Face Recognition with Independent Component Based Super-resolution

Osman Gokhan Sezer[†,a], Yucel Altunbasak[b], Aytul Ercil[a]

[a]Faculty of Engineering and Natural Sciences, Sabanci Univ., Istanbul, Turkiye, 34956
[b]School of Elec. and Comp. Eng. , Georgia Inst. of Tech., Atlanta, GA, USA, 30332-0250

ABSTRACT

Performance of current face recognition algorithms reduces significantly when they are applied to low-resolution face images. To handle this problem, super-resolution techniques can be applied either in the pixel domain or in the face subspace. Since face images are high dimensional data which are mostly redundant for the face recognition task, feature extraction methods that reduce the dimension of the data are becoming standard for face analysis. Hence, applying super-resolution in this feature domain, in other words in face subspace, rather than in pixel domain, brings many advantages in computation together with robustness against noise and motion estimation errors. Therefore, we propose new super-resolution algorithms using Bayesian estimation and projection onto convex sets methods in feature domain and present a comparative analysis of the proposed algorithms with those already in the literature.

Keywords: Face recognition, super resolution, independent component analysis, projection onto convex sets, bayesian estimation.

1- INTRODUCTION

Due to many challenges such as pose difference, illumination and expression changes, accessories, aging and distance from the image acquisition device which brings about low resolution images, face recognition is still an unsolved problem. Compared with other challenges, face recognition from low resolution data (video or image) reduces the performance of the existing systems significantly. Various techniques have been proposed to obtain a single high resolution image from many low resolution images to enhance the face recognition performance [1-4]. However, since in the previous works those low resolution images are generated by their high-resolution counterparts synthetically by downsampling, the true effect of super-resolution in a real face recognition problem actual has not been studied yet. The previous methods can be classified as pixel and face-subspace domain techniques. We can further divide pixel domain techniques into two main categories as: those methods using face specific constraints [2, 3, 4] and without any face specific constraints [1]. As it is well known, the face images are special class of data that have a fixed configuration (i.e, relative locations of the mouth, nose or eyes are fixed). Hence the idea of using face-subspace methods which incorporates configuration information of face into a reduced dimensional subspace model not only reduces computational cost but also enables more robust representation against the noise and the motion estimation errors [5].

Independent component analysis (ICA), which finds linear transformation of the data that maximize the statistical independence, appeared in the last two decade as a new data analysis tool. ICA found itself many application areas and applied to various problems successfully ranging from analysis of the brain signals [6] to the image and speech processing [7,8]. Quite recently, ICA has been applied to face recognition problem too, and face recognition performance of ICA and comparative analysis of it with different recognition algorithms have been studied exhaustively in many research papers [9,10]. In these researches, it is demonstrated that although ICA gives more discriminative features than PCA, face recognition performances of these two methods are found to be close to each other. However, it is argued in [11] that the metric induced by ICA is superior to PCA by providing a representation that is more robust to the effects of the noise. In addition to this, it is shown in [9] that the linear reconstruction based on ICA gives better SNR than the reconstruction based on PCA in noisy or limited precision environments. For our case, similar to the effect of the variation in the expressions and illumination, reduced resolution can be viewed as noisy version of the canonical representation of the reference (or high resolution) face image in the face-subspace. Therefore, representing face images in the independent component face-subspace may facilitate robustness against noise compared to eigenface-subspace

[†] This research was partially supported FP6-2004-ACC-SSA-2 SPICE project. Osman Gokhan Sezer is currently a Ph.D. student in LEMS Laboratory at Brown University, Providence, RI; sezer@brown.edu ; phone 1 401 451-9763

representations. Hence, it is reasonable to expect better face recognition performance when the super-resolution algorithm is applied in the independent component face-subspace to recover the features of the high resolution image. In this paper, we introduce the independent component analysis based super-resolution techniques using Bayesian estimation and the method of projection onto convex sets (POCS). We observed that modeling the noise processes in low-dimensional subspace and then the extraction of the statistics of these processes from face images for estimating canonical representation of high resolution face provide robustness against the noise and the motion estimation errors when it is judged against pixel domain estimation algorithms. Besides, the independent component face-subspace based super-resolution is found to be superior to eigenface-based super-resolution for robustness against the noise. Two different experimental setups are used for evaluating recognition performance of the proposed algorithms. In the first setup, low-resolution face videos are generated by downsampling real face-video sequences with different blur kernels. These low and high-resolution video sequences are later used to compare the aspects of the independent and the principal component based super-resolution techniques together with pixel domain super-resolution. On the other hand, second experimental setup is employed to verify that the proposed algorithms works in a real scenario where the low resolution videos are obtained by appropriate camera setup rather than generating by downsampling the high resolution videos. We have created a new database (VPA Super-resolution Face Database) to test the effect of the proposed super-resolution techniques and those ones that are already available in the literature for the face recognition problem.

The organization of the paper is as follows: In Sec. 2 mathematical overview of subspace methods (principal and independent component analysis) used in the proposed algorithms is given. Super-resolution techniques both in pixel and subspace domain is provided with details in Sec. 3. Sec. 4 contains experimental procedures and results. Finally conclusions are drawn at the end of the paper.

2- SUBSPACE METHODS

2.1 Principle Component Analysis

PCA technique, also known as Karhunen-Loeve transform (KLT), finds dimensionality reducing linear projections that maximizes the scatter of all projected samples. If the total scatter is defined by S_T;

$$S_T = \sum_{k=1}^{N} (x_k - \mu)(x_k - \mu)^T \tag{2.1}$$

where $x_k \in \mathbb{R}^{N^2}$ are face images ordered lexicographically, N is the total number of sample images and μ is the mean image of all sample images. After applying a linear transformation, we will obtain transformed features (or PCA coefficients) $\overline{x}_k \in \mathbb{R}^n$ in the reduced dimensional subspace;

$$\overline{x}_k = E_n{}^T x_k \tag{2.2}$$

where E_n is a $N^2 \times n$ matrix with orthonormal columns containing corresponding eigenvectors of the scatter matrix having the largest n eigenvalues [12,13].

2.2 Independent Component Analysis

ICA is a method that can perform blind source separation. Since both the source signals and how these signals are mixed are unknown, separation is named as blind. ICA algorithm finds a linear coordinate system such that resulting signals will be statistically independent. ICA not only makes signals uncorrelated like PCA does, but also reduces higher order dependencies between the signals. Compared with the classical methods, ICA is a powerful method for finding the factors that are mutually independent with the non-Gaussian distributions. In the ICA model, linear or nonlinear mixtures of the hidden factors or independent components constitute the observed data. Basic linear mixture model of ICA can be expressed mathematically as [14]:

$$x = As \tag{2.3}$$

where x is the $N^2 \times 1$ observation vector containing the lexicographically ordered observed data, s is the $n \times 1$ source vector and A is the $N^2 \times n$ mixing matrix ($N^2 \gg n$). The aim is to estimate the unknown A and s from the observation vector x. Our only assumption is non-Gaussianity and statistical independence of the sources. The goal of ICA is to find an orthogonal $n \times N^2$ transformation matrix W such that statistical dependencies between the estimated sources are minimized.

$$s_x = W_x x = W\hat{y} = WD_n^{-1/2} E_n{}^T x \tag{2.4}$$

where the $n \times 1$ vector $\hat{y}$ denotes the whitened data, D_n denotes the diagonal matrix containing the n largest eigenvalues of the covariance matrix of x and E_n denotes the corresponding matrix whose columns are the eigenvectors corresponding to the n largest eigenvalues as described before. In the estimation process, sphering is generally an optional stage which enables faster converges. However, sphering makes estimation vulnerable against noise due to the trailing eigenvalues which tend to capture noise in the data appear as denominator. Observe the relative magnitude of the eigenvalues extracted from CMU PIE Database [15] in Fig. 2.1. When the eigenvalue index is greater than 10, the corresponding eigenvalues have relatively small magnitudes, and if they were included in the whitening transformation these small eigenvalues will lead to decreased performance in ICA face recognition by amplifying the effects of noise, hence we exclude sphering in our experiments. But for the sake of completeness, we stick to the convention used in the literature through out the estimation of mixing matrix. The next step, therefore, is to apply ICA algorithms to create face subspace. We have used symmetric fixed point algorithm with $f(x)=tanh(x)$ nonlinearity in order to obtain a fast solution with a simple algorithm. The algorithm starts from a random orthogonal matrix W and in each iteration rows of it, (w^T), is updated by;

$$w_i := E\{\hat{y} f(w_i^T \hat{y})\} - E\{f'(w_i \hat{y})\} w_i \tag{2.5}$$

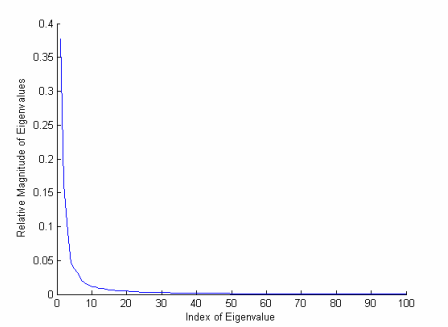


Figure 2.1. Relative magnitude of eigenvalues $\left(\lambda_i / \sum_k \lambda_k\right)$

followed by orthonormalization of the matrix through

$$W := (WW^T)^{-1/2} W. \tag{2.6}$$

Finally, after convergence is achieved, the estimated basis is constructed as

$$A_x = E_n D_n^{1/2} W^T \tag{2.7}$$

where A_x is the $N^2 \times n$ mixing matrix in the ICA model and each column of A_x corresponds to a basis image (or vector). Since ICA is an orthogonal projection into a reduced dimensional space, we expect to have error when we want to reconstruct the original image. This reconstruction error arises due to dimension reduction in the whitening process. As we have shown in eq.(2.4), PCA is widely used in the whitening process of ICA which enables reduction of dimensionality. Dimension reduction and sphering are two consecutive stages in the whitening process. Projecting the observed data x onto eigenvectors of the n largest eigenvalues has been written in eq.(2.2) as;

$$\bar{x} = E_n^T x \tag{2.8}$$

The projection of the observed data to the reduced dimensional subspace causes loss of information, which in return brings about reconstruction error when we want to retrieve the original observed data x.

$$x = E_n \bar{x} + e_x \tag{2.9}$$

In the above equation, e_x corresponds to the reconstruction error of x. Using eq. (2.4) we can write following equations:

$$s_x = W D_n^{-1/2} \bar{x} \Rightarrow \bar{x} = (W D_n^{-1/2})^{-1} s_x \Rightarrow \bar{x} = D_n^{1/2} W^T s_x \tag{2.10}$$

Substitution of eq.(2.10) into eq.(2.9) will give,

$$x = E_n D_n^{1/2} W^T s_x + e_x \tag{2.11}$$

Finally using equality in eq.(2.7), we will get,

$$x = A_x s_x + e_x \tag{2.12}$$

The above model in eq.(2.12) looks like the noisy ICA model. However, together with the general assumption of Gaussian reconstruction error, estimation of the mixing matrix, A_x, and sources, s_x turns into estimation of noise-free ICA model's mixing matrix. Note that if we denote noise-free data as:

$$v = A_x s_x \tag{2.13}$$

We can write the observed data as $x=v+e_x$. In ICA, our aim is to find projections that maximize non-Gaussianity, and such a projection w^T will give us $w^Tx=w^Tv+w^Te_x$. Since we assume e_x to be a Gaussian noise, w^Te_x will be zero (e.g., kurtosis of a Gaussian random variable will be zero). Thus, the measure of non-Gaussianity for w^Tx (noisy data) will be equal to the measure of non-Gaussianity for w^Tv (noise-free data). Beside that, since the noise term comes from the whitening process, we do not need bias removal techniques for estimation of the mixing matrix [14].

After estimation of the mixing matrix, the de-mixing matrix is found by pseudo-matrix inversion.

$$W_x = (A_x{}^T A_x)^{-1} A_x{}^T = WD_n^{-1/2} E_n{}^T \tag{2.14}$$

Hence, sources s_x are estimated by multiplying both sides of eq.(2.12) with the de-mixing matrix W_x,

$$s_x = W_x x \tag{2.15}$$

where,

$$W_x e_x = WD_n^{-1/2} E_n{}^T e_x = 0 \;\Leftrightarrow\; E_n{}^T e_x = 0 \tag{2.16}$$

3- METHODOLOGY

Super-resolution can be thought as a signal restoration problem where the original form of the signal is assumed to be the high-resolution image and it is estimated from its degraded and noise corrupted low-resolution versions. Since signal restoration is a common problem for various fields in signal processing including image processing, speech processing, and system identification, there are various techniques in the literature to solve the restoration problem under different degradation models. In our case, we have used generative imaging model as our distortion mechanism where degradation operator and noise processes are separated to have a better analysis. The imaging model can be written as:

$$y^{(i)} = H^{(i)} x + n^{(i)}, \text{ for i=1...K} \tag{2.17}$$

where x in $\mathbb{R}^{N^2}$ and y in $\mathbb{R}^{M^2}$ are lexicographically ordered high-resolution and low-resolution image, respectively ($M=\ell N$ for $0<\ell<1$). The superscript (i) denotes the observation number, $H^{(i)}$ is a linear degradation operator which incorporates motion, blurring, and downsampling processes, and $n^{(i)}$ is the noise process where both are for the i'th observation, and there are K such observations. Dimensions of $H^{(i)}$ and $n^{(i)}$ are $\ell^2N^2xN^2$ and ℓ^2N^2x1, respectively.

The degradation matrix $H^{(i)}$, integrates effects of motion, blur, and distance from the camera into the imaging model while the noise vector $n^{(i)}$ represents the observation noise that incorporates the quality of the camera into model. The degradation matrix $H(i)$ can be written as:

$$H^{(i)} = DBW^{(i)} \tag{2.18}$$

where D is the $\ell^2N^2xN^2$ decimation matrix, B is the N^2xN^2 blur matrix, and $W^{(i)}$ is N^2xN^2 motion warping matrix. Here we assume that decimation and blurring matrices are the same for every observation (refer to the Appendix of [16], for a practical blur computation method), hence only the motion warping matrix changes depending on the observed low-resolution image. Further information about the imaging model can be found in [16, 17, 18]. In the Sec. 4.4, the details of the registration algorithm are provided. Moreover, the motion warping matrix, $W^{(i)}$, is obtained by using motion vectors coming from this registration phase of the algorithm.

Since recovering the high-resolution image, x, by using the linear set of equations in eq.(2.17) is an ill-posed problem, solutions like the inverse filtering will just amplify the noise in the observations. Hence, regularization methods are needed to incorporate a priori information about the distortion mechanism and the noise process to get a better estimate of the high-resolution image. In this paper, we have used the method of POCS and regularization by Bayesian Estimation to recover high-resolution images. POCS is an iterative method which enables one to employ a priori information about the degradation operator, the noise statistics and the actual high-resolution image distribution together with measured data to find a feasible solution consistent with the number of constraints. For each constraint, a closed convex constraint is defined such that the members of the set satisfy the given constraint and the actual high-resolution image is also a member of the set. Moreover, if appropriate constraint sets are defined, high-resolution image will be a member of the intersection set, i.e., a member of feasible region. A feasible solution, on the other hand, can be found by successive projection of an initial estimate onto the constraint sets set. The fundamental mathematical concepts for POCS are given in [19, 20].

We have used the POCS algorithm for both pixel-domain and face subspace super-resolutions, i.e., for 2-D and 1-D signal restoration with different constraint sets. In both cases, a priori information about the degradation operator and the noise statistics are used to employ constraints on the residual, defined by

$$r = y^{(i)} - H^{(i)}\hat{x} \tag{2.19}$$

where $\hat{x}$ is an estimate of the high-resolution image (i.e., original signal) in the model given in eq. (2.17). A priori information about the degradation operator $H^{(i)}$ can be estimated from the low-resolution images. Since B and D are assumed to be known only the motion warping matrix $W^{(i)}$ should be computed. Referring to eq.'s (2.17) and (2.18), we observe that in order to calculate the motion warping matrix $W^{(i)}$ by using registration algorithms we need to have high-resolution images that are unknown to us. So, it is practically impossible to have actual $W^{(i)}$, however, we can still estimate the motion warping matrix $W^{(i)}$ by using low-resolution images or frames. If the motion vector of a pixel found by application of registration algorithm to low-resolution images is multiplied by a constant $1/\ell$ and mapped to the motion vector of the corresponding $1/\ell \times 1/\ell$ block of pixels in the high-resolution image where ℓ is the downsampling factor set in the imaging model (eq.(2.17)), we will have a reasonable estimate of $W^{(i)}$. A priori information about the noise statistics, on the other hand, can be estimated from training images (signals) or simply by trial-and-error method till satisfactory results are obtained as the output of the POCS algorithm.

In theory, if statistics of noise and residual are approximately equal, then we can say that the true solution is achieved. So, our aim is to constrain the residual in order to have the same statistical characteristics as the noise $n^{(i)}$ given in imaging model. Therefore, using the confidence limits derived from the sample statistics will enable us to determine the limits of approximation. In the Trussell and Civanlar's paper [21] such constraints on the statistics of the residual are defined thoroughly.

3.1 POCS based super-resolution in the pixel-domain

For the pixel-domain super-resolution we have used constraints defined for outliers of residual and amplitude of the high-resolution image estimate. The constraints on the outliers of the residual are performed by projecting the outlier values of the residual which deviate an unlikely amount from the mean. For the most of the time Gaussian noise is assumed, hence the appropriate confidence limits are easily found from the tables. The convex set will simply be defined as,

$$C_0 = \left\{ x \,\middle|\, \left| y_j^{(i)} - \left[H^{(i)} x \right]_j \right| \le \delta_0 \right\} \tag{2.20}$$

where C_0, here represents the intersection of many single convex sets which are defined for each pixel in the image (or point in the signal). In order to achieve the point x (high-resolution image) which is a member of C_0, the projection is made by again applying the sequential projection method outlined before. Therefore, we project any point in the image whose residual lies outside the specified limit, hence be forced to lie within the limit. The residual of each point (or pixel) in the low-resolution image can be defined as,

$$r_i = y_j^{(i)} - \left[H^{(i)} x \right]_j \tag{2.21}$$

The projection formulation for correcting the residual at this pixel in the low-resolution image is

$$P_0 x = \begin{cases} x + \dfrac{(r_i - \delta_0)}{\|h_i\|^2} h_i & , \quad \text{if } r_i > \delta_0 \\ x & , \quad \text{if } -\delta_0 < r_i < \delta_0 \\ x + \dfrac{(r_i + \delta_0)}{\|h_i\|^2} h_i & , \quad \text{if } r_i < -\delta_0 \end{cases} \tag{2.22}$$

where h_i is the column vector containing the i'th row of the matrix $H^{(i)}$. After sequential application of projection for every pixel locations in the low-resolution image, additional constraints can be applied. As an additional constraint, we have employed amplitude constraint defined as,

$$C_A = \left\{ x \,|\, \alpha \le x_j \le \beta \right\} \tag{2.23}$$

to ensure appropriate gray-level images with amplitude bounds $\alpha = 0$ and $\beta = 255$. The projection operator of constraint C_A, P_A, will be simply a clipping algorithm.

Finally we will get an estimate of the high-resolution image by successive projection onto C_0 and C_A which can be written as:

$$\hat{x}_{\kappa+1} = P_A P_0 \hat{x}_\kappa \qquad \text{for } \kappa=0,1,2.. \tag{2.24}$$

The initial estimate of the high-resolution image for iterative projections is obtained by bilinearly interpolating one of the low-resolution images which is selected as the reference image. Please refer to [16] and [21] for further details.

3.2 POCS based super-resolution in face-subspace

The PCA algorithm enables one to represent a face image as a linear combination of orthonormal vectors, called eigenfaces. These eigenfaces are actually eigenvectors of the scatter matrix given in eq. (2.1) that corresponds to the largest *n* eigenvalues. Hence, using the eigenfaces one can represent a face image with minimum reconstruction error in the least square sense which can be written as:

$$x = \phi a + e_x \tag{2.25}$$

where ϕ is $N^2 \times L$ linear transformation matrix containing eigenfaces in its column, a is $L \times 1$ coefficient vector of eigenfaces, and e_x is reconstruction error which is orthogonal to the linear space defined by eigenfaces.

Using these definitions together with ICA model given in eq.(2.12), we propose to apply the method of POCS in the face-subspace to estimate the feature vectors of the high-resolution face image. In order to derive an efficient super-resolution algorithm in the face-subspace, we need to obtain an observation model for the reconstruction of the feature vectors of the high-resolution face images. For the POCS-based super-resolution on face subspace, we have shown the PCA and ICA representations together since they have parallel derivation steps. The observation model will not neglect the spatial-domain observation noise given in eq.(2.17) and the subspace representation error (or the reconstruction error). Since we have two different resolutions, we need two principal component subspaces, one for high-resolution face images and the other for low-resolution face images (cf. left part of eq.(2.26)). Similarly, we have two independent component subspaces for high-resolution and low-resolution face images too (cf. right part of eq.(2.26)). We will stick to the notion given in [5] for PCA representation.

PCA	**ICA**
$x = \phi a + e_x$	$x = A_x s_x + e_x$
$y^{(i)} = \psi a^{(i)} + e_y^{(i)}$ for i=1...K	$y^{(i)} = A_y s_y{}^{(i)} + e_y{}^{(i)}$ for i=1,....,K

(2.26)

where x is $N^2 \times 1$ high resolution face image $y^{(i)}$ is the *i*'th observation of the low-resolution face image which is $\ell^2 N^2 \times 1$. For PCA part, ψ and $e_y{}^{(i)}$ are $\ell^2 N^2 \times L$ eigenface matrix and $\ell^2 N^2 \times 1$ reconstruction error, respectively, for low-resolution face-subspace. The coefficients of eigenfaces are a and $a^{(i)}$ for high and low dimension face-subspaces which have same dimensionalities, $L \times 1$. For ICA part, A_x is an $N^2 \times L$ and A_y is an $\ell^2 N^2 \times L$ matrix containing independent component faces in their columns for high-resolution and low-resolution images, respectively. The coefficients of independent component basis vectors are s_x and $sy^{(i)}$ for high and low dimension face-subspaces which have same dimensionalities, $L \times 1$.

Substituting the subspace representations of low and high resolution face images given in eq. (2.26) into imaging model of eq.(2.17), we will obtain

$\psi a^{(i)} + e_y^{(i)} = H^{(i)}(\phi a + e_x) + n^{(i)}$	$A_y s_y{}^{(i)} + e_y{}^{(i)} = H^{(i)}(A_x s_x + e_x) + n^{(i)}$
$\psi a^{(i)} + e_y^{(i)} = H^{(i)}\phi a + H^{(i)} e_x + n^{(i)}$	$A_y s_y{}^{(i)} + e_y{}^{(i)} = H^{(i)} A_x s_x + H^{(i)} e_x + n^{(i)}$

(2.27)

We know that if we project eq. (2.27) into lower-dimensional face subspace, we will eliminate the reconstruction error $e_y{}^{(i)}$ using the fact that it is orthogonal to ψ for PCA, and is orthogonal to W_y for ICA as defined in eq.(2.16).

$\psi^T e_y^{(i)} = 0$	$W_y e_y^{(i)} = 0$

(2.28)

and since the eigenface matrix is orthonormal for PCA, and de-mixing matrix W_y is the inverse of mixing matrix A_y for ICA, we have the following equalities,

$\psi^T \psi = I$	$W_y A_y = I$

(2.29)

If we multiply eq. (2.27) with ψT for PCA and W_y for ICA we will get,

$a^{(i)} = \psi^T H^{(i)} \phi a + \psi^T H^{(i)} e_x + \psi^T n^{(i)}$	$s_y{}^{(i)} = W_y H^{(i)} A_x s_x + W_y H^{(i)} e_x + W_y n^{(i)}$

(2.30)

Observe that the model in eq. (2.30) resembles the imaging model in the pixel-domain given in eq. (2.17). Both equations explain degraded or "inaccurate" vector in terms of the unknown original or "true" vector plus a noise term. Moreover, quite similar to the pixel domain formulation given in eq. (2.19) we can write the residuals in principal component and independent component subspace as

PCA	**ICA**	
$r_{PCA}^{(i)} = a^{(i)} - \psi^T H^{(i)} \phi \hat{a}$	$r_{ICA}^{(i)} = s_y^{(i)} - W_y H^{(i)} A_x \hat{s}_x$	(2.31)

where $\hat{a}$ and $\hat{s}_x$ are the estimate of the eigenface coefficients and independent component coefficients of the high-resolution face image. Furthermore, we will incorporate a priori information about the noise process into the algorithm by means of defining constraints on the residual in eq. (2.31) so as to find the POCS estimate for this ill-posed problem. We have used two constraint set with respect to the outliers of residual and the variance of the residual. First, the convex set for outliers of the residuals are defined as,

$$C_0 = \left\{ a \,|\, \left| a^{(i)} - \left[\psi^{\mathrm{T}} H^{(i)} \phi a \right] \right|_j \le \delta_0 \right\} \quad \Big| \quad C_0 = \left\{ s_x \,|\, \left| s_y^{(i)} - \left[W_y H^{(i)} A_x s_x \right] \right|_j \le \delta_0 \right\} \tag{2.32}$$

where subscript j denotes the j'th element of the vector and δ_o is the a priori bound reflecting the statistical confidence with which the "true" feature vector a or s_x is a member of the set C_0. The bound, δ_o, is determined from noise statistics where we can write the noise in two subspace observations as,

$$n_{subspace} = \psi^T H^{(i)} e_x + \psi^T n^{(i)} \quad \Big| \quad n_{subspace} = W_y H^{(i)} e_x + W_y n^{(i)} \tag{2.33}$$

The general assumption for the reconstruction error e_x and the pixel-domain representation error $n^{(i)}$ is that they have a Gaussian IID distribution. Hence, since ψ, W_y and $H^{(i)}$ are linear operators, we will expect that the subspace representation error $n_{subspace}$ to have a Gaussian distribution for both cases too. Therefore, if we select our bound as,

$$\delta_0 = 3\sigma_{subspace} \tag{2.34}$$

where $\sigma_{subspace}$ is the mean standard deviation of the components of the noise vector $n_{subspace}$, we will get a 99% confidence. The projection operators therefore will be,

$$P_{0_j}^{(i)} a = \begin{cases} a + \dfrac{(r_j^{(i)} - \delta_0)}{\left\| \left[\psi^T H^{(i)} \phi \right]_j \right\|^2} \left[\psi^T H^{(i)} \phi \right]_j & , \quad r_j^{(i)} > \delta_0 \\ a & , \quad -\delta_0 < r_j^{(i)} < \delta_0 \\ a + \dfrac{(r_j^{(i)} + \delta_0)}{\left\| \left[\psi^T H^{(i)} \phi \right]_j \right\|^2} \left[\psi^T H^{(i)} \phi \right]_j & , \quad r_j^{(i)} < -\delta_0 \end{cases} \quad \Bigg| \quad P_{0_j}^{(i)} s_x = \begin{cases} s_x + \dfrac{(r_j^{(i)} - \delta_0)}{\left\| \left[W_y H^{(i)} A_x \right]_j \right\|^2} \left[W_y H^{(i)} A_x \right]_j & , \quad r_j^{(i)} > \delta_0 \\ s_x & , -\delta_0 < r_j^{(i)} < \delta_0 \\ s_x + \dfrac{(r_j^{(i)} + \delta_0)}{\left\| \left[W_y H^{(i)} A_x \right]_j \right\|^2} \left[W_y H^{(i)} A_x \right]_j & , \quad r_j^{(i)} < -\delta_0 \end{cases} \tag{2.35}$$

where $[\psi^T H^{(i)} \phi]_j$ and $[W_y H^{(i)} A_x]_j$ are the column vectors containing the j'th row of $\psi^T H^{(i)} \phi$ and $W_y H^{(i)} A_x$ matrices respectively, and $r_j^{(i)}$ is the j'th component of the residual of the i'th observation.

As mentioned before, we have also employed variance constraints with following convex sets,

$$C_V = \left\{ a \,|\, \left\| a^{(i)} - \left[\psi^{\mathrm{T}} H^{(i)} \phi a \right] \right\|^2 \le \delta_v \right\} \quad \Big| \quad C_V = \left\{ s_x \,|\, \left\| s_y^{(i)} - \left[W_y H^{(i)} A_x s_x \right] \right\|^2 \le \delta_v \right\} \tag{2.36}$$

where δ_V is the priori bound reflecting statistical confidence and can be determined by using the formulation defined. As we have mentioned earlier, Gaussian noise is assumed for residual of each point (or pixel) given in eq.(2.21) therefore the sample variance has a chi square distribution. Nevertheless, since the number of points in the signal is likely to be large, the Gaussian approximation to the chi square is valid and the confidence limit, δ_V, can be calculated by the following formulation:

$$\delta_V = \sigma^2 \left[\pm \lim_{0.95} + \sqrt{2(N-1)} \right]^2 / 2N \tag{2.37}$$

where N is the number of points in the signal, $\lim_{0.95}$ is the 95 percent confidence limit for the standard normal distribution, and σ^2 is the mean sample variance of the residual which has chi square distribution. The projection operator onto this constraint set will be

$$\begin{aligned}
&\textbf{PCA: } P_V^{(i)}a = \begin{cases} a + (\mathrm{M}^{\mathrm{T}}\mathrm{M} + \frac{1}{\lambda}I)^{-1}\mathrm{M}^{\mathrm{T}}(a^{(i)} - \mathrm{M}a) & , \text{ if } a \notin C_V \\ a & , \text{ if } a \in C_V \end{cases} \\
&\textbf{ICA: } P_V^{(i)}s_x = \begin{cases} s_x + (\mathrm{M}^{\mathrm{T}}\mathrm{M} + \frac{1}{\lambda}I)^{-1}\mathrm{M}^{\mathrm{T}}(s_y^{(i)} - \mathrm{M}s_x) & , \text{ if } s_x \notin C_V \\ s_x & , \text{ if } s_x \in C_V \end{cases}
\end{aligned} \tag{2.38}$$

where M is equal to $[\psi^T H^{(i)} \phi]$ for PCA and $[W_y H^{(i)} A_x]$ for ICA case, and λ is the Lagrange multiplier coming from the optimization formulation (refer to Appendix of [21] for derivation of the projection operators)

3.3 Subspace-based Super-resolution Using Bayesian Estimation

Different methods have been suggested in the literature to solve the super-resolution problem. In [5] rather than the POCS based reconstruction algorithms which is delineated in Sec. 3.1 and 3.2 a Bayesian estimation based method is suggested for reconstruction of the high-resolution face images in principal component subspace and it is found to be a robust tool against noise and motion estimation errors compared to pixel-domain super-resolution. In this paper, we extended this approach to independent component subspace, aiming to utilize the findings delineated in [16] that the metric induced by ICA is superior to PCA by providing a representation that is more robust to the effects of noise. Hence we expect ICA to give us a better representation that is more robust to noise and motion estimation errors for applying super-resolution. Details of eigenface-domain super-resolution can be found in [5].

3.3.1 Independent Component Subspace

We will use Bayesian estimation for estimating the feature vector of the high-resolution face image. Here, we will try to maximize posterior probability of the feature vector of the high-resolution face image, s_x, by using prior probability of s_x and conditional probability of feature vectors of low-resolution images (i.e., observations) $p(s_y^{(1)},\ldots,s_y^{(M)}|s_x)$. Hence, the MAP estimator will become

$$\hat{s}_x = \arg\max_{s_x}(p(s_y{}^{(1)},\ldots,s_y{}^{(M)} \mid s_x)\,p(s_x)) \tag{2.39}$$

In order to construct our MAP estimator we need to model the prior probability $p(s_x)$ and the conditional probability $p(s_y^{(1)},\ldots,s_y^{(M)}|s_x)$. Since we have used the second architecture of ICA which was defined in [9], modeling prior probability with a super-Gaussian density function is reasonable. Here we assume that the prior probability is Laplacian:

$$p(s_x) = \lambda \exp(-\lambda(s_x - \mu_x)) \tag{2.40}$$

where μ_x is the mean of the feature vector s_x, and λ is the density parameter which can be computed by its relation to the covariance of s_x , Λ, as :

$$\lambda = \sqrt{2}\Lambda^{-1/2} \tag{2.41}$$

It is important to note that, since components of the s_x vector are independent, the covariance matrix is diagonal. In order to model the joint conditional probability of feature vectors of low-resolution images $p(s_y^{(1)},\ldots,s_y^{(M)}|s_x)$, we will use ICA part of eq.(2.30). The first step is to divide eq.(2.30) into two parts as signal and noise. Thus, the noise part will be;

$$n_{subspace} = W_y H^{(i)} e_x + W_y n^{(i)} \tag{2.42}$$

The noise model contains terms coming from the reconstruction error and the image modeling error in the pixel domain which are generally assumed to be IID Gaussian [18]. Therefore, our noise model is also IDD Gaussian since H is simply a linear operator and $W_y W_y^T$ is nonsingular. Hence, we have:

$$p(n_{subspace}) = \mathcal{N}(\mu_n{}^{(i)}, K) \tag{2.43}$$

where $\mu_v^{(i)}$ is the mean vector and K is the covariance matrix of the noise. Since the noise is IID Gaussian we can express the joint conditional probability as the multiplication of marginals,

$$p(s_y{}^{(1)},\ldots,s_y{}^{(M)} \mid s_x) = p(s_y{}^{(1)} \mid s_x) \times \ldots \times p(s_y{}^{(M)} \mid s_x) \tag{2.44}$$

Using eq.(2.30) we can write the marginal conditional probabilities. Since s_x is given, the marginal probabilities will have the same statistical properties as noise, except for the mean. Hence we have:

$$p(s_y^{(i)} \mid s_x) = \mathcal{N}(W_y H^{(i)} A_x s_x + \mu_n^{(i)}, K) \tag{2.45}$$

Since we can write the joint probability as the product of marginal probabilities, by defining $z^{(i)}=s_y^{(i)}-W_y H^{(i)} A_x s_x - \mu_n^{(i)}$ we obtain:

$$p(s_y^{(1)},...,s_y^{(M)} \mid s_x) = \frac{1}{L}\exp(-\sum_{i=1}^{M} z^{(i)T} K^{-1} z^{(i)}) \tag{2.46}$$

where L is a normalization constant. Substituting eq.(2.46) and eq.(2.40) into eq.(2.39) will give us the MAP estimator for ICA feature vector of the high-resolution image.

$$\hat{s}_x = \arg\min(\sum_{i=1}^{M}\left[z^{(i)T} K^{-1} z^{(i)} \right] + \sqrt{2}\Lambda^{-1/2} \mid s_x - \mu_x \mid) \tag{2.47}$$

The solution can be obtained by an iterative steepest descent algorithm.

4- EXPERIMENTS

4.1 Face Video Databases

In our experiments we have tried to double the resolution of video frames by applying the proposed methods on two different video databases. The first database is M2VTS database, created by UCL Laboratoire de telecommunications et télédétection [22]. The other one is a new database that is particularly designed to test the performance of super-resolution algorithms for face recognition problem and created by Computer Vision and Pattern Analysis (VPA) Laboratory group at Sabancı University.

4.2 M2VTS Face Video Database

We have used 36 face videos of different subjects obtained from the M2VTS database. Each frontal face video was shot when the subject was counting from zero to nine. Since the videos in the M2VTS database are in high-resolution, we need to create low-resolution video sequences synthetically. In order to simulate the effects of the quality of the camera, and its distance to the face, we have convoluted frames with Gaussian kernels having differing variance and downsampled them.

Theoretically, 4 or more images are needed to double the resolution of a reference image. In this experiment, we have used five consecutive images where the center frame, shown in the Fig. 4.1-a, is selected as the reference one. In the same way, we have randomly extracted five consecutive frames from each face video of different subjects. Later, we manually aligned these face images according to the locations of eyes, cropped and scaled them in order to fit into 68 x 84 pixel reference high resolution face image model so as to enable the application of the principal and independent component analysis. After obtaining these face frame sequences, we have created their low-resolution counterparts by first convolving with a 15x15 pixels Gaussian kernel and downsampling them by a factor of two as described before. In the Fig. 4.1-b, five low-resolution frames of the same reference frame obtained by Gaussian kernels with differing variances is illustrated.

(a)

(b)

Figure 4.1: Five consecutive frames extracted from a face video of M2VTS database (a),low-resolution face video frames obtained by convolving with Gaussian kernels having variances: 1, 10, 20, 30, 40, and 50 pixels and by downsampling (b).

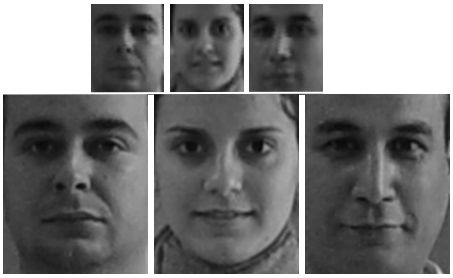


Figure 4.2: Low-resolution face video frames (at the top), corresponding high-resolution face images (at the bottom)

4.3 VPA Super-resolution Face Database

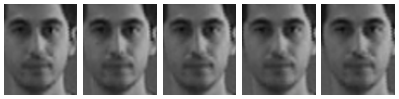
The second face database that we have used in our experiments is VPA Super-resolution (SR) Face Database. VPA SR Face Database consists of frontal face images and videos of 32 people and is particularly designed to test super-resolution techniques. It contains low-resolution and blurry face video data together with the corresponding high resolution face image of each person. Face videos have just translational movement of each face shot by a commercial SONY DVR camera from a distance in ambient light and uncontrolled environment. The high resolution face images are taken by SONY DCS F707 Digital Still Camera with closer distance again in ambient light so as to acquire face images having higher (double) resolution than those faces in the video frames (refer to Fig. 4.2). The high-resolution images are manually aligned according to the locations of eyes, cropped and scales to fit into 68 x 84 pixel reference face model. For low resolution face videos, a reference frame is selected from each video sequence randomly and these reference frames are manually aligned according to eye locations and cropped and scales to fit into 34 x 42 pixel reference model for low-resolution face images, as described before.

Again, since our aim is to increase the resolution by a factor of two, using four or more frames from the face videos is theoretically enough. Hence, we have used five frames, two previous and two next consecutive frames together with the reference one. Other neighboring face video frames are scaled and cropped according to the aligned reference frame shown in Fig. 4.3.

4.4 Pixel-domain Super-resolution

The pixel-domain super-resolution algorithm can be decomposed into three sub-problems: (i) Registration, (ii) calculating blur, and (iii) employing a priori information for reconstruction. Previously, we have explained how a priori information is incorporated into the reconstruction problem and we have also noted that a fixed blur is assumed in the derivations. Therefore, we have to clarify the registration algorithm that is used to estimate displacements of the pixels between consecutive frames.

It is a relatively difficult task to make an accurate displacement estimation for the face videos where there are multiple rigid motions, which we call as the pose of the head, as well as non-rigid ones (e.g., motion of lips and cheeks). For our application two-level hierarchical block matching algorithm (HBMA) appeared to be an appropriate method for displacement estimation due to its computational advantages and ease of use. Even though theoretically it is not possible to capture non-rigid motion by using block matching algorithm, if appropriate block size and search region are chosen, satisfactory displacement estimation performance will be achieved. The HBMA that we have applied for the displacement estimation is devised to detect sub-pixel motion information and it is quite similar to HBMA described in [23] but rather than assigning same motion vector for all pixel in the block, we have assigned the estimated displacement vector to the center pixel of each block and repeat this for every pixel in the frame. We assume that there are no large displacements (since the databases we have used in our experiments have enough time resolution this assumption is reasonable) and make displacement estimation sufficiently locally adaptive (side note: refer to [26] for a discussion on not using validity maps in case of non-rigid motion).



Frame #: *1* *2* *3* *4* *5*

Figure 4.3: Five consecutive face video frames from VPA SR Face Database, third frame is the reference one. Note: Head is moving to right side of the page.

The result obtained using M2VTS data base given in Fig 4.4-ii appears to be very encouraging, however if we list the reasons for such a good performance, we should note that the data has very low noise level and quite high time resolution which enables accurate displacement estimation. In a more realistic case, where the noise level is high and the time resolution hinders accurate displacement estimation, we have observed that the pixel-domain super-resolution does not improve the quality of the video and even degrades the initial estimate due to the distortion effects of the inaccurate motion estimations coming from each frame cumulates onto the reference face image. We have employed the same pixel-domain super-resolution algorithm to VPA SR Face Database (refer to Fig. 4.4-iii and -iv for sample results). Some resolution improvement can be seen in Fig. 4.4-iii, but due to accumulation of the effects of inaccurate motion estimations in Fig. 4.4-iv we observe highly degraded and distorted face image which can disable proper face recognition.

4.5 Subspace-domain Super-resolution

The treatment of the problem in Fig 4.4-iv resides in either increasing the time resolution of the video to have more accurate motion estimation (which is not possible if you do not have control over the acquisition device) or developing a more robust method that handles this problem in a different way. Henceforth, in the following section we will delineate how we implemented proposed subspace based super-resolution algorithms.

4.5.1 Obtaining Face Subspaces

In our experiments, we have the Illumination and Expression databases of CMU PIE database [15] for constructing face subspaces both for principal and independent component domain. We did not used M2VTS and VPA-SR databases while building face subspaces, in order to make super-resolution experiments more realistic. CMU Illumination and Expression databases encompass frontal face images of 68 people. In the CMU Illumination databases face images are taken in four different illumination conditions while faces are in fixed position, so as to enable investigating just the effects of illumination over face recognition systems. On the other hand, CMU Expression database contains face images of these 68 people in similar illumination conditions with four different facial expressions.

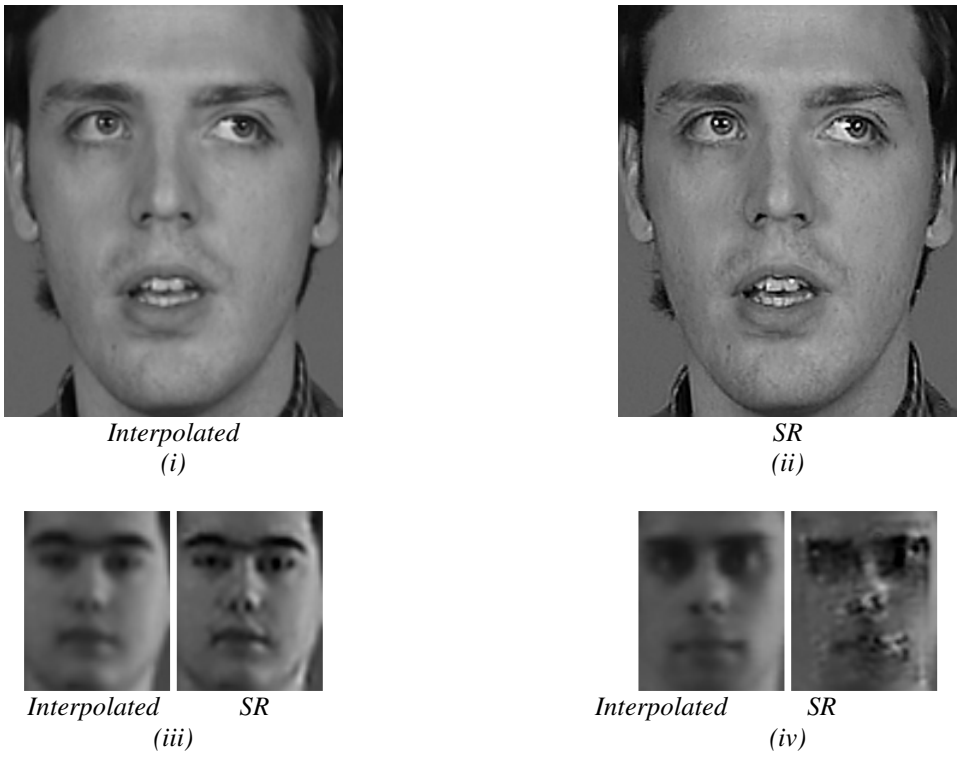


Figure 4.4: Interpolated face image *(i)* and when pixel domain super-resolution (SR) applied to it *(ii)* from M2VTS database. Sample results from VPA SR Face Database when interpolation and SR is applied *(iii)* and *(iv)*.

(a) (b)

Figure 4.5: Eigenfaces extracted from CMU Illumination and Expression databases (a), and independent component basis images estimated by using CMU Illumination and Expression databases (b).

In the estimation of independent components and extraction of principal components, we have used both illumination and expression databases where each one contains 68x4=272 face images that totally sum up to 544 face images. We manually aligned these face images according to the locations of the eyes, cropped and scaled them in order to fit into a 68 x 84 pixel reference high resolution face image model (same procedure applied to both M2VTS and VPA databases). Later, we downsampled these images into 34 x 42 pixel, so as to have low resolution face images that we have used to construct principal and independent component subspaces. We have reduced the dimensionality to 100 in both principal and independent component analysis while constructing subspaces of high-resolution and low-resolution face images (refer to Fig. 4.5 for basis vectors that span the subspaces).

In the estimation of mixing and de-mixing matrices of ICA, we have used the so called second architecture defined in [9]. This architecture uses ICA to find a representation, in which the coefficients used to code images are statistically independent, i.e., a factorial face code (refer to Fig. 4.6). For encoding objects that are characterized by high-order combinations of features, Barlow and Attick have discussed advantages of factorial codes [26, 27].

To achieve such a factorial coding of face images, we organized the data matrix X so that each column of X represents a different face image where each one is normalized to zero mean and unit variance. This corresponds to treating the columns of the mixing matrix A_x as the set of basis images. The ICA coefficients for a single face image, s_x, are obtained by,

$$S_x = W_x X \tag{4.1}$$

where W_x and X are de-mixing and data matrices, respectively, and each column of S_x contains the ICA coefficients of the basis images (i.e., s_x vector). As given in eq. (2.14) the de-mixing matrix can be written as

$$W_x = W D_{100}^{-1/2} E_{100}^{T} \tag{4.2}$$

where E_{100}^{T} reduces the dimensionality of the input data to 100 by projecting it onto the principal component subspace and the $D_{100}^{-1/2}$ matrix spheres the input projected data. We know that sphering is an optional stage which just fastens the convergence speed of the algorithm. However, it makes ICA estimation vulnerable to noise because the trailing eigenvalues which tend to capture noise in the data appear as denominator, hence we exclude sphering in our experiments, as we have mentioned before. The W matrix in eq. (4.2) projects the data onto the independent component face subspace and we obtain the ICA representation of the face images.

4.5.2 Reconstruction of Feature Vectors:

For each LR face video, one frame is selected as the reference frame then bilinearly interpolated by a factor of two and projected onto the principal component subspace ϕ or the independent component subspace W_x as an initial estimate of the high resolution feature vectors. Two previous and two next consecutive frames are projected onto the principle component subspace ψ or the independent component subspace W_y together with the low resolution reference frame to extract independent and principal component feature vectors in low resolution face subspaces. The $H^{(i)}$ matrices contains the decimation matrix $D^{(i)}$, the blur matrix $B^{(i)}$, and the motion warping matrix $W^{(i)}$. Here, the decimation matrix maps mean value of pixels of the high resolution image in 2x2 block of pixels to one pixel in the low resolution image; hence it is a fixed matrix for all observations.

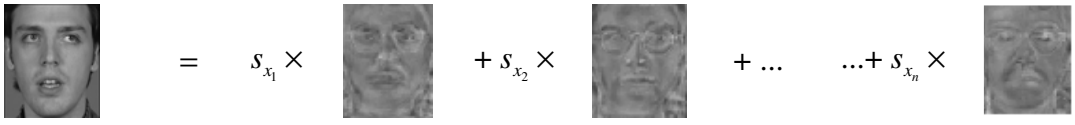


Figure 4.6: Factorial code representation attained by second architecture of ICA, where s_x's are statistically independent coefficients and basis images are columns of mixing matrix A_x.

For blurring, we have used a fixed blur matrix which convolves the image with a 5x5 Gaussian kernel with zero mean and one pixel variance. On the other hand, the motion warping matrix $W^{(i)}$ is needed to be calculated for all observations. However, it is practically impossible to have the actual $W^{(i)}$, but we can still estimate the motion warping matrix $W^{(i)}$ by using low-resolution images or frames. If the motion vector of a pixel found by using low-resolution images is multiplied by a constant 2 and mapped to the motion vector of the corresponding 2x2 block of pixels in the high-resolution image where 2 is the downsampling factor that is used in our experiments, we will have a reasonable estimate of $W^{(i)}$. Using the model parameters estimated for independent and principal component analysis, we estimated the high resolution feature vectors as described in Sec. 3.2 and 3.3.

4.6 Face Recognition Performances

Our face recognition scenario resembles a possible security system that tries to recognize faces coming from surveillance cameras which contains low quality face images due to distance between the face and the camera and the ambient illumination conditions. Here, we assume that the security system has high-resolution face images of the possible suspects and wants to recognize him/her from the surveillance videos for some security concerns. We have presented four different face recognition techniques for M2VTS database experiments that are implemented for enabling comparisons. We have used three different distance metrics in the decision phase of the algorithms: (i) L1 norm, (ii) L2 norm, and (iii) cosine similarity (or normalized correlation coefficient, CC).

The aim of using the M2VTS database is to emphasize the differentiation of PCA and ICA representations in case of a noisy environment. By using Gaussian blur kernels with varying variance, we have simulated this effect with the assumption that Gaussian blur degrades the quality of the low-resolution frames which appears as noise in the canonical representations of the faces in subspaces. Observe Fig. 4.7. Note the cases when super-resolution is not applied (point-and-dashed lines in the Fig. 4.7, just bilinearly interpolated frames are used), figures supports the claims given in [9] and [11] that ICA offers a better representation than PCA for face recognition in L1 norm but they provide nearly same recognition performance in L2 norm and CC distance metrics .

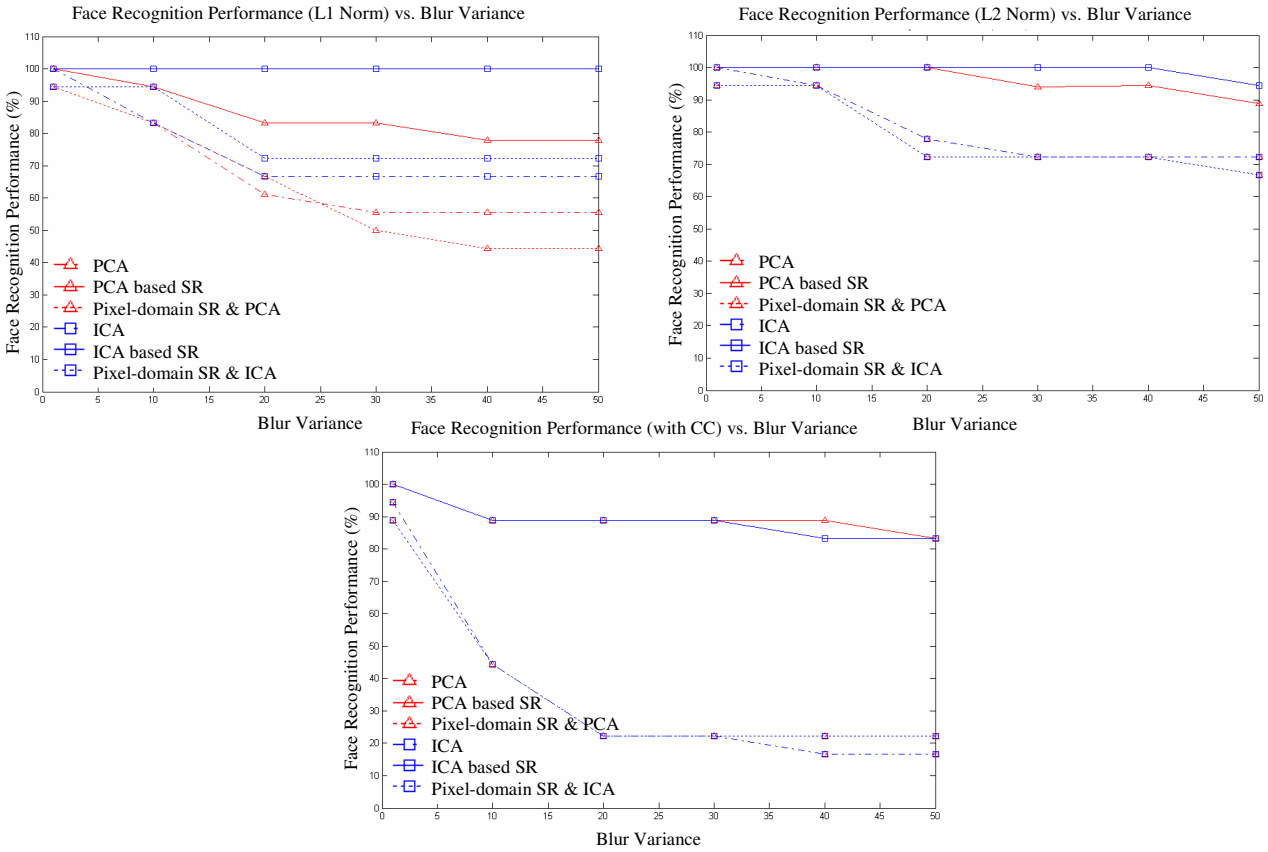


Figure 4.7: Comparison of the face recognition performances with L1, L2, and CC as distance metrics

If we compare the proposed independent component based super-resolution algorithm (with legend ICA based SR) with the principal component based one (with legend PCA based SR), in L1 and L2 norm space, we observe that as noise level increases, the proposed method outperforms PCA based SR with 100% recognition rate in all noise levels. Note that the blur in pixel-domain is regarded as noise in canonical representation, in other words in subspace representation.

The face recognition scheme for M2VTS database is formed by selecting a frame randomly as training frame in each face video sequence. Afterwards for testing, five consecutive frames are chosen 2 seconds after the training frame, thus training and testing frames have different expression and slightly different pose. Later, these five consecutive frames are convolved with Gaussian blur and downsampled by a factor of 2; hence low-resolution frames are attained for testing. In the Fig. 4.7, the face recognition is done by classifying the output feature vectors of low-resolution center frame after the application of subspace super-resolution methods (PCA based SR and ICA based SR legends in the Fig. 4.7), after the application of the pixel domain super-resolution methods (legends with Pixel-domain SR) and after the application of the bilinear interpolation (single ICA and PCA legends in the figures), separately. Observe that results of the pixel domain super-resolution are close to the results of the interpolated ones.

Using synthetically created low-resolution video frames brings about the question whether these methods can work in a real situation where neither high resolution images nor low-resolution video frames are created by some downsampling operation. In the generative imaging model given in eq.(2.17), we observe that the low-resolution image, y, is formed by blurring and downsampling the high-resolution image. Creating low-resolution video frames by blurring and downsampling and then reconstructing them with the same generative model appears to work due to the formulation, but would this model work when there is no synthetic relation of formation between the high and low resolution images? VPA Face Database described in Sec. 4.3 is formed to test whether this generative imaging model works in a more realistic scenario. Table 1 shows results of the five different face recognition approaches with cosine similarity metric. Observe that pixel domain super-resolution algorithm, which works pretty well when the noise level is low and motion estimation errors are small, fails to improve the recognition rate compared with the straight forward application of bilinear interpolation shown in the first column. However, subspace techniques works equally well in this level of noise and nearly 20% improvement in the recognition rate is achieved.

Table 1. Recognition performance (%) for, Column # 1: Bilinear interpolation, Column # 2: Pixel-domain super-resolution, Column # 3: Subspace-based Bayesian (MAP) estimation (the result in the ICA row is the result of the proposed method), Column # 4: Subspace-based POCS with outliers of residual constraint, Column # 5: Subspace-based POCS with variance of the residual constraint.

#	*1*	*2*	*3*	*4*	*5*
PCA	*56.25*	*56.25*	*75*	*62.5*	*62.5*
ICA	*56.25*	*56.25*	*75*	*62.5*	*62.5*

CONCLUSIONS

This paper aims to improve the face recognition performance of existing systems that use static images by means of improving their resolution via incorporating information coming from multiple video frames. A general framework that enables comparison of different super-resolution techniques on face recognition problem is given for the first time in the literature. One of the outcomes of this paper is to implement and demonstrate that super-resolution formulations derived from generative imaging model works in a non-synthetic experimental setup. Previously in the literature low-resolution images were generally created synthetically from high resolution images; therefore there was a need to demonstrate performance of super-resolution techniques in a real setup. For this a new database, called VPA SR Face Database, is created. This paper also presents how pixel-domain super-resolution can be achieved from face videos where there are non-rigid motions together with rigid ones. In preceding works, pixel-domain super-resolution is just applied in case of single or multiple objects having rigid motions. By extending studies in pixel-domain super-resolution into non-rigid motion circumstances, this paper is the first effort. Besides, POCS based signal reconstruction algorithm in the subspace domain is formulated for super-resolution and it is found to improve face recognition performance by 6.25 % in VPA SR Face Database compared to pixel-domain super-resolution. Application of eigenface-based super-resolution method proposed by Gunturk et al [5] was limited to synthetic low resolution still face images. Here, we applied this method to real video sequences and demonstrated that it improves face recognition performance. Also the independent component based super-resolution technique using Bayesian estimation is revealed to be superior to eigenface-based one as noise level increases. Moreover, an increase in recognition rate of 18.75% is achieved by using Bayesian estimation based subspace super-resolution compared to applications without super-resolution. In conclusion,

we can say that subspace based super-resolution methods together with computational advantages, provides robustness against noise compared to pixel-domain super-resolution algorithms. Consequently, super-resolution on face subspace appears to be a promising approach to improve the face recognition performance of the currently existing face recognition systems.

ACKNOWLEDGEMENTS

Authors acknowledge Assist. Prof. Hasan Ates's useful discussions on set-theoretic approach. Authors also express their gratitude to the members of Computer Vision and Pattern Analysis (VPA) and Telecommunications Laboratories and graduate students in Sabancı University for participating VPA Super-resolution Face Database.

REFERENCES

[1] T. E. Boult, M.-C. Chiang, and R. J. Micheals, *Super-Resolution via Image Warping*, Super-Resolution Imaging, Ed. S. Chaudhuri, Boston: Kluwer Academic Publishers, 2002, pp. 131-169.

[2] S. Baker and T. Kanade, *Hallucinating Faces*, Fourth International Conf. Automatic Face and Gesture Recognition, March 2000.

[3] C. Liu, H.-Y. Shum, and C.-S. Zhang, *A Two-Step Approach to Hallucinating Faces: Global Parametric Model and Local Nonparametric Model*, Proc. IEEE Conf. Computer Vision and Pattern Recognition, December 2001.

[4] D. Capel and A. Zisserman, *Super-Resolution from Multiple Views Using Learnt Image Models*, Proc. IEEE Conf. Computer Vision and Pattern Recognition, December 2001.

[5] B. K. Gunturk, A.U. Batur, Y. Altunbasak, M. H. Hayes III, and R. M. Mersereau, *Eigenface-based Super-resolution for Face Recognition*, Proc. International Conf. on Image Processing, 2002

[6] Makeig, S., Bell, A. J., Jung, T., & Sejnowski, T. J. (1996). *Independent component analysis of electroencephalographic data.* In D. Touretzky, M. Moser, & M. Hasselmo (Eds.), Advances in neural information processing systems, 8(pp. 145–151). Cambridge, MA: MIT Press.

[7] P. O. Hoyer and A. Hyvarinen*, Feature Extraction from Colour and Stereo Images*, Network: Computation in Neural Systems 11: 191-210,August, 2000

[8] J. Lee, H. Jung, T. Lee, and S. Lee, *Speech feature extraction using independent component analysis*, IEEE ICASSP, June 2000, vol. 3, pp. 1631–1634.

[9] M. S. Bartlett, J. R. Movellan, and T. J. Sejnowski, *Face Recognition by Independent Component Analysis*, IEEE Trans. On Neural Networks, vol. 13, no. 6, pp.1450-1464, 2002

[10] C. Liu, and H. Wechsler, *Comparative Assessment of Independent Component Analysis (ICA) for Face Recognition*, Proc. Second International Conf. on Audio- and Video-based Biometric Person authentication, March 1999.

[11] J.-P. Nadal and N. Parga, *Non-linear neurons in the low noise limit: A factorial code maximizes information transfer*, Network, vol. 5, pp.565–581, 1994.

[12] M. Turk and A. Pentland, *Eigenfaces for Recognition*, J. Cognitive Neuroscience, vol. 3, no. 1, 1991

[13] Sirovich, L. and Kirby, M., *Low dimensional procedure for the characterization of human faces*, J. of the Optical Society of America, vol. 4, no. 3, pp. 519–524, 1987.

[14] A. Hyvarinen, *Survey on independent component analysis*, Neural Computing Surveys, 2, (1999), 94-128.

[15] T. Sim, S. Baker, and M. Bsat, *The CMU Pose, Illumination, and Expression (PIE) Database*, Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition, 2002.

[16] A.J. Patti, M. I. Sezan, and A. M. Tekalp, *Super-resolution video reconstruction with arbitrary sampling lattice and nonzero aperture time*, IEEE Trans. Image Processing, vol. 6, Aug. 1997, pp:1064-1076.

[17] Schultz, R. R. and Stevenson, R. L., "Extraction of high-resolution frames from video sequences," IEEE Trans. Image Processing, vol. 5, pp. 996–1011, June 1996.

[18] Elad, M. and Feuer, A., "Restoration of a single superresolution image from several blurred, noisy and undersampled measured images," IEEE Trans. Image Processing, vol. 6, pp. 1646–1658, December 1997.

[19] M.I. Sezan, *An overview of convex projections theory and its application to image recovery problem*, Ultramicroscopy, vol. 40, 1992, pp:55-67.

[20] H. Stark and Y. Yang, *Vector Space Projections: A numerical Approach to Signal and Image Processing, Neural Nets, and Optics*, Wiley series in Telecommunications and Signal Processing,1998

[21] H. J. Trussell and Mehmet R. Civanlar, *The feasible solution in signal processing*, IEEE Trans. Acous., Speech, and Signal Process., vol. ASSP-32, no. 2, April 1984, pp:201-212.

[22] M2VTS Multimodal Face Database Release 1.00, http://www.tele.ucl.ac.be/M2VTS

[23] M. Bierling and R. Thoma, *Motion Compensating Field Interpolation Using A Hierarchically Structured Displacement Estimator*, Signal Processing, 11, 1986, pp: 387-404.

[24] P.E. Eren, M. I. Sezan, and A. M. Tekalp, *Robust, Object-Based High-Resolution Image Reconstruction from Low-Resolution Video*, IEEE Trans. on Image Processing, vol.6, no. 10, October 1997, pp: 1446-1452.

[25] M. Irani, and S. Peleg, *Motion Analysis for Image Enhancement: Resolution, Occlusion, and Transparency*, J. Visual Commun. Image Represent., vol. 4, Dec. 1993, pp:324-335.

[26] O. G. Sezer, *Superresolution Techniques for Face Recognition from Video*, M.Sc. Thesis, Sabanci University, Istanbul, Turkiye, 2005

[27] H. B. Barlow, *Unsupervised learning*, Neural Comput., vol. 1,1989, pp.295–311.

[28] J. J. Atick, *Could information theory provide an ecological theory of sensory processing?*, Network, vol. 3, 1992, pp. 213–251.