, \theta) = R_{nm}(\rho) e^{jnm\theta}, \quad \rho \leq 1 \tag{4}
\]

where

\[
\begin{align*}
R_{nm}(\rho) &= \sum_{s=0}^{\min(n-m, m, n)} (-1)^s \frac{(n-s)!}{s!(n+|m|)/2-s!(n-|m|)/2-s!} \rho^{n-2s} \quad \text{Radial polynomial defined as:} \\
Z_{nm} &= \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 f(\rho, \theta) V_{nm}(\rho, \theta) \rho d\rho d\theta \\
&= \frac{n+1}{\pi} \int_0^{2\pi} \int_0^1 f(\rho, \theta) R_{nm}(\rho) e^{jnm\theta} \rho d\rho d\theta \tag{6}
\end{align*}
\]

The rotation invariant property of the Zernike Moments, \( |Z_{nm}^{\text{rotated}}| = |Z_{nm}| \), has been proven in [5]. Here, we prove the reflection invariant property of the Zernike moments. Since a reflection against the \( y \) axis can be decomposed into a rotation, and a reflection against the \( x \) axis, we only have to prove the Zernike moments are reflection invariant against the \( x \) axis.
If the reflected image against the $x$ axis is denoted by $f_{RF}$, the relationship between the reflected image and the original one in the same polar coordinates is

$$f_{RF}(\rho, \theta) = f(\rho, -\theta)$$  \hspace{1cm} (7)

By a change of variable $\theta_1 = -\theta$,

$$Z_{nm}^{RF} = \frac{n + 1}{\pi} \int_0^{2\pi} \int_0^1 f(\rho, -\theta) R_{nm}(\rho) e^{-jm\theta} \rho d\rho d\theta$$  

$$= -\frac{n + 1}{\pi} \int_0^{-2\pi} \int_0^1 f(\rho, \theta_1) R_{nm}(\rho) e^{jm\theta_1} \rho d\rho d\theta_1.$$  \hspace{1cm} (8)

And,

$$|Z_{nm}^{RF}| = \frac{n + 1}{\pi} \int_0^{2\pi} \int_0^1 |f(\rho, \theta_1) R_{nm}(\rho) \rho| d\rho d\theta_1$$  

$$= \frac{n + 1}{\pi} \int_0^{2\pi} \int_0^1 |f(\rho, \theta) R_{nm}(\rho) \rho| d\rho d\theta = |Z_{nm}|.$$  

The first 36 Zernike moments of up to order 10, are used as the proposed ICAZMSD. As we can observe from Fig.2, the ICAZMSDs, (c1) and (c2), which corresponds to the affine related shape images in (a1) and (a2) are almost the same, while they are different from their counterpart in (c3) that corresponds to the shape image in (a3). The steps of extracting the proposed ICAZMSD are summarized in Fig.3.

**Fig. 3.** Diagram of the ICAZMSD extraction

### 4. EXPERIMENTAL RESULTS

#### 4.1. Experiment on traffic sign database

In the first experiment, we took 5 pictures of a "Stop" sign from different viewpoints and distances, the ICAZMSDs was then extracted. The FastICA algorithm[4] was used in performing ICA. As we can observe from Fig.4, although the shapes of the "Stop" sign taken from different viewing angle and distance are different, their corresponding ICAZMSDs are almost the same. At the same time, they are different from the ICAZMSD extracted from the "No Left Turn" sign. Difference was also observed in comparison with the ICAZMSDs extracted from other traffic signs in our experiment. Next, 5 pictures of the "Stop" sign from different viewpoints and distances, and 30 pictures of other traffic signs were used as test database for shape-based retrieval. Using any one of the five pictures of the "Stop" sign, the retrieval system, using the proposed ICAZMSD, was able to retrieve all the other four related pictures in the top four matches without any error. Jeffery divergence was used as the similarity measure, as it showed better performance than other similarity measures in our experiments.

![Fig. 4](image_url)

(a) and (c) Pictures of a "Stop" sign, taken from different viewing angle and distance; (b) and (d) the corresponding ICAZMSDs of the shapes of the "Stop" sign in (a) and (c), respectively; (e) Pictures of a "No Left Turn" sign; (f) and the corresponding ICAZMSD of (e)

#### 4.2. Experiment on complex shape image database

In the second experiment, we further tested the ICAZMSD as feature for shape-based silhouette retrieval, using a larger shape image database. Each of the 50 different complex shape silhouette images, shown in Fig.5, was affine transformed using 80 transformation matrices, creating a $50 \times 80 = 4000$ image test database. The parameters of the transformation in the experiment are: $(s_x = s_y = 1), (s_x = 0.7; s_y = -0.7), (s_x = s_y = 2), (s_x = 1; s_y = 2), \theta \in [0, \pi/6, 3\pi/5, 6\pi/7], $ and $\alpha \in [0, 0.25, 0.5, 0.75, 1]$. Thus, we have 50 classes of images and each of them has 80 affine related silhouette images. Ideally, all the 80 images, affine related to the query, should be retrieved and ranked at the top 80 matches, when a query image is given. The retrieval rate is defined as the percentage of the correctly retrieved images in the top 80 matches. We used each of the 4000 images as query and got an overall retrieval rate of 99.80%, using the proposed ICAZMSD. In comparison, the Invariant Zernik Moment Shape Descriptor(IZMSD) [3], which is only scale and rotation invariant, but not affine invariant, had an overall retrieval rate...
of 39.78%. The average precision vs. recall graph (APRG) is another measure, often used in measuring the performance of a retrieval system. Fig.6(a) shows the APRG of the retrieval, using the ICAZMSD. The APRG indicates very good performance. For large parts of the APRG curve, the average precision rates are equal to 100%, and for the rest parts, they are very close to 100%. In comparison, the APRG of the retrieval system using the IZMSD, goes downward quickly, when the recall value increases.

Fig. 5. The original 50 complex shape images that generate the 4000 image test database used in the second experiment.

4.3. Experiment on simple shape image database

In the third experiment, we compare the ICAZMSD with the Affine Invariant Fourier Shape Descriptor (AIFSD)[1], a contour-based affine invariant shape descriptor, using simple shape image database. 70 different simple shape silhouette images of different classes, were selected from the MPEG-7 CE Shape-1 Part-B data set[6]. Each of them are affine transformed using the same 80 transformation matrices used in the second experiment. On this $70 \times 80 = 5600$ image test database, the overall retrieval rates, using the ICAZMSD, AIFSD and IZMSD, are 92.25%, 79.74%, and 36.01%, respectively. Also, as we can observe from their APRGs in Fig.6(b), the proposed ICAZMSD overperform both of its counterparts.

Comparing the performance of the proposed ICAZMSD in the second and the third experiments, we found that it works better on complex shape images than on simple shape images. Our interpolation is that the complex shapes, containing the information of contour, holes, and disjoint parts inside them, can provide far more information for the ICAZMSD to distinguish different shapes than the simple shapes containing only the contour information.

5. CONCLUSIONS

In this paper, we propose a new affine invariant region-based shape descriptor, the ICAZMSD. ICA is used to transform shape into its canonical form, Zernike moments are then extracted from the canonical shape as the affine invariant region-based shape descriptor. We tested the descriptor by using it as feature in shape-based silhouette image retrieval. Using the proposed ICAZMSD, experiments on a 4000 complex shape database and on a 5600 simple shape database show promising retrieval rates of 99.80% and 92.25%, respectively. The proposed ICAZMSD fill the need for an affine invariant region-based shape descriptor, which can describe both simple and complex shapes.

6. REFERENCES


