2011 International Conference on Mechatronic Science, Electric Engineering and Computer
August 19-22, 2011, Jilin, China

An Effective Image Shape Feature Detection and Description Method

Xiuxin Chen, Shiang Wei, Suxia Xing
College of Computer and Information Engineering
Beijing Technology and Business University
Beijing, China
cxx1305@sina.com

Kebin Jia
College of Electronic Information & Control Engineering
Beijing University of Technology
Beijing, China

***Abstract*—To solve the problems of complexity and low robustness that exist in present image shape feature detection and description methods, a new image shape feature detection and description method is proposed. First, salient regions are detected using a method based on connected component. Then, a descriptor composed of Hu's moment invariants and eccentricity is used to descript the detected regions. Experiments show its effectiveness and robustness to translation, rotation and scaling.**

Keywords- salient region detection; Hu's moment invariants; shape feature exaction

I. INTRODUCTION

The detection and description of image content features is very important in content based image/video retrieval, object tracking, wide baseline stereo image matching and image registration. The content features in an image includes color feature, texture feature and shape feature in which shape feature is an important one. The extraction of image shape feature includes shape feature detection and its description.

Some shape feature extraction methods have been proposed. In shape feature detection, Matas[1] introduced Maximally Stable Extremal Regions(MSER) method and Tuytelaars[2] introduced a method based on edges and Kadir[3] introduce a salient region detector. Mikolajczyk[4] compared 6 commonly used region detectors and concluded that MSER is better than others under most circumstances. In shape feature description, moment invariant is an important method and has been widely used. Hu[5] introduced 7 moment invariants which have been used in many applications.

In this paper, a new image shape detection and description method is proposed. First, shapes are detected using a method based on connected component. Then, Hu's moment invariants and eccentricity are used to descript the detected shapes. Experiments show that the proposed method is robust to translation, rotation and scaling.

II. IMAGE SHAPE FEATURE DETECTION

The shape feature detection method is designed for gray images. If the original image is a color image, it is first transformed to a gray one. The shape feature detection process is as follows: first, connected components of the same gray value are detected. Then, neighbor connected components whose pixel value difference is less than a threshold *delta* are merged and the relative change rate of the shape's area is computed. All neighbor regions are processed in this way. The ultimate merging result regions are candidates of shapes we want. To improve performance, image histogram equalization is performed before shape detection.

A. Shape Feature Detection and Its Realization

The tree structure is used in the detection of shapes. First, all pixels in the image are sorted in ascending order using bin sorting method. Then, each pixel in the image is initialized to a connected component tree. All pixels in the image form a forest. The data structure of the component tree is defined as:

```
struct Reg
{
  unsigned int parent;   /*parent tree index             */
  unsigned int height;  /*the height of the component */
  unsigned int area; /*the area of the connected component*/
};
```

After initialization, the parent of each component tree refers to itself and the height and area are both 1. Then, starting from the pixel with minimum value, neighbor pixels with the same value are merged together. In the realization of merging, the UNION method of Union-Find Set is used. To avoid the degenerated-tree, different merging methods are used according to their height. If A and B are two neighbor regions with the same value, the merging process is as follows:

1. If the root of A and B is the same one, it means that they have been merged to a tree and nothing will be done.

2. If the roots of A and B are different and they have the same pixel value, merge them according to the following process:

(1) If $A.height>B.height$, A becomes parent of B;

(2) Else, B becomes parent of A.

3. If the roots of A and B are different and their values are different too, the one that has larger value becomes parent of the other one and forms another connected region.

In this way, after merging all pixels in the image form a connected-region tree and connected regions with larger values is in the upper part of the tree and connected regions with smaller values are in the lower part of the tree.

To easy the following process, another data structure is defined to save connected regions:

```
struct StableReg
```

978-1-61284-722-1/11/$26.00 ©2011 IEEE

```
{
 int     root;   /*root of the connected region           */
 int     num;    /**number of the connected region    */
 byte  value;    /*pixel value of the connected region*/
 unsigned int  area;  /* area of the connected region */
 unsigned int max_stable;/*whether is a stable region*/
} ;
```

B. Decision of Stable Regions

Neighbor regions whose pixel value difference is little than the threshold *delta* are merged. In this paper, the value of *delta* is 5. And also the area change rate after merging is computed.

$$\mathrm{var}iation = \frac{R_\Delta}{R}$$

Where R_Δ is the area changed after merging and *R* is the smaller area of the two merged regions. The regions whose area change rate after merging is less than a threshold are considered as the candidates of stable regions. In this paper, the threshold is 0.1.

C. Eliminating of Bad Regions.

Regions detected through above method usually include background regions and noise regions which are not what we want, so they must be eliminated.

Background regions usually have large areas, so they can be eliminated according to their areas. In this paper, regions whose areas are larger than 35% of the image's total area are considered background. Noise regions are normally small, so we consider regions with less than 20 pixels as noise areas and eliminate them. Moreover, repeated regions are likely to occur in the detection process. If a region's area is almost the same as its parent's, we consider them as repeated regions and reserve parent region only.

III. THE DESCRIPTION OF REGION FEATURES

A. Hu Moment Invariants

Image Moments. For a 2 dimensional function $f(x,y) \in L(R^2)$ which is defined in *o-xy*, its (*p*+*q*) order origin moment is defined as follows:

$$m_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} x^p y^q f(x,y) \tag{1}$$

Since origin moment is not translation invariant, central moment which is translation invariant is defined:

$$\mu_{pq} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (x-\bar{x})^p (y-\bar{y})^q f(x,y) \tag{2}$$

Where $\bar{x} = m_{10}/m_{00}$ and $\bar{y} = m_{01}/m_{00}$.

For a discrete image $f(x,y)$, its (*p*+*q*) order origin moment and central moment are defined as follows:

$$m_{pq} = \sum_x \sum_y x^p y^q f(x,y) \tag{3}$$

$$\mu_{pq} = \sum_x \sum_y (x-\bar{x})^p (y-\bar{y})^q f(x,y) \tag{4}$$

Wher $\bar{x} = m_{10}/m_{00}$ and $\bar{y} = m_{01}/m_{00}$. Normalized central moment is defined as:

$$\eta_{pq} = \frac{\mu_{pq}}{\mu_{00}^r}$$

Where $r = \frac{p+q}{2} + 1$.

Hu[5] constructed 7 moment invariants using tow order and three order normalized central moments which are invariant to translation, rotation and scaling. They are:

$$\phi_1 = \eta_{02} + \eta_{20}$$

$$\phi_2 = (\eta_{20} - \eta_{02})^2 + 4\eta_{11}^2$$

$$\phi_3 = (\eta_{30} - 3\eta_{12})^2 + 3(\eta_{21} - \eta_{03})^2$$

$$\phi_4 = (\eta_{30} + \eta_{12})^2 + (\eta_{21} + \eta_{03})^2$$

$$\begin{aligned}\phi_5 = {} & (\eta_{30} - 3\eta_{12})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 \\ & - 3(\eta_{21} + \eta_{30})^2] + (3\eta_{21} - \eta_{03})(\eta_{21} + \eta_{03}) \\ & [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]\end{aligned}$$

$$\begin{aligned}\phi_6 = {} & (\eta_{20} - \eta_{02})[(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{30})^2] \\ & + 4\eta_{11}(\eta_{30} + \eta_{12})(\eta_{21} + \eta_{03})\end{aligned}$$

$$\begin{aligned}\phi_7 = {} & (3\eta_{21} - \eta_{03})(\eta_{30} + \eta_{12})[(\eta_{30} + \eta_{12})^2 \\ & - 3(\eta_{21} + \eta_{30})^2] + (3\eta_{12} - \eta_{30})(\eta_{21} + \eta_{30}) \\ & [3(\eta_{30} + \eta_{12})^2 - (\eta_{21} + \eta_{03})^2]\end{aligned}$$

Discrete Hu moments can be constructed in the same way. Experiments show that the first 3 Hu moments are good enough in description of region features, so they are adopted in this paper.

B. Eccentricity

Eccentricity is defined as follows:

$$E = \frac{(\eta_{20} - \eta_{02})^2 + 4\eta_{11}{}^2}{(\eta_{20} + \eta_{02})^2} \tag{5}$$

Eccentricity is invariant to translation, rotation and scaling.

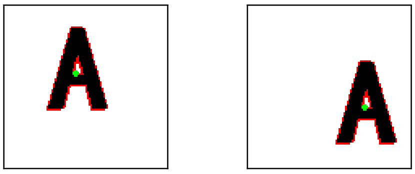
This paper uses $F = (\phi_1, \phi_2, \phi_3, E)$ as the descriptor of the regions detected.

IV. EXPERIMENTS

Experiments are performed on two groups of images. Each group includes 10 original images. The first group is composed of images of upper letters of A to J and the second group is composed of images of various types of symbols. Each original image in two groups is translated, rotated and scaled to get 5 transformed images. Altogether 120 images are used to test the method. In the experiment, ϕ_1, ϕ_2, ϕ_3 are multiplied by 1000.

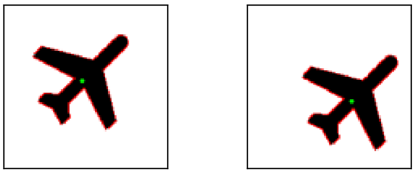
A. Translation Invariant Testing

Translate the original images to various directions and compare the region detection results and the descriptors of the regions detected. Fig. 1 shows the region detection results of two original images and the images after translation. The green dots in the figure are the centers of the regions detected and red edges are the edges of the regions.



(a1) The original image (a2) Image after translation

(a) An image in group one



(b1) The original image (b2) Image after translation

(b) An image in group two

Figure 1. Testing results of detector for translation

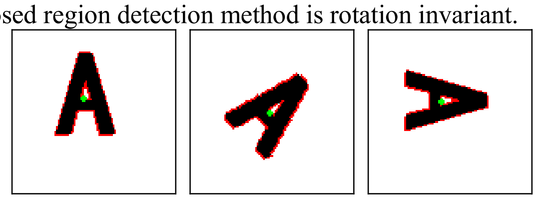
Fig.1 shows that the proposed detection method is robust to translation. Table 1 shows the values of descriptors for images in Fig.1. From table 1 we can see that the proposed descriptor is robust to translation.

TABLE I. VALUES OF DESCRIPTORS FOR IMAGES IN FIG.1

	image (a1)	image (a2)	image (b1)	image (b2)
ϕ_1	292.702	292.702	303.895	303.895
ϕ_2	14.917	14.917	0.125	0.125
ϕ_3	10.005	10.005	4.525	4.525
E	0.1741	0.1741	0.0014	0.0014

B. *Rotation Invariant Testing*

Rotate each original image 45 degrees and 90 degrees clockwise to get two rotated images. Fig.2 shows the region detection results. From fig.2 we can see that the proposed region detection method is rotation invariant.



(a1) The original image (a2) 45 degrees (a3) 90 degrees

(a) An image in group one

(b_1) The original image (b_2) 45 degrees (b_3) 90 degrees

(b) An image in group two

Figure 2. Testing results of detector for rotation

Table 2 shows the values of descriptors for images in Fig.2.

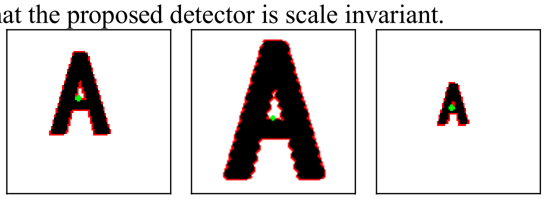
TABLE II. VALUES OF DESCRIPTORS FOR IMAGES IN FIG.2

	image (a1)	image (a2)	image (a3)	image (b1)	image (b2)	image (b3)
ϕ_1	292.702	275.244	292.702	303.895	289.849	303.894
ϕ_2	14.917	12.890	14.917	0.125	0.197	0.125
ϕ_3	10.005	8.509	10.005	4.525	4.150	4.525
E	0.1741	0.1701	0.1741	0.0014	0.0023	0.0013

Table 2 shows that after rotation the values of descriptors change a little. Experiments show that when the image is rotated 90 degrees, 180 degrees or 270 degrees, the values of descriptors remain unchanged. When the image is rotated 45 degrees, 135 degrees or 315 degrees the values of descriptors change the most and the amount remains very little. We can see that the proposed descriptor is rotation invariant.

C. *Scaling Invariant Testing*

Scale the original image to its double size and half size respectively to get two scaled images. Fig. 3 shows the detection results of proposed detector. From fig.3 we can see that the proposed detector is scale invariant.



(a1) The original image (a2) Double size (a3) Half size

(a) An image in group one

(b1) The original image (b2) Double size (b3) Half size

(b) An image in group two

Figure 3. Testing results of detector for scaling

Table 3 shows the values of descriptors for images in Fig.3.

TABLE III. VALUES OF DESCRIPTORS FOR IMAGES IN FIG.3

	image (a1)	image (a2)	image (a3)	image (b1)	image (b2)	image (b3)
ϕ_1	292.702	279.844	259.152	303.895	291.063	283.900
ϕ_2	14.917	13.640	11.251	0.125	0.135	0.153
ϕ_3	10.005	8.758	6.646	4.525	4.024	3.914
E	0.1741	0.1742	0.1675	0.0014	0.0016	0.0018

Table 3 shows that the proposed descriptor is robust to scaling.

The experiments are performed under computer of Intel Core 2 Duo CPU, 2.4GHz, 2G RAM and with operating system of Windows XP. The algorithm is realized using Microsoft Visual C++ 6.0. The detection and description time of the regions is relevant to the size of images. For images of 128*128, the average time is 0.8 second which is very fast.

V. CONCLUSION

An effective region detector and descriptor is proposed in this paper. The detector is based on connected component and the descriptor is composed of Hu's moment invariants and eccentricity. Experiments show that the method is robust to translation, rotation and scaling and can be used in content based image/video retrieval and object tracking.

ACKNOWLEDGMENT

This paper is sponsored by the National Natural Science Foundation of China under Grant No. 30970780 and Beijing Municipal Education Commission research projects under Grant No. KW201010011002.

REFERENCES

[1] Matas, J. Chum, O., Urban, M., and Pajdla, T. "Robust wide-baseline stereo from maximally stable extremal regions," Proceedings of the British Machine Vision Conference, Cardiff, UK, 2002: 384–393.

[2] T. Tuytelaars, L. Van Gool. "Matching widely separated view based on affine invariant regions," International Journal of Computer Vision, 2004, 59(1):61-85.

[3] T. Kadir, M. Brady, A. Zisserman. "An affine invariant salient region detector," Proceedings of Eighth European Conference on Computer Vision, 2004:345-457.

[4] K. Mikolajczyk, T. Tuytelaars, C. Schmid et al. "A comparison of affine region detectors," International Journal of Computer Vision, 2005,65(1/2): 43-72.

[5] Hu M K. "Visual pattern recognition by moment invariants," IRE Transactions on Information Theory, 1962,IT-8:179-187.