Intelligent Three Step Sampling Design for Face Recognition

Yanjun Yan and Lisa Ann Osadciw

Department of Electrical Engineering and Computer Science
Syracuse University

Motivations

- Scalablility: huge user group ⇒ worse accuracy
- *Learning curve*: performance(all data) ≈ performance(enough data)
- Over-training ⇒ overfitting ⇒ worse generalizability and scalability
- Clustering: a big multi-array recognition problem ⇒ smaller ones [LT07]

Procedure

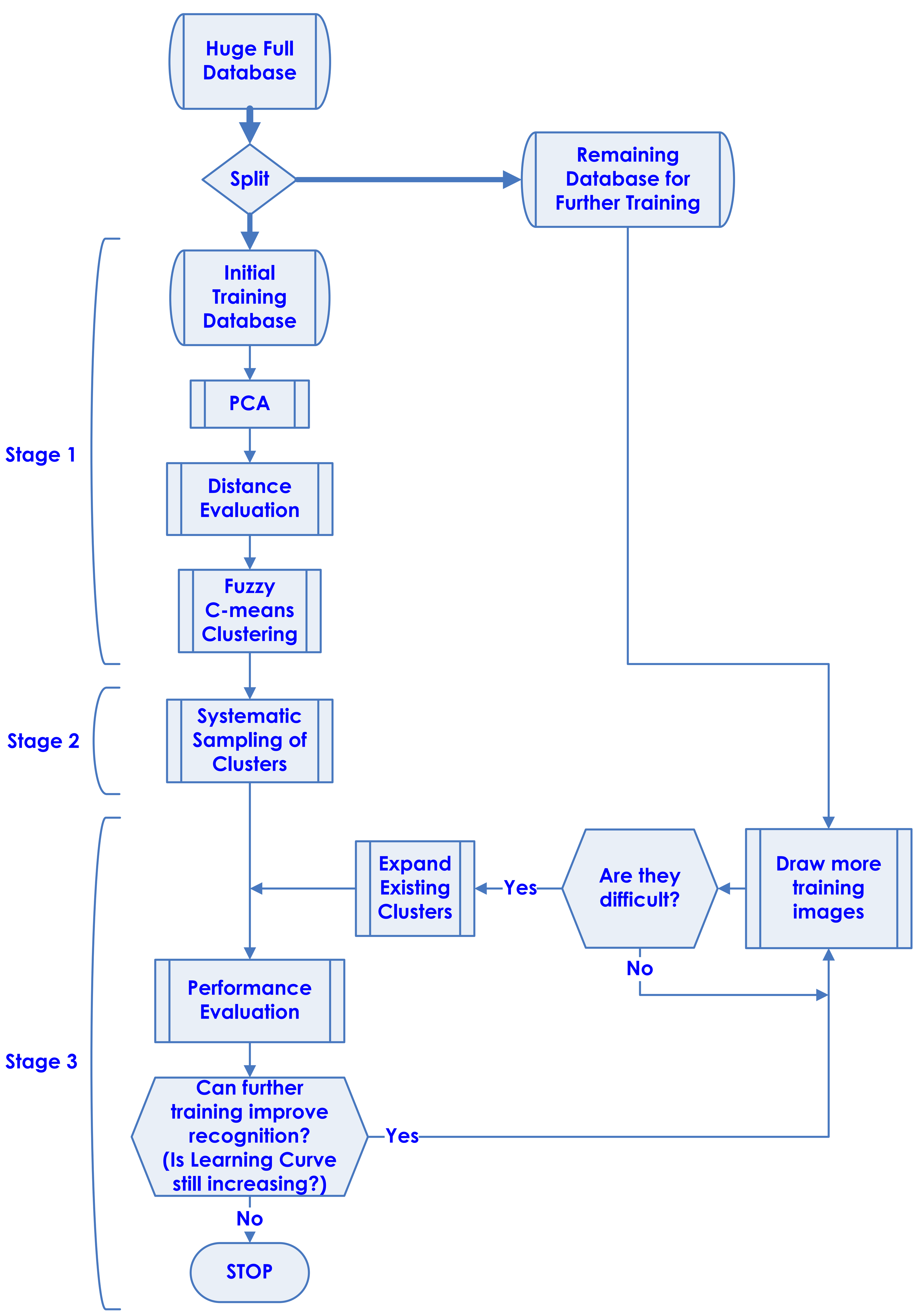

Figure 1: Flowchart of the three step intelligent sampling

Details

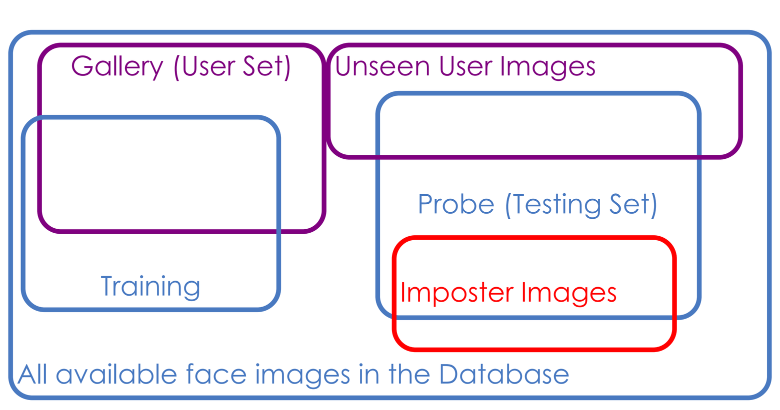

Figure 2: FERET protocol in database partition

A face image, $f_{i,s,c}$, is the ith image ($i \in [1, n_{s,c}]$) of subject s ($s \in [1, S_c]$) in cluster c ($c \in [1, C]$). The difficulty to recognize an image, $D(f_{i,s,c})$, is defined as

$$D(f_{i,s,c}) = \frac{\sum_{\substack{t=1\\t\neq s}}^{S_c} \sum_{j=1}^{n_{t,c}} exp[-Dist(f_{i,s,c},\ f_{j,t,c})]}{\sum_{\substack{t=1\\t\neq s}}^{S} \sum_{j=1}^{n_{t,c}} 1} + \frac{\sum_{\substack{d=1\\d\neq c}}^{C} \sum_{j=1}^{n_{s,d}} \{1 - exp[-Dist(f_{i,s,c},\ f_{j,s,d})]\}}{\sum_{\substack{d=1\\d\neq c}}^{C} \sum_{j=1}^{n_{s,d}} 1}. \quad (1)$$

The first term is caused by all the images from other subjects (index t) within the same cluster c. The denominator is the number compared with current image $f_{i,s,c}$. The second term is caused by the images from the same subject s but in other clusters (index d), rare but more detrimental.

The true derivative of recognition accuracy is d, with empirical value $\hat{d}$. By Chebyshev inequality, we have $Prob(|\hat{d} - d| \geq \varepsilon) \leq \frac{var(\hat{d})}{\varepsilon^2}$, where $var(\hat{d})$ is estimated by accuracy variance σ^2/m (m is the number of images in current set). If the plateau is reached, namely, $d = 0$, and the probability that $\hat{d}$ is ε away from d is limited to be no larger than δ, then

$$Prob(|\hat{d}| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2 m} \leq \delta. \quad (2)$$

When $|\hat{d}| \geq \varepsilon$ and $\frac{\sigma^2}{\varepsilon^2 m} \leq \delta$, the expansion is stopped. Otherwise when $|\hat{d}| < \varepsilon$ and $\frac{\sigma^2}{\varepsilon^2 m} > \delta$, the learning curve is not converged, and more samples should be added. Rewrite the second half of Inequality (2), $m \geq \frac{\sigma^2}{\varepsilon^2 \delta}$, and the number of images used in next training should be m. If current set contains m' images, then $m - m'$ images are added. At inconsistent instances the expansion is continued for now.

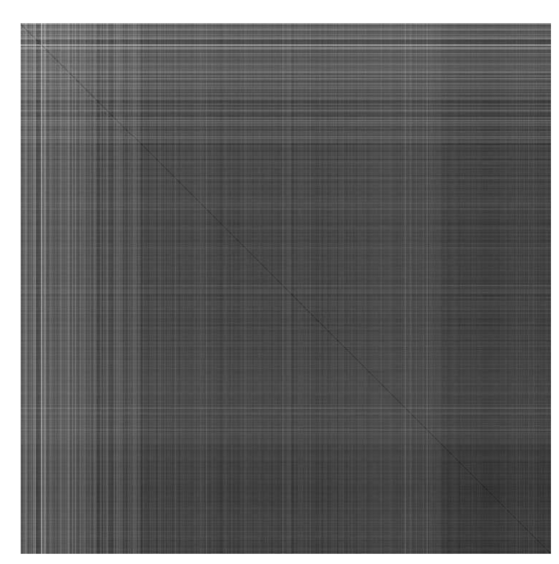

Figure 3: Distance matrix of full training dataset

Results

Recognition Rate (%)			
		by PCA	by DFLDA
Without Fusion	Face	79.375	94.286
	Eyes	61.875	70.714
	Nose	58.125	67.123
	Mouth	51.875	64.286
	Forehead	73.75	73.571
With Fusion	Score Fusion	**90.625**	**99.286**
	Decision Fusion	**88.75**	**97.857**

Table 1: Face Recognition Rate

The face recognition algorithm used in experiment is subspace based uni-modal modular processing. However, the proposed intelligent sampling scheme works at the training set construction phase, and it works for any learning algorithm without system overhaul.

References

[LT07] Zhifeng Li and Xiaoou Tang. Using support vector machines to enhance the performance of bayesian face recognition. *Information Forensics and Security, IEEE Transactions on*, 2(2):174–180, June 2007.