

Human Shape Recognition Algorithm Design Based on Hu Moments and Zernike Moments

Weiguo Chen, Jinchuan Lu

Guangxi technological college of machinery and electricity, Nanjing 530007, China.

1097795279@qq.com

Keywords: Hu moments, Zernike moments, human shape recognition, invariant moments.

Abstract. The invariant moment feature of target is taken as an important approach to human shape recognition. This paper gives a general review of Hu moments and Zernike moments in human shape recognition, and utilizes the minimum distance classifier to classify some moving targets. By comparing their characteristics in specific applications, this paper provides some basis for the selection of invariant moments in human shape recognition algorithm.

Introduction

Human shape recognition is an important application of dynamic video target detection technology and belongs to the scope of image recognition in terms of pattern recognition. It extracts the features of moving targets in images to classify the moving targets. Invariant moment is a feature of image with translation, rotation or scale invariance. The invariant moments commonly used in human shape recognition are Hu moments and Zernike moments. Introduced by M.K. Hu in 1962, Hu moments [1] have 7 invariant moment expressions for translation, rotation or scale invariance. Zernike moments are a set of orthogonal moments originated from the orthogonal moment theory put forth by Teague[2]. For its good rotation invariance, orthogonal moments can be used to construct any higher-order moment, and this integral operation is insensitive to noise.

Principles of Invariant Moments

Hu Invariant Moments

It is assumed that the gray scale distribution in target zone D is $f(x, y)$, $(x, y) \in D$. To describe the target, the gray scale distribution of the zone \bar{D} outside target zone D is regarded as 0. Hence, the origin moment and central moment of $p+q$ -order zone in the target are transformed to [3,4]:

$$m_{pq} = \iint_D x^p y^q f(x, y) dx dy \quad p, q = 0, 1, 2, \dots \quad (1)$$

$$\mu_{pq} = \iint_D (x - \bar{x})^p (y - \bar{y})^q f(x, y) dx dy \quad p, q = 0, 1, 2, \dots \quad (2)$$

There is the conversion relationship between central moment and origin moment as follows:

$$\mu_{pq} = \sum_{k=0}^p \sum_{l=0}^q \binom{p}{k} \binom{q}{l} (-1)^{k+l} m_{p-k, q-l} m_{10}^k m_{01}^l m_{00}^{-(k+l)} \quad (3)$$

The normalized central moment is defined as follows:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^r} \quad (4)$$

In which, $r = \frac{p+q+2}{2}$, $p+q = 2, 3, \dots$.

For convenience and generality, it is assumed that image is converted into a binary image after edge extraction. In other words, the pixel gray scale is 1 at the boundary of target zone, but 0 at other areas. Hence, boundary moment can be defined as follows:

$$m_{pq} = \oint_C x^p y^q ds \quad p, q = 0, 1, 2, \dots \quad (12)$$

In which, \oint_C stands for the integral of the curve C along the boundary of target, i.e.

$$ds = \sqrt{(dx)^2 + (dy)^2}.$$

The centralized boundary moment is as follows:

$$\mu_{pq} = \oint_C (x - \bar{x})^p (y - \bar{y})^q ds \quad p, q = 0, 1, 2, \dots \quad (13)$$

In which, $\bar{x} = m_{10} / m_{00}$ and $\bar{y} = m_{01} / m_{00}$.

To digital image, there will be:

$$m_{pq} = \sum_{(x,y) \in C} x^p y^q \quad (14)$$

$$\mu_{pq} = \sum_{(x,y) \in C} (x - \bar{x})^p (y - \bar{y})^q \quad (15)$$

μ_{pq} is invariant to translation, rotation and scale.

The normalized boundary moment is defined as follows:

$$\eta'_{pq} = \mu_{pq} / \mu_{00}^{p+q+1} \quad p+q = 2, 3, \dots \quad (16)$$

It is assumed that the boundary C of target zone is a smooth plane curve, x, y coordinates are scaled down at the same factor $r > 0$, and the curve C changes to C' in the new coordinate system, at random $r > 0$, there is:

$$\frac{\mu'_{pq}}{(\mu'_{00})^{p+q+1}} = \frac{r^{p+q+1} \mu_{pq}}{r^{p+q+1} |C|^{p+q+1}} = \frac{\mu_{pq}}{\mu_{00}^{p+q+1}} \quad (19)$$

Then, η'_{pq} is substituted into the above seven moment invariant functions in place of η_{pq} , so seven normalized boundary central moment functions are obtained to be translation, rotation and scale invariant.

Zernike Invariant Moments

In a practical problem, image is often digital, so it is necessary to make real and imaginary numbers of Zernike moment discrete. Also, Zernike polynomial is orthogonal inside unit circle, so it is necessary to convert the image involved into the polar coordinates inside the unit circle. For this reason, Mukundan and Ramakrishnan put forward the following transformation: for the generality of any $N \times N$ image $f(x, y)$, let the origin of coordinates be the center of image, so there is $-N/2 \leq x, y \leq N/2$ [5-6]. For any pixel (x, y) , 2 parameters are introduced, i.e. ρ and σ , which are exclusively corresponding to the pixel defined as follows:

$$\rho = \max(|x|, |y|) \quad (25)$$

If $|x| = \rho$, there is:

$$\sigma = 2(\rho - x)y / |y| + xy / \rho \quad (26)$$

If $|y| = \rho$, there is:

$$\sigma = 2y - xy / \rho \quad (27)$$

It is easy to see that ρ ranges from 1 to $N/2$ and σ ranges from 1 to 8ρ . The corresponding polar coordinates are defined based on the parameters ρ and σ as follows:

$$r = 2\rho / N, \theta = (\pi\sigma) / (4\rho) \quad (28)$$

After this conversion, the expressions for real and imaginary numbers of Zernike moment A_{nm} (complex number) are obtained as follows:

$$C_{nm} = \frac{2n+2}{N^2} \sum_{\rho=1}^{N/2} R_{nm}(2\rho/N) \sum_{\sigma=1}^{8\rho} \cos \frac{\pi m \sigma}{4\rho} f(\rho, \sigma) \quad (29)$$

$$S_{nm} = -\frac{2n+2}{N^2} \sum_{\rho=1}^{N/2} R_{nm}(2\rho/N) \sum_{\sigma=1}^{8\rho} \sin \frac{\pi m \sigma}{4\rho} f(\rho, \sigma) \quad (30)$$

Zernike moment is a shape descriptor based on image zone, and its base is orthogonal radial polynomial, so it is rotation invariant, easy to construct higher-order moment and lowly redundant. Additionally, transformation is conducted based on the translation and scale invariance of geometric moments, so Zernike moments can be also translation and scale invariant. The specific procedure is as follows:

(1) Target image is partitioned from the original image to find the minimum bounding rectangle in the zone where the target belongs to. Let the rectangle have the length a and the breadth b , the center of gravity of image is translated to the origin, and the image is scaled down by $2/\sqrt{a^2+b^2}$ times and placed inside the unit circle.

(2) Calculate the area of target in the image, e.g. zero-order geometric moment of target image:

$$m_{00} = \iint f(x, y) dx dy \quad (32)$$

(3) Calculate the Zernike moment in each order inside the unit circle:

$$A_{nm} = \frac{n+1}{\pi} \sum_x \sum_y f(x, y) V_{nm}^*(\rho, \theta) \quad (33)$$

(4) Normalize the Zernike moments, and take the modulus value $|A_{nm}|$ of Zernike moments as the eigenvector of image retrieval:

$$|A_{nm}|' = \frac{|A_{nm}|}{m_{00}} \quad (34)$$

Process and Analysis of Human Shape Recognition

Human shape recognition involves video frame image extraction, frame image preprocessing, moving target invariant moment extraction, classification and recognition.

(1) Video image frame extraction employs the method based on sampling, whose interval is determined as the case may be. The sampling interval is analyzed with the video used in this paper to draw the following conclusions:

① Taking the speed of moving target as the reference, the sampling interval should be as short as possible to prevent any omission. This paper uses the video of about 30s with the frame rate of 18FPS. In the video, the moving object is a human that appears for about 10s. The video has the indoor stairway as the environment. Generally speaking, a human should walk normally in this environment, at the speed of 1.1~1.4m/s. In other words, the time needed by a human to complete a stride is in second. Hence, from the perspective of human's speed of walk, when the sampling interval is 1s, omission can be basically avoided.

② To prevent frame redundancy, sampling interval should be longer than the time needed for subsequent frame processing. The test shows that the time needed for subsequent processing of frame images in the video ranges 423~497ms, so sampling interval should be larger than 497ms to avoid frame redundancy.

Above all, sampling interval is taken to be 500ms, in order to avoid omission and frame redundancy. For the video at 18FPS in this paper, two frames each second are taken as the key frames.

(2) Frame image preprocessing includes several steps, such as, background removal, image segmentation, denoising, and contour extraction.



(1) Image with moving object (b) Background image (c) Image after differentiation

Fig. 1 Background Differentiation

① This paper removes the background based on the background differentiation. After differentiating the backgrounds between frame image containing moving object and background image to highlight the moving object in the image as shown in Fig. 1.

② Maximum between-cluster variance is an approach of self-adaptive threshold determination put forth by a Japanese scholar Otsu, so it is also known as Otsu's method, abbreviated as OTSU. The moving target image obtained after background differentiation receives the threshold segmentation with OTSU to gain the binary foreground image of moving object, as shown in Fig. 2.



Fig. 2 Foreground Image of Moving Target after Threshold Segmentation with OTSU

③ Considering the noise distribution of foreground image obtained after threshold segmentation with OTSU, this paper employs median filtering method to select the 5×5 medium filter window, as presented in Fig. 3.



Fig. 3 Image Obtained after Processing with Median Filtering Method

④ Roberts edge operator makes use of local difference operator to find the edge, so it achieves the high edge locating accuracy and the best response to low-noise images with sharp edge. Roberts edge operator is employed to extract the contour of image after denoising, as presented in Fig. 4.



Fig. 4 Contour Extraction with Roberts Operator

(3) For different invariant moments to be extracted, different preprocessed images are utilized.

① Hu invariant moment extracted in this paper for the moving object in the image is boundary moment, so the output image is the contour image of moving target after preprocessing. Hu invariant moment is structurally uncomplicated, but higher order of central moments for constructing Hu invariant moment will lead to worse anti-interference of Hu invariant moment, and significantly increases the complexity of operation. Hence, seven Hu invariant moments are all constructed by central moments of less than 3 orders. Fig. 5 presents the flow chart of extracting Hu invariant moments.

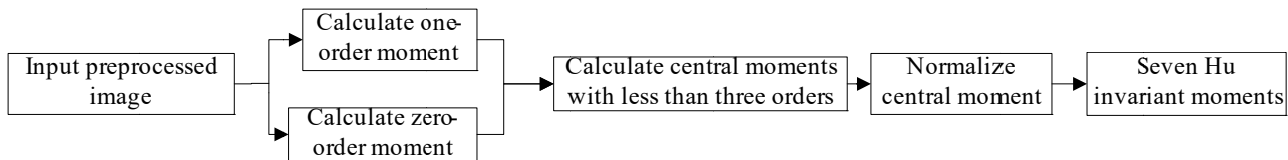


Fig. 5 Flow Chart of Extracting Hu Invariant Moments

② Based on the definition of Zernike invariant moment, it is very easy to obtain the Zernike invariant moments of target image. On this basis, transformation is carried out correspondingly to keep the favorable rotation invariance and add the scale and translation invariance. Additionally, it must be noted that Zernike invariant moments are zone moments, so the input image is a zone image of moving target after preprocessing. Fig. 6 presents the flow chart of extracting Zernike invariant moments.

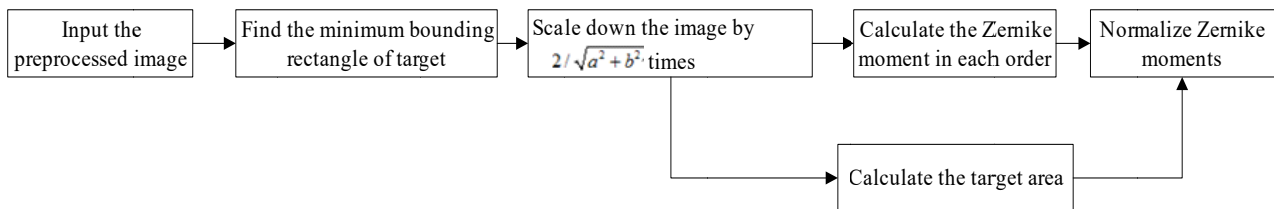


Fig. 6 Flow Chart of Extracting Zernike Invariant Moments

(4) Minimum distance classifier is employed to classify a sample. It is assumed that the sample can be classified into the category m , the key step of designing the minimum distance classifier is to calculate the minimum distance from the i^{th} attribute x_i of the element in the category k to the central value $\overline{x_{ki}}$ of the attribute. This paper chooses Euclidean distance pattern to calculate the minimum distance with Equation (39).

$$d_{ki}^2 = (x_i - \overline{x_{ki}})^2 \quad (39)$$

It is assumed that there is data tuple X , its element attribute is $x_i (i=1,2,\dots,n)$, k stands for category, so the central value of each category is $\overline{x_{ki}}$.

Training procedure:

(1) All data tuples in the training set are divided into m sets based on their category.

(2) The central value $\overline{x_{ki}}$ of each attribute representing each category is generated for m sets.

(3) Calculate the variance of each attribute for each category $\sigma_{ki}(k = 1, 2, \dots, m; i = 1, 2, \dots, n)$, that

is, d_{ki}^2 .

The classification steps include:

(1) For each data tuple X to be classified, calculate its distance from $\overline{x_{ki}}(i = 1, 2, \dots, n)$, i.e.

$d_k^2(k = 1, 2, \dots, m)$.

(2) Determine which closest category X belongs to.

The specific calculation is as follows:

All attributes in data tuples presented in this paper are digital, so there are the following calculations:

(1) Calculation of $\overline{x_{ki}}$

The mean value of the i^{th} attribute in the category k is the arithmetic mean value of the attribute in this category.

(2) Calculation of d_k^2

$$d_k^2 = \sum_{i=1}^n (x_i - \overline{x_{ki}})^2 \quad (40)$$

This paper involves 268 samples, including 201 human shape samples and 67 animal shape samples. Human shape samples include 67 images for three postures each, i.e. stand, stoop and squat, while animal shape samples involve 67 images for two commonest animals each, i.e. cat and dog. The Hu invariant moments and Zernike invariant moments are extracted from all these samples, and then classified with minimum distance classifier. The results are shown in Table 1 and Table 2.

Table 1 Classification of Hu Invariant Moments in Samples with Minimum Classifier

Category	Actual Number	Number of Accurate Recognition	Accuracy
Human (stand)	67	65	97%
Human (stoop)	67	60	89.6%
Human (squat)	67	56	83.6%
Animal	67	60	89.6%

Table 2 Classification of Zernike Invariant Moments in Samples with Minimum Classifier

Category	Actual Number	Number of Accurate Recognition	Accuracy
Human (stand)	67	66	98.5%
Human (stoop)	67	62	92.5%
Human (squat)	67	60	89.6%
Animal	67	65	97%

As revealed in the sample test, human shape recognition achieves better results based on Zernike invariant moments than Hu invariant moments. Obviously, Zernike invariant moments have better clustering effect than Hu invariant moments with regard to human shape image.

Conclusion

This paper introduces the comparative analysis of Hu moments and Zernike moments, and employs the minimum distance classifier to classify and recognize such two types of moments for moving target. These moments have translation, rotation and scale invariance, but they are still different from each other in some aspects. Hu moments do not contain comprehensive information on the target, but have more redundant information. Zernike moments are orthogonal, able to construct any higher-order moment, containing comprehensive information and insensitive to noise. This paper provides some basis for selection of invariant moments for moving target in the human shape recognition algorithm, and utilize the classification results of invariant moments to prove the feasibility of invariable moments in human shape recognition.

Acknowledgement

This research was financially supported by The Project of The 2016 middle-aged and young teachers in colleges and universities in guangxi basic ability promotion project of Based on intelligent video monitoring research of human abnormal behavior detection at ATM (No.KY2016YB651). Corresponding author: Jinchuan Lu. E-mail: lujinchuan99@163.com

References

- [1]. HU M K. Visual Pattern Recognition by Moment Inv.ariants. IEEE Trans on Information Theory, 1962, (8): 179~187.
- [2]. Kulkarni S A, Borde P L, Manza R R, et al. Offline Handwritten MODI Character Recognition Using HU, Zernike Moments and Zoning. Computer Science, 2014.
- [3]. Guruprasad P, Majumdar J. Achieving Premier Invariance to Scale and Rotation for Nandinagari Character Recognition by Comparing Multi Moment Features. Procedia Computer Science, 2016, 93:564-570.
- [4]. Oluleye B, Leisa A, Leng J, et al. Zernike Moments and Genetic Algorithm: Tutorial and Application. British Journal of Mathematics & Computer Science, 2014, 4(15):2217-2236.
- [5]. Sadanand A K, Prashant L B, Ramesh R M, et al. Impact of zoning on Zernike moments for handwritten MODI character recognition. IEEE International Conference on Computer Communication and Control. 2015.
- [6]. Wang T, Liao S. Chinese character recognition by Zernike moments. International Conference on Audio, Language and Image Processing. IEEE, 2014:771-774.