Prototype Setting for Elastic Matching-Based Image Pattern Recognition

Naoki Matsumoto   Seiichi Uchida   Hiroaki Sakoe
Dept. of Intelligent Systems, Kyushu University,
6-10-1, Hakozaki, Higashi-ku, Fukuoka-shi, Japan
uchida@is.kyushu-u.ac.jp

Abstract

The purpose of this paper is to emphasize the importance the consistency between the distance measures on prototype setting and discrimination in elastic matching (EM)-based recognition. Specifically, this paper focuses on the following points: (i) confirmation of performance degradation when Euclidean distance is used on prototype setting whereas EM-distance is used on discrimination, and (ii) proposal of new prototype setting algorithm where this inconsistency is avoided. Through an experiment of handwritten character recognition, the effectiveness of the proposed algorithm was quantified.

1. Introduction

In image pattern recognition, elastic matching (EM) has been employed (e.g., [1]) to fit a prototype image to an input image pattern. Theoretically, the distance of the two patterns after this fitting, hereafter called EM distance, is invariant to the geometric deformations of the input pattern. Thus, EM is promising for image pattern recognition robust to the deformations.

The setting of prototypes is generally a very important task for image pattern recognition. The prototypes should be set to satisfy the following two conflicting requirements as possible: (i) they should be large enough to cover all input patterns of their class for higher recognition accuracy and (ii) they should be small enough for less computational requirements. In EM-based image recognition, each prototype is further required to be an image.

For the automatic setting of prototypes, clustering has been utilized widely [2]. Simple clustering algorithms, such as k-means, are based on the iteration of two steps; partitioning step and updating step. In the partitioning step, a set of training patterns of a class is partitioned into several subsets, called clusters, according to the distance between the training patterns and (temporary) prototypes. Then in the updating step, the center of gravity of each cluster, called centroid, is selected as a new prototype of the class. In the clustering algorithms specialized for recognition problems, such as GLVQ [3], the centroids are further modified considering the centroids of neighboring classes.

Those conventional clustering algorithms will not be appropriate for setting the prototypes of EM-based recognition. This is due to the *inconsistency* between the distance measures on prototype setting and discrimination. Namely, the above conventional clustering algorithms provide the prototypes optimized under some criterion based on the Euclidean distance. In other words, the conventional algorithms provide the prototypes optimized not for EM distance-based discrimination, but for Euclidean distance-based discrimination.

The purpose of this paper is summarized as follows: (i) confirmation of the degradation of the performance of EM-based recognition due to the above inconsistency, and (ii) proposal of new clustering algorithm for EM-based recognition. In addition, the effectiveness of the proposed algorithm is shown through an experiment of handwritten character recognition.

Most of conventional EM-based image pattern recognition techniques do not pay strict attention to the prototype setting task. For example, in the EM-based character recognition technique of [4], all training patterns are directly used as prototypes regardless of computational complexity. On the other hand, there are a few EM-based recognition techniques considering the task, such as [5]. Those techniques, however, do not emphasize the solution of the inconsistency, and therefore sufficient investigation to reveal the effect of the inconsistency has not been made.

2. Image pattern recognition using elastic matching

Before describing the proposed algorithm for prototype setting, a typical EM-based image pattern recognition procedure is outlined in this section. Let $\boldsymbol{X} = \{x(i,j) \mid i,j = 1,2,\ldots,N\}$ denote an unknown input pattern and $\boldsymbol{R}_k = \{r_k(u,v) \mid u,v = 1,2,\ldots,N\}$ denote the kth prototype

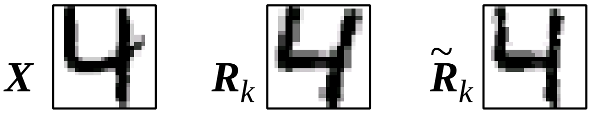


Figure 1. Input X, prototype R_k, and the prototype fitted to the input by EM, $\tilde{R}_k$.

of a class, where N is image size. Then the EM distance $D_{\rm EM}(\boldsymbol{X},\boldsymbol{R}_k)$ is obtained by solving a model-constrained pixel-to-pixel correspondence optimization problem, i.e.,

$$\begin{aligned} & D_{\rm EM}(\boldsymbol{X},\boldsymbol{R}_k) \\ &\quad = \min_{\{(u_{i,j},v_{i,j})\}\in\mathcal{M}} \sum_{i,j} \|x(i,j)-r_k(u_{i,j},v_{i,j})\|, \end{aligned} \tag{1}$$

where $(u_{i,j},v_{i,j})$ denotes the pixel on $\boldsymbol{R}_k$ corresponding to the pixel (i,j) on $\boldsymbol{X}$ and $\mathcal{M}$ denotes the deformation model assumed in EM. Theoretically, the distance $D_{\rm EM}(\boldsymbol{X},\boldsymbol{R}_k)$ is invariant to the geometric deformation compensable by $\mathcal{M}$ and therefore the discrimination using the EM distance as its criterion is expected to provide better recognition performance than the discrimination using the simple Euclidean distance

$$D(\boldsymbol{X},\boldsymbol{R}_k)=\sum_{i,j}\|x(i,j)-r_k(i,j)\|. \tag{2}$$

Let $\tilde{\boldsymbol{R}}_k$ denote the prototype $\boldsymbol{R}_k$ optimally fitted to $\boldsymbol{X}$, i.e., $\tilde{\boldsymbol{R}}_k=\{r_k(\tilde{u}_{i,j},\tilde{v}_{i,j})\}$ where $(\tilde{u}_{i,j},\tilde{v}_{i,j})$ denotes $(u_{i,j},v_{i,j})$ which minimizes (1) . Clearly, $D_{\rm EM}(\boldsymbol{X},\boldsymbol{R}_k)=D(\boldsymbol{X},\tilde{\boldsymbol{R}}_k)$. **Figure 1** shows an example of EM.

3. Anisotropic property of EM distance

The Euclidean distance $D(\boldsymbol{X},\boldsymbol{R}_k)$ is isotropic and therefore the patterns equidistant from prototype $\boldsymbol{R}_k$ form a hyper-sphere in image pattern space. On contrast, the EM distance $D_{\rm EM}(\boldsymbol{X},\boldsymbol{R}_k)$ is anisotropic (i.e., not isotropic), and the equidistant patterns form an irregular (and sometimes disconnected) surface. The shape of this surface depends on the deformation model $\mathcal{M}$.

Figure 2 experimentally ensures the anisotropic property of the EM distance. Each dot in this figure represents a 256 (=16×16) dimensional handwritten character pattern mapped into the two-dimensional subspace spanned by their first two principal axes. The small triangle represents the centroid (the center of gravity) which attains the minimum sum of the distances to the character patterns. When the Euclidean distance is used as the measure, the centroid is located at the center of the dots (**Fig. 2**(a)). In contrast, when the EM distance (given by the EM technique in **Section 6.1**) is used, the centroid is located near the boundary of the distribution (**Fig. 2**(b)).

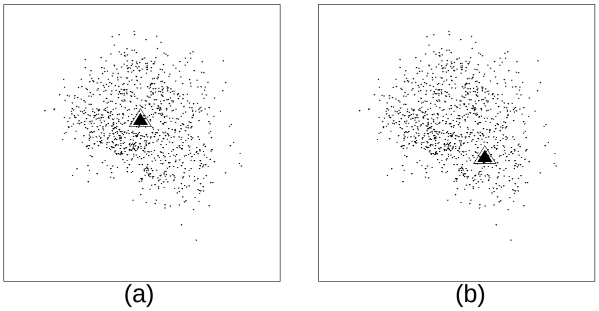
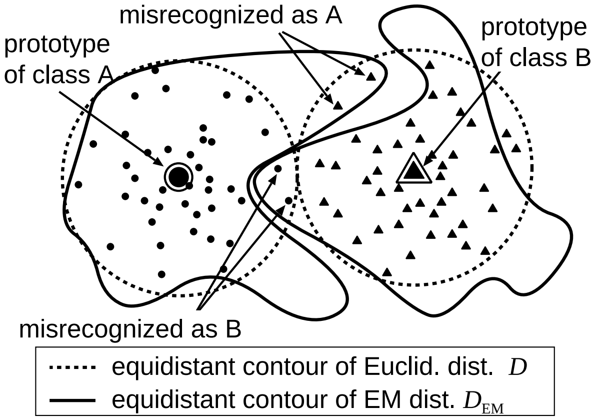


Figure 2. The centroids of handwritten character patterns under (a) Euclidean distance and (b) EM distance. Note that all patterns are displayed in two-dimensional subspace.



misrecognized as A
prototype of class A
prototype of class B
misrecognized as B
equidistant contour of Euclid. dist. D
equidistant contour of EM dist. $D_{\rm EM}$

Figure 3. Equidistant contours showing the reason of misrecognition when prototypes are set under Euclidean distance and discrimination is performed under EM distance.

4. Problem on conventional clustering algorithms

Clustering is one of the most popular strategies for prototype setting. Let $T=\{\boldsymbol{T}_l|l=1,\ldots,L\}$ denote the set of the training patterns of a class, and C_k denotes the kth cluster ($k=1,\ldots,K$), where $\bigcup_k C_k=T$ and $C_k\cap C_{k'}=\emptyset$ for $k\neq k'$. Then the purpose of clustering is to optimize clusters $\{C_k\}$ and prototypes $\{\boldsymbol{R}_k\}$ as a representative pattern of cluster C_k.

In the conventional clustering techniques, the Euclidean distance has been utilized in their criteria. For example, k-means clustering algorithm is based on the minimization of the following criterion (called the sum-of-squared error criterion [6]) with respect to $\{\boldsymbol{R}_k\}$ and $\{C_k\}$:

$$J=\sum_k\sum_{\boldsymbol{T}_l\in C_k} D(\boldsymbol{T}_l,\boldsymbol{R}_k). \tag{3}$$

[Step1: Initialization]
Choose K initial centroids $\boldsymbol{R}_1, \ldots, \boldsymbol{R}_K$ from a training pattern set $\{\boldsymbol{T}_l \mid l = 1, \ldots, L\}$.

[Step2: Partitioning]
For each training pattern $\boldsymbol{T}_l$, find its nearest centroid $\boldsymbol{R}_{k'}$, where $k' = \mathrm{argmin}_k D_{\mathrm{EM}}(\boldsymbol{T}_l, \boldsymbol{R}_k)$. Then assign $\boldsymbol{T}_l$ to cluster $C_{k'}$.

[Step3: Updating]
Each $\boldsymbol{R}_k$ is updated by replacing $r_k(u, v)$ by the average value of (u, v)'s corresponding pixels on $\boldsymbol{T}_l$ in C_k.

[Step4: Convergence check]
If $\boldsymbol{R}_k$ is changed by Step3 go to Step2; otherwise, terminate the algorithm.

Figure 4. The proposed clustering algorithm for setting prototypes $\{\boldsymbol{R}_k\}$.

Thus, the prototype $\boldsymbol{R}_k$ will be set around the center of cluster C_k as shown in **Fig. 2**(a). In more sophisticated clustering algorithms, such as GLVQ, the Euclidean distance is utilized in some manner.

Those conventional clustering algorithms are not appropriate as prototype setting technique for EM-based recognition. This is because of the difference between the Euclidean distance and the EM distance, which is revealed in the previous section. **Figure 3** illustrates the degradation of recognition performance when prototypes are set under the Euclidean distance and discrimination is performed under the EM distance. Namely, the prototypes optimized under the Euclidean distance is optimal for Euclidean distance-based discrimination and not optimal for EM distance-based discrimination.

5. The proposed clustering algorithm

In the proposed clustering algorithm, the EM distance is incorporated on setting prototypes for the discrimination based on the same EM distance. Although many clustering algorithms have the potential to incorporate the EM distance, the k-means clustering algorithm is picked out here because of its simplicity.

Our problem for setting the prototypes $\{\boldsymbol{R}_k\}$ based on the EM distance can be formulated as the minimization problem of the following criterion J_{EM}:

$$J_{\mathrm{EM}} = \sum_k \sum_{\boldsymbol{T}_l \in C_k} D_{\mathrm{EM}}(\boldsymbol{T}_l, \boldsymbol{R}_k). \quad (4)$$

Since this criterion is based on the EM distance, the misrecognition due to the situation of **Fig. 3** can be reduced with the prototypes optimized under this criterion.

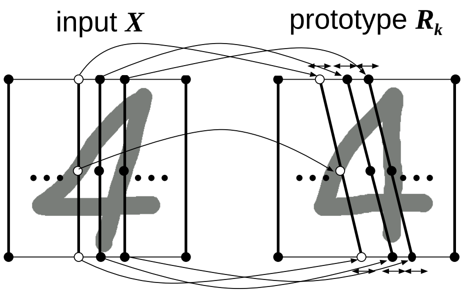


Figure 5. Elastic matching in the experiment.

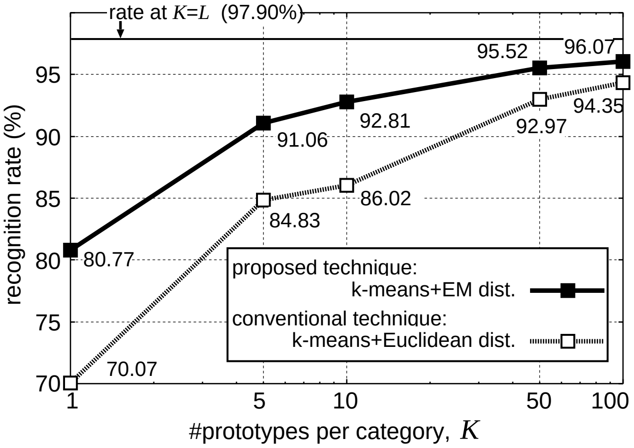


Figure 6. Recognition results of MNIST.

The minimization of J_{EM} is more difficult than that of J, because the calculation of $D_{\mathrm{EM}}(\boldsymbol{T}_l, \boldsymbol{R}_k)$ itself includes the optimization of the pixel-to-pixel correspondence between $\boldsymbol{T}_l$ and $\boldsymbol{R}_k$ as defined in (1). **Figure 4** is a practical algorithm to obtain approximate solution, where centroids $\{\boldsymbol{R}_k\}$, clusters $\{C_k\}$, and pixel correspondences are optimized sequentially. Specifically, in Step 2, clusters $\{C_k\}$ and pixel correspondences are optimized under fixed prototype $\boldsymbol{R}_k$. (Recall that the pixel correspondences are optimized in the calculation of D_{EM}.) In Step 3, prototypes $\boldsymbol{R}_k$ are then updated through the optimization under fixed clusters and pixel correspondences. The validity of the updating procedure of Step 3 is shown in **Appendix A**.

6. Experimental result

6.1. Experimental setup

The standard handwritten numeral database called MNIST (60,000 training patterns and 10,000 test patterns) was used in the experiment to evaluate the prototypes provided by the proposed algorithm. As preprocessing, each sample was scaled into 14×14 and then one-pixel-wide margin was placed around it so that $N = 16$.

A simple EM technique shown in **Fig. 5** was employed, while the proposed clustering algorithm can employ to any EM technique. As its deformation model $\mathcal{M}$, it is assumed that the prototype $\boldsymbol{R}_k$ is fitted to the input pattern $\boldsymbol{X}$ by linearly interpolating the pixel correspondences of the top side and the bottom side. Thus, only $2N$ variables $\{u_{i,1}, u_{i,N} \mid i = 1, \ldots, N\}$ are to be optimized. Those variables can be globally optimized by the DP algorithm of [7].

6.2. Evaluation in recognition performance

Using the prototypes provided by the proposed and the conventional k-means algorithms, a recognition experiment was conducted. The discrimination was based on the minimum-distance discrimination using the EM distance even when the conventional (Euclidean) k-means was used at prototype setting. Note that if all training patterns are used as the prototypes, namely, if $K = L \sim 6,000$, the recognition rate was 97.90% [1].

Figure 6 shows the recognition rates as the function of K. Each recognition rate is the average over 10 trials with different initial patterns. This result shows that the prototypes by the proposed algorithm (indexed as "k-means+EM dist") can provide higher recognition rates than those by the conventional technique (indexed as "k-means+Euclidean dist") [2]. Thus, it is indicated that (i) the consistency of distance measures on prototype setting and discrimination is necessary and (ii) the proposed algorithm is useful for setting the prototypes for EM-based recognition.

The above results will be improved by using the EM distance in more sophisticated clustering algorithms instead of k-means. In fact, the recognition rates by GLVQ with the EM distance were 95.48% and 96.24% at $K = 5$ and 10, respectively, whereas the recognition rates by GLVQ with the Euclidean distance were 90.87% and 93.08%. (The detail of those experiments will be published somewhere.) The use of more sophisticated EM is also promising.

7. Conclusion

A clustering-based prototype setting algorithm was proposed for elastic matching (EM)-based image pattern recognition. In the proposed algorithm, EM distance is newly employed instead of the conventional Euclidean distance in the criterion of prototype optimization in order to avoid the inconsistency between distance measures on prototype setting and discrimination. From handwritten numeral recognition experiments, it was shown that better recognition rates can be attained with the prototypes by the proposed algorithm.

[1] If the Euclidean distance was used as the discrimination measure, this rate remains at 95.70%. This fact shows the usefulness of EM.

[2] As shown in **Fig. 6**, this superiority decreases as K increases. This is quite natural because for a large K, any clustering algorithm becomes less meaningful (most of the clusters would contain only one training pattern) and its recognition rate will converge at the rate on $K = L$.

References

[1] L. G. Brown, "A survey of image registration techniques," *ACM Computing Surveys*, vol. 24, no. 4, pp. 325-376, 1992.

[2] C. -L. Liu and M. Nakagawa, "Evaluation of prototype learning algorithms for nearest-neighbor classifier in application to handwritten character recognition," *Pattern Recognition*, vol. 34, pp. 601-615, 2001.

[3] A. Sato and K. Yamada, "A formulation of learning vector quantization using a new misclassification measure," *Proc. ICPR* , vol. 1, pp. 322-325, 1998.

[4] A. K. Jain and D. Zongker, "Representation and recognition of handwritten digits using deformable templates," *IEEE Trans. PAMI*, vol. 19, no. 12, pp. 1386-1391, 1997.

[5] T. Hastie, et al., "Learning prototype models for tangent distance," *Advances in Neural Information Processing Systems*, vol. 7, pp. 999-1006, 1995.

[6] R. O. Duda and P. E. Hart, Pattern Classification and Scene Analysis, Wiley Interscience, 1973.

[7] S. Uchida and H. Sakoe, "Piecewise linear two-dimensional warping," *Proc. ICPR*, vol.3, pp.538–541, 2000.

A. Updating centroids

The purpose of the Step 3 in **Fig. 4** is to derive centroid $\boldsymbol{R}_k$ which minimizes (4) under fixed clusters $\{C_k\}$ and pixel correspondences. Such $\boldsymbol{R}_k$ should satisfy the following equation:

$$\partial \left(\sum_{\boldsymbol{T}_l \in C_k} \sum_{i,j} \| t_l(i,j) - r_k(u^l_{i,j}, v^l_{i,j}) \| \right) \Big/ \partial r_k(u,v) = 0,$$

where $(u^l_{i,j}, v^l_{i,j})$ represents the pixel correspondence between $\boldsymbol{T}_l$ and $\boldsymbol{R}_k$, provided during the calculation of D_{EM} in Step 2. The above equation can be rewritten as

$$\sum_{\boldsymbol{T}_l \in C_k} \sum_{i,j} \left(t_l(i,j) - r_k(u^l_{i,j}, v^l_{i,j}) \right) \frac{\partial r_k(u^l_{i,j}, v^l_{i,j})}{\partial r_k(u,v)} = 0.$$

Using the relation

$$r_k(u^l_{i,j}, v^l_{i,j}) = \sum_{u,v} r_k(u,v) \delta(u^l_{i,j} - u, v^l_{i,j} - v),$$

we finally obtain

$$r_k(u,v) = \frac{\displaystyle\sum_{\boldsymbol{T}_l \in C_k} \sum_{i,j} t_l(i,j) \delta(u^l_{i,j} - u, v^l_{i,j} - v)}{\displaystyle\sum_{i,j} \delta(u^l_{i,j} - u, v^l_{i,j} - v)}.$$

This equation means that pixel value $r_k(u,v)$ should be updated as the average value of its corresponding pixels on $\boldsymbol{T}_l \in C_k$. Thus, the validity of Step 3 is shown.