http://ansinet.com/itj

ITJ

ISSN 1812-5638

INFORMATION TECHNOLOGY JOURNAL

ANSInet
Asian Network for Scientific Information
308 Lasani Town, Sargodha Road, Faisalabad - Pakistan

Information Technology Journal 12 (12): 2342-2349, 2013
ISSN 1812-5638 / DOI: 10.3923/itj.2013.2342.2349

2-D Cartoon Character Detection based on Scalable-Shape Context and Hough Voting

[1]Tiejun Zhang, [1]Qi Han, [1,2]Ahmed A. Abd El-Latif, [3]Xuefeng Bai and [1]Xiamu Niu
[1]School of Computer Science and Technology, Harbin Institute of Technology, 150080 Harbin, China
[2]Department of Mathematics, Faculty of Science, Menoufia University, Shebin El-Koom 32511, Egypt
[3]Harbin Institute of Technology, Shenzhen Graduate School, 518055 Guangdong, China

Abstract: Cartoon pirate uploading is a very serious problem for the image and video-sharing website. In this study, we propose a new method to detect the characters in 2D-cartoon images, aiming at rejecting pirate uploading automatically. We extract the curve in the cartoon image as the main content and then design a local shape feature named Scalable-Shape Context (SSC) to present the local shape of cartoon. Firstly, we use the Harris-Laplace corner detector to find the key points at multi-scale in the cartoon image, most of which are localized at the junctions of curves. Secondly, the scale of each key point is used as a reference scale for Shape Context (SC) to describe the curvilinear structure around the key points. Then, the matching problem between the key points extracted from the input model and testing image is solved as an optimal assignment problem. Finally, a Hough-voting scheme is employed to find the location of the character in the testing image. The experimental results show that the proposed SSC-based detection method is effective in the detection of 2D-cartoon characters.

Key words: Cartoon character detection, scalable-shape context, hough-voting

INTRODUCTION

Cartoon is well-liked by both children and adults for its comic characters and the funny drawing style. In history, lots of classic cartoon characters had been produced by artists and shown to readers which formed a big industry. The economic value of cartoon attracts both the cartoon fans and pirate, who may get and upload the cartoon media copy to sharing website to make a profit. Some pirates sometimes even re-edit the cartoon for advertisement, create new cartoon product based on the famous cartoon characters and publish as their own works.

A method to solve the pirate problem of cartoon is to block pirate's way to the sharing website, so that the sharing website can be protected from legal liability for spreading pirate cartoons. Therefore, it is necessary for the sharing web sit to detect and then reject the cartoons with copyright statement.

The chief value of cartoons is the characters in them and the characters are also the most frequently pirated content. To the best of our knowledge, there is no previous proper method for detection the characters in the 2-D cartoon image. The central idea of the general object detection problem lies in finding matches between object features and the image features (Mikolajczyk and Schmid, 2005). Global features usually contain too much noise because the character usually localize in part of the whole picture (Szeliski, 2010; Deselaers *et al*., 2008). Plenty of local patch based feature have been proposed and successfully applied in object detection, such as Scale-Invariant Feature Transform (SIFT) (Lowe, 1999), Speeded Up Robust Features (SURF) (Bay *et al*., 2008), Maximally Stable Extremal Regions (MSER) (Matas *et al*., 2002) and scale and affine invariant interest point (Mikolajczyk and Schmid, 2004). These features are, to some extent, invariant to illumination, projective transformations and other common object variations. Most of the features work well in the natural images, producing abundant features with scale or rotation invariance; however, they are inadequate for the cartoon character detection. Compared to the natural image, 2D cartoon have simpler, more artificial contents and is composed of regions with simple coloring and explicit curves to separate different regions, as shown in Fig. 1. There is little texture information to localize the key points and descript them using the gradient based local feature, especially in the flash cartoon. The classic local features mentioned above will not perform well for the cartoon images.

Corresponding Author: Tiejun Zhang, School of Computer Science and Technology, Harbin Institute of Technology, 150080 Harbin, China

In the cartoon image, curve is the essential element to present the content information, such as shape, motion and expression. It provides human with most of the information about the cartoon and can be extracted fast and accurately (Zhang *et al*., 2009, 2013). Modeling the curve information as the feature coincides with the human instinct. Shape Context (SC), proposed by Belongie *et al*. (2002), is one of the most widely used features for curvilinear structure description. SC allows for measuring shape similarity between curvilinear structures and is used in digit recognition, silhouette similarity-based retrieval and 3D object recognition.

Fig. 1: Classic 2D cartoon image from a cartoon TV series (the main content is delivered by explicit curve and the large uniform regions)

The basic idea of SC is to model the distribution of other curve pixels relative to the selected pixel. Diagram of log-polar histogram bins is used in computing the shape context, 12 equal divided bins for the relative angle and five exponential increasing radiuses for the relative distance as shown in the Fig. 2c. The amount of pixels falling into each bin forms the SC descriptor of the reference pixel. Taking the diagram into a 2-D axis with θ as the horizontal axis and log(r) as vertical axis, the feature extracted in Fig. 2a at the red square and b at the triangle can be shown as Fig. 2d and e. For the shape detection, SC provides a compact, rotation invariant, yet highly discriminating descriptor but it is very sensitive to the changing scale. If the radius of the diagram changes, the descriptor will change, as shown in the Fig. 2f.

Thought extracted from different handwriting, d-e are quite similar to each other. Due to the changing in scale, d-f are quite different, thought extracted from the same location on the same handwriting.

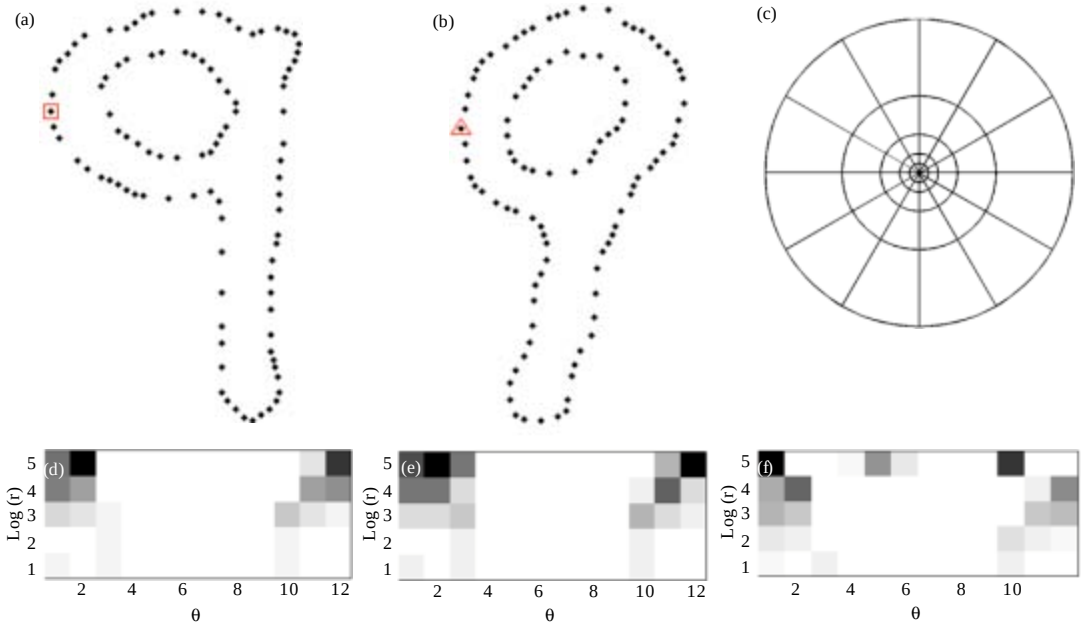

Fig. 2(a-f): SC feature extraction and the sensitivity to the scale changing, (a) Outline of a handwriting of digit 9 by some person, (b) Outline of handwriting of digit 9 by another person, (c) Bin partition using the diagram of log-polar histogram bins, (d) Feature extracted on the red square in (a) at scale σ_t, (e) The feature extracted on the red triangle in (b) at scale σ_t and (f) The feature extracted on the red square in (a) at scale $2\sigma_t$

In this study, we propose a new local invariant features, namely Scalable-Shape Context (SSC), based on the Harris-Laplace corner detector (Mikolajczyk and Schmid, 2004) and the Shape Context descriptor for the cartoon character detection. To solve the scale sensitive problem of Shape Context, we first use the Harris-Laplace corner detector to localize the key points and corresponding scale in the cartoon image. Then, the scale of each key point is used as a reference scale by Shape Context to describe the curvilinear structure around the key points, so that the feature can be extracted at a consistent scale among images. Given the features, we formulate the matching between the cartoon character and the testing cartoon image as a weighted bigraph matching problem and then find the optimal matching. Then a Hough voting scheme is introduced to localize the character. The experimental results show the effectiveness of the proposed SSC-based detection method.

SCALABLE SHAPE CONTEXT (SSC) BASED OBJECT DETECTION

The cartoon character detection method is shown in Fig. 3.

We first extracted the SSC features and match the features extracted from the cartoon character and the testing image. Then the Hough voting rules are learned from the cartoon image. Finally, the matched features in the testing image vote for the center and the scale of the character based on the voting rule.

Scale shape context

Multi-scale harris corner detection: Multi-scale Harris corners are detected by the scale-adapted Harris function and selected in scale-space by the Laplacian-of-Gaussian operator, also called Harris-Laplace corner. It shows high repeatability and localization accuracy inherited from the Harris detector. The Harris detector which is one of the most reliable interest point detectors, is based on the second moment matrix (the auto-correlation matrix). To detect corners in multi-scale, this matrix must be adapted to scale changes to make it independent of the image resolution. The scale-adapted second moment matrix can be defined by Eq. 1 (Mikolajczyk and Schmid, 2004):

$$\mu(x,\sigma_I,\sigma_D)=\begin{bmatrix}\mu_{11} & \mu_{12}\\ \mu_{21} & \mu_{22}\end{bmatrix} = \sigma_D^2 g(\sigma_I)\times\begin{bmatrix}L_x^2(x,\sigma_D) & L_x(x,\sigma_D)L_y(x,\sigma_D)\\ L_y(x,\sigma_D)L_x(x,\sigma_D) & L_y^2(x,\sigma_D)\end{bmatrix} \quad (1)$$

With:

$$L_x(x,\sigma_D)=\frac{\partial}{\partial x}g(\sigma_D)\times I(x) \quad (2)$$

$$g(\sigma)=\frac{1}{2\pi\sigma^2}e^{\frac{x^2+y^2}{2\sigma^2}} \quad (3)$$

where, σ_I is the smooth filter scale and the σ_D is the differentiation scale. I(x) is the given image, x = (x, y), L_x, L_y is the derivative computed in the x and y direction at scale σ_D, respectively. The Matrix describes the gradient distribution in a local neighborhood of a point. The Harris corner of certain scale can be localized by finding the local maxima in the Harris measure (Harris and Stephens, 1988), as shown in Eq. 4:

$$R = \det(\mu(x, \sigma_I, \sigma_D)) - \alpha \operatorname{trace}^2(\mu(x, \sigma_I, \sigma_D)) \quad (4)$$

Then, we verify for each of the scales if the local maxima of certain scale is also the maxima in the scale-space. Harris corners of different scale may shift more or less while the scale changing, so a search around the maxima of R is needed. This search is time consuming.

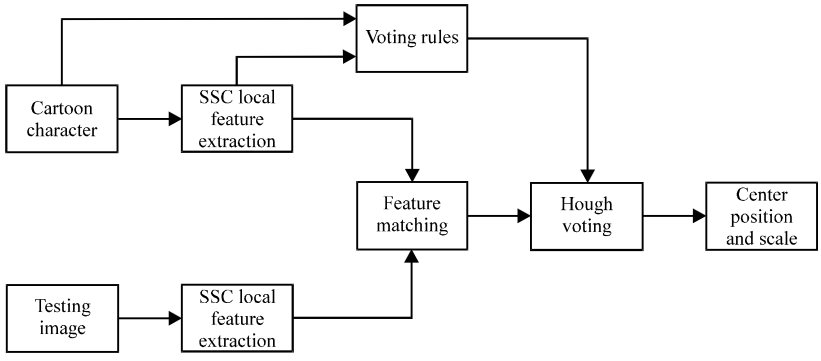

Fig. 3: Scalable Shape Context-based cartoon character detection

There is also a fast approximate solution based on the observation that the scale Harris corner arise together with the local maxima of the Laplacian-of-Gaussian over scale. The local maxima of the Laplacian-of-Gaussian are computed by Eq. 5 to verify the scale of the corner:

$$|LoG(x, \sigma_n)| = \sigma^2_n |L_{xx}(x, \sigma_n)+L_{yy}(x, \sigma_n)| \quad (5)$$

The cartoon corner detected using Harris-Laplace corner detector can be shown in the Fig. 4b. The center of the red circle is the location of the Harris-Laplace corner and the radius r is the scale at which Harris measure R reaches its maxima over scales.

Descript using the shape context: Once given the key point location and the corresponding scale, we use shape context to describe the local feature. We extract all the curves, including edge and decoration lines in the cartoon image, using the method by Zhang *et al.* (2013). Afterwards, we use the SC to describe the key points.

There are several differences between the Shape Context and our SSC. Firstly, our method uses the Harris Corner, rather than the curve pixel, as the reference point which is more stable and much less than the curve pixel. Secondly, we use the Harris Corner scale as reference scale, rather than the size of the image as in Shape Context.

Consider a reference point (a Harris corner) as p_i, $i\in(1, 2...m)$ and the curve pixels as $pixel_l$, $l\in(1, 2...M)$. The Harris corner's scale is s, we use bins that are uniform in log-polar[2] space, with 12 bins for the angular direction and 5 bins for the polar direction (Fig. 2c), with radius $s\times2^t$, $t\in(0, 1...4)$, as shown in Fig. 5b. s is relative stable over

Fig. 4(a-b): (a) Cartoon image for testing and (b) Red circles are the Harris corners detected by the Harris-Laplace corner detector on the cartoon image

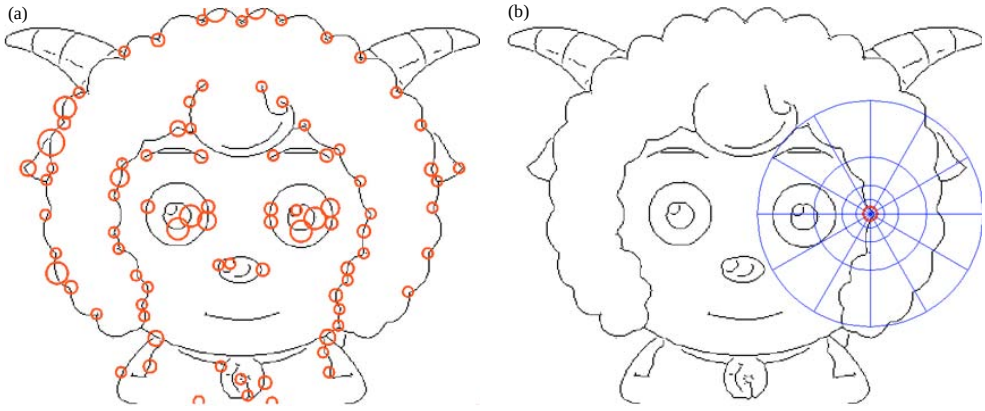

Fig. 5(a-b): (a) Curve and the corner (centered at red circle) extracted from Fig. 1 and (b) SSC diagram locating at one of the corners. The radius of the inner circle diagram is equal to the corner's scale and radius of the ex-circle is 2^4 times of the scale

images, so that shape context can be computed in a stable scale to be invariant to the scale changing from image to image. We compute the histogram h_i as follows:

$$h_i(k) = \{\text{pixel} \neq p_i: (\text{pixel}-p_i) \in \text{bin}(k)\} \quad (6)$$

Let $C_{ij} = C(p_i, q_j)$ denote the matching cost between two feature points. We use the χ^2 test statistic to compute C^{ij}, as shown in Eq. 7:

$$C_{ij} = C(p_i, q_j) = \frac{1}{2}\sum_{k=1}^{K}\frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)} \quad (7)$$

where, $h_i(k)$ and $h_j(k)$ denote the K-bin normalized histogram at p_i and q_j, respectively. χ^2 test is a measure for the independence. If the features of p_i and q_j are different, C_{ij} will be relatively large and if the features of p_i and q_j are the same, C_{ij} will reach a minimum of 0.

Bipartite graph matching: Given the set of costs C_{ij}, between all pairs of feature points p_i from the cartoon character and q_j from the testing image, we want to find an optimal matching that minimizes the cost:

$$H(\pi) = \sum_i C(p_i, q_{\pi(i)}) \quad (8)$$

where, π is a permutation of j. This is a weighted bipartite matching problem which can be solved efficiently by Jonker and Volgenant (1987). While matching, parts of the feature points may not be matched because outliers or occupation, so dummy key points are added, for both the data balance and the anchoring the mismatched feature point. The match cost between dummy and the feature points are set to a low value, so that the outliers are tend to match with the dummy.

Hough voting for character detection: Here, we present the cartoon character model as a set of scalable shape context feature points. The matching between the character model and the testing image induces a translation and scale transformation, we let each match pair vote for the presence of a cartoon character at a certain image location and scale (Leibe *et al*., 2004; Maji and Malik, 2009). Let $_{if}$ denote the features at location l_i and S(O, x) denotes score that object O arise at location x:

$$\begin{aligned} S(O,x) &= \sum_j p(O,x,f_j,l_j) \\ &= \sum_j p(f_j,l_j)p(O,x \mid f_j,l_j) \end{aligned} \quad (9)$$

While the $_{if}$ and I_i is independent to each other and Eq. 9 is equal to Eq. 10:

$$\begin{aligned} S(O,x) &\propto \sum_j p(O,x \mid f_j,l_j) \\ &= \sum_{i,j} p(C_i \mid f_j,l_j)p(O,x \mid C_i,f_j,l_j) \end{aligned} \quad (10)$$

C_i is the match feature in the model. Because the matching has no relation with the position, so $p(C_i|f_j, l_j) = p(C_i|f_j)$. So the Eq. 10 can be written as Eq. 11:

$$\begin{aligned} S(O,x) &\propto \sum_{i,j} p(C_i \mid f_j)p(O,x \mid C_i,l_j) \\ &= \sum_{i,j} p(C_i \mid f_j)p(x \mid O,C_i,l_j)p(O \mid C_i,l_j) \end{aligned} \quad (11)$$

where, $p(C_i|f)$ is the feature matching probability, between the feature f from the testing image and C_i from the cartoon character model. $p(x|O, C_i, I_j)$ is the probability of the location if C_i is matched to a feature in location I_j, can be written as a. $p(O|C_i, I_j)$ is the probability that C_i for the existence of the object. In this study, we define $p(C_i|f)$ as Eq. 12:

$$p(C_i \mid f) = \begin{cases} \frac{1}{Z}\exp(-\gamma d(C_i,f)) & \text{if } d(C_i,f) \le t \\ 0 & \text{otherwise} \end{cases} \quad (12)$$

where, Z is the constant for $p(C_i|f)$ to make the integral equal to 1, t and γ are positive real number. Then Eq. 11 can be written as Eq. 13:

$$\begin{aligned} S(O,x) &\propto \sum_{i,j} p(C_i \mid f_j)\cdot p(x \mid O,C_i,l_j)\cdot p(O \mid C_i,l_j) \\ &= \sum_i \frac{1}{Z}\exp(-\gamma d(C_i,f))\cdot\frac{1}{\sqrt{2\pi}\alpha}\exp\left(-\frac{|x - l_{\pi(i)} - s_r P|^2}{2\alpha^2}\right)\cdot\frac{1}{N} \\ \text{where} \quad d(C_i,f) &= \begin{cases} C(p_i,p_{\pi(i)}) & \text{if } C(p_i,p_{\pi(i)}) \le t \\ 0 & \text{otherwise} \end{cases} \end{aligned} \quad (13)$$

where, s_r is the relative scale between testing image and the model and is computed by the relative ration of the matched features, P is the relative positive from center of the model to the feature C_i.

Local maxima in the voting space, S(O, x), give the center location of the object and the relative scale defines the corresponding window's size. The score is computed by the following equation:

$$S = \alpha S_{\text{Maxima}} + (1-\alpha)\frac{M}{m} \quad (14)$$

where, M is the number of the matched feature points which fall into the location window; m is the number of the model feature points. S_{maxima} stands for the local maxima of S(O, x). Herein, $\alpha = 0.9$. The more feature points vote for the maximum of the S(O, x), the bigger are the two terms of Eq. 14.

EXPERIMENTAL RESULTS

We extract the scaling shape context feature from the cartoon image and find the optimal matching between the cartoon character model and the testing images. A Hough-style voting scheme is used to locate the cartoon character in the testing image. The simulation is carried on a cartoon images database, containing three famous cartoon characters, such as doraemon, xiyangyang, snoopy etc., collected from cartoon TV series and internet. Each class of the character contains at least 40 images with different positions, angles or different backgrounds. The dataset also includes one character image as model for each class of character. During the detection, we use all images in the database as test set for each class which make for a large negative test set. It allows getting a reliable value for the incidence of false-positives generated by the detection algorithm. We counts a detection to be correct if the overlap of the detected and ground-truth bounding box is greater than 50%, according to the PASCAL criterion. Figure 6 shows the feature matching between the given model and one of the test image, as well as the Hough voting space generated in the voting. Figure 7 shows some of the detection result and the corresponding Hough space. It is clear that the maximum value in the Hough space indicate the existence of a character.

The results in Fig. 8 shows that our method gains a much higher detection rate than the SIFT and SC. SIFT

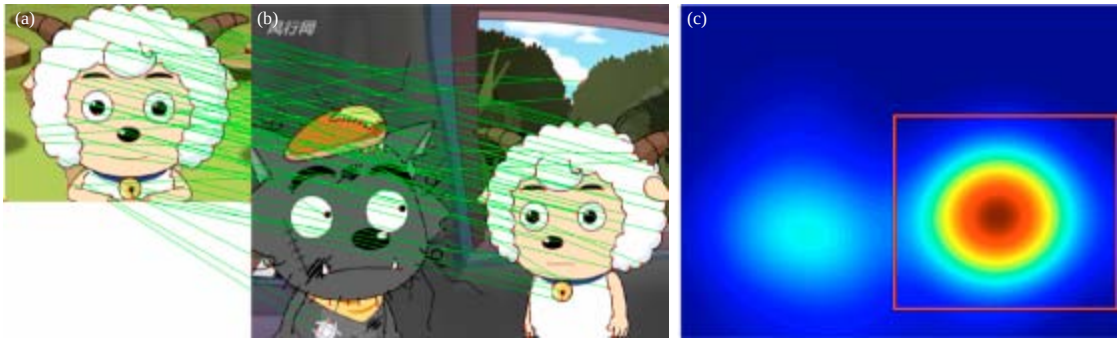

Fig. 6(a-c): (a) Matches between the model of Xiyangyang and the testing image, (b) Red points in are the feature location, the blue line connect matching points and (c) Hough voting space and the red square is the detected object window

Fig. 7(a-b): Result of detection the cartoon character Xiyangyang (a) Original images with the red square marking the detection results and (b) Hough voting space of each image in (a)

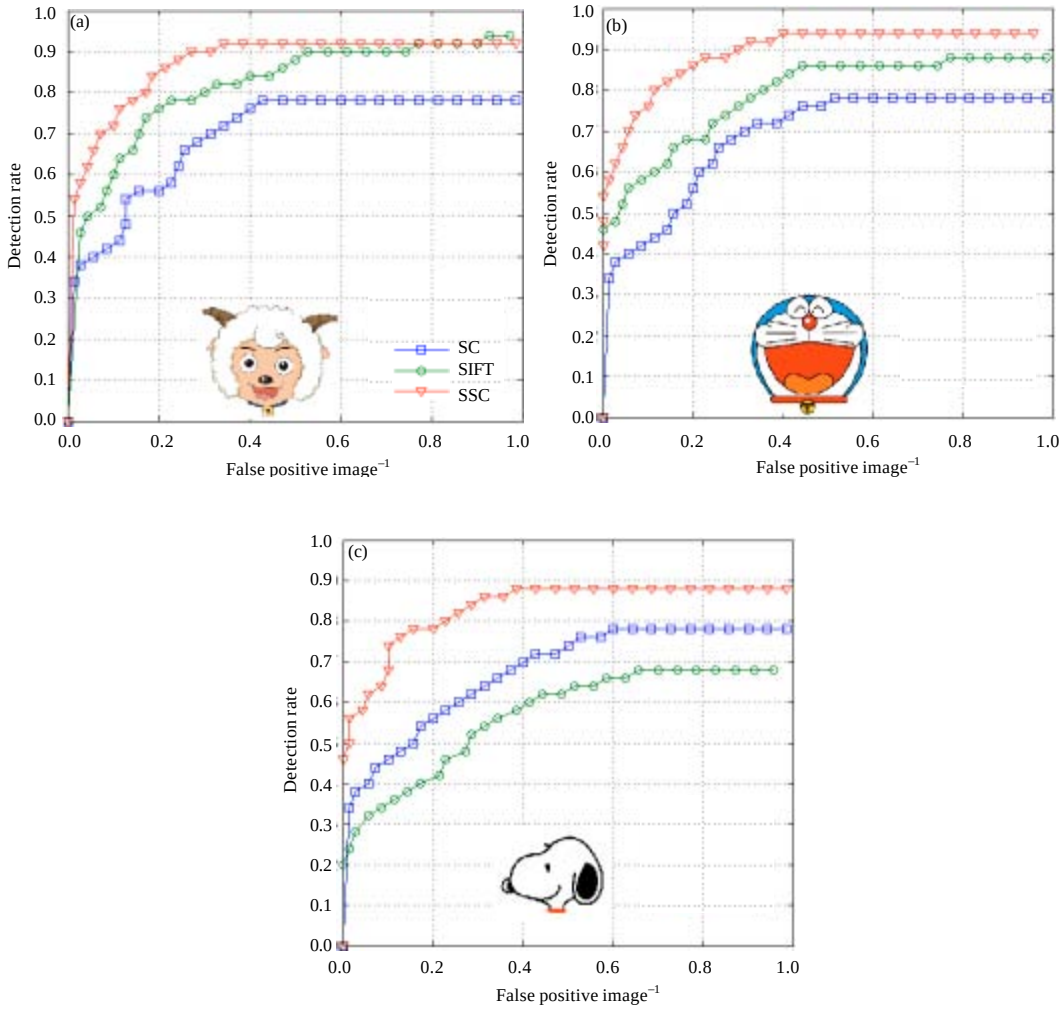

Fig. 8(a-c): Detection rate (DR) vs. False Positive per Image (FPPI) curve for evaluating the cartoon character detection (a) Detection rate vs. False positive for xiyangyang, (b) Detection rate vs. False positive for doraemon and (c) Detection rate vs. False positive for snoopy

Table 1: Comparison of detection rate over methods

	Xiyangyang	Doraemon	Snoopy
SSC	0.90/0.92	0.90/0.94	0.84/0.88
SC	0.68/0.76	0.70/0.74	0.64/0.70
SIFT	0.80/0.84	0.76/0.74	0.54/0.58

get a low detection rate is that the cartoon image is lack of texture and number of SIFT feature is much less than that in the natural images. The mismatching between different cartoon images is very serious. SC feature is sensitive to the scale changing, so that it performs well if the object scale is similar to the model but perform much worse if the scale is not agree.

Table 1 shows the numerical differences between features.

CONCLUSION

In this study, we proposed a local feature, named Scalable Shape Context (SSC), for 2-D cartoon characters detection based on the Multi-scale Harris Corner detector and Shape Context. The feature uses the Multi-scale Harris Corner as key point location and then describes the neighbor curve structure using Shape Context on the scale at which the Harris corner gets its maximum. The proposed SSC method has the following advantages (1) SSC is invariant to both rotation and scale-changing, (2) It reduces the number of the feature points, so as to accelerate the matching and (3) It is more likely to model meaningful curve structure and the relationship between

the curves, by locating corner-like structures. Experiments conducted show that the proposed SSC detection method is effective and gains better results compared to some other notable local features in the cartoon character database.

ACKNOWLEDGMENT

This study is supported by the National Natural Science Foundation of China (61100187), the Fundamental Research Funds for the Central Universities (Grant No. HIT. NSRIF. 2010046) and the China Postdoctoral Science Foundation (2011M500666).

REFERENCES

Bay, H., A. Essa, T. Tuytelaarsb and L.V. Goola, 2008. Speeded-Up Robust Features (SURF). Comput. Vision Image Understand., 110: 346-359.

Belongie, S., J. Malik and J. Puzicha, 2002. Shape matching and object recognition using shape contexts. IEEE Trans. Patt. Anal. Mach. Intell., 24: 509-522.

Deselaers, T., D. Keysers and H. Ney, 2008. Features for image retrieval: An experimental comparison. Inform. Retrieval, 11: 77-107.

Harris, C. and M. Stephens, 1988. A combined corner and edge detector. Proceedings of the 4th Alvey Vision Conference, September 2, 1988, Manchester, UK., pp: 147-151.

Jonker, R. and A. Volgenant, 1987. A shortest augmenting path algorithm for dense and sparse linear assignment problems. Computing, 38: 325-340.

Leibe, B., A. Leonardis and B. Schiel, 2004. Combined object categorization and segmentation with an implicit shape model. Proceedings of the ECCV Workshop on Statistical Learning in Computer Vision, May 2004, Prague, pp: 17-32.

Lowe, D.G., 1999. Object recognition from local scale-invariant features. Proc. 7th IEEE Int. Conf. Comput. Vision, 2: 1150-1157.

Maji, S. and J. Malik, 2009. Object detection using a max-margin Hough transform. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, June 20-25, 2009, Miami, FL., pp: 1038-1045.

Matas, J., O. Chum, M. Urban and T. Pajdla, 2002. Robust wide baseline stereo from maximally stable extremal regions. Br. Machine Vision Conf., 1: 384-393.

Mikolajczyk, K. and C. Schmid, 2004. Scale and affine invariant interest point detectors. Int. J. Comput. Vision, 60: 63-86.

Mikolajczyk, K.I. and C. Schmid, 2005. A performance evaluation of local descriptors. IEEE Trans. Pattern Anal. Mach. Intell., 27: 1615-1630.

Szeliski, R., 2010. Computer Vision: Algorithms and Applications. Springer, New York pp: 680-700.

Zhang, S.H., T. Chen, Y.F. Zhang, S.M. Hu and R.R. Martin, 2009. Vectorizing cartoon animations. IEEE Trans. Visualization Comput. Graph., 15: 618-629.

Zhang, T., Q. Han, X. Bai and X. Niu, 2013. Decorative line and edge extraction in cartoon images. Res. J. Applied Sci. Eng. Technol., 5: 4013-4017.