Directional Blurred Shape Model Descriptor for Sketched Symbol Recognition

Wei Deng, Lingda Wu, Zhonghua Yao, and Yashuai Lü

***Abstract*—A new descriptor is proposed for sketched symbol recognition. It incorporates local direction information of sampling points into the BSM (Blurred Shape Model) descriptor, which represent only the spatial distribution of a symbol. In the proposed directional BSM (DBSM), a symbol is firstly decomposed into four directional sub-symbols. Then in each sub-symbol, the nrBSM (Non-Rigid BSM) descriptor is extracted. The four nrBSMs form the final DBSM descriptor. It inherits from the traditional BSM the invariance to stroke order and number. Moreover, the new descriptor results in better performances in sketched symbol recognition.**

***Index Terms*—Sketched symbol recognition, blurred shape model (BSM), directional information, classification.**

I. Introduction

Sketching is a natural form of human communication and has become an increasingly popular tool for interacting with user interfaces. It is a fast and efficient means of capturing information by automatically interpret hand-drawn sketches. With the growing popularity of digital input devices, there is increasing interest in building sketch-based user interfaces, whose key technology is sketch recognition.

Sketch recognition refers to recognition of pre-defined symbols or free-form drawings (e.g., an unconstrained circuit drawing); in the latter case, the recognition task is generally preceded by segmentation in order to locate individual symbols [1]. However, this paper focuses on the recognition of hand-drawn isolated symbols.

However, many challenges remain in terms of intra-class compactness and inter-class separation due to the variability of sketching. Because it is likely that different users have different drawing styles, such as the stroke-order and –number, as well as symbol size and complex local deformation. Moreover, the styles may differ even the same individual at different times. A good recognition algorithm should take these challenges in consideration and place few drawing constraints on users.

Statistical method can accomplish this task well. It requires the robust descriptors in order to obtain rich information of the data. Escalera *et al*. [2] present the Blurred Shape Model descriptor (BSM). It encodes the spatial probability of appearance of the shape pixel and their context information. Almazan *et al*. [3] then improve the BSM descriptor by combining shape deformations with appearance variability and propose the modified version named nrBSM (Non-Rigid Blurred Shape Model). The modification permits to overcome the rigidity of the original BSM, adapting it to the shape to be represented. We present in this paper a new descriptor called DBSM (directional BSM) for sketched symbol recognition. DBSM represents not only the spatial distribution of sample points, but also their local directional information. As a result, it is more informative and consequently more effective than the traditional nrBSM descriptor in sketched symbol recognition. The DBSM-based method inherits from the traditional BSM and nrBSM the invariance of stroke order and number.

II. Related Work

According to a widely accepted taxonomy, the methods of sketched symbol recognition are classified into two main categories: structural and statistical [4].

Structural methods focus on building structural shape descriptions. Its basis step is stroke segmentation and primitive recognition using temporal and spatial features. Then a sketched symbol can be represented as a tree or graph and the similarity between two sketches can be calculated by structural matching [5]. Hammond and Davis [6] developed a hierarchical language to describe how diagrams are drawn, displayed, and edited. Then they used this language to perform automatic symbol recognition. Attributed relational graph (ARG) is an excellent statistical model to describe both geometry and topology of a symbol [7], and is insensitive to orientation, scaling, and drawing order. The advantage of structural methods is distinguishing similar shapes. But the disadvantage is their sensitivity to results of primitive recognition. So high accuracy in stroke segmentation and primitive recognition is necessary.

Statistical methods look at the visual appearance of shapes and symbols instead of the relation between geometric primitives. Ouyang and Davis [8] proposed a visual approach to sketched symbol recognition. It used a set of visual features that captured on-line stroke properties like orientation and endpoint location. Eitz *et al*. [9] described a large scale exploration of human sketches. They analyzed the distribution of non-expert sketches of everyday objects and developed a bag-of-features sketch representation. Willems *et al.* [10] explored a large number of on-line features, which were sorted in three feature sets due to different levels of details. Delaye and Anquetil [11] presented a set of 49 features, called HBF49, for the representation of hand-drawn symbols to be used as a reference for evaluation of symbol recognition systems.

Manuscript received December 20, 2014; revised March 28, 2015. This work was supported in part by the "HGJ" National Major Technological Projects under Grant No. 2013ZX01045-004-002.

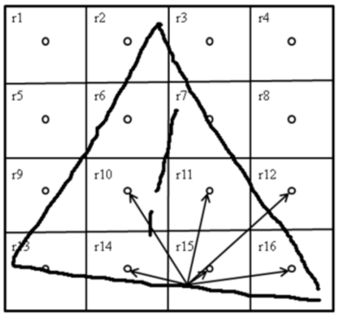
The authors are all with the Equipment Academy, Beijing 101416, China (e-mail: dengwei@whu.edu.cn, wld@nudt.edu.cn, visworker07@163.com, freelancer_lys@163.com).

DOI: 10.7763/IJIEE.2015.V5.581

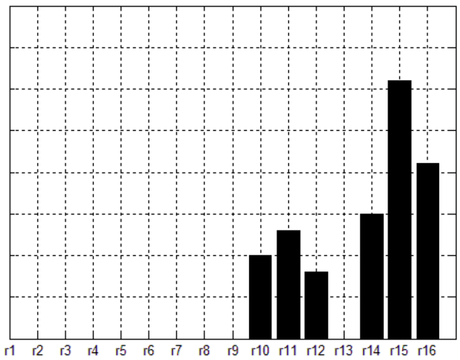
III. Original BSM and nrBSM

A. *Blurred Shape Model (BSM)*

The main idea of the BSM descriptor [2] is to describe a given shape by a probability density function encoding the probability of pixel densities of a certain number of image sub-regions. Given a set of points forming the shape of a particular symbol, each point contributes to compute the BSM descriptor. This is done by dividing the given image in a $k \times k$ grid with equal-sized sub-regions (cells). Then, each cell receives votes not only from the shape pixels located inside its corresponding cell, but also from those located in the adjacent cells. Thereby, every pixel contributes to the density measure of its sub-region cell, and its neighboring ones. For each cell, there is a weighted sum value of the points in itself and neighbors. The weight is inversely proportional to the square of the distance between the point and centroid of the cell. In Fig. 1(a), shape description is shown for an apple symbol. The output of BSM is a vector histogram, where each position contains the accumulated value of each sub-region, and contains the spatial distribution in the context of the sub-region and its neighbors.



(a) Distances of a contour point SP to its centroids.



(b) Vector descriptor update using the distance of (a)

Fig.1. BSM density estimation example [2].

B. *Non-Rigid Blurred Shape Model (nrBSM)*

The nrBSM [3] is almost the same as BSM. However, it has the different grid. In original BSM descriptor the uniform grid is used. The nrBSM is based on the region partitioning procedure of the adaptive hierarchical density histogram, which consists in iteratively producing regions of the symbol using the geometrical centroid estimation. Specifically, firstly a horizontal line and a vertical line are drawn. They both cross the centroid of the symbol. So the rectangular region is divided into four sub-regions by the two lines. Next, in each sub-region we iteratively do the same produce until the $k \times k$ grid is gained.

With the adaptive hierarchical grid, the nrBSM can be used for non-rigid symbol recognition, such as sketched symbols.

IV. Proposed DBSM Description Method

We propose a new descriptor DBSM, inspired by the nrBSM descriptor. The new descriptor consists for nrBSM, correspond to four sub-symbols generated by directional decomposition. The proposed method mainly contains three steps. The flowchart is shown in Fig. 2.

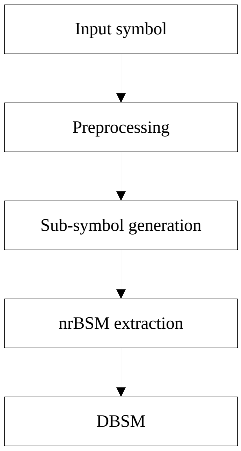


Fig. 2. The flowchart of the proposed method.

A. *Preprocessing*

In most statistical method, the first step is preprocessing. It directly facilitates pattern description and affects the quality of description. The preprocessing tasks of sketch include resampling, noise elimination and shape scaling.

Prior to the feature extraction itself, we chose to normalize patterns in scale and translation, and to apply a trajectory resampling strategy. These operations are simple to perform and guarantee a better stability of extracted features, for any type of input pattern [12].

Since on-line strokes are typically sampled at a constant temporal frequency, the distance between neighboring points varies based on the pen speed. This produces more sampled points where the pen is typically slower. In order to make feature extraction more reliable, we resample each stroke at a constant spatial distance. In our experiments the sampling interval is set to 1.

Next we remove differences of scale and translation. The coordinates of stroke points are simply shifted and scaled such that all points are enclosed in a standard box. In our experiments we set $x, y \in [0,100]$. It means translating maximal dimension of a symbol to 100 with aspect ratio preserved. It means the linear scaling, keeping the aspect ratio. It has been shown to have good results [13]. We do not translate the on-line trajectory to image. So direction features are accurate with the absence of jagged edge.

Finally the noise reduction technique is carried out. It contains three operations: smoothing, filtering, wild point correction, which is often used in the handwritten character recognition [14].

B. *Sub-Symbol Generation*

The symbol is divided into four sub-symbols by directional decomposition. For each resampled point p, calculate its local line as the local line (p_{i-1}, p_i). The local stroke line is assigned to four directional planes, corresponding to four chaincode directions, i.e., 0, 45, 90 and 135 degrees. One major advantage is the independence of local stroke direction, e.g. the decomposition of (p_i,p_{i+1}) is the same as that of (p_{i+1},p_i). Each local line, defined by two consecutive points, is decomposed into two components in two neighboring chaincode directions, as shown in Fig. 3.

In each directional plane there is a sub-symbol. The pixel of the sub-symbol is set as the response of the corresponding resampled point. So we can get more information for sketched symbol recognition with the four directional sub-symbol.

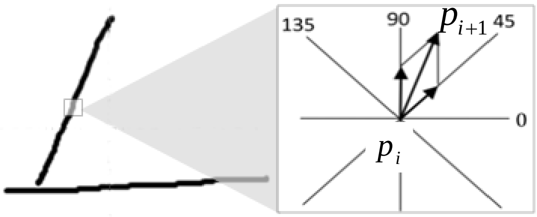


Fig. 3 Directional decomposition of a local line segment.

C. nrBSM Extraction

In each sub-symbol, the nrBSM descriptor can be extracted. The only difference with the original nrBSM in [3] is that the pixel value in sub-symbol is in [0, 1], while in [3] the pixel value is always 1. The pseudo-code is shown in Fig. 4.

Input: A grid of $k \times k$ cells and $S = \{x_i, y_i\}$, where i = 1,2,…,n
Output: the DBSM descriptor
1. Compute the four directional response values of each resampling point, denoted as dir_1,dir_2, dir_3 and dir_4, respectively.
2. Denote grid as $R = \{r_1, ..., r_{k^2}\}$, with c_j the centroid for grid cell r_j , $j = 1, 2, ..., k^2$
3. Denote $N(r)$ as the neighbor cells of cell r, Let r^p be the cell which contains the point p. Initialize the probability vector v as $v_g(j)$=0, $j = 1, 2, ..., k^2$, g=1, 2, 3, 4
(Corresponding to four sub-symbols)
4.For each point $p(x_i, y_i)$
  D=0
  For each $r_h \in N(r^p)$, $(h \in \{1, 2, ..., k^2\})$
    $d_h = d(p, r_h) = \| p - c_h \|^2$
    $D = D + 1/d_h$
  End for
  Update the probability vector v as
  $v_1(r_h) = v(r_h) + dir_1/(Dd_h)$,
  $v_2(r_h) = v(r_h) + dir_2/(Dd_h)$,
  $v_3(r_h) = v(r_h) + dir_3/(Dd_h)$,
  $v_4(r_h) = v(r_h) + dir_4/(Dd_h)$.
End for
5.Normalize $v_g(h) = v_g(h) / \sum_{j=1}^{k^2} v_g(j)$, $\forall h$ = 1, 2,…, k^2 , g=1, 2, 3, 4.
6. Combine the for sub-descriptors into the final DBSM descriptor as DBSM = { v_1, v_2, v_3, v_4 }

Fig. 4. The pseudo-code of DBSM descriptor.

In our experiments the k is set to 8. So the DBSM descriptor is a 8 × 8 × 4 = 256 dimensional vector. To modify the feature distribution for improving the classification performance, all the measurements in the feature vectors are transformed by variable transformation $y = x^p$. In our experiments the performance is optimal when p=3.

V. Experiments and Results

A. The Datasets

We have evaluated our method in two datasets. The first one is NicIcon database [15]. It is composed of 26,163 sketched symbols of 14 classes in the domain of crisis management and incident response systems. An example is shown in Fig. 5. The database comes from 34 different writers and it is commonly used for on-line symbol recognition, but off-line data is also available. The database is already divided into three subsets (Training, Test and Evaluation sets) for both writer dependent and independent settings. We have selected the on-line data which is writer-independent. Because it is the most similar to human-computer interaction.

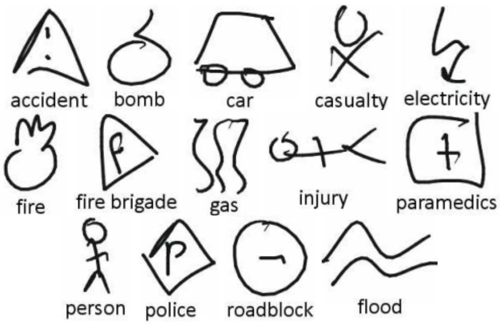


Fig. 5. Samples from each class in the NicIcon dataset.

The second dataset is the symbols named COAD [16]. It contains the symbols from Course of Action Diagrams database (COAD). It has 20 classes of symbols and Fig. 6 shows a set of examples. It is composed of 620 sketched symbols in total. The database comes from 8 different users.

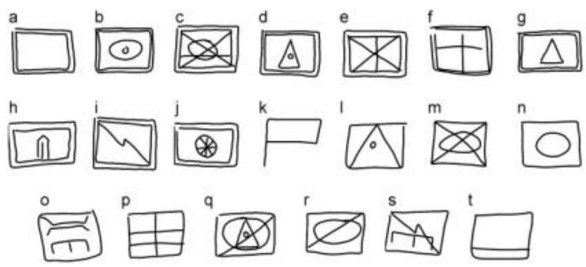


Fig. 6. Samples from each class in the COAD dataset.

B. Classification

We use the k-NN and SVM as two classifiers. For k-NN the optimal parameter k is get with the training samples. For SVM we evaluated the performance using the LibSVM toolbox with a RBF kernel.

C. Recognition Results

The recognition accuracy with original nrBSM descriptor is shown in Table I, while the result of our DBSM descriptor is shown in Table II. The new DBSM is better than the original nrBSM in both classifier and both datasets.

TABLE I: THE RECOGNITION ACCURACY WITH NRBSM

Dataset	NicIcon	COAD
k-NN	81.7%	87.2%
SVM	85.3%	90.3%

TABLE II: THE RECOGNITION ACCURACY WITH OUR DBSM

Dataset	NicIcon	COAD
k-NN	94.8%	96.3%
SVM	96.6%	97.6%

D. Runtime Performance

Finally, we evaluate the runtime performance of our approach. We use Matlab to program this method on a 3.10×2 GHz machine (Inter Core™ i5-2400). The average runtime of recognizing one symbol is 13ms more than the nrBSM. So this new descriptor is worthy to be used with small time cost to achieve better accuracy.

VI. CONCLUSION

This paper presents a novel descriptor for recognizing sketched symbols. The representation method decomposes a symbol into four componential patterns according to the direction of sample points before computing BSM on each componential pattern, respectively. In this way, the local directional information is integrated into the BSM features, which mainly describe the spatial distribution information. In our experiments it achieves better accuracy than the original nrBSM features. Moreover, it a statistical method, which uses the visual feature of shape points, instead of regarding the shape as an image or temporal sequences of geometric primitives. This method is invariant to stroke order and number. In future we will use the descriptor to segment symbols in a full diagram, i.e., the so-called stroke groping.

REFERENCES

[1] C. Tirkaz, B. Yanikoglu, and M. Sezgin. "Sketched symbol recognition with auto-completion," *Pattern Recognition*, vol. 45, pp. 3926-3937, 2012.

[2] S. Escalera, A. Fornes, O. Pujol *et al*., "Blurred shape model for binary and grey-level symbol recognition," *Pattern Recognition Letters*, vol. 30, pp. 1424-1433, 2009.

[3] J. Almazan, A. Fornes, and E. Valveny, "A non-rigid appearance model for shape description and recognition," *Pattern Recognition*, vol. 45, pp. 3105-3113, 2012.

[4] M. de Rosa, "New methods, techniques and applications for sketch recognition," PhD dissertation, University of Salerno, Via Ponte Don Melillo, Fisciano (SA), Italy, 2014.

[5] G. Costagliola, M. de Rosa, and V. Fuccella, "Recognition and autocompletion of partially drawn symbols by using polar histograms as spatial relation descriptors," *Computers & Graphics*, vol. 39, pp. 101-116, 2014.

[6] T. Hammond and R. Davis, "LADDER, a sketching language for user interface developers," *Computer & Graphics*, vol. 29, pp. 518-532, 2005.

[7] W. Lee, L. Kara *et al*., "An efficient graph-based recognition for hand-drawn symbols," *Computers & Graphics*, vol. 31, pp. 554-567, 2007.

[8] T. Y. Ouyang and R. Davis, "A visual approach to sketch symbol recognition," in *Proc. the 21st International Joint Conference on Artificial Intelligence*, 2009, pp. 1463-1468.

[9] M. Eitz, J. Hays, and M. Alexa, "How do humans sketch object?" *ACM Transactions on Graphics*, vol. 31, 2012.

[10] D. Willems, R. Niels, M. van Gerven *et al*., "Iconic and multi-stroke gesture recognition," *Pattern Recognition*, vol. 42, pp. 3303-3312, 2009.

[11] A. Delaye and E. Anquetil, "HBF49 feature set: A first unified baseline for online symbol recognition," *Pattern Recognition*, vol. 46, pp. 117-130, 2012.

[12] C. L. Liu, K. Nakashima, H. Saho *et al*., "Handwritten digit recognition: investigation of normalization and feature extraction techniques," *Pattern Recognition*, vol. 37, pp. 265-279, 2004.

[13] C. L. Liu, "Handwritten Chinese character recognition: Effects of shape normalization and feature extraction," *Arabic and Chinese Handwriting Recognition*, pp. 104-128, Springer Berlin Heidelberg, 2008.

[14] C. L. Liu, S. Jaeger, and M Nakagawa, "Online recognition of Chinese character: The state-of-the-art," *IEEE Transaction on Pattern Analysis and Machine Intelligence*, vol. 26, pp. 198-212, 2004.

[15] R. Niels, D. Willems, and L. Vuurpijl, "The NicIcon database of handwritten icons," in *Proc. 11th International Conference on frontiers in Handwriting Recognition*, Montreal, Canada, August 19-21, 2008.

[16] R. Arandjelovic and T. M. Sezgin. "Sketch recognition by fusion of temporal and image-based features," *Pattern Recognition*, vol. 44, pp. 1225-1234, 2011.

Wei Deng was born in 1986. He is now a Ph.D. candidate in Equipment Academy, Beijing, China. His research work is mainly focused on sketch recognition and multimedia technology.

Lingda Wu was born in 1962, China. She is a professor in Equipment Academy, Beijing, China. Her research interests are mainly in virtual reality technology and milti-media technology.

Zhonghua Yao was born in 1988, China. He is now a Ph.D. candidate in Equipment Academy, Beijing, China. He is graduated from National University Of Defense Technology in 2013, major in multimedia information system, visual analysis, human and computer interaction. His research work is mainly focused on data visualization.

Yashuai Lü, born in 1981, China. He received his Ph.D. degree from the National University of Defense Technology in 2009. His major field of study is computer architecture. Now he works at the Academy of Equipment, China. His main research interests include processor architecture and computer graphics.