Face Recognition on Low Quality Surveillance Images, by Compensating Degradation

Shiva Rudrani and Sukhendu Das

Department of Computer Science and Engineering,
Indian Institute of Technology Madras, Chennai-600036, India.
{srudrani,sdas}@cse.iitm.ac.in

Abstract. Face images obtained by an outdoor surveillance camera, are often confronted with severe degradations (e.g., low-resolution, low-contrast, blur and noise). This significantly limits the performance of face recognition (FR) systems. This paper presents a framework to overcome the degradation in images obtained by an outdoor surveillance camera, to improve the performance of FR. We have defined a measure that is based on the difference in intensity histograms of face images, to estimate the amount of degradation. In the past, super-resolution techniques have been proposed to increase the image resolution for face recognition. In this work, we attempt a combination of partial restoration (using super-resolution, interpolation etc.) of probe samples (long distance shots of outdoor) and simulated degradation of gallery samples (indoor shots). Due to the unavailability of any benchmark face database with gallery and probe images, we have built our own database[1] and conducted experiments on a realistic surveillance face database. **PCA** and **FLDA** have been used as baseline face recognition classifiers. The aim is to illustrate the effectiveness of our proposed method of compensating the degradation in surveillance data, rather than designing a specific classifier space suited for degraded test probes. The efficiency of the method is shown by improvement in the face classification accuracy, while comparing results obtained separately using training with acquired indoor gallery samples and then testing with the outdoor probes.

Keywords: Degradation, Face database, Face recognition, Image quality, Surveillance, Super-resolution.

1 Introduction

The goal of an intelligent surveillance systems is to accurately identify "Who is Where?". Face recognition has become more attractive than ever because of the increasing need for security. In a typical surveillance scenario, images used for training a face recognition (FR) system might be available beforehand from sources such as passport, identity card, digital record etc., these snaps are taken under well controlled environment in indoor setup (laboratory, control room) whereas testing images are available when a subject comes under a

[1] This database will be made available in public for research and academics purposes.

surveillance scene. Images obtained by surveillance security cameras are often confronted with degradations (e.g., low-resolution, low-contrast, blur and noise). These degradations are due to environmental conditions, interface circuitry (IP, analog camera) or camera's hardware/software limitations. Recognition accuracy of current intensity-based FR systems significantly drop off, if facial images are of low quality. Most face recognition systems [2][13][1][3], have been shown to work in controlled environments where both training as well as testing samples are acquired in similar controlled illumination conditions in indoor environments.

With ever increasing demands to combine "security with surveillance" in an integrated and automated framework, it is necessary to analyze samples of face images of subjects acquired by a surveillance camera from a long distance (≥ 50 yards).

Estimating the degradation parameters is an important problem because it allows better estimate of lost information from the observed image data. Once these parameters are estimated, it can be useful in recognition in two ways: we can either simulate the degradation on good quality images or apply an inverse process on low quality images to enhance them. We have adapted the former approach and have used the estimated blur parameter to simulate the degradation on good quality (i.e. training) images.

The rest of the paper is organized as follows: A brief on data acquisition and description is given in Section 2. The proposed framework is presented in Section 3. In this section, we describe the steps involved in the parameter estimation process. In Section 4, we give experimentation details and present the results of two baseline classifiers, namely **PCA** and **FLDA** [2][4] for different cases of training and testing.

2 Outdoor Surveillance Setup Used for Data Acquisition

In an outdoor surveillance scenario, samples used for training the classifier are called gallery images whereas those used for testing are probe images. Therefore, gallery is the term used for a part of the database of faces, obtained in a controlled (indoor) environment for different persons (subjects), whereas probe images are the face image samples obtained from video frames acquired in an uncontrolled (outdoor) environment using a surveillance camera. The outdoor images are captured from a distance of 50m-100m, placing the camera at around 20m-25m of elevation. The face regions were extracted from the video frames using the popular Viola-Jones face detector (**V-J-F-D**) [10], which is considered as a benchmark in the area of work on face detection. Figure 1 shows a typical example of indoor and outdoor scenes where face templates are enclosed with rectangles. The enclosed face templates clearly depict the difference in resolution of the two face images. It is clear from the figure that, the acquired probe images have severe degradation besides low resolution. The complexity of the problem is evident from the degradation (large change in resolution and contrast along with blur and noise) of the outdoor (probe) with respect to that in indoor (gallery) shots.

Fig. 1. Sample frames from indoor and outdoor videos. The rectangular template around the face indicating the spatial extent of the face, as detected using **V-J-F-D**[10]: (a) Frame from indoor video, (b) Frame from outdoor video of the same subject.

To our knowledge, it is an unique database to provide face images acquired from long distance (between subject and sensor), for biometric applications in surveillance scenario. Data acquisition is done to simulate a typical surveillance system and no special equipment is used to magnify or enhance the outdoor images. Thus, this database is an useful resource to research community.

3 Proposed Framework

The proposed frame work has two stages. In the first stage we estimate the degradation parameter and in the second stage we do recognition for different cases of training and testing. By different cases of training and testing we mean the different combinations of types of face data used for training and testing. Figure 2 shows the proposed framework, where videos obtained from the cameras (indoor for gallery and outdoor for probes) are fed to the **V-J-F-D**. As described in the previous section, face images are stored in gallery or probe database depending on whether it is detected from indoor or an outdoor scene. These stored gallery and probe images are used to estimate degradation parameters by proposed technique described in the next section. After these parameters are estimated, the gallery images are degraded with these parameters to produce (simulated) degraded images of different resolutions which are used for training the FR system in second stage. To define different experimental cases used in later stage of the proposed frame work, we first introduce some symbols to denote the type of training and testing samples. These symbols and the corresponding sample details are presented in Table 1. We also attempt to solve the ill-posed problem of image enhancement and restoration for probe images. We have used the gray level images for our experimentation. This decision is well motivated by the fact that outdoor images are of low contrast which barely contains any useful color information.

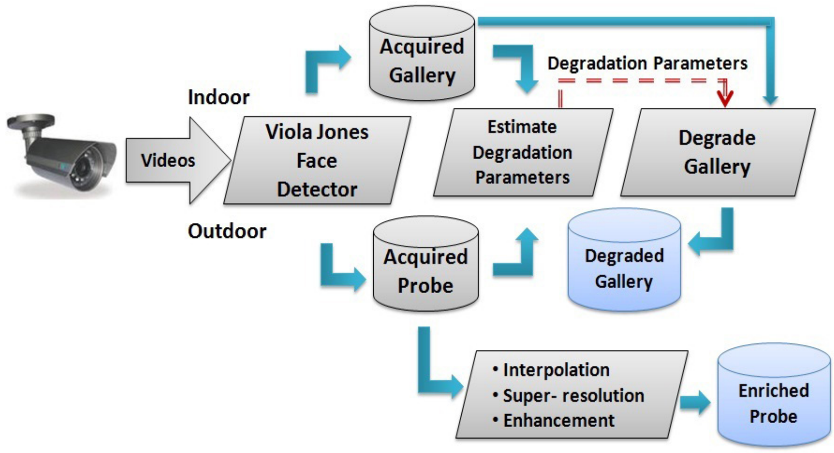

Fig. 2. The proposed framework for estimating and compensating the degradation in face images acquired under surveillance conditions.

Table 1. List of acronyms for face data at different resolutions and intermediate stages of processing

Data	***Abbreviation***	***Resolution***	***Sample description***
Gallery	AG	250x250	Acquired gallery
Downsampled Gallery	LRG	45x45	Low resolution gallery
	LRDG		Low resolution degraded gallery
	MRG	90x90	Medium resolution gallery
	MRDG		Medium resolution degraded gallery
Probe	AP	45x45	Acquired probe
Up-Sampled Probe	INTP	90x90	Interpolated probe
	SRP		Super-resolved probe

3.1 Estimating and Simulating the Degradation

This section deals with the detailed description of the degradation estimation process. To estimate the degradation, we have defined a measure that is simple, intuitive and based on gray level intensity value of images. This leads to the easiness of implementation and thus depicts a strong characteristic of the proposed methodology. In this work, we have consider the degradation due to blurring. A typical formulation of the degraded image $p(x, y)$ in the spatial domain and its relation with the ideal image $g(x, y)$ is given by the following [5]:

$$p(x, y) = h(x, y) * g(x, y) + n(x, y) \tag{1}$$

where, $h(x, y)$ is the point-spread function (**PSF**), '*' denotes the 2D convolution and $n(x, y)$ is the additive noise. Our objective is to obtain an estimate of

$h(x, y)$ and then use them for improving the accuracy of face recognition. In this direction, we have proposed an empirical method to estimate the degradation parameter for blur **PSF**. Later, this estimated parameter is used to degrade the acquired gallery images so that they appear close to the corresponding probe images. In this way, we obtain a set of (simulated) degraded gallery images at different resolutions (details are given in Table 1) that are later used for training in face recognition. In our experiment, we have assumed that, the nature of blur is Gaussian and parameter to be estimated is standard deviation (σ) for the Gaussian function.

We start with downsampling the gallery images. Downsampling step is shown in Fig. 3(a). This step is required in order to compensate for the difference in the resolution of detected gallery and probe faces - gallery faces having more resolution. This difference is due to the fact that gallery images are taken in an indoor environment (close range) while the probe images are taken from a distant outdoor surveillance camera. We mention again that we have also tried the super-resolution technique to obtained higher resolution probe images. Although, use of super-resolution in an automated way (without human intervention) on the free form face images (**V-J-F-D** output) is difficult. Successive frames of a video might not be available from outdoor data, due to acquisition conditions of data capture and camera properties. In addition, due to poor lighting and low resolution, **V-J-F-D** failed to detect the face template in a few cases.

In Fig. 3(b), the first row shows the blurred downsampled gallery images with different σ values for a chosen gallery image of a particular subject. Probable σ value for blur is expected to lie in the range 0.25-2.5 (detected empirically based on visual observation from a large set of test cases). The second row in Fig. 3(b), shows an example of probe image for the same subject. Probe images are filtered using Wiener filter [5] in order to minimize the noise and smooth the aliasing effect (due to digitization in sensor and acquisition hardware). On an average, resolution of probe images lie in the range (40-50) while that of gallery images is (200-300).

Next, the blurred downsampled gallery images (as shown in first row of Fig. 3(b)) are normalized with respect to the chosen probe image. The corresponding normalized images are shown in the last row of Fig. 3(b). It is clear from the figure that the Normalized gallery images is visually closer to the probe image, for which the histograms are depicted in Figs. 3(c) and 3(d). Figure 3(c) shows the histograms corresponding to chosen downsampled gallery and probe images. Qualitatively the histograms differs a lot, due to the difference in global illumination of the pair of images. In Fig. 3(d), histograms for the Normalized blurred gallery images (for some σ values) are shown. We can clearly observe in Fig. 3(d) that the dynamic range of the histogram is altered to make it appear close to the histogram of the probe image. The above process is repeated for different combinations of available gallery and probe images for a particular subject, i.e., if there are 10 gallery images and 10 probe images for a subject then we use 100 (10*10) combinations. This process is repeated for all the subjects. It is to be noted that, in each combination we have used gallery and probe images for the

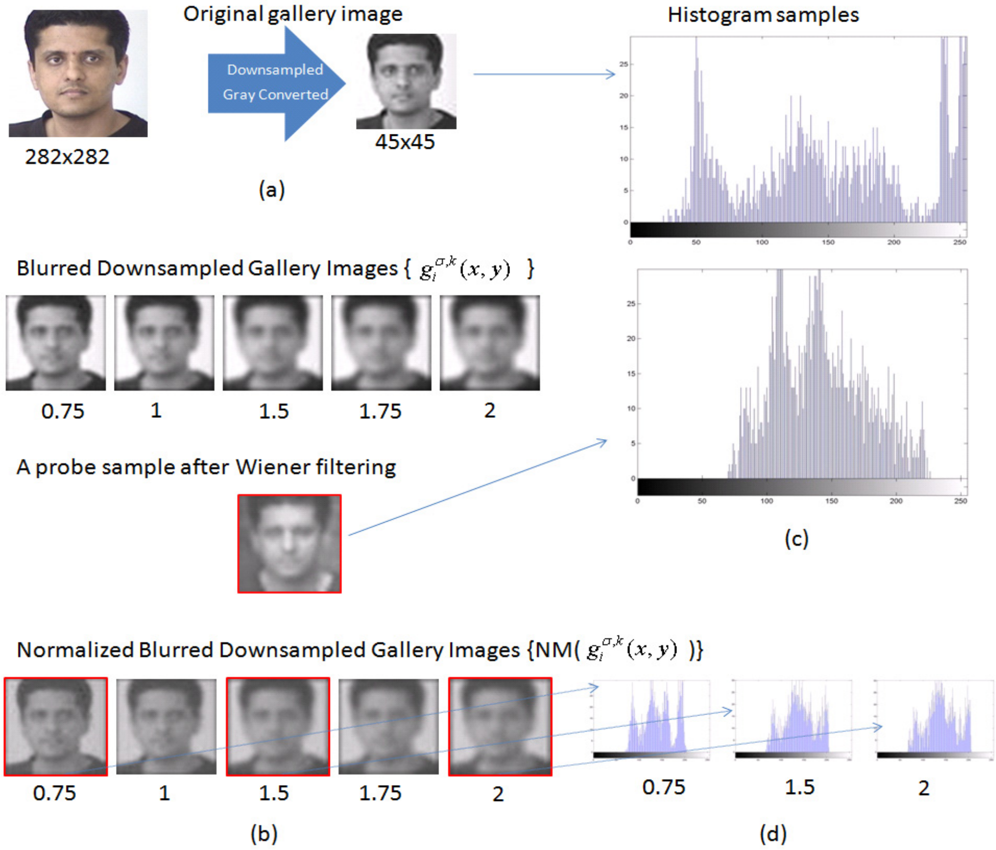

Fig. 3. Sample images of faces and the intensity histograms of a few, showing the process of estimating the degradation, with different parameter values.

same subject. This should not be confused with the use of class label information of probe (testing) images, because at this stage we are estimating the degradation parameter for the surveillance system with the help of pre-acquired data (both gallery and probe for a few subject). Once the degradation parameter is obtained we can use it to simulate degradation for any future data.

Next, we present the mathematical formulation for the process described above and define a measure used in estimating the degradation parameter σ_{blur} (for blur **PSF**).

3.2 The Measure

The measure namely, **SoHD** (Sum of Histogram Difference) is defined based on the intensity histograms of Normalized blurred downsampled gallery images and

Wiener filtered probe images. **SoHD** is given as:

$$\mathbf{SoHD} = \frac{1}{(m_k * n_k)} \sum_{j=1}^{m_k} \sum_{i=1}^{n_k} sum\{|HD_i^{\sigma,k} - HP_j^k|\} \tag{2}$$

Here,

$$HD_i^{\sigma,k} = Hist(NM(g_i^{\sigma,k}(x,y), p_j^k(x,y))) \tag{3}$$

where, $g_i^{\sigma,k}(x,y)$ is the degraded gallery image obtained by convolution of the i^{th} gallery image for k^{th} subject with a Gaussian function with standard deviation σ. Also, $Hist()$ computes the histogram and NM denotes a Normalization operation [7]. Similarly, for the histogram of a probe image we have;

$$HP_j^k = Hist(p_j^k(x,y)) \tag{4}$$

Also, m_k in Eqn.2 denotes the number of probe images used for estimating the degradation for the k^{th} subject and n_k is the number of gallery images used for estimating the degradation for the k^{th} subject.

Notice that the operator $sum\{\ \}$ in Eqn. 2 denotes the algebraic sum of the elements of the vector $|HD_i^{\sigma,k} - HP_j^k|$. **SoHD** is the measure obtained for each subject separately. The plot of the measure **SoHD**, averaged over 51 subjects is shown in Fig. 4. We observe that the measure **SoHD** saturates after some value of σ. To find such an optimal value of σ for k^{th} subject, we use the following condition;

$$\sigma_{SoHD_k} = \{\sigma : |\frac{d(\mathbf{SoHD}_k)}{d\sigma}| < Th_{SoHD}(\simeq 0)\} \tag{5}$$

where, Th_{SoHD} is some low threshold value. To obtain an average measure over all subjects, we define σ_{blur} as follows:

$$\sigma_{blur} = \frac{1}{K} \sum_{k=1}^{K} (\sigma_{SoHD_k}) \tag{6}$$

where, K is the total number of subjects. **SoHD**, when observed with increasing values of σ, saturate after some value of σ, which is determined using Eqn. 5. This is the point at which the degraded blurred (and Normalized) gallery image appears qualitatively similar to that in probe. With these estimated parameters, we first blur the gallery images. Empirical observation over 51 subjects produced optimal values of the parameter as: $\sigma_{blur} = 1.25$ for the value of Th_{SoHD} as 0.2. In this way, we obtain the simulated degraded gallery images, which are comparable with the low quality probe images. Later, these degraded gallery images are used to train the classifier for recognition. When a probe is detected, it's face image is extracted and tested for recognition. Training with the acquired gallery images produce a low value of accuracy because of the large difference in the quality of gallery and probe images. In the next section, we show how degradation improves the classification accuracy.

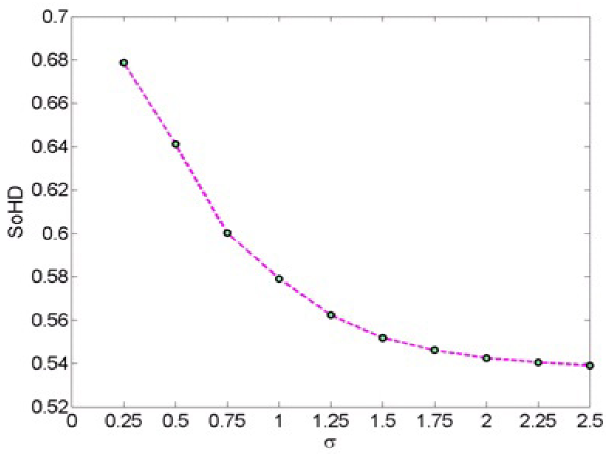

Fig. 4. Plot of the measure **SoHD**, averaged for 51 subjects.

4 Experimental Results

A face recognition system is claimed to be efficient, robust and reliable, only after it has gone through rigorous testing and verification. A real-world database like that of ours would be the most preferable for this purpose. Many face database are available to the research community, but they are still far from real-world conditions for surveillance applications. The proposed database is very challenging because of its large variation over training and testing samples. There are 20 samples in gallery as well as in probe per subject which are near frontal faces. Different experimental cases used in the second stage of the proposed frame work, are listed below:

1. Training:-LRG; Testing:-AP
2. Training:-LRDG; Testing:-AP
3. Training:-MRDG; Testing:-INTP
4. Training:-MRDG; Testing:-SRP
5. Training:-MRG; Testing:-MRG

We have obtained Cumulative match score(CMS) curves for the above experimental cases which is shown in Fig. 5. For each of the curve, training and testing case are denoted by the abbreviation from Table 1. The efficiency of the estimated degradation parameter is presented with the help of two baseline methods: **PCA** and **FLDA** [2]. The performance of the green curves (LRG-AP combination) shows the worst performance. This situation does not involve any processing on the face samples before they are fed to the classifier. The RED curve (MRG-MRG) is an ideal situation, where both the training-testing pairs are based on the indoor acquired gallery samples. For the rest of the curves, either the training gallery has been degraded with the estimated blur parameter (Eqn. 6), or the probe has been enhanced or both done.

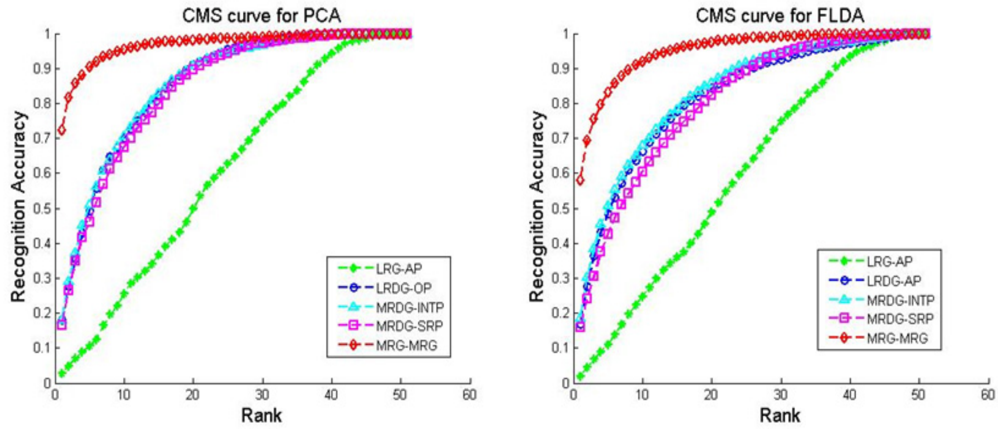

Fig. 5. CMS curves for comparing the performance of the system, when trained and tested as per different experimental cases, using; (a) **PCA** [9] and (b) **FLDA** [2] for face recognition.

Experimental results shows that training with degraded gallery images provides a much improved performance compared to training with acquired (downsampled to medium resolution) gallery. This improvement is significant given the complexity of face samples as probes in our database. We have obtained these performances by taking a 10-fold study of the classifier output. In each fold, 10 training samples per subject is selected randomly from the set of 20 from gallery. Similarly, for testing 10 training samples per subject is selected randomly from the set of 20 from probe in each fold. Total number of subjects for which the curves are obtained is 51. As it is clear from the results that, **PCA** performs better than **FLDA** in this scenario, because **PCA** features are expected to perform better in case of noise and degradation.

5 Conclusions

The work proposed in this paper concerns a face recognition application under surveillance conditions. It is focused on estimating degradation due to out-of-focus blur, low contrast and low resolution. We define a measure- **SoHD** which is quite intuitive, simple and easy to implement. From this measure, we obtain the parameter σ_{blur} for out-of-focus blur. Next, we simulate the degradation on the gallery images. Finally, we train the classifier with degraded gallery instead of original gallery to obtain significantly improved recognition accuracy.

As part of future work, we intend to test our method on a larger database of subjects. A combination of partial restoration or enhancement of probe samples using filters or more robust super-resolution techniques along with partial simulation of degradation on gallery, is to be explored for better results. State of the art methods - **K-PCA**, **K-LDA** [12][8], dual-space [11] and **SVM** [6][8] based face recognizer may be used with our proposed method to improve the classification accuracy further.

References

1. M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski. Face recognition by independent component analysis. *IEEE Transactions on Neural Networks*, 13(6):1450–1464, November 2002.
2. P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman. Eigenfaces vs. fisherfaces: Recognition using class specific linear projection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 19(7):711–720, July 1997.
3. H. Cevikalp, M. Neamtu, M. Wilkes, and A. Barkana. Discriminative common vectors for face recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(1):4–13, January 2005.
4. K. Fukunaga. *Introduction to Statistical Pattern Recognition*. Academic Press, San Diego, 1990.
5. R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Prentice Hall, Upper Saddle River, N.J., 2008.
6. B. Heisele, P. Ho, and T. Poggio. Face recognition with support vector machines: global versus component-based approach. In *Proc. of the Eighth IEEE International Conference on Computer Vision*, pages 688–694, July 2001.
7. L. Hong, Y. Wan, and A. Jain. Fingerprint image enhancement: Algorithm and performance evaluation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):777–789, August 1998.
8. B. Scholkopf and A. J. Smola. *Learning with Kernels*. The MIT Press, Massachusetts, 2002.
9. M. Turk and A. Pentland. Eigenfaces for recognition. *Journal of Cognitive Neurosicence*, 3(1):71–86, 1991.
10. P. Viola and M. J. Jones. Robust real-time face detection. *International Journal of Computer Vision*, 57(2):137–154, 2004.
11. X. Wang and X. Tang. Dual-space linear discriminant analysis for face recognition. In *Proc. Computer Vision and Pattern Recognition*, pages 564–569. IEEE Computer Society, 2004.
12. M.-H. Yang. Kernel eigenfaces vs. kernel fisherfaces: Face recognition using kernel methods. In *Proc. of the Fifth IEEE International Conference on Automatic Face and Gesture Recognition*, pages 215–220. IEEE Computer Society Washington, DC, USA, May 2002.
13. H. Yu and J. Yang. A direct LDA algorithm for high-dimensional data with application to face recognition. *Pattern Recognition*, 34:2067–2070, October 2001.