BIOMETRIC RECOGNITION USING ENTROPY-BASED DISCRETIZATION

Ajay Kumar[1,2], *David Zhang*[1]

[1] Department of Computing
The Hong Kong Polytechnic University
Hung Hom, Kowloon, Hong Kong

[2] Department of Electrical Engineering
Indian Institute of Technology Delhi
Hauz Khas, New Delhi, India

Email: ajaykr@ieee.org, csdzhang@comp.polyu.edu.hk

ABSTRACT

The biometrics based recognition systems proposed in the literature have not yet exploited user-specific dependencies in the feature level representation. This paper suggests and investigates the performance improvement of the existing biometric systems using the discretization of extracted features. The performance improvement due to the unsupervised and supervised discretization schemes is compared on verity of classifiers; KNN, naïve Bayes, SVM and FFN. The experimental results on the hand-geometry database of 100 users achieve significant improvement in the recognition accuracy and confirm the usefulness of discretization in biometrics systems.

Index Terms**—** Biometrics, Hand Geometry, Personal Recognition, Feature Representation, Feature Discretization.

1. INTRODUCTION

The feature representation in biometric literature has received little attention and the prior work has been quite limited to the usage of normalization schemes for performance improvement. A survey on biometrics literature [1] suggests that there has not been any effort to exploit user-specific dependencies in the feature level representation. The usefulness of discretization schemes is yet to be investigated in biometric based user identification. This paper therefore suggests and investigates the performance improvement using both supervised and unsupervised discretization schemes. The way features should be discretizaed is highly dependent on the classifiers. Therefore the improvement in the recognition accuracy, using different discretization approaches, is ascertained on variety of classifiers.

The discretization of biometric features can offer several advantages. The discrete features are closer to knowledge-level representation than the continuous (nominal) values which may be unstable due to noise or inaccuracies in the feature extraction or image normalization algorithms. The problems due to such perturbations is likely to be smaller for those hand-geometry systems [3] that use user-pegs to constrain the rotation and translation of hand than those systems [4] employing unconstrained peg-free imaging, which highly relies on the efficiencies of the algorithm to achieve illumination, translation and rotation invariant features.

The idea of discretization is to project continuous feature values into discrete ones such that the projection preserves important distinction among different users. Figure 1 illustrates the transformation of continuous feature F into discrete feature F^* with values $\{V_1, V_2, V_3, \ldots, V_n\}$ for $n = 4$. Each value V_i of the new feature F^* represents certain range of numeric values in the original feature F. Discretization methods in the machine learning literature have been categorized into supervised and unsupervised categories. The unsupervised discretization in its simplest form, also known as *equal-interval width*, divides an observed feature value range into k equal size bins where the parameter k is provided by user. Another unsupervised approach for the discretization is to use *equal-frequency intervals*. This approach sorts the available values of a feature and then assigns them to $1/k$ of the values in each bin. The supervised approaches also examine the distribution of class labels (users) and are more likely to give higher accuracies. The supervised discretization allows inter-class feature dependencies to be captured in the feature discretization and thus indirectly promoting accuracy.

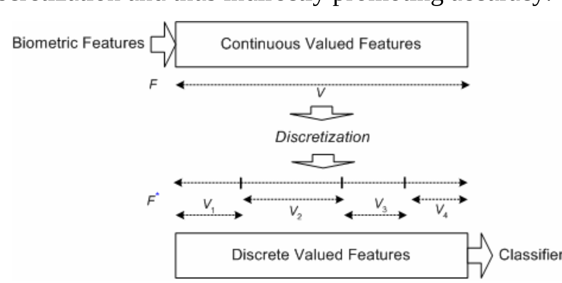

Figure 1: Discretization of biometric features to exploit class-specific dependencies.

2. ENTROPY-BASED DISCRETIZATION

The potential problems with the unsupervised discretization methods is the loss of classification information because of the resulting discretized feature values that are strongly associated with different classes in the same interval [5]. The supervised discretization uses sorted feature values to locate the potential interval boundaries, *i.e*. cut point T, such that the resulting interval has strong majority of one particular class. The cut point for discretization is selected by evaluating favorite disparity measure, *i.e* class entropies, of candidate partitions. The multiple intervals of a feature are computed by recursively applying this algorithm on two intervals of previous split until some stopping criteria is satisfied.

The set S of instances, *i.e*. training samples, of a sorted feature array is firstly partitioned into subset S_1 and S_2. The class entropy of subset S is defined as;

$$\mathrm{Ent}(S) = -\sum_{i=1}^{Z} p(C_i, S)\log_2(p(C_i, S)) \tag{1}$$

where $p(C_i, S)$ is the proportion of samples/instances lying in a class C_i and Z is the total number of classes. The resulting class entropy, due to partition of S into S_1 and S_2 is estimated by weighted average of resulting individual entropies. The class information entropy of the partition induced by a cut point T, for a feature F, is computed as follows [8]:

$$E(F,T;S) = \frac{|S_1|}{S}\mathrm{Ent}(S_1) + \frac{|S_2|}{S}\mathrm{Ent}(S_2) \tag{2}$$

The cutpoint for which $E(F,T;S)$ is minimum amongst all the candidate cutpoints is taken as best cutpoint T_F and determines the binary discretization of feature F. The splitting procedure is recursively applied unless a stopping criterion is reached. The stopping criteria prescribe to accept a partition induced by cutpoint T only if there is any gain after splitting. Thus a partition due to cutpoint T is accepted only if;

$$\mathrm{Gain}(F,T;S) > \frac{\log_2(M-1)}{M} + \frac{\Delta(F,T;S)}{M} \tag{3}$$

where

$$\mathrm{Gain}(F,T;S) = \mathrm{Ent}(S) - E(F,T;S)\text{, and}$$

$$\Delta(F,T;S) = \log_2(3^Z - 2) - [z\mathrm{Ent}(S) - z_1\mathrm{Ent}(S_1) - z_2\mathrm{Ent}(S_2)] \tag{4}$$

The number of samples in set S is denoted as M, and the number of classes present in S_1 and S_2 are z_1 and z_2 respectively.

Another supervised approach for evaluating the worth of the features is to measure the average compression (per sample) of the class afforded by an attribute. Kononenko [6] has shown that this criterion is the most promising on multivalued features among a number of other simple impurity-based measures. This measure, commonly referred as minimum description length (MDL), *i.e*.,

$$MDL(F_i) = \frac{1}{n}\left(\log_2\binom{n}{n_1,\ldots,n_Z} + \log_2\binom{n+Z-1}{Z-1} - \sum_j \log_2\binom{n_{.j}}{n_{1j},\ldots,n_{Zj}} - \sum_j \log_2\binom{n_{.j}+Z-1}{Z-1}\right) \tag{5}$$

where n is the number of training samples, Z is the total number of classes, n_i is the number of training samples from class z_i, $n_{.j}$ is the number of training instances with j-th value of a given feature, and n_{ij} is the number of training samples of class z_i having j-th value of the feature. The first two terms in (5) represent the description length of the class labels prior to partitioning on the values of a feature, while the remaining two terms represents description length after partitioning. The cutpoint for discretization is selected by evaluating gain in MDL, instead of entropy in (3) and (4), from the candidate partitions.

3. EXPERIMENTS

The experimental results reported in this paper investigated the performance gain for the hand geometry biometrics. We acquired the right hand images of 100 users using a digital camera within an interval of 3 months. Each of these users contributed about 5 images in one session and only 10 images from every user were employed in our experiments. The hand images were acquired using simple peg-free imaging setup as detailed in [7]. The acquired images were binarized and employed for feature extraction. The thresholding limit was automatically computed, once for each acquisition setup, using Otsu's approach, and used in subsequent images. We extracted 23 hand geometry features and used in our experiments: 4 finger length (h_1-h_4), 8 finger width (h_5-h_{12}), palm width (h_{13}), palm length (h_{14}), hand area (h_{15}), hand length (h_{16}), perimeter (h_{17}), solidity (h_{18}), extent (h_{19}), convex area (h_{20}), eccentricity (h_{21}), and *x-y* position of centroid relative to shape boundary (h_{22}-h_{23}). The details of these features can be found in [4], [7]. The 5 hand images from each of the users were used for the training and remaining were employed for the testing.

The training samples from the 100 users were subjected to discretization as detailed in section 3. The performance for unsupervised discretization using *equal interval width* and *equal frequency interval* was investigated on four classifiers; *KNN*, *Naïve Bayes*, *SVM and FFN*. The *k*-nearest neighbors were obtained from the minimum Euclidean distance between the query feature vector and those from training samples. The parameters of *SVM* and *FFN* employed in the experiments were empirically selected. The *SVM* using polynomial kernel achieved much better results than those from radial basis function. Therefore to conserve the space only results from polynomial kernel are reported. The *SVM* training was achieved with *C-SVM*, a commonly used *SVM* classification algorithm [4]. The training parameter γ and ε were empirically fixed at 1 and 0.001 respectively. Similarly the number of input nodes in *FFN* were also empirically selected for the best performance; 80. The *FFN* neuron weights were updated using resilient backpropagation algorithm and the training was aborted if the maximum number of training steps reached to 1000. The confusion matrix resulting from the experiments on 500 test samples is quite large to be reproduced in this paper. Therefore we selected following few performance indices, *i.e*. *Kappa Statistic* [9], *Accuracy* and *Precision*, to ascertain the performance improvement.

$$Kappa = \frac{P(A) - P(E)}{1 - P(E)}, \; Accuracy = \frac{TP + TN}{TP + TN + FP + FN} \; \Pr ecison = \frac{TP}{TP + FP} \quad (6)$$

where $P(A)$ is the observed proportion of true positive (TP) and true negative (TN), $P(E)$ is the expected proportion of TP and TN, FN and FP respectively represents the false negative and false positive matches from the test data.

4. RESULTS

The *k*-Nearest Neighbor (*k-NN*) classifier is a simple nonparametric classifier that does not require training and hence most commonly preferred in biometric recognition. Figure 4(a) illustrates its gain in recognition accuracy with the increase in number of bins used to discretize the features in unsupervised approach. We can observe the initial increase in performance and its stabilization in subsequent stages/increase. The equal frequency approach achieves better performance for smaller number of bins while equal width approach outperforms for higher number (crossover of about 16) of bins. Table 1 illustrates comparative performance for *k-NN* classifier using unsupervised and supervised discretization schemes. The average *precision* result in this table closely follows the recognition *accuracy*, except those in the last row which suggest the presence of large false positive matches in absence of any feature discretization. The overall performance indicator *kappa* closely follows the results from recognition accuracy, suggesting the significant increase in performance with the feature discretization. The results in Table 1 suggest that the feature discretization using entropy based heuristics outperforms those based using MDL representation of hand geometry features.

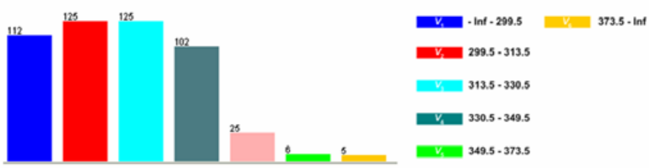

Figure 2: Supervised discretization of feature h_1 from the training samples.

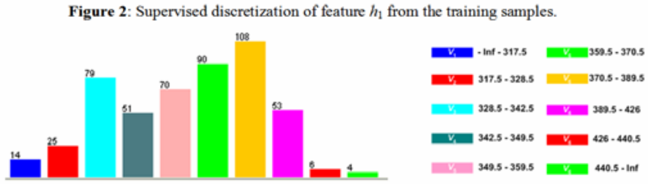

Figure 3: Supervised discretization of feature h_2 from the training samples.

The supervised discretization of features using entropy gain discretized 23 hand geometry features into 181 discrete levels. The discretization of some of these features from the training samples is illustrated in Figure 2. The first feature (h_1, the left finger length) required only 7 discrete levels which can be observed from Figure 2. This Figure also illustrates the distribution of these 7 discrete levels among 500 training samples. Similar partitioning of continuous feature values into discrete ones for the feature h_2 is illustrated in Figure 3. The discretization scheme using entropy-based heuristic, on an average, required 7.87 discrete levels per feature. However the supervised discretization using *equal interval width* does not offer any gain in recognition accuracy if less than 12 bins are employed for discretization. Even with the increase in number of bins the maximum recognition accuracy that can be achieved from *equal interval width* is 92.6% (22 bins) while those from *equal frequency interval* is 92.4% (12 bins), *i.e.*, smaller than that can be achieved from supervised entropy-based approach requiring only 7.87 bins on an average. The feature discretization requirements using MDL were huge, *i.e.* requiring an average of 23 discrete levels per feature, while achieving maximum accuracy of only 89.2%.

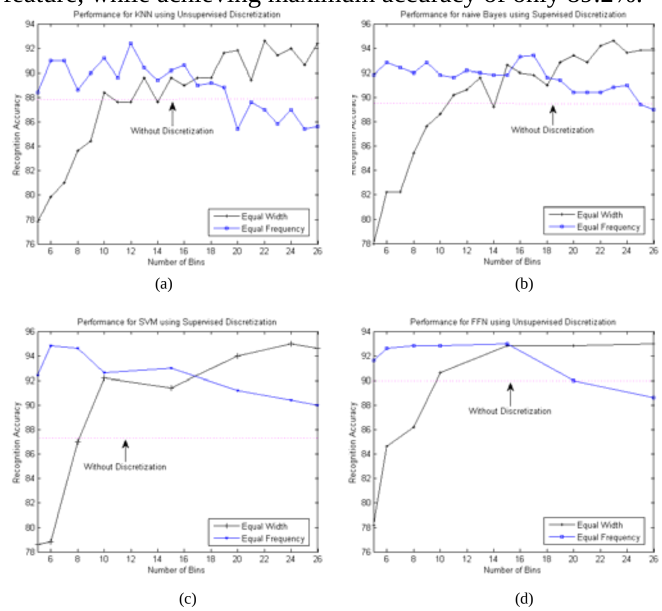

Figure 4: Performance improvement using discretization for (a) KNN, (b) naïve Bayes, (c) SVM, and (d) FFN classifier.

Table 2 illustrates summary of comparative performance achieved from the *Naïve Bayes* classifier. The usage of unsupervised discretization can deliver 5.2% improvement in recognition accuracy for *equal interval width* and 3.4% for *equal frequency interval* discretization. The entropy-based discretization achieves the best performance, *i.e.* recognition accuracy of 94.6%, while employing minimum average number (7.87) of bins for feature representation. The performance indices from the naïve Bayes are slightly better than those from *k-NN* but this comes with the added cost of increased classifier complexity. Table 3 similarly presents comparative results from SVM while Table 4 summarizes results from FFN classifiers. Comparison of Table 1, 2, 3, and 4 suggest that the discretization of features achieves significant increase in performance for these four classifiers while SVM performing best of all with recognition accuracy of 95%.

5. CONCLUSIONS

The experimental results illustrated in previous section suggests that (i) the discretization of hand biometric features achieves significant improvement in the performance (6.1% for *k-NN*, 5.2% for naïve Bayes, 7% for SVM, 4% for FFN), (ii) gradual increase (decrease) in performance for *equal frequency* (*width*) unsupervised discretization with the increase in number of bins and the crossover is between 14-17 bins, (iii) the performance of *equal frequency* interval is much better for smaller number of bins which is computationally attractive for online biometric devices, and (iv) supervised discretization scheme using entropy based heuristics achieves the best overall performance, *i.e*. highest recognition accuracy with smallest average number of bins, and is highly recommended for its usage.

The discretization biometric features can significantly reduces the number of possible values of the acquired continuous features and can be useful for several reasons; the classifier operating on discretized data investigates narrow space of possible hypotheses and thus reduces the likelihood of overfitting, *i.e*. chances of finding complex hypotheses that fits well for the training samples just by chance. Secondly, the discretization accelerates learning because discrete features processed faster, than continuous ones, assuming that the time required for the discretization of continuous features is negligible. Thus in addition to higher recognition accuracy, the discretization significantly reduces the complexity of classifiers than those directly operating on normalized biometric data.

The supervised discretization requires recomputation of discrete intervals (levels) every time a new user is added to the biometric system and therefore highly suitable for those biometric systems in which number of users is fixed, *e.g*. access in buildings and offices. In situations where the number of users varies dynamically, unsupervised discretization can be better alternative to avoid recomputation of discretization intervals with each new user addition. This work has illustrated the benefits of discretization for hand geometry system and its exploitation for other biometric traits, *i.e*. ear, palmprint, face, *etc*, is expected/suggested for performance improvement. The cost effective discretization of continuous biometric features, based on some performance indices (*EER*, *FAR* or *FRR*), can be highly useful in dynamically controlling the performance of biometric systems and is suggested for future work.

Table 1: Comparative performance for *k-NN* with and without feature discretization.

	KNN		
	Accuracy	*Kappa*	*Precision*
Supervised - Entropy	**93.8**	**0.9374**	**93.94**
Supervised - MDL	89.2	0.8909	89.07
Unsupervised - *Equal Frequency*	92.4	0.9232	92.90
Unsupervised – *Equal Width*	92.6	0.9232	92.92
Without Discretization	87.6	0.8747	79.57

Table 2: Comparative performance for **Naïve Bayes** with and without feature discretization.

	Naïve Bayes		
	Accuracy	*Kappa*	*Precision*
Supervised - Entropy	**94.6**	0.9455	94.83
Supervised - MDL	91.2	0.9111	90.68
Unsupervised - *Equal Frequency*	93.4	0.9333	93.11
Unsupervised – *Equal Width*	94.6	0.9455	94.86
Without Discretization	89.4	0.8929	95.20

Table 3: Comparative performance for SVM with and without feature discretization.

	SVM		
	Accuracy	*Kappa*	*Precision*
Supervised - Entropy	**95**	0.9495	95.83
Supervised - MDL	92.2	0.9212	94.81
Unsupervised - *Equal Frequency*	94.8	0.9475	95.82
Unsupervised – *Equal Width*	95	0.9495	95.83
Without Discretization	87.2	0.8707	77.65

Table 4: Comparative performance for FFN with and without feature normalization.

	FFN		
	Accuracy	*Kappa*	*Precision*
Supervised - Entropy	**94**	0.9394	94.29
Supervised - MDL	91.2	0.9111	90.99
Unsupervised - *Equal Frequency*	92.8	0.9273	92.95
Unsupervised – *Equal Width*	93	0.9293	94.21
Without Discretization	90	0.8990	91.45

6. REFERENCES

[1] J. Wayman, A. K. Jain, D. Maltoni, and D. Maio (*Eds*.), *Biometric Systems*: *Technology, Design, and Performance Evaluation*, Springer, 2005.

[2] I. H. Witten and E. Frank, *Data Mining*: *Practical Machine Learning Tools and Techniques with Java Implementations*, Morgan Kaufman, 1999.

[3] R. Sanchez-Reillo, C. Sanchez-Avila, and A. Gonzales-Marcos, "Biometric identification through hand geometry measurements," *IEEE Trans. Patt. Anal. Machine Intell.*, vol. 22, no. 10, pp. 1168-1171, 2000.

[4] A. Kumar and D. Zhang, "Personal recognition using shape and texture," *IEEE Trans. Image Process*, vol. 15, no. 8, pp 2454-2461, Aug. 2006.

[5] D. Dougherty, R. Kohavi, and M. Sahami, "Supervised and unsupervised discretisation of continuous features," *Proc. 12th Inlt. Conf Machine Learning*, pp. 194-202, Tahoe City, CA, Morgan Kaufmann, 1995.

[6] I. Kononenko, "On biases in estimating multi-valued attributes," *Proc. 14th Intl. Joint Conf Artificial Intelligence, IJCAI*, Morgan Kaufmann, pp. 1034-1040, 1995.

[7] A. Kumar, D. C. M. Wong, H. C. Shen, and A. K. Jain, "Personal authentication using hand images," *Pattern Recognition Letters*, vol. 27, pp. 1478-1486, Oct. 2006.

[8] U. M. Fayyad and K. B. Irani, "On the handling of continuous-valued attributes in decision tree generation," *Machine Learning*, vol. 8, pp. 87-102, 1992.

[9] J. Carletta, "Assessing agreement on classification tasks: the kappa statistic," *Computational Linguistics*, vol. 22. no. 2, pp. 249-254, 1996.