

# Entropy based Biometric Template Clustering

Michele Nappi<sup>1</sup>, Daniel Riccio<sup>1</sup> and Maria De Marsico<sup>2</sup>

<sup>1</sup>Biometric and Image Processing Laboratory, University of Salerno, Via Ponte don Melillo, Fisciano, Italy

<sup>2</sup>Department of Computer Science, Sapienza University of Rome, Rome, Italy

Keywords: Biometrics, Clustering, Entropy.

Abstract: Though speed and accuracy are two competing requirements for large scale biometric recognition, they both suffer from large database size. Clustering seems promising to reduce the search space. This can improve accuracy, but may even contrarily affect it by a poor selection of the candidate cluster for the search. We present a novel technique that exploits gallery entropy for clustering. The comparison with K-Means demonstrates that we achieve a better clustering result, yet without fixing the number of clusters a-priori.

## 1 INTRODUCTION

Most research results on biometric identification rely on relatively small datasets. However, in massive applications not only accuracy but also scalability and response time are important. Low response times and high accuracy seldom agree. Moreover, researches have shown that false positives increase geometrically with database size (Maltoni et al., 2003). False Acceptance Rate (FAR) depends on algorithms, and on a trade-off with False Rejection Rate (FRR), so it is impossible to reduce it indefinitely. Different possibilities of performance improvement are rather related to the database, and to the size of the search space. Feature space reduction aims at faster matching operations (e.g. see (Singh, 2009)). On the other hand, clustering (or binning) aims at reducing the search space. Unfortunately, biometric databases do not lend themselves to a natural grouping/ordering of templates, so that the latter is a challenging problem.

## 2 ENTROPY AND BIOMETRICS

Some recent works (Bhatnagar and Kumar, 2009) demonstrate that biometrics can exploit models from Information Theory. Capture and feature extraction modules can be modeled as signal noisy sources, while a matcher /classifier can be considered as a decoder on a noisy channel. In the specific case of a biometric system, Shannon entropy can measure the difference of a subject from a population using

features extracted by a Feature Extraction Technique (FET). The easiest way to integrate entropy into a biometric systems, is to use it as an estimate of the degree of randomness of pixels in the image  $I$  of a single sample. However, a more profitable entropy-based analysis can allow relating the discriminant power of the templates of a subject with those of other subjects (De Marsico et al., 2012), or to find subsets with low informative variation, which in our approach correspond to clusters of similar templates.

We consider a gallery  $G$  of templates, a feature extraction technique  $F$ , a template similarity measure  $d$ .  $F$  takes a sample image  $I$  as input, and produces a template  $v$  as output, i.e.  $v=F(I)$ . The similarity measure  $d$  associates a real scalar value to a pair of templates. We first compare a probe template  $v$  with a gallery template  $g_i$ . We get  $d(v, g_i)$  and denote it as  $s_{i,v}$ . After a possible score normalization,  $s_{i,v}$  is a real value in the interval  $[0,1]$ .

We assume that an oracle (e.g. a matching algorithm) has already decided that the template  $v$  belongs to a specific subject  $k$ . Therefore, we can assume a probability distribution over the sub-gallery  $G_k$  such that the score  $s_{i,v}$  can be interpreted as the probability that template  $v$  conforms to  $g_i$ :

$$s_{i,v} = p(v \approx g_i) \quad (1)$$

In order to represent such a probability,  $s_{i,v}$  must range in  $[0,1]$ , and is normalized with respect to  $\sum_i(s_{i,v})$  so that the sum over  $G_k$  is 1.

The entropy definition can be now applied to the gallery  $G_k$  with respect to probe  $v$  as follows:

$$H(G_k, v) = -\frac{1}{\log_2(|G|)} \sum_{i=1}^{|G|} s_{i,v} \log_2(s_{i,v}) \quad (2)$$

We can next compute a measure of entropy for the whole gallery  $G$ , by considering each gallery template  $g_j$  in turn as a probe. Given  $Q$  the set of pairs  $q_{i,j}=(g_i, g_j)$  of elements in  $G$  such that  $s_{i,j}>0$ :

$$H(G) = -\frac{1}{\log_2(|Q|)} \sum_{q_{i,j} \in Q} s_{i,j} \log_2(s_{i,j}), \quad (3)$$

The proposed formulation provides values in the range  $[0,1]$  irrespective of the size of the gallery.

$H(G)$  represents a measure of heterogeneity for  $G$  that can be used to order the gallery according to the informative power of the samples. The proposed procedure takes a gallery  $G$  as input; it computes an all-against-all similarity matrix  $M$ , where  $M(i,j)=d(g_i, g_j)$ ,  $\forall g_i$  and  $g_j \in G$ , and the value for  $H(G)$ . For each  $g_i \in G$ ,  $M$  is used to compute  $H(G \setminus \{g_i\})$  that would be obtained by considering  $g_i$  as a new sample  $v$ , not already in  $G$ . The  $g_i$  achieving the minimum  $f(G, g_i)=H(G) - H(G \setminus \{g_i\})$  is selected;  $M$  is updated by deleting the  $i$ -th row and column, and the process is repeated, until all elements of  $G$  have been selected. In practice, we select the templates in descending order of representativeness. The inhomogeneity of the set progressively reduces, and we use this to identify clusters of templates with similar information content. An example of the ordering of a set of templates is given in Figure 1.

### 3 E-AC CLUSTERING

Our clustering problem is nontrivial. Representation used in face recognition algorithms such as eigenfaces (Turk and Pentland, 1991) or graph-based approaches (Wiskott et al., 1996) do not explicitly encode suitable information. Moreover, the templates in each cluster (bin) have to be significantly less than the total number. This involves a trade-off between search space reduction and bin-miss errors.

Among clustering techniques, K-Means Clustering is very popular. However, among its drawbacks, it needs to fix  $k$  a priori, which is a problem in itself. We propose a method based on the entropy of a set of templates, which does not require to fix the number of classes. The dataset is partitioned according to the information content of the single template with respect to the gallery, so that templates with similar information are grouped together. The proposed technique performs better

than K-Mean, when compared with a *gold standard*, as we will show through experimental results.

The robustness of the characteristics that a Feature Extraction Technique (FET) extracts from a biometric sample, can significantly influence the performance of a biometric classifier. A *Distance Matrix* (or symmetrically, a *Similarity Matrix*) contains the distance between each pair of templates in a set. A classifier puts two templates in the same class if their distance is below a given threshold.

We call the algorithm for biometric clustering Entropy based Aggregative Clustering (E-AC). It applies to the input biometric templates the entropy-based function  $f$  in Section 2, to sort them in a list  $L$  according to representativeness. The templates in the terminal part of the list (the last selected ones) are very similar to each other, and they can possibly fit into a single cluster. The procedure creates a new cluster and the last  $m$  elements of  $L$  are moved into it. The remaining elements of  $L$  are then considered backwards. The last one is compared with those already in the new cluster, using the Pearson's correlation index ( $-1$  = maximum negative correlation;  $0$  = no correlation;  $1$  = maximum direct correlation). If at least 30% of the comparisons provide a correlation greater than 0.8, the element is moved into the cluster too, and the new last one is considered. The insertion stops, when an item does not meet the condition to be inserted.

The procedure is repeated, after that residual templates are reordered, since template removal generally changes the entropy of the set. Neither the number of clusters nor the number of items in each cluster are fixed a priori. The problem of templates of the same class which are placed in different clusters is solved by a further aggregation phase.

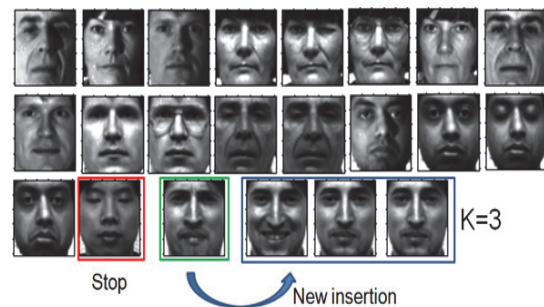


Figure 1: An example E-AC clustering. We here create a cluster of size  $m=3$ . The immediately preceding template satisfies the insertion criterion, while the further preceding one does not and therefore causes a stop, the creation of a new ordered list  $L'$  and the initialization of a new cluster.

The aggregation process computes a (Pearson) correlation matrix for each pair of clusters  $C_h$  and  $C_k$

from the previous phase. The *Correlation Matrix*  $CM$ , size  $|C_h| \times |C_k|$ , is such that  $CM(i, j) = \text{corr}(t_i, t_j)$ ,  $\forall t_i \in C_h$  and  $t_j \in C_k$ . Given a positive threshold  $\tau$  (here 0.2, 0.3, and 0.4 have been used), fixed in advance, E-AC evaluates the percentage of entries in  $CM$  with a correlation coefficient of at least 0.8. If this percentage exceeds the threshold  $\tau$ , the two clusters are merged. The process stops when no pair of clusters can be further merged.

Figure 1 shows the starting phase of the described process on an ordered list of templates.

## 4 EXPERIMENTAL RESULTS

Clustering algorithms were tested with a subset of faces from FERET (Phillips et al., 1998). We selected 35 subjects with at least 6 frontal images labeled  $fa$  or  $fb$ , according to the FERET protocol. The subset contains 366 images with  $256 \times 384$  resolution and 8 bits depth.

The samples were segmented and normalized through an automated process, which cuts and scales it with respect to the interocular distance.

We considered Principal Component Analysis (Kirby and Sirovich, 1990), Linear Discriminant Analysis (Zhao and Yuen, 2008), and Face Analysis for Commercial Entities (FACE) (De Marsico et al., 2012). These three FETs were chosen for their different robustness to pose, illumination and expression (PIE) distortions. For feature vectors produced by PCA and LDA,  $d$  measure was the city block distance (norm 1), and was also used in the calculation of clusters with the K-Means technique. For FACE biometric templates,  $d$  was a localized version of the index of correlation (more details about this in (De Marsico et al., 2012)).

Each subject in FERET database is identified by a label. Therefore, the benchmark set can be considered as a sort of *gold standard*. This allows implementing standard procedures for *external evaluation* in testing the considered clustering algorithms. Notice that, given the nature of the clustering used for comparison, achieving a good similarity with it also implies to be less affected by the bin miss problem. We compared the performance of our clustering algorithm with that of K-Means (adopting its MATLAB implementation) on the considered FETs. Besides the number of clusters, we based the comparison on *Rand Measure* (RM) (Rand, 1971), and on *Fowlkes–Mallows Index* (FMI) (Fowlkes and Mallows, 1983) defined as the geometric average of Precision and Recall:

$$FMI = \sqrt{P \cdot R}. \quad (4)$$

Precision and recall concentrate the evaluation on the true positives, asking what percentage of the relevant elements have been correctly classified and how many false positives have also been returned. When comparing two clusterings  $C_1$  and  $C_2$ , we can consider as true positives the points that are present in the same cluster in both clusterings, false positives as the points that are present in the same cluster in  $C_1$  but not in  $C_2$ , false negatives as the points that are present in the same cluster in  $C_2$  but not in  $C_1$ , and as true negatives the points that are in different clusters in both  $C_1$  and  $C_2$ .

The first experiment evaluates the performance of K-Means, when the face samples are points in the space generated by PCA or LDA. This experiment was not carried out with FACE, due to the different distance function and of the particular comparison algorithm. K-Means requires the number of clusters, which has been set to the number of subjects from FERET, i.e. 35. The comparison was performed with the clustering induced by the database labeling, which is the true correct one (our *gold standard*). Table 1 shows that the robustness of the FET heavily influences the performance of the K-Means algorithm. This phenomenon is confirmed by the different values of Precision / Recall with PCA and LDA, which reach 0.3846 and 0.3796 in the first case, against 0.4822 and 0.4309 in the second.

Table 1: Comparison of K-Means, with different FETs, and the database clustering induced by included labels.

FET	RM	FMI
PCA	0.9667	0.3821
LDA	0.9717	0.4558

The second experiment evaluated the algorithm E-AC. Also for this experiment, the three FETs have been considered as a starting point for the generation of the initial set of clusters, which were then input to the merging procedure. Even in this case, the comparison is with the database labelling. Results of this experiment are summarized in Table 2.

Table 2: Comparison of E-AC clustering using different FETs with the database labeling.

FET	RM	FMI
PCA	0.9729	0.4121
LDA	0.9736	0.4087
FACE	0.9777	0.5227

We can observe how a more robust FET allows entropy to provide a better ordering of templates in the input set. As a consequence, it is possible to

detach from the tail of the ordered list a series of template sequences which actually belong to the same class. This increases the uniformity and consistency of initial clusters, allowing a better result after the merging procedure. In fact, there is no implementation of a process of cluster split or deletion of an item from a cluster: when a template is added to a cluster, it is never removed from it. If the initial clusters, produced by E-AC are very heterogeneous (i.e. contain templates belonging to different classes), the final result, will be hopelessly affected by this. In a final experiment, the merging step was reapplied to the final clustering with increasing thresholds, in an iterative way. With LDA, the procedure initially generated 104 clusters, to which merging was applied with threshold 0.2 to obtain the first final clustering (the procedure used for Table 2). The merging procedure was applied again to the set of clusters obtained so far, with a higher threshold 0.3, and again to the set of clusters obtained with a threshold 0.4. Table 3 shows the performance in terms of number of clusters and performance indices for the various iterations. The same procedure was applied with FACE obtaining the results in Table 4.

Table 3: Performance indices for the sequence of iterations of the merging step when E-AC is applied with LDA FET.

FET	Cl.s	RM	FMI	P	R
it-0	104	0.9736	0.4087	0.84	0.20
it-1	84	0.9729	0.3927	0.79	0.19
it-2	60	0.9771	0.5077	0.78	0.33
it-3	45	0.9806	0.5856	0.71	0.48

Table 4: Performance indices for the sequence of iterations of the merging step when E-AC is applied with FACE FET.

FET	Cl.s	RM	FMI	P	R
it-0	62	0.9777	0.5227	0.81	0.34
it-1	50	0.9813	0.6237	0.79	0.49
it-2	39	0.9833	0.6790	0.71	0.65

Table 3 and Table 4 show that the different applications of the merging procedure consistently fuse together the clusters with similar templates, as indicated by the growth of the value of Recall. However, the reduction of Precision shows that merging may put elements of different classes within the same cluster.

E-AC is slightly slower than K-Means, but this can be fixed by suitable computation optimizations.

## 5 CONCLUSIONS

Clustering is a promising solution to address the problem of biometric recognition with a large scale database. K-means Clustering is a very popular technique to address the problem, but needs the parameter  $k$  a-priori. Our technique achieves better results even without this information.

## REFERENCES

- Bhatnagar J., Kumar A.(2009). On Estimating Some Performance Indices for Biometric Identification. *Pattern Recognition*, vol. 42( 5), pp. 1805-1818.
- De Marsico M., Nappi M., Riccio D. (2012). Entropy in Biometric Face Template Analysis. In: Campilho, Aurélio and Kamel, Mohamed (eds.) *Proceedings of International Conference on Image Analysis and Recognition – ICIAR 2012. Lecture Notes in Computer Science*, Vol. 7325, pp. 72-79.
- De Marsico M., Nappi M., Riccio D., Wechsler H. (2012). Robust Face Recognition for Uncontrolled Pose and Illumination Changes. *IEEE Trans. on Systems, Man and Cybernetics, Part A: Systems and Humans*, vol.PP, no.99, pp.1-15, doi: 10.1109/TSMCA.2012.2192427
- Fowlkes E. B., Mallows C. L. (1983). A Method for Comparing Two Hierarchical Clusterings. *J. of the American Statistical Association*, 78(383), 553–569.
- Kirby M., Sirovich L. (1990). Application of the Karhunen Loeve procedure for the characterization of human faces. *IEEE Trans. on Pattern Analysis and Machine Intelligence* Vol.12, pp.103–108.
- Maltoni D., Maio D., Jain A. K., Prabhakar S. (2003). *Handbook of Fingerprint Recognition*. Springer.
- Phillips P. J., Wechsler H., Huang J., Rauss P.(1998). The FERET Database and Evaluation Procedure for Face Recognition Algorithms. *Image and Vision Computing Journal*, Vol. 16(5), pp. 295-306.
- Rand W. M. (1971). Objective criteria for the evaluation of clustering methods. *J. of the American Statistical Association* Vol, 66(336), 846–850.
- Singh, J. K. (2009). A Clustering and Indexing Technique suitable for Biometric Databases. MSc Thesis, Indian Institute Of Technology Kanpur, Kanpur, India.
- Turk M., Pentland A. (1991) Eigen Faces for Recognition. *J. of Cognitive Neuroscience*, Vol. 3(1), pp. 71-86
- Wiskott, L., Fellous, J. M., Krüger, N., von der Malsburg, C. (1996). Face Recognition by Elastic Bunch Graph Matching. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol. 19(7,) pp.775-779
- Zhao H., Yuen P.C. (2008). Incremental linear discriminant analysis for face recognition. *IEEE Trans. on Systems, Man and Cybernetics – Part B: Cybernetics*. Vol. 38, pp. 210–221