

A New Shape Matching Measure for Nonlinear Distorted Object Recognition

S. Srisuk[†], M. Tamsri[†], R. Fooprateepsiri^{*}, P. Sookavatana[†] and K. Sunat[†]

Department of Computer Engineering[†], Department of Information Technology^{*},
Mahanakorn University of Technology, Bangkok, THAILAND, 10530
E-mail: s.srisuk@mut.ac.th

Abstract. In this paper, we present a new approach for hand-written character and digit recognitions based on shape descriptor and the Hausdorff Context. We start at finding the corresponding points between two shapes by using a modified shape context. We then use these correspondences as key geometric points for shape alignment with the Thin Plate Spline (TPS) model. After the transformation has been applied completely, the distance between two shapes is computed by a new distance measure, the Hausdorff Context. We achieve a very high recognition rate of 98.7% on 268 images.

1 Introduction

Recognizing of an object in an unconstrained condition is a hard problem. One of the examples of the distorted object recognition is the hand written character/digit recognition. In order to handle the situations we may deal with during test phase, a robust method that is insensitive to variations of an image should be constructed. This includes effects due to varying in writing styles and sizes of characters. It is known that the same characters written by different users may be seen as different classes of the object due to the local distortion, size variation and rotation of character. Fu *et al.* [4] proposed an adaptive learning system for hand written Chinese character recognition. They mainly addressed the user adaption issues using self-growing probabilistic decision-based neural networks (SPDNNs). The user adaptation of the parameters of SPDNN was formulated as incremental reinforced and antireinforced learning procedures. In summary, they have developed 1) an SPDNN based handwriting recognition system; 2) a two-stage recognition structure; and 3) a three-phase training methodology for a) a global coarse classifier; b) a user independent hand written character recognizer; and c) a user adaptation module on a personal computer. The SPDNN has some nice properties because of its discrimination function obeys a probability constraint so that a lower false acceptance and false rejection rates can be achieved. Torres-Mendez *et al.* [5] presented a new method for object recognition under translation, rotation and scaling conditions that has been applied to character recognition. In the preprocessing step, they take into account the invariant properties of the normalized moment of inertia. They also proposed a new coding of an object which describes its topological characteristics. The vectors obtained in the preprocessing step were then used as inputs to a holographic nearest-neighbor (HNN), a fast algorithm for object recognition. The algorithm was

tested on character data set in which the test images were different in 17 sizes and 14 rotations without addressing the problem of local geometrical distortion of such characters. They achieved 98% correct recognition on their own data set.

The organization of this paper is as follows. Section 2 describes a method for finding corresponding points between two shapes using standard and a modified shape context. Shape alignment using TPS model is presented in section 3. In section 4, we introduce a new shape similarity measure that integrates spatial as well as structural information on the shape by which a more accurate shape matching measure can be achieved. Section 5 presents our experimental results. Finally, we conclude in section 6.

2 The Shape Context

In our approach, a shape is represented by a discrete set of points obtained from an edge detector. Let us denote by $\mathcal{P} = \{p_1, \dots, p_n\}, p_i \in \mathbb{R}^2$ the set of n edge pixels. The shape context at a reference pixel is computed from the set of the remaining $n - 1$ points [7]. In other words, at an edge pixel p_i , we compute a coarse histogram h_i of the relative coordinates of all other edge points in the image, with respect to the pixel p_i :

$$h_i(k) = \#\{q \neq p_i | (q - p_i) \in \text{bin}(k)\}. \tag{1}$$

The histogram h_i is therefore defined as the shape context of p_i . Each shape context is a log-polar histogram of the coordinates of the rest of the point set measured using the reference point as the origin. More practically, we count the number of points that fall inside each bin k . The log-polar histogram bins are used in computing the shape context as shown in Fig. 1 (a).

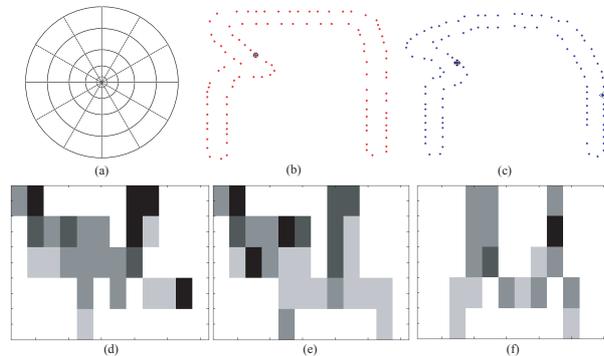


Fig. 1. (a) Diagram of log-polar histogram. (b) and (c) Example of edge pixels generated from two hand-written characters. (d) Shape context example of the point marked by \circ in (b). (e) Shape context example of the point marked by \diamond in (c). (f) Shape context example of the point marked by \blacktriangleleft in (c), (dark=large value). Here we use 5 bins for $\log r$ and 12 bins for θ .

As shown in Figs. 1 (d), (e) and (f) the shape context is a compact representation of the distribution of points relative to each point. This log-polar histogram is more

sensitive to positions of nearby sample points than to those of points farther away. The points marked by \circ and \diamond in Figs. 1 (b) and (c) are a correct corresponding point between two shapes, achieving by comparing its shape context histogram with a similar one as shown in Figs. 1 (d) and (e). It should be noted that the shape context histogram in Fig. 1 (f) is significantly different from the ones in Figs. 1 (d) and (e) which is not a correct corresponding point. The cost of shape context matching between two points, p_i and q_j , is denoted by $\mathcal{C}_{\text{sc}}(p_i, q_j)$ and computed by using the χ^2 test statistics:

$$\mathcal{C}_{\text{sc}}(p_i, q_j) = \frac{1}{2} \sum_{k=1}^K \frac{[h_i(k) - h_j(k)]^2}{h_i(k) + h_j(k)}, \quad (2)$$

where p_i is a point on shape \mathcal{P} and q_j a point on shape \mathcal{Q} , $h_i(k)$ and $h_j(k)$ denote the K-bin normalized histogram at p_i and q_j , respectively.

2.1 Shape Context with Second Order Derivative and Distance Measure

The shape context is a rich descriptor providing an image representation which describes a shape characteristic around a specific point. It is however not powerful enough to use the shape context alone in order to find the correspondences in distorted objects, such as hand-written character and digit. We propose here a new approach to improve the performance by incorporating a new features to the shape context. Let us define by $\mathcal{C}_k = \mathcal{C}_{\text{sc}}(p_i, q_j)$ the shape context cost at position $k = (i, j)$, then the previous position $\mathcal{C}_{k-1} = \mathcal{C}_{\text{sc}}(p_{i-1}, q_{j-1})$ and the next position $\mathcal{C}_{k+1} = \mathcal{C}_{\text{sc}}(p_{i+1}, q_{j+1})$ are the closest points to position k . In an image processing, a second-order derivative of the one-dimensional function $f(x)$ is the difference, i.e. $\frac{\partial^2 f}{\partial x^2} = f(x+1) + f(x-1) - 2f(x)$. We then define a second-order derivative as the difference between the closest points at position k .

$$\mathcal{C}_{\text{diff}}(p_i, q_j) = \mathcal{C}_{k+1} + \mathcal{C}_{k-1} - 2\mathcal{C}_k, \quad (3)$$

where the assumption comes from the fact that the closest points $k-1$ and $k+1$ should also have the similar shape context. In addition, the corresponding points between two similar shapes, e.g. two points on the original and distorted versions of the same object, tend to close to each other. The distance between two points can then be defined as

$$\mathcal{C}_{\text{dist}}(p_i, q_j) = \begin{cases} 0 & \text{if } \|p_i - q_j\| \leq \tau \\ \|p_i - q_j\| & \text{Otherwise,} \end{cases} \quad (4)$$

where $\|\cdot\|$ is a norm between two points, and τ is some threshold. Finally, the shape context cost is a weighted summation

$$\mathcal{C}(p_i, q_j) = w_1 \mathcal{C}_{\text{sc}}(p_i, q_j) + w_2 \mathcal{C}_{\text{diff}}(p_i, q_j) + w_3 \mathcal{C}_{\text{dist}}(p_i, q_j), \quad (5)$$

where w_1, w_2 and w_3 are weighting parameters with $\sum w_i = 1$. These additional features can be used to improve the performance of the shape context in which the more correct corresponding points can be achieved. The better the corresponding points are known, the better the transformation between two shapes can be accomplished.

3 Thin Plate Spline Model

The points on one shape can be warped to another points on the other shape by Thin Plate Spline (TPS) model. This can be achieved by finding the corresponding points between two shapes by shape context cost matching, and then using these correspondences as the key points for geometric transformation. These points may be transformed by affine model

$$T(x) = \mathcal{H}x + b, \tag{6}$$

where \mathcal{H} is a matrix of homogeneous coordinates and b the translational offset vector. The least squares solution $\hat{T} = (\hat{\mathcal{H}}, \hat{b})$ is then computed by

$$\hat{\mathcal{H}} = (Q^+P)^t, \tag{7}$$

$$\hat{b} = \frac{1}{n} \sum_{i=1}^n (p_i - q_{\pi(i)}), \tag{8}$$

where P and Q contain coordinates on shape \mathcal{P} and \mathcal{Q} , respectively, i.e.,

$$P = \begin{bmatrix} 1 & p_{11} & p_{12} \\ \vdots & \vdots & \vdots \\ 1 & p_{n1} & p_{n2} \end{bmatrix}, Q = \begin{bmatrix} 1 & q_{11} & q_{12} \\ \vdots & \vdots & \vdots \\ 1 & q_{n1} & q_{n2} \end{bmatrix}.$$

Let us define by L the matrix of this system, i.e.,

$$L = \begin{bmatrix} K & P \\ P^t & 0 \end{bmatrix},$$

where L is a $(n + 3) \times (n + 3)$ matrix. In a two-dimensional homogeneous coordinate representation, points are transformed from coordinate $p_i = (x_i, y_i)$ to $v_i = (x'_i, y'_i)$ with the matrix operation

$$\begin{bmatrix} v \\ 0 \end{bmatrix} = \begin{bmatrix} K & P \\ P^t & 0 \end{bmatrix} \cdot \begin{bmatrix} w \\ a \end{bmatrix},$$

where

$$K = \begin{bmatrix} 0 & U(r_{12}) & \cdots & U(r_{1n}) \\ U(r_{21}) & 0 & \cdots & U(r_{2n}) \\ \cdots & \cdots & \cdots & \cdots \\ U(r_{n1}) & U(r_{n2}) & \cdots & 0 \end{bmatrix}, \tag{9}$$

with $r_{ij} = \|(x_i, y_i) - (x_j, y_j)\|$, w and v are column vector constructed from w_i and v_i , a is column vector with elements a_1, a_x, a_y , and 0 is a 3x3 matrix of zeros. Let us define by v_i the target function values at corresponding point $p_i = (x_i, y_i)$ in the plane, where v_i is usually set to (x'_i, y'_i) . We can write TPS interpolant $f(x, y)$ to minimize the bending energy as

$$I_f = \int \int_{\mathbb{R}^2} \left(\frac{\partial^2 f}{\partial x^2} \right)^2 + 2 \left(\frac{\partial^2 f}{\partial x \partial y} \right)^2 + \left(\frac{\partial^2 f}{\partial y^2} \right)^2 dx dy, \quad (10)$$

which has the form

$$f(x, y) = a_1 + a_x x + a_y y + \sum_{i=1}^n w_i U(\|(x_i, y_i) - (x, y)\|), \quad (11)$$

where $U(r)$ is a kernel function and defined by $U(r) = r^2 \log r^2$ [3], with the following constrain solutions:

$$\sum_{i=1}^n w_i = 0 \quad \text{and} \quad \sum_{i=1}^n w_i x_i = \sum_{i=1}^n w_i y_i = 0.$$

Finally, the matrix v of target point is obtained at each iteration, we then use the matrix v as the initial point for the next iteration by which the bending energy function in (10) is minimized.

4 Shape Distance

Within the last several years, numerous algorithms have been proposed for shape matching. The most successful shape matching algorithms can be categorized into two main parts: spatial shape matching and structural shape matching. The former operate on the spatial information of the shape, for example, the distance between two shapes. The latter, on the other hand, make use of the structural information of the shape, e.g. curvature maxima, axis of the shape, pairwise geometric histograms [1], shape context [7], etc. One of the most successful spatial shape similarity measures is the Hausdorff distance [2]. This measure is robust to data uncertainty and perturbations in the locations of pixels, since it measures the proximity of the two shapes rather than their exact superposition. The shape context [7], on the other hand, is a local descriptor which captures the distribution of the remaining points relative to it, thus offering a globally discriminative characterization. We present here a new shape similarity measure to integrate the structural information as well as spatial information of the shapes extracted from an edge detector.

4.1 The Hausdorff Distance

Given two point sets A and B , the Hausdorff distance [2] between A and B is defined as

$$H(A, B) = \max(h(A, B), h(B, A)), \quad (12)$$

$$h(A, B) = \max_{a \in A} \min_{b \in B} \|a - b\|, \quad (13)$$

$$h(B, A) = \max_{b \in B} \min_{a \in A} \|a - b\|, \quad (14)$$

where $\|\cdot\|$ denotes some norm of points of A and B . This measure indicates the degree of similarity between two point sets. It can be calculated without an explicit pairing

of points in their respective data sets. The conventional Hausdorff distance, however, is not robust to the presence of noise. Dubuisson *et. al.* [6] have studied 24 different variations of the Hausdorff distance in the presence of noise. A modified Hausdorff distance (MHD) using an average distance between the points of one set to the other set gives the best result. This measure is the most widely used in the task of object recognition and defined as

$$h(A, B) = \frac{1}{n} \sum_{a \in A} \min_{b \in B} \|a - b\|, \tag{15}$$

with $h(B, A)$ defined similarly. This modified Hausdorff distance is less sensitive to noise than the conventional one. It is possible, however, to endow the Hausdorff distance with even more attractive features as it is shown in the next section.

4.2 The Hausdorff Context

In this section, we propose a novel shape similarity measure, the ‘‘Hausdorff context’’, based on the combination of the Hausdorff distance and the shape context. This new measure is designed to integrate the spatial information with the structural information of the shape. Generally, the Hausdorff distance uses only the spatial information of edge pixels without considering the inherent local structural characteristics of such points. The shape context, on the other hand, is a very discriminative point descriptor, facilitating easy and robust point matching by incorporating global shape information into a local descriptor. Let us define by $h(A, B) = \max_{a \in A} d(a, B)$ the direct Hausdorff distance, where $d(a, B) = \min_{b \in B} \|a - b\|$. Consider point a on the first shape and a set B of the second shape as shown in Fig. 2. Let us suppose that A and B are the same shape, except that B is distorted by some noise and is broken at some points because of segmentation and edge detection errors.

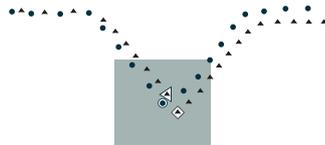


Fig. 2. The integration of minimum Hausdorff distance and shape context. The grey shade indicates the neighborhood area. The point marked by \circ is a sample point a of the first shape A . The points marked by \triangleleft and \diamond are the candidate matching points of the second shape B .

The Hausdorff distance measures the distance from point a to all points of set B , $d(a, B)$, then selects the one at the minimum distance among them. In this case, the candidate point marked by \triangleleft is selected and, then, the distance between them is used as the result. The minimum distance is therefore based only on the spatial information. This is not useful when using the Hausdorff distance in the presence of noise, when we have to deal with the broken point problem caused by segmentation and edge detection errors,

etc. To the best of our knowledge, there is no work in the point matching Hausdorff distance with structural point information. We propose an alternative way to find the minimum distance between point a and set B to overcome the above problem. Instead of finding the nearest distance, in our approach, the point descriptor, shape context, is used to find the best matching between point a and set B . We, therefore, call this shape similarity measure as ‘‘Hausdorff context’’.

$$h_{HC}(A, B) = \sum_{a \in A} \underbrace{w(a, b')}_{\text{spatial information}} \min_{b \in B} \overbrace{\mathcal{C}_{SC}(a, N(b))}^{\text{structural information}}, \quad (16)$$

$$w(a, b') = \frac{\mathcal{D}(a, b')}{\sum_{a \in A} \mathcal{D}(a, b')} \text{ and } \sum w(a, b') = 1, \quad (17)$$

where $b' = \arg \min_{b \in B} \mathcal{C}_{SC}(a, N(b))$, and $\mathcal{C}_{SC}()$ is the χ^2 test statistic as defined in (2). In the example shown in Fig. 2 the candidate point b' is the one marked by \diamond which is the correct corresponding point between point a and a point in set B . The cost of matching between two points a and b , $\mathcal{C}_{SC}(a, N(b))$, is weighted by their distance, $\mathcal{D}(a, b')$. Therefore $w()$ is a normalized distance between points a and b' over the entire distance between sets A and B . Furthermore, the neighborhood $N()$ is designed to reduce the computation time of the shape matching, since it finds the best point matching only in the neighborhood area. Thus faster performance improvement can be achieved. The $h_{HC}(B, A)$ is defined in a similar way. The shape similarity measure in (16) with the maximum Hausdorff matching in (12) is defined to be a confidence level of matching:

$$\text{dist}_{HC}(A, B) = 1 - H(A, B). \quad (18)$$

5 Experimental Results

The hand written data set was collected from student and staff at Mahanakorn University of Technology. The data set contained 196 characters and 72 digits. We used 120 characters and 30 digits for training set, the remaining characters and digits were used to test the algorithm. In order to reduce the computational time, we used 100 points generated from the shape of characters and digits. The edge pixels were produced from Canny edge detector. The characters and digits in the training set were different from the test set. Fig. 3 shows examples of the hand-written characters and digits from the training and test sets. It should be noted that the characters and digits in the training and test sets were used to test the algorithm without the preprocessing step, i.e. the normalization step was not used so that they are different in size, rotation and local distortion.

Our proposed method has three-step approach: 1) find the corresponding points between two shapes using a modified shape context described in subsection 2.1, 2) map points on one shape to the other by thin plate spline, and 3) perform shape matching measure using our new proposed Hausdorff Context. As discussed in [7], the whole

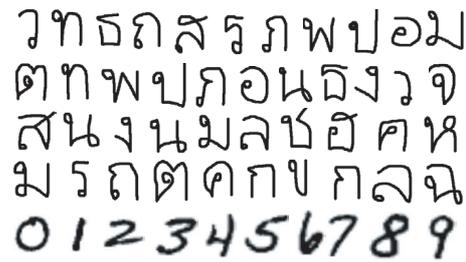


Fig. 3. Examples of the hand-written characters and digits with different styles and size variations.

points on one shape are too much detailed since the shape may contain noise and may vary from shape to shape. In Fig. 4, the shapes are generated by randomly selected with uniform distance between points. This helps us to control the distribution on the shape by which the matching confidence can be maximized. Figs. 4 (a) and (b) show a comparison result of the corresponding points between two shapes using the standard and a modified shape context. As a result, the corresponding points can be correctly discovered by employing the shape context with second order derivative and distance measure, while using the standard shape context fails to find the correct corresponding points at some points. The shape matching measure can then be improved by using our new features which are added to the shape context.

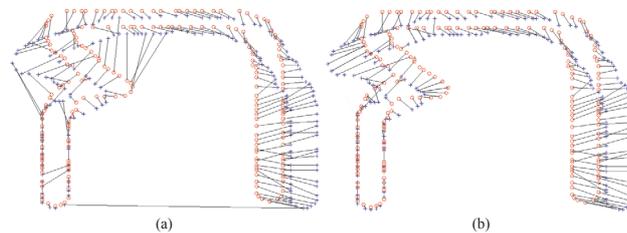


Fig. 4. A comparison result of the corresponding points between two shapes using (a) standard shape context and (b) a modified shape context.

Figs. 5 (a) and (b) show the sampling points on the shapes \mathcal{P} and \mathcal{Q} obtained from training set and test set. It should be noted that the variations between shapes \mathcal{P} and \mathcal{Q} are significantly difference by local geometrical distortion, rotation, scaling and translation. One can also see from Figs. 5 (c) and (d) the difference between shapes \mathcal{P} and \mathcal{Q} in position, size, orientation and bending of the characters in which it was superimposed on one another. Fig. 5 (c) show the corresponding points measured using standard shape context, while in Fig. 5 (d) the better corresponding points measured using a modified shape context were illustrated. In Fig. 5 (e), we show the results of the transformation of shape \mathcal{Q} to the shape \mathcal{P} using TPS model with a modified shape context by which

the matching confidence between two shapes should be maximized. It is worth stressing that the shape matching measure, the Hausdorff Context, will be applied after the shape Q has been aligned to shape P so that the recognition is simplified because the minimized within-class variance. On the test phase, an unknown character or digit was input into the system. The corresponding points between two shapes were then computed using a modified shape context. An unknown character or digit was transformed to the shape in training set by TPS model. Finally, the distance between two shapes was measured by the Hausdorff Context. We have tested the standard shape context and our proposed method on 268 images, and the results were shown in Table 1. As a result, we achieve a very high recognition rate of 98.7%.

Table 1. Performance comparison of distorted object recognition.

Methods	Recognition Rates (%)
standard SC	95.6
standard SC+HC	97.4
modified SC+HC	98.7

*** SC-Shape Context, HC-Hausdorff Context

6 Summary and Conclusions

We have presented a general framework for hand-written character and digit recognitions based on the shape context with second order derivative and distance measure, and a modified Hausdorff Distance. The corresponding points between two shapes were firstly computed using a modified shape context. The corresponding points on one shape were then aligned to the other shape by TPS model which helps us minimize the matching error. Finally, the distance between two shapes was computed by a new shape similarity measure, the Hausdorff Context. We have shown the results on hand-written character and digit recognitions on 268 images.

References

1. A. C. Evans, N. A. Thacker and J. E. W. Mathew, "Pairwise Representation of Shape," *In Proc. of International Conference on Pattern Recognition*, pp. 133-136, 1992.
2. D. P. Huttenlocher, G. Klanderman and W. Rucklidge, "Comparing Images using the Hausdorff Distance," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 15, No. 9, pp. 850-863, 1993.
3. F. L. Bookstein, "Principal Warps: Thin-Plate Splines and Decomposition of Deformations," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 6, pp. 567-585, 1989.
4. H. C. Fu, H. Y. Chang, Y. Y. Xu and H. T. Pao, "User Adaptive Handwriting Recognition by Self-Growing Probabilistic Decision-Based Neural Networks," *IEEE Trans. on Neural Networks*, Vol. 11, No. 6, pp. 1373-1384, 2000.

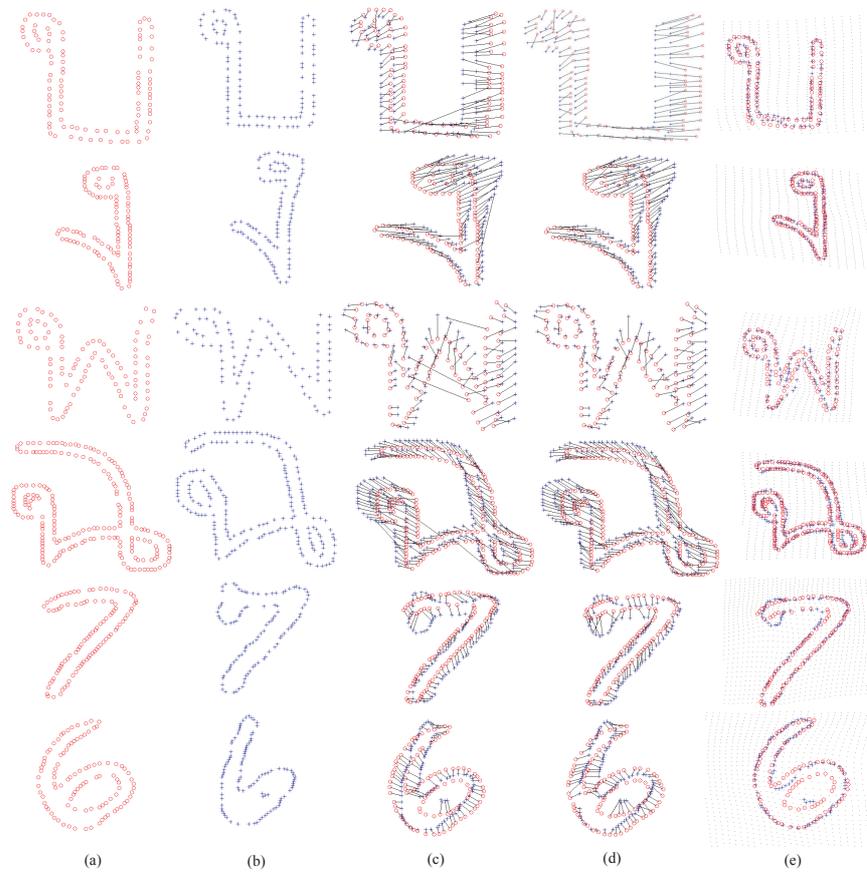


Fig. 5. (a) Sampling points on shape \mathcal{P} in training set. (b) Sampling points on shape \mathcal{Q} in test set. The corresponding points between \mathcal{P} and \mathcal{Q} measured using (c) the standard shape context and (d) a modified shape context. (e) Results after shape \mathcal{Q} has been transformed to shape \mathcal{P} using the regularized TPS model with a modified shape context.

5. L. A. Torres-Mendez, J. C. Ruiz-Suarez, L. E. Sucar and G. Gomez, "Translation, Rotation, and Scale-Invariant Object Recognition," *IEEE Trans. on Systems, Man, and Cybernetics-Part C: Applications and Reviews*, Vol. 30, No. 1, pp. 125-130, 2000.
6. M. Dubuisson and A. K. Jain, "A Modified Hausdorff Distance for Object Matching," *In Proc. of International Conference on Pattern Recognition*, pp. 566-568, 1994.
7. S. Belongie, J. Malik and J. Puzicha, "Shape Matching and Object Recognition using Shape Context," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 24, No. 24, pp. 509-522, 2002.