SUPERRESOLUTION TECHNIQUES
FOR FACE RECOGNITION FROM VIDEO

by
Osman Gökhan Sezer
B.S., E.E., Boğaziçi University, 2003

Submitted to the Graduate School of Engineering
and Natural Sciences in partially fulfillment of
the requirement for the degree of
Master of Science

Graduate Program in Electronics Engineering and Computer Science
Sabancı University
Spring 2005

SUPERRESOLUTION TECHNIQUES
FOR FACE RECOGNITION FROM VIDEO

APPROVED BY:

Prof. Aytül Erçil …………………………
(Thesis Supervisor)

Assoc. Prof. Yücel Altunbaşak …………………………
(Thesis Co-supervisor)

Assist. Prof. Hasan Ateş …………………………

Assoc. Prof. Mustafa Ünel …………………………

Assist. Prof. Berrin Yanıkoğlu …………………………

DATE OF APPROVAL: …………………………

ACKNOWLEDGEMENTS

I would like to thank Prof. Aytül Erçil for her kindness, understanding, and support throughout my academic life together with being the one who introduced me computer vision. I am grateful to Assoc. Prof. Yücel Altunbaşak for his important contributions to my thesis. I also acknowledge Assist. Prof. Hasan Ateş's useful discussions on set-theoretic approach. I owe many thanks to Assoc. Prof. Mustafa Ünel for his encouragements and sharing his scientific vision and, also to Assist. Prof. Berrin Yanıkoğlu for participating in my thesis jury.

I wish to express my gratitude to the members of VPA and Telecommunications Laboratories and graduate students in Sabancı University for participating our superresolution face database.

I would like to thank to the founders of this wonderful university, especially Sakıp Sabancı (RIP), for providing us such a comfortable and creative atmosphere.

Finally, many many thanks to my family, Ayfer and İsmail Sezer …

ABSTRACT

SUPERRESOLUTION TECHNIQUES
FOR FACE RECOGNITION FROM VIDEO

Performance of current face recognition algorithms reduces significantly when they are applied to low-resolution face images. To handle this problem, superresolution 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 and those already in the literature.

ÖZET

ÇÖZÜNÜRLÜLÜK ARTIRICI YÖNTEMLERLE VİDEODAN YÜZ TANIMA

Mevcut yüz tanıma algoritmalarinin başarımı düşük çözünürlüklü yüz imgelerine uygulandıklarında önemli ölçüde azaltmaktadır. Bu problemin çözmek için çözürlülük arttırma yöntemleri piksel alanında yahut yüz alt-uzayında uygulanabilmektedir. Yüz imgeleri coğu yüz tanıma işlevi açısından gereksiz yüksek boyutlu verilerden oluşur, bu da boyut düşüren öznitelik çıkarma yöntemlerini yüz analizinde standart uygulama haline getirmiştir. Dolayısıyla çözünürlülük artırma yöntemlerini piksel alani yerine öznitelik alanında, bir başka deyişle yüz alt-uzayında, uygulamanın hesaplamalar açısından yararları olduğu gibi gürültüye ve hareket kestirimi hatalarına karşı gürbüzlüğü de sağlamıştır. Bu nedenle, biz Bayesçi kestirim ve dışbükey kümelere izdüşüm yöntemleriyle öznitelik tabanlı çözünürlülük arttırıcı yeni algoritmalar önermekte ve önerilen yöntemleri literatürde mevcut olanlar ile karşılaştırmalı analizini sunmaktayız.

Table of Contents

List of Figures

List of Tables

1. INTRODUCTION

Face recognition has received significant attention during the last decade and many researchers study various aspects of it and still face recognition studies dominate computer vision and pattern recognition conferences. There are at least two reasons for this trend; the first one is wide range of commercial and law enforcement applications and the second is the availability of feasible computer technology to develop and implement applications that demand strong computational power. Today, automatic recognition of human faces became a field that gathers many researchers from different disciplines such as image processing, pattern recognition, computer vision and graphics, and psychology.

1.1 Applications of Face Recognition

Every day, we are facing new products of technology, prompting us to enter our PIN code or passwords such as money transactions in the internet or to get cash from ATM, even to use our cell phone SIM card, a dozen of others to access internet and so on. Therefore, the need for reliable methods of biometric personal identification is obvious. In fact, there are such reliable methods like fingerprint analysis, retinal or iris scans, however these methods rely on the cooperation of the participant. Face recognition systems, on the other hand, can perform person identification without the cooperation or knowledge of participant which is advantageous in some applications such as surveillance, suspect tracking and investigation. Typical application of face recognition systems can be listed in four main categories:[1]

(i) Entertainment: Video Game, Virtual Reality, Training Programs, Human-Robot-Interaction, Human-Computer-Interaction.

(ii) Smart Cards: Driver's Licenses, Entitlement Programs, Immigration, National ID, Passports, Voter Registration, Welfare Fraud

(iii) Information Security: TV Parent Control, Personal Device Logon, Desktop Logon, Application Security, Database Security, File Encryption, Intranet Security, Internet Access, Medical Records, Secure trading Terminals

(iv) Law Enforcement and Surveillance: Advance Video Surveillance, CCTV Control, Portal Control, Post-Event Analysis, Shoplifting, Suspect Tracking and Investigation.

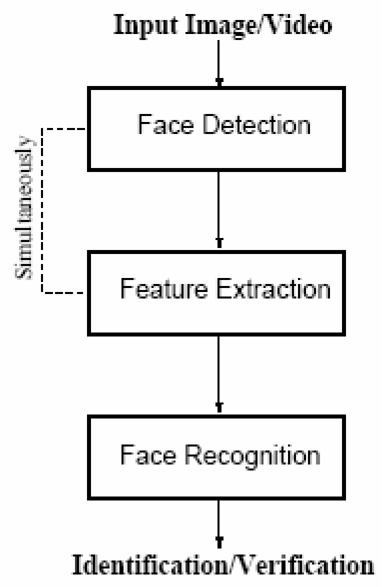


Figure 1.1. General Configuration of Face Recognition Systems [1]

These commercial and law enforcement applications of face recognition systems vary according to the format of the imagery such as static images, controlled-format face images to uncontrolled video sequences.

We can classify these face recognition systems into two categories; those using (i) static images and (ii) video frames. Independent of this categorization, we can give the general configuration of face recognition systems, as depicted in Fig 1.1.

There are advantages and disadvantages of using a static image or video frames depending on the specific application and the availability of the data. Moreover, challenges posed by the face detection, extraction and recognition algorithms differ for static images and video frames. For applications such as recognizing one from his/her driver's license ID photograph, a system using static face image would be enough due to controlled nature of the image acquisition process. However for recognizing a face from a scene image of an airport where there are many faces using video information for detecting faces can be more helpful than a single static image.[1]

1.2 Face Recognition

Research on automatic face recognition started in the 1970's after the seminal work of Kanade [2,3] and Kelly [4], till then more than thirty years extensive research has been conducted in this field by people coming from different backgrounds from engineers to neuroscientists. Feature-based methods appear in these first efforts to build automatic face recognition systems. In this approach, different geometric parameters are used to describe facial properties such as eyebrow thickness, eyebrow's arches and width of nose. Later in 1990's, template matching-based methods were accepted and investigated by the research community. In template-matching methods general practice was to compare intensity values of the whole or part of the faces with those faces in the database, refer to [5] for comparison of these first popular methods. In these comparisons, template-matching algorithms were found to attain higher correct recognition rates than feature-based methods, with the expense of high computational demand due to dimensionality of lexicographically ordered face image data. This problem was later eliminated by making use of intrinsic dimensionality of the face space which is known to be much smaller due to similar topology they have. The appearance-based face recognition methods that capture or approximate these face manifolds are later developed such as principal component analysis (PCA) [6,7], Fisherface [8], and independent component analysis [9] and successfully applied to face recognition problem.

1.3 Scope of the Thesis

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 [10-13]. These pixel domain techniques can be divided into two main categories as: those methods using face specific constraints [11, 12, 13] and without any face specific constraints [10]. Different from applying superresolution to any arbitrary image, face images have a fixed configuration (i.e, relative location of the mouth, nose or eyes are fixed) which can be utilized to obtain a better super-resolution performance. Therefore, using holistic methods, such as principle component analysis (PCA), which incorporates configuration information of face into a reduced dimensional subspace model, enables faster computation and robustness against noise and motion estimation errors [14].

Face recognition performance of independent component analysis (ICA) and comparative analysis of it with different recognition algorithms have been studied exhaustively in many research papers [9,15]. 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 [16] that the metric induced by ICA is superior to PCA by providing a representation that is more robust to the effects of noise. In addition to this, it is shown in [9] that linear reconstruction based on ICA gives better SNR than 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 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 thesis, we present the fusion of independent component analysis with superresolution techniques using Maximum A Posteriori (MAP) estimate. We observed that modeling noise processes in low dimensional subspace and then extraction of statistics of these processes from face images for estimating canonical representation of high resolution face provide robustness against noise and motion estimation errors when it is judged against pixel domain estimation algorithms. Besides, the independent component face-subspace based superresolution is found to be superior to eigenface-based super-resolution for robustness against noise. Two different experimental setups are used for evaluating recognition performance of the proposed algorithm. In the first setup real video sequences with different blur kernels are used to compare the independent and principal component based superresolution techniques. Second experimental setup is employed to verify that the proposed algorithm works in a real scenario. We have created a new database (VPA Superresolution Face Database) to test the effect of the developed superresolution technique and those techniques that are already available in the literature to the face recognition problem.

1.4 Outline of the Thesis

The organization of the thesis is as follows: In Chapter 2, mathematical overview of subspace methods (principal and independent component analysis) used in the proposed algorithms is explained. Superresolution techniques both in pixel and subspace domain is provided with details in Chapter 3. In Chapter 4, experimental procedure and results are presented and conclusions are given.

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 $x_k \in \mathbb{R}^{\mathrm{N}^2}$ 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) $\overline{x}_k \in \mathbb{R}^{\mathrm{n}}$ in the reduced dimensional subspace;

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

where E_n is a $\mathrm{N}^2 \times n$ matrix with orthonormal columns containing corresponding eigenvectors of the scatter matrix having the largest n eigenvalues [6,7].

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 [17]:

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

where x is the $\mathrm{N}^2 \times 1$ observation vector containing the lexicographically ordered observed data, s is the $n \times 1$ source vector and A is the $\mathrm{N}^2 \times n$ mixing matrix ($\mathrm{N}^2 \gg n$). 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 $n \times \mathrm{N}^2$ 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 since the trailing eigenvalues which tend to capture noise in the data appear as denominator, observe relative magnitude of eigenvalues extracted from CMU PIE Database [18] 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 ICA face recognition performance 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 one of the ICA algorithms available in the literature. We have used symmetric fixed point algorithm with *f*(x)=*tanh(x)* nonlinearity in order to obtain a fast solution using a simple algorithm.

The algorithm starts from a random orthogonal matrix *W* and in each iteration rows of it, (w^T), is updated by;

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

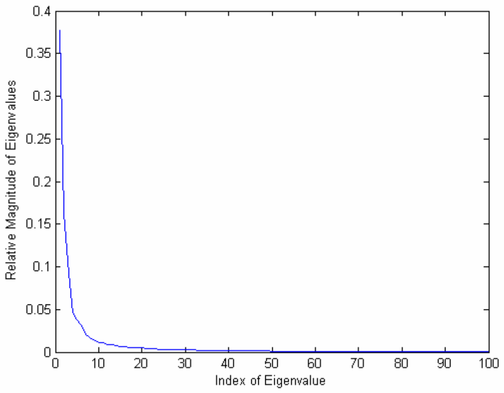


Figure 2.1. Relative magnitude of eigenvalues $\left(\lambda_i / \sum_k \lambda_k\right)$

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 $\mathrm{N}^2 \times n$ 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;

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

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.

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

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

$$\begin{aligned} s_x &= W D_n^{-1/2} \overline{x} \\ \overline{x} &= (W D_n^{-1/2})^{-1} s_x \\ &\mathit{and} \\ \overline{x} &= D_n^{1/2} W^T s_x \end{aligned} \tag{2.10}$$

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 $w^T x=w^T v+w^T e_x$**.** Since we assume e_x to be a Gaussian noise, $w^T e_x$ will be zero (e.g., kurtosis of a Gaussian random variable will be zero). Thus, the measure of non-Gaussianity for $w^T x$ (noisy data) will be equal to the measure of non-Gaussianity for $w^T v$ (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 [17].

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

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

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,

$$W_x e_x = W D_n^{-1/2} E_n^T e_x = 0 \quad \Leftrightarrow \quad E_n^T e_x = 0 \tag{2.16}$$

3. SUPERRESOLUTION

3.1 Overview

Superresolution algorithm is formulated as a signal restoration problem where the original form of the signal is assumed to be the high-resolution image and is estimated from degraded and noise corrupted versions of this high-resolution image, in other words, from its low-resolution versions. The signal restoration problem is a common problem of various fields in signal processing including image processing (as in our case), speech processing, and system identification. Therefore there are various techniques in the literature to solve the restoration problem under different degradation models. The most common model can be written as:

$$y = O[x] \tag{3.1}$$

where x in $\mathbb{R}^{N^2}$ and y in $\mathbb{R}^{M^2}$ are high-resolution and low-resolution image, respectively, in lexicographical order, and $O[.]$ is an operator. The M and N are related by the following equation

$$\mathrm{M}=\ell.\mathrm{N} \text{ and } 0<\ell<1 \tag{3.2}$$

The straightforward solution to this problem is to apply inverse filtering to recover x by finding $O^{-1}[.]$ which is the least square inverse. However, since M<N the inverse filtering is ill-conditioned [19,20], it amplifies the noise present in the observations. Information about the degradation operator, $O[.]$, is just used to obtain the inverse filter. Nevertheless, priori information about the original high resolution image and the distortion mechanism should somehow be incorporated into the model so as to obtain better restoration methods. The estimate of the high resolution image, $\hat{x}$, can be written as:

$$\hat{x} = R[y] \tag{3.3}$$

where $R[.]$ is the operator derived from the degradation model and a priori knowledge employed and y is the low-resolution image. Due to the dimensionality of the data the amount of a priori knowledge is very high and it is not possible to obtain $R[.]$ in one step. Therefore, iterative techniques which are very flexible for employing various types of a priori information are needed to find reasonable high-resolution image estimates. Projection onto Convex Sets (POCS), which will be described shortly, is such an iterative technique where a priori information is defined as convex constraint sets.

3.2 Pixel-domain Imaging Model

To have better analysis of the distortion mechanism, degradation operator and noise process are separated in the imaging model which can be given as:

$$y^{(i)} = H^{(i)}x + n^{(i)}, \text{ for i=1...K} \tag{3.4}$$

where 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:

$$H^{(i)} = DBW^{(i)} \tag{3.5}$$

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 (*cf.* a practical blur computation method can be found in the Appendix of [21]), hence only the motion warping matrix changes depending on the observed low-resolution image. Further information about the imaging model can be found in [21, 22, 23]. In the section 4.2.1, the registration algorithm used in the pixel-domain superresolution is given in detail. Moreover, the motion warping matrix, $W^{(i)}$, is obtained by using motion vectors coming from this registration phase of the algorithm.

In eq. (3.4) what we observe is a set of linear equations where, due to ill-posed nature of the problem, inverse filtering does not help recovering the high-resolution image $\hat{x}$. Hence, regularization methods are needed which refer to finding an "acceptable" estimate of the ideal solution to this ill-posed problem. Two significant aspects of regularization are (i) quantitatively defining what is an "acceptable" estimate, and (ii) making use of a priori information and constraints about the degradation operator $H^{(i)}$, the high-resolution image *x*, and the low-resolution image *y* in determining the estimate [24].

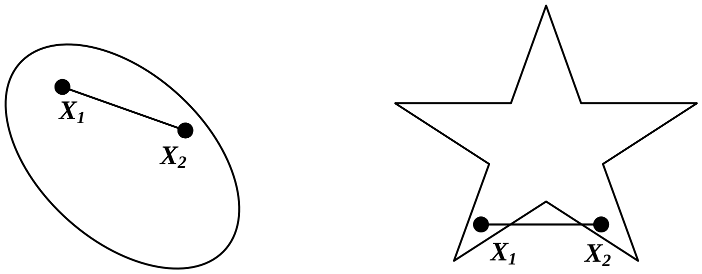


Figure 3.1. Examples of convex and non-convex sets (left and right diagrams)

According to how an estimate of the actual image is defined, image recovery methods can be categorized into three groups. An estimate can be defined on the basis of (i) optimality criterion (minimum mean square error and Bayesian methods); (ii) an optimality criterion subject to constraints (as in the case of constrained least square and maximum entropy methods); (iii) a priori information and constraints (as in the case of set-theoretic methods such as projection onto convex sets (POCS) method, method of generalized projections)[25]. There are papers in the literature which investigates the relation between different regularization methods such as [26] and [27]. In this thesis, for pixel domain superresolution our focus will be on the method of POCS.

3.3 Projection onto Convex Sets (POCS)

POCS is an iterative method which enables 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 set (refer to Fig. 3.1) 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 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. The fundamental mathematical concepts for POCS are given in [25, 28], here we will give short definitions necessary for understanding the method of POCS.

The theorem of POCS is limited to Hilbert spaces, such as N-dimensional Euclidean space $\mathbb{R}^N$ with usual inner product and norm definitions. If we denote X_1 and X_2 as two vectors in Hilbert space (i.e., in $\mathbb{R}^N$), we can give a formal definition of convexity for the set S as: The set S in the $\mathbb{R}^N$ is said to be convex if and only if for any two members of the set X_1 and X_2, the vector $X = \delta X_1+(1-\delta)X_2$ is a member of the set S for 0<δ<1 (refer to Fig. 3.1, left diagram). In addition to that, in order to have a closed set S in the Hilbert space, every convergent sequence $\{X_n\}$in S should converge to a vector in S. Let C be a closed convex set and x be a vector outside the set C both in a Hilbert space $\mathcal{H}$, then these definitions ensure the existence of a unique element x^* in C as the closest point to the vector x.

$$\min_{y\in C}\|x-y\| = \|x-x^*\| \triangleq \|x-\mathrm{P}x\| \tag{3.6}$$

where P: $\mathcal{H} \rightarrow C$ is the projection operator which projects onto the set C.

There are three essential features that play a key role in the convergence theory of POCS; the projection operator has to be (i) nonexpansive, (ii) asymptotically regular, and (iii) the projection operator has to have a fixed point. First, an operator O: $\mathcal{H} \rightarrow \mathcal{H}$ is said to be nonexpansive if for any X_1 and X_2 in $\mathcal{H}$, the following inequality is satisfied;

$$\|\mathrm{O}X_1 - \mathrm{O}X_2\| \le \|X_1 - X_2\|. \tag{3.7}$$

Similarly, the operator O is said to be asymptotically regular if and only if for any $X\in \mathcal{H}$, we have,

$$\lim_{n\to\infty}\left\|\mathrm{O}^{\mathrm{n}}X - \mathrm{O}^{\mathrm{n+1}}X\right\| = 0. \tag{3.8}$$

Finally, the operator O has a fixed point if successive applications of the operator O on an arbitrary vector X_0 converge to the same point. Hence, a fixed point X_{fp} should satisfy the equation,

$$\mathrm{O}X_{fp} = X_{fp}. \tag{3.9}$$

These features are used by Youla in [29] to prove the convergence criteria of the method of POCS for the first time to the engineering literature. Therefore, on the basis of these essential features the POCS algorithm converges to a feasible point in the intersection of the constraint sets $C_0 = \cap_i C_i$ where C_0 is a nonempty and C_i is a closed convex set in the Hilbert space.

Let P_i be the projection operator which projects a vector onto the close convex set C_i which satisfies the essential features described above. Then, the iteration given by

$$X_{k+1} = P_1 P_2 ... P_M X_k \tag{3.10}$$

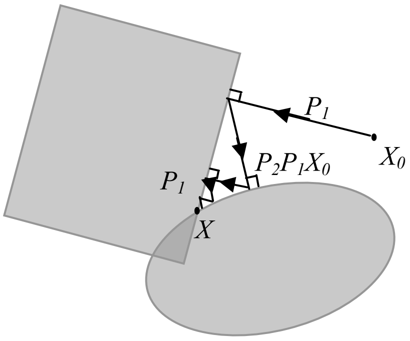


Figure 3.2. Illustration of the mechanism of POCS in case of two constraint sets. Initial point X_0 converges to the member of the intersection set X.

converges strongly to a point in C_0, if X_k's are finite dimensional vectors. Points in the intersection set are called feasible solutions. Thus, by this sequential projection method, a feasible solution can be reached. However, due to selection of the initial estimate and the order of projection at the beginning of the iterative process, uniqueness of the solution is not guarantied [25].

In this thesis, we have used the POCS algorithm for both pixel-domain and face subspace superresolutions, 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 noise statistics are used to employ constraints on the residual, defined by

$$r = y^{(i)} - H^{(i)}\hat{x} \tag{3.11}$$

where $\hat{x}$ is an estimate of the high-resolution image (i.e., original signal) in the model given in eq. (3.4). A priori information about the degradation operator $H^{(i)}$ can be estimated from 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 (3.4) and (3.5), 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 ℓ and mapped to the motion vector of the corresponding $\ell \times \ell$ block of pixels in the high-resolution image where ℓ is the downsampling factor set in the imaging model (eq.(3.4)), 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 noise $n^{(i)}$ given in imaging model. Therefore, using confidence limits derived from sample statistics will enable us to determine the limits of approximation. In the Trussell and Civanlar's paper [30] such constraints on the statistics of the residual are defined thoroughly. In this thesis, for pixel-domain super-resolution we have used constraints defined for outliers of residual and amplitude of the high-resolution image estimate.

3.4 POCS based superresolution in the pixel-domain

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,

$$C_0 = \left\{ x \,|\, \left| y_j^{(i)} - \left[H^{(i)} x \right]_j \right| \leq \delta_0 \right\} \tag{3.12}$$

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,

$$r_i = y_j^{(i)} - \left[H^{(i)} x \right]_j \tag{3.13}$$

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

$$P_0 x = \begin{cases} x + \dfrac{(r_i - \delta_0)}{\left\| h_i \right\|^2} h_i & , \quad \text{if } r_i > \delta_0 \\ x & , \quad \text{if } -\delta_0 < r_i < \delta_0 \\ x + \dfrac{(r_i + \delta_0)}{\left\| h_i \right\|^2} h_i & , \quad \text{if } r_i < -\delta_0 \end{cases} \tag{3.14}$$

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,

$$C_A = \left\{ x \mid \alpha \leq x_j \leq \beta \right\} \tag{3.15}$$

to ensure appropriate gray-level images with amplitude bounds $\alpha = 0$ and β=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:

$$\hat{x}_{\ell+1} = P_A P_0 \hat{x}_{\ell} \qquad \text{for } \ell=0,1,2.. \tag{3.16}$$

The initial estimate of the high-resolution image for iterative projections is obtained by bilinearly interpolating one of the low-resolution image which is selected as the reference image. Please refer to [21] and [30] for further details. In the Fig. 3.3, application of pixel-domain POCS for increasing the resolution of a dollar image is illustrated. As a side note, like in the most of the application, where there are more than two constraints, pure projection operator with unity relaxation parameters is used in the superresolution applications in this thesis.

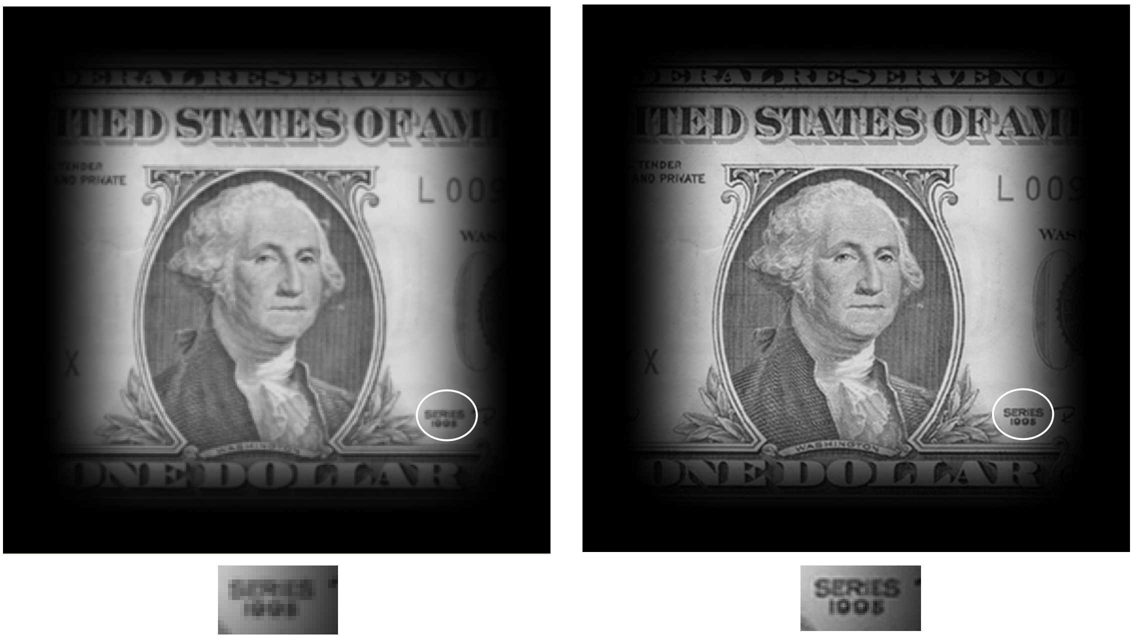


Figure 3.3. Result of bilinearly interpolating a LR dollar image by a factor of two (left), and result of superresolution algorithm with pixel-domain POCS method using 4 LR images (right).

3.5 POCS based superresolution using subspace methods

It is also possible to apply superresolution in the feature domain (i.e., face subspace domain) rather than implementing a pixel domain superresolution algorithm and then extracting features of the high resolution face image for face recognition. It is discussed in [14] that this approach not only brings computational advantages but also robustness against noise and motion estimation error. Hence in this section we will derive the necessary tools for enabling superresolution in the feature or subspace domain by making use of POCS algorithm.

Together with the outlier of residual constraints defined before, we have also used the variance of residual constraints for face subspace superresolution. For this case, variance of the residual is forced to be limited by the variance of the noise. The convex constraint set for the variance of the residual is defined as,

$$C_V = \left\{ x \mid \left\| y^{(i)} - \mathrm{H}^{(i)} x \right\|^2 \le \delta_V \right\}. \tag{3.17}$$

The projection operator of C_V, P_V can be formulated as,

$$P_V x = \begin{cases} x + (\mathrm{H}^{\mathrm{T}}\mathrm{H} + \frac{1}{\lambda} I)^{-1} \mathrm{H}^{\mathrm{T}} (y^{(i)} - \mathrm{H}^{(\mathrm{i})} x) & , \ \text{if}\, x \notin C_V \\ x & , \ \text{if}\, x \in C_V \end{cases} \tag{3.18}$$

where λ is the Lagrange multiplier coming from the optimization formulation which gives an idea about the modification done on the signal (in face-subspace). Derivation of the projection operators can be found in [30]. As we have mentioned earlier, Gaussian noise is assumed for residual of each point (or pixel) given in eq.(3.13) 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:

$$\delta_V = \sigma^2 \left[\pm \lim_{0.95} + \sqrt{2(N-1)} \right]^2 / 2N \tag{3.19}$$

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.

3.5.1 Principal Component Subspace

As it is mentioned before, the PCA algorithm extracts orthonormal linear projections, called eigenvectors that maximize the scatter of all projected samples. In the face recognition problem; first Sirovich and Kirby [7] showed that KLT-based (or PCA-based) dimensionality reduction can be used efficiently to represent face images by projecting face images onto low-dimensional linear subspace that is computed using the KLT. Later, Turk and Pentland utilized this idea and implemented a very successful face recognition system using PCA [6]. After these pioneering works, PCA is widely accepted in face recognition studies and became a standard procedure for dimensionality reduction.

The PCA algorithm enables one to represent a face image as 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:

$$x = \phi a + e_x \tag{3.20}$$

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 $N^2 \times 1$ reconstruction error which is orthogonal to the linear space defined by eigenfaces.

Using these definitions, in order to derive an efficient superresolution algorithm in this subspace, we need to obtain an observation model for the reconstruction of the eigenface coefficients of the high-resolution face images. The observation model will not neglect the spatial-domain observation noise given in eq.(3.4) 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. We will stick to the notion given in [14].

$$\begin{gathered} x = \phi a + e_x \\ y^{(i)} = \psi a^{(i)} + e_y^{(i)} \quad \text{for i=1...K} \end{gathered} \tag{3.21}$$

where x is $N^2 \times 1$ lexicographically ordered high-resolution face image, ϕ and e_x are $N^2 \times L$ eigenface matrix and $N^2 \times 1$ reconstruction error, respectively, for high-resolution face-subspace. Similarly, $y^{(i)}$ is the *i*'th observation of the low-resolution face image which is $\ell^2 N^2 \times 1$, 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 a and $a^{(i)}$ are for high and low dimension face-subspaces which have same dimensionalities, $L \times 1$.

Substituting the subspace representation of low and high resolution face images given in eq. (3.21) into imaging model, we will obtain

$$\begin{aligned} \psi a^{(i)} + e_y^{(i)} &= H^{(i)}(\phi a + e_x) + n^{(i)} \\ \psi a^{(i)} + e_y^{(i)} &= H^{(i)}\phi a + H^{(i)} e_x + n^{(i)} \end{aligned} \tag{3.22}$$

We know that if we project eq. (3.22) into lower-dimensional face subspace, we will eliminate the reconstruction error $e_y^{(i)}$ using the fact that it is orthogonal to ψ.

$$\psi^T e_y^{(i)} = 0, \quad \text{for i=1,....,K} \tag{3.23}$$

and since the eigenface matrix is orthonormal we have

$$\psi^T \psi = I\,. \tag{3.24}$$

and by multiplying both sides of eq. (3.22) with ψ^T we will get,

$$a^{(i)} = \psi^T H^{(i)} \phi a + \psi^T H^{(i)} e_x + \psi^T n^{(i)} \tag{3.25}$$

Observe that the model in eq. (3.25) resembles the imaging model in the pixel-domain given in eq. (3.4). 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. (3.11) we can write the residual in principal component subspace as

$$r^{(i)} = a^{(i)} - \psi^T H^{(i)} \phi \hat{a} \tag{3.26}$$

where $\hat{a}$ is the estimate of the eigenface 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. (3.26) so as to find the POCS estimate for this ill-posed problem. We have used outliers of residual and variance of the residual constraints defined in section 3.4. First, the convex set for outliers of the residuals are defined as,

$$C_0 = \left\{ a \mid \left| a_j^{(i)} - \left[\psi^{\mathrm{T}} \mathrm{H}^{(i)} \phi a\right]_j \right| \le \delta_0 \right\} \tag{3.27}$$

where subscript *j* denotes the *j*’th element of the vector and δ_0 is the a priori bound reflecting the statistical confidence with which the “true” feature vector a is a member of the set C_0. The bound, δ_0, is determined from noise statistics where we can write the noise in subspace observation as,

$$n_{subspace} = \psi^T H^{(i)} e_x + \psi^T n^{(i)} \tag{3.28}$$

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 and $H^{(i)}$ are linear operators, we will expect that the subspace representation error $n_{subspace}$ to have a Gaussian distribution too. Therefore, if we select our bound as,

$$\delta_0 = 3\sigma_{subspace} \tag{3.29}$$

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 operator therefore will be,

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

(3.30)

where $\left[\psi^T H^{(i)} \phi\right]_j$ is the column vector containing the j’th row of $\left[\psi^T H^{(i)} \phi\right]$ matrix and $r_j^{(i)}$ is the *j*’th component of the residual of the *i*'th observation.

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

$$C_V = \left\{ a \mid \left\| a^{(i)} - \left[\psi^{\mathrm{T}} H^{(i)} \phi a\right] \right\|^2 \le \delta_v \right\} \tag{3.31}$$

where δ_V is the priori bound reflecting statistical confidence and can be determined by using the formulation defined in eq. (3.19). The projection operator onto this constraint set will be

$$P_V^{(i)}a = \begin{cases} a + (\mathrm{M}^{\mathrm{T}}\mathrm{M} + \frac{1}{\lambda}I)^{-1}\mathrm{M}^{\mathrm{T}}(a^{(i)} - \mathrm{M}a) & , \text{ if } a \notin C_V \\ a & , \text{ if } a \in C_V \end{cases} \tag{3.32}$$

where M is equal to $\left[\psi^T H^{(i)}\phi\right]$, and λ is the Lagrange multiplier coming from the optimization formulation.

3.5.2 Independent Component Subspace

Same principles delineated in section 3.5.1 hold for independent component subspace too, but with different subspace notations. Independent component face subspace representation of high-resolution images and the low-resolution image can be written as:

$$x = A_x s_x + e_x \tag{3.33}$$

$$y^{(i)} = A_y s_y{}^{(i)} + e_y{}^{(i)} \text{ for i=1,....,K} \tag{3.34}$$

where 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. If we substitute face space representation equations (eq.(3.33) and eq.(3.34)) into our pixel-domain imaging model given in eq.(3.4), we will get a observation model in the face space as follows,

$$\begin{aligned} A_y s_y{}^{(i)} + e_y{}^{(i)} &= H^{(i)}(A_x s_x + e_x) + n^{(i)} \\ A_y s_y{}^{(i)} + e_y{}^{(i)} &= H^{(i)} A_x s_x + H^{(i)} e_x + n^{(i)} \end{aligned} \tag{3.35}$$

In order to make our new imaging model for face space similar to the pixel domain model, we multiply both sides of eq.(3.35) with W_y which enables us to represent the low-resolution image in the face space as a linear mapping of the high-resolution image in the face subspace plus a noise term (refer to eq.(3.39)).

$$W_y A_y s_y{}^{(i)} + W_y e_y{}^{(i)} = W_y H^{(i)} A_x s_x + W_y H^{(i)} e_x + W_y n^{(i)} \tag{3.36}$$

Since,

$$W_y A_y = I \tag{3.37}$$

and from eq.(2.16) we know that

$$W_y e_y = 0 \tag{3.38}$$

Then eq. (3.36) turns into

$$s_y{}^{(i)} = W_y H^{(i)} A_x s_x + W_y H^{(i)} e_x + W_y n^{(i)} \tag{3.39}$$

Using this observation model, we can form our constraint set similar to those given in section 3.5.1. The convex constraint sets for outliers of the residual for this case will be

$$C_0 = \left\{ s_x \,|\, \left| s_y^{(i)} - \left[W_y H^{(i)} A_x s_x \right] \right|_j \leq \delta_0 \right\} \tag{3.40}$$

where j subscript denotes the j'th component of the residual vector. We can calculate the statistical confidence bound, δ_0, as given in eq. (3.29) using the noise process

$$n_{subspace} = W_y H^{(i)} e_x + W_y n^{(i)} . \tag{3.41}$$

The projection operator for independent component subspace then becomes

$$P_{0j}^{(i)} s_x = \begin{cases} s_x + \dfrac{(r_j^{(i)} - \delta_0)}{\left\| \left[W_y H^{(i)} A_x \right]_j \right\|^2} \left[W_y H^{(i)} A_x \right]_j & , \quad \text{if } r_j^{(i)} > \delta_0 \\ s_x & , \quad \text{if } -\delta_0 < r_j^{(i)} < \delta_0 \\ s_x + \dfrac{(r_j^{(i)} + \delta_0)}{\left\| \left[W_y H^{(i)} A_x \right]_j \right\|^2} \left[W_y H^{(i)} A_x \right]_j & , \quad \text{if } r_j^{(i)} < -\delta_0 \end{cases}$$

(3.42)

where $\left[W_y H^{(i)} A_x \right]_j$ is the column vector containing the j'th row of $\left[W_y H^{(i)} A_x \right]$ matrix and $r_j^{(i)}$ is the j'th component of the residual of i'th observation.

Finally, the convex constraint sets for the variance of the residual defined in independent component subspace can be given as

$$C_V = \left\{ s_x \,|\, \left\| s_y^{(i)} - \left[W_y H^{(i)} A_x s_x \right] \right\|^2 \leq \delta_v \right\} \tag{3.43}$$

where a priori bound for statistical confidence, δ_V, can be calculated from eq. (3.19). The corresponding projection operator onto these convex constraint sets is found as

$$P_V^{(i)} s_x = \begin{cases} s_x + (\mathrm{M}^{\mathrm{T}}\mathrm{M} + \frac{1}{\lambda} I)^{-1} \mathrm{M}^{\mathrm{T}} (s_y^{(i)} - \mathrm{M} s_x) & , \quad \text{if } s_x \notin C_V \\ s_x & , \quad \text{if } s_x \in C_V \end{cases} \tag{3.44}$$

where M is equal to $\left[W_y H^{(i)} A_x \right]$. In the chapter 4, we have given experimental results using the derived methods for POCS based subspace superresolution. In the Appendix, we have shown that the necessary condition for convergences of the projections, P_V's, is to have a positive λ. Hence, rather than calculating λ for each constraint set, it is also possible to assign a fixed λ which can be determined by trial-and-error to achieve the fastest convergence.

3.6 Subspace-based Superresolution Using Bayesian Estimation

Different methods have been suggested in the literature to solve the superresolution problem. In [14] rather than the POCS based reconstruction algorithm which is delineated in section 3.5, a Bayesian estimation method is suggested for reconstruction of the high-resolution face images in principal component subspace. In this thesis, we extended this approach to independent component subspace, aiming to have a better representation that is more robust to noise and motion estimation errors. Details of principal component subspace reconstruction can be found in [14]; here we just present our proposed methodology in the independent component subspace.

3.6.1 Independent Component Subspace

It is logical to expect a similar performance on robustness and computational advantages for independent component model face subspace when compared with pixel domain super-resolution, besides, the independent component feature vectors are found to be more robust to noise than eigenface feature vectors (refer to [9]). Hence estimating true feature vectors of ICA directly from feature vectors of low-resolution images should enable better recognition performance together with advantages obtained in robustness and computation due to face subspace representation.

Bayesian estimation will be used for estimating the true feature vector. Here, we will try to maximize posterior probability of the true feature vector s_x by using prior probability of $\boldsymbol{s_x}$ and conditional probability of feature vectors of low-resolution images $p(s_y^{(1)},...,s_y^{(M)}|s_x)$. Hence, the MAP estimator will become

$$\tilde{s}_x = \arg\max_{s_x}(p(s_y{}^{(1)},...,s_y{}^{(M)} \mid s_x)p(s_x)) \qquad (3.45)$$

In order to construct our MAP estimator we need to model the prior probability $p(s_x)$ and the conditional probability $p(s_y^{(1)},...,s_y^{(M)}|s_x)$. 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:

$$p(s_x) = \lambda \exp(-\lambda(s_x - \mu_x)) \qquad (3.46)$$

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 :

$$\lambda = \sqrt{2}\Lambda^{-1/2} \qquad (3.47)$$

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 $p(s_y^{(1)},...,s_y^{(M)}|s_x)$, we will use eq.(3.39). The first step is to divide eq.(3.39) into two parts as signal and noise. Thus, the noise part will be;

$$n_{subspace} = W_y H^{(i)} e_x + W_y n^{(i)} \qquad (3.48)$$

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 [23]. 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:

$$p(n_{subspace}) = \mathcal{N}(\mu_n^{(i)}, K) \qquad (3.49)$$

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 multiplication of marginals,

$$p(s_y^{(1)},...,s_y^{(M)} \mid s_x) = p(s_y^{(1)} \mid s_x) \times ... \times p(s_y^{(M)} \mid s_x) \qquad (3.50)$$

Using eq.(3.39) we can write marginal conditional probabilities, note that since s_x is given, marginal probabilities will have the same statistical properties as noise, except for the mean.

Hence we have:

$$p(s_y^{(i)} \mid s_x) = \mathcal{N}(W_y H^{(i)} A_x s_x + \mu_n^{(i)}, K) \qquad (3.51)$$

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

$$p(s_y^{(1)},...,s_y^{(M)} \mid s_x) = \frac{1}{L}\exp(-\sum_{i=1}^{M} z^{(i)T} K^{-1} z^{(i)}) \qquad (3.52)$$

where L is a normalization constant. Substituting eq.(3.52) and eq.(3.46) into eq.(3.45) will give us the MAP estimator for ICA feature vector of the high-resolution image.

$$\tilde{s}_x = \arg\min(\sum_{i=1}^{M}\left[z^{(i)T} K^{-1} z^{(i)}\right] + \sqrt{2}\Lambda^{-1/2} \mid s_x - \mu_x \mid) \qquad (3.53)$$

The solution can be obtained by an iterative steepest descent algorithm. The *(k+1)*'th step of this iteration can be written as:

$$s_x^{k+1} = s_x^k - \alpha \nabla C(s_x^k) \tag{3.54}$$

where α is the step size parameter and $\nabla C(s_x^k)$ is the gradient of the cost function calculated at s_x^k. We have selected the original MAP estimator as our cost function.

$$\begin{aligned} C(s_x) = \frac{1}{4}\sum_{i=1}^{M}\left(s_y^{(i)} - W_y H^{(i)} A_x s_x - \mu_n\right)^T K^{-1}\left(s_y^{(i)} - W_y H^{(i)} A_x s_x - \mu_n\right) \\ + \frac{1}{4}\sqrt{2}\Lambda^{-1}(s_x - \mu_x) \end{aligned} \tag{3.55}$$

Taking the derivative of cost function with respect to s_x, will give us the gradient of the cost function as:

$$\begin{aligned} \nabla C(s_x) = -\frac{1}{2}\sum_{i=1}^{M} A_x^T H^{(i)^T} W_y^T K^{-1}\left(s_y^{(i)} - W_y H^{(i)} A_x s_x - \mu_n\right) \\ + \frac{1}{4}\sqrt{2}\Lambda^{-1} \end{aligned} \tag{3.56}$$

As the step size of the iteration, we selected α to be inversely proportional to the Hessian of the cost function *C(s$_x$)* rather than choosing a fixed constant. The formulation used for updating the step size α is given as:

$$\alpha = \frac{(\nabla C(s_x^k))^T (\nabla C(s_x^k))}{(\nabla C(s_x^k))^T (Hessian)(\nabla C(s_x^k))}, \tag{3.57}$$

where *Hessian* is the Hessian matrix found by

$$Hessian = \frac{1}{2}\sum_{i=1}^{M} A_x^T H^{(i)T} W_y^T K^{-1} W_y H^{(i)} A_x \tag{3.58}$$

4. EXPERIMENTS

4.1 Face Video Databases

We have tested the proposed methods using two different video databases. The first database is M2VTS database, created by UCL Laboratoire de telecommunications et télédétection [31]. The other one is a new database that is especially designed to test the performance of superresolution algorithms for face recognition problem and created by VPA group at Sabancı University.

4.1.1 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 high-resolutional, 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, the original videos are convolved by 15x15 pixel Gaussian blur kernels with varying variances and downsampled by a factor of two.

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, is the reference one. In the same way, we have 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 Gaussian kernel and downsampling them by a factor of two as described before.

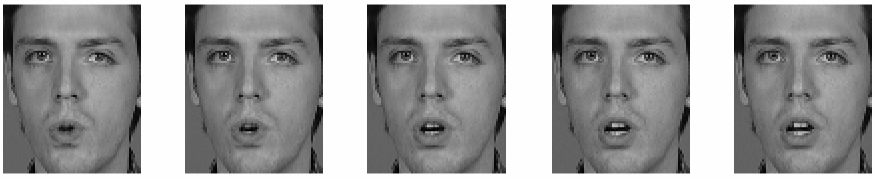


Figure 4.1. Five consecutive frames extracted from a face video of M2VTS database

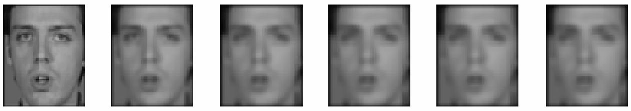


Figure 4.2. Low-resolution face video frames obtained by convolving with Gaussian kernels having variances: 1, 10, 20, 30, 40, and 50 pixels and by downsampling.

In the Fig. 4.2, five low-resolution frames of the same reference frame obtained by Gaussian kernels with differing variances is illustrated.

4.1.2 VPA Superresolution Face Database:

The second face database that we have used in our experiments is VPA Superresolution Face Database which consists of frontal face images and videos of 32 people to test our proposed super-resolution techniques. VPA SR Face Database contains low-resolution and blurry face video data together with the corresponding high resolution face image of each person. Face videos contain just translational movement of each face shot by SONY DVR camera from a distance in ambient light while 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-resolution (double) than those faces in the video frames (refer to Fig. 4.3). 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 for each one 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.

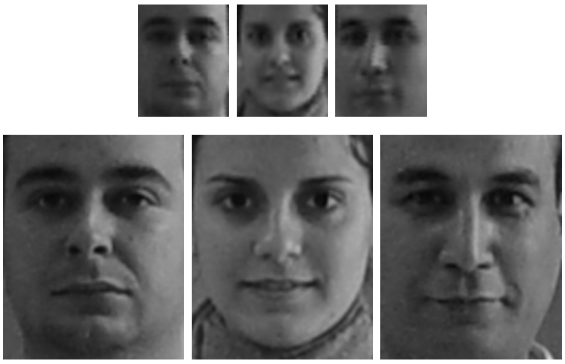


Figure 4.3. Low-resolution face video frames (at the top) , corresponding high-resolution face images (at the bottom)

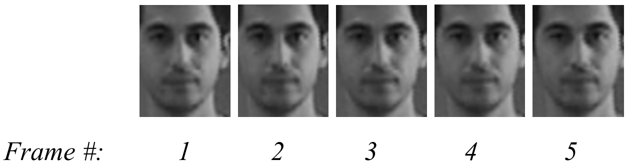


Figure 4.4. Five consecutive face video frames from VPA SR Face Database, third frame is the reference one. Note: Head is moving to right side of the page.

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 face video frames are scaled and cropped according to the aligned reference frame shown in Fig. 4.4.

4.2 Pixel-domain Superresolution

The pixel-domain superresolution 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.

4.2.1 Registration Algorithm

It is quite a difficult task to make an accurate displacement estimation for 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). Hence, one straightforward strategy is to use an appropriate mesh model and divide face frames into triangles where each one captures single rigid motion. Later, using motion vector of the nodes of the triangles, the motion vector of each pixel inside each triangle can be approximated using bilinear interpolation. Hexagonal matching algorithm [32] (HMA) seems to be a possible solution for this problem, however a uniform mesh model with very small triangle size is needed for this algorithm in order to decompose a non-rigid motion into rigid ones in each triangle. Therefore, we need to use a very dense mesh which is computationally inefficient. Another choice for displacement estimation will be polygonal matching algorithms (PMA).

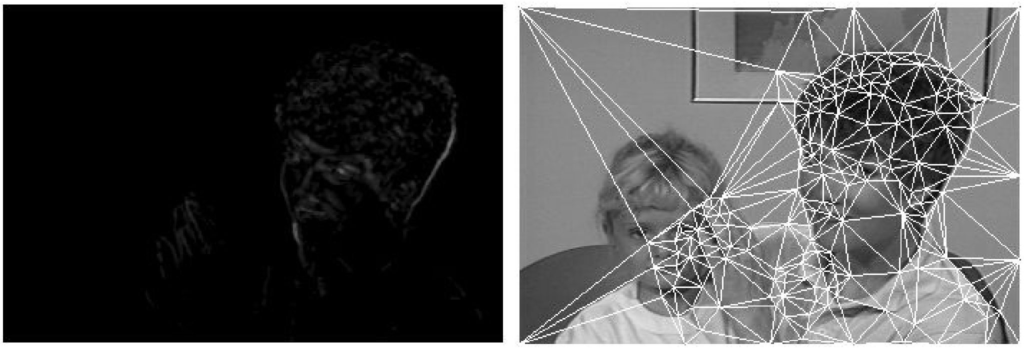


Figure 4.5. Frame distance image obtained by subtracting two consecutive video frames to detect motion fields (left), non-uniform content-based mesh is laid onto reference video frame (right)

PMA has the same principle as HMA but the mesh model is non-uniform and is designed according to the content based mesh model described in [33], so as to have dense triangular mesh around the motion fields and coarse triangles in motionless regions (illustrated in Fig. 4.5). Although this algorithm reduces the computational cost compared to the HMA, we observed that it is vulnerable to motion estimation errors. In other words, if one of the nodes of the triangular mesh is estimated with error, nonlinear deformations are created in the reconstructed face image which may affect the face recognition performance in return.

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.

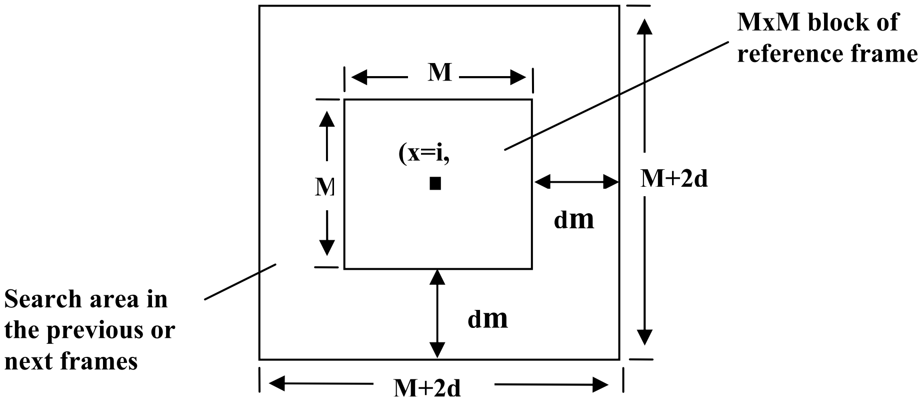

MxM block of reference frame
M
(x=i,
M
M+2d
dm
dm
Search area in the previous or next frames
M+2d

Figure 4.6. Geometry of MxM block and search region

The HBMA that we have applied for displacement estimation is devised to detect subpixel motion information and it is different than HBMA described in [34]. Moreover, we first applied full search BMA to face frames that are bilinearly interpolated by a factor of two with the geometry shown in Fig.4.6 where we selected M as seven and dm as one. Hence, 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. Afterwards, the magnitude of the motion vector which is found for the block in the interpolated frame is divided by two and assigned as the motion vector for a NxN block (N is selected as three in our experiments) in the original frame which shares the same center pixel with the search block in the interpolated frame (refer to Fig. 4.7).

Super-resolution algorithm with POCS in pixel domain needs very accurate displacement information to work successfully without degrading the quality of the image; otherwise it is common practice to disable POCS algorithm according to a validity map proposed in [35]. The validity map is a type of indicator which is formed by finding motion estimation errors and determining places where estimation error exceeds some pre-determined threshold value or values. Subsequently, inaccurate motion estimates during the reconstruction process are discarded. This is easily achieved within the POCS framework by defining the data consistency constraint sets and performing associated projections only for those pixel locations for which the motion vectors are accurate.

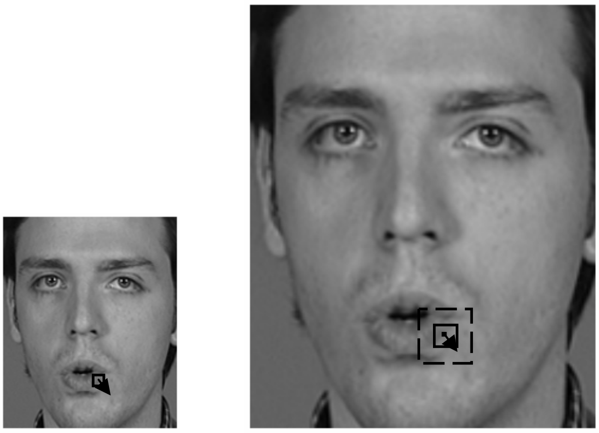


Figure 4.7. Illustration of the displacement estimation algorithm. Motion vector found in bilinearly interpolated face frame (right), and motion vector of the corresponding block in the original frame (left).

Validity maps and similar methods that disable POCS when motion is estimated inaccurately (which occurs mostly around the object boundaries and occluded parts in the image) have previously been applied to videos when there is single or multiple objects having rigid motions in [21, 35, 36]. However, for our case we have to deal with non-rigid motion in face frames where the application of segmentation or validity maps degrades the quality of the reconstructed face image. Observe Fig. 4.8, where POCS is applied to pixel locations determined by validity maps, an example of validity map is given in Fig. 4.8-(b). Since motion is non-rigid, superresolution with POCS method causes artificial speckles which clearly reduce the quality of the face image (refer to Fig. 4.8-(c)). Moreover, when we use a validity map since we disable POCS according to the validity map, we do not perform the projection operation to all the pixel locations in each frame, which in return increases the number of frames required to achieve higher resolution face images.

Due to disadvantages like increase in the number of required frames and artificial speckles, we did not use validity maps to disable POCS in some pixel locations with the expense of having motion estimation errors in and around the occluded regions or the boundaries of the face. Nevertheless, we achieved perceptually convincing and satisfactory results, with very negligible motion estimation error around the occluded regions. As we have proposed before, since we have used holistic face recognition techniques such as principal and independent component analysis, we observed that, in fact, these errors do not affect the canonical representation of the face (hence face recognition performance) significantly.

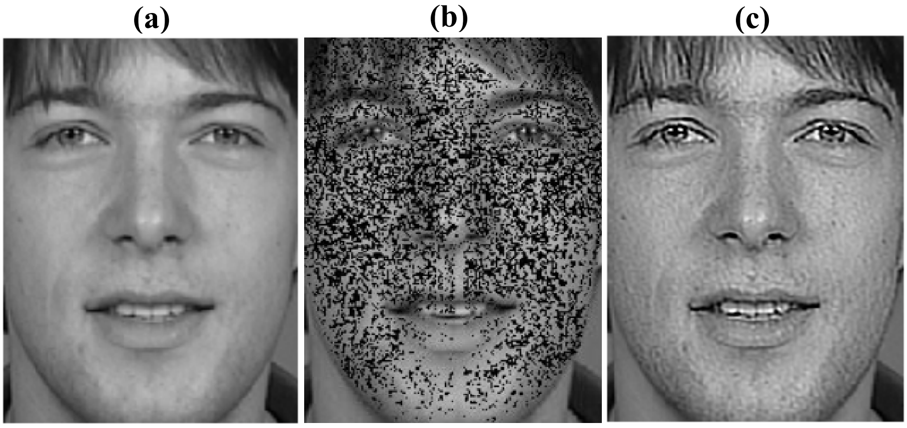
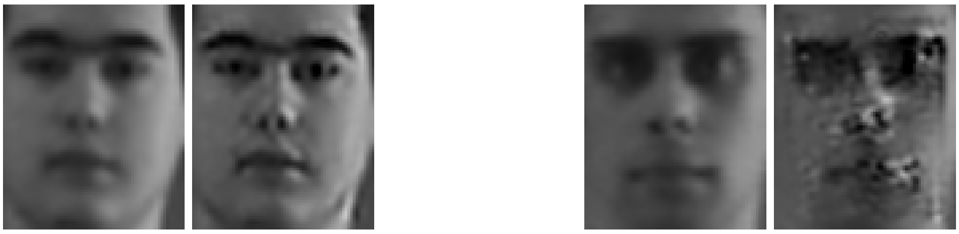


Figure 4.8. The application of validity maps. Bilinearly interpolated reference frame (a), validity map obtained from next video frame (motion est. done accurately in the black region) (b), SR in pixel-domain applied to the reference frame using validity maps (c).

4.2.2 Pixel-domain Superresolution Results

It is common practice to use imaging charts to analyze the performance of the superresolution algorithms. Since we have used face databases in our experiments, we don't have such opportunity to evaluate the resolution increment the algorithm provides. However, we expect better perceptual results from the superresolution algorithm which exploits information coming from many frames than the result of a sharpening filter. In the Fig. 4.10, we have given results when the sharpening filter is applied to the bilinearly interpolated face image in (c) and when superresolution algorithm based on pixel-domain POCS is applied in (d). Observe that edges are emphasized in (c) compared to (b), nevertheless, blockiness appeared around the edges of the face (e.g. around mouth, eyes and chin) due to sharpening. Superresolution, on the other hand, gives better perceptual performance around the edges and obviously increases resolution by incorporating subpixel information coming from different frames into the reference frame.

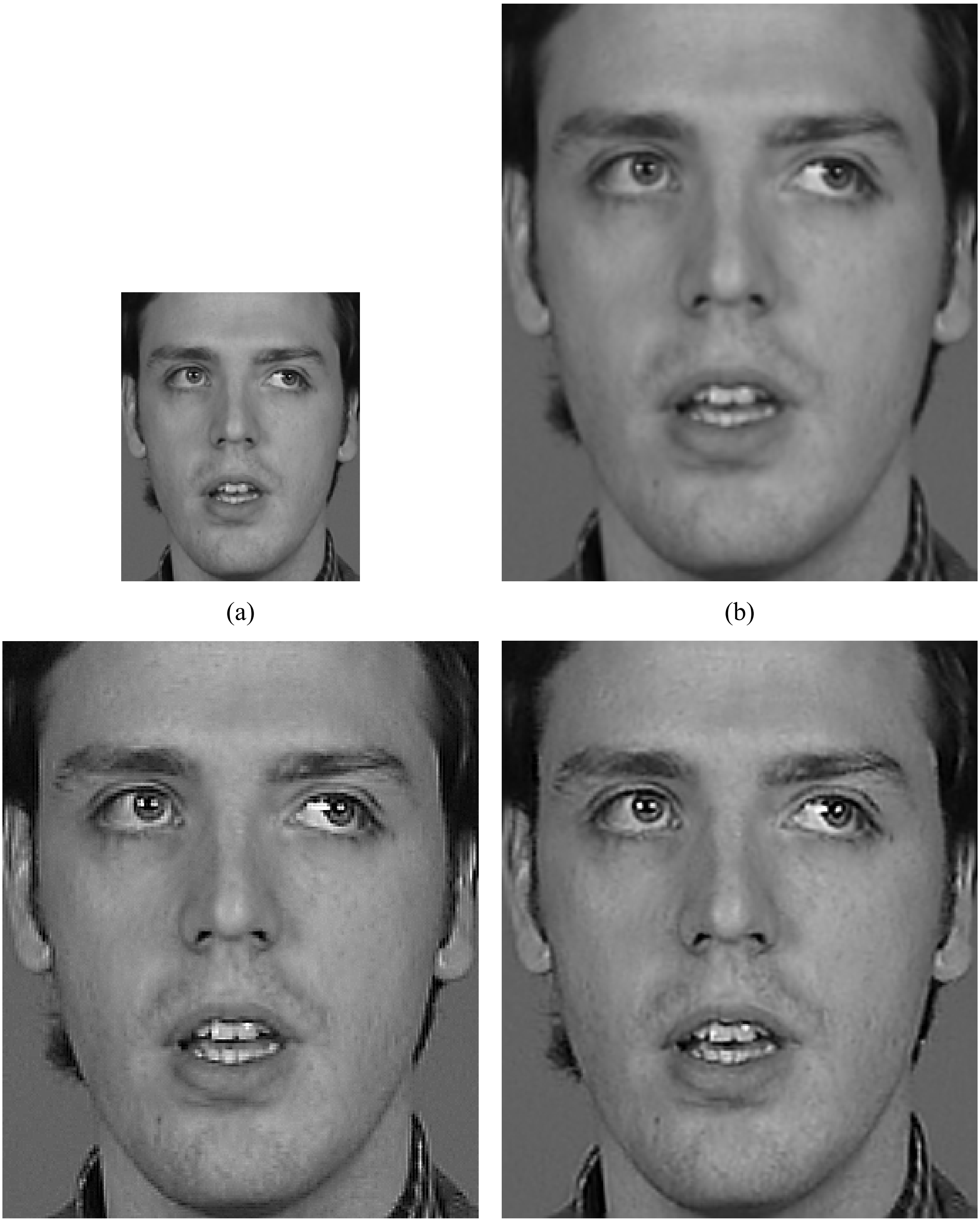
The result obtained using M2VTS data base given in Fig 4.10-(d) appears to be very encouraging, however if we list the reasons for such a good performance, we should note that the data has very low noise level and quite high time resolution which enables accurate displacement estimation. In a more realistic case, where noise level is high and time resolution hinders accurate displacement estimation, we have observed that pixel-domain superresolution does not improve the quality of the video and even degrades the initial estimate since distorting effects of inaccurate motion estimation coming from each frame cumulates onto the reference face image. We have employed the same pixel-domain superresolution algorithm to VPA SR Face Database (refer to Fig. 4.9 for sample results).



Interpolated *SR*
(1)

Interpolated *SR*
(2)

Figure 4.9. Samples of the results of pixel domain superresolution for VPA SR Face Database



(a) (b)

(c) (d)

Figure 4.10. Original reference face image (a) , reference image bilinearly interpolated by a factor of two (b), sharpening filter is applied to the bilinearly interpolated face image (c), pixel-domain superresolution with HBMA described above applied to reference image (d).

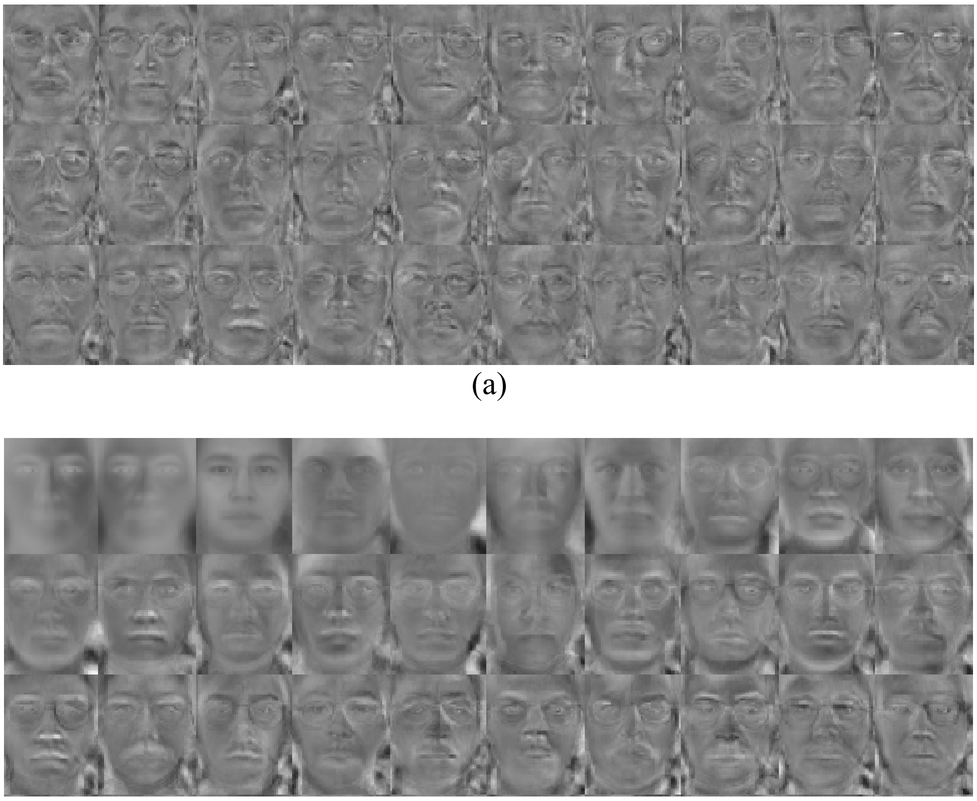
Some resolution improvement can be seen in Fig. 4.9-(1), but due to accumulation of the effects of inaccurate motion estimation in Fig. 4.9-(2) we observe highly degraded and distorted face image which can disable proper face recognition.

4.3 Subspace-domain Superresolution

The cure for the problem in Fig 4.9-(2) 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. The proposed subspace methods in this thesis suggest new approaches to this problem which can be thought of as an extension of Gunturk et. al.'s work [14].

4.3.1 Obtaining Face Subspaces

In our experiments, we have used CMU Illumination and Expression [18] databases for constructing face subspaces both for principal and independent component domain. 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. 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. 4.11 for basis vectors that span the subspaces).



(a)

(b)

Figure 4.11. Independent component basis images estimated by using CMU Illumination and Expression databases (a), and eigenfaces extracted from CMU Illumination and Expression databases (b).

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. For encoding objects that are characterized by high-order combinations of features, Barlow and Attick have discussed advantages of factorial codes [37, 38].

To achieve such a factorial coding of face images, we organize 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).

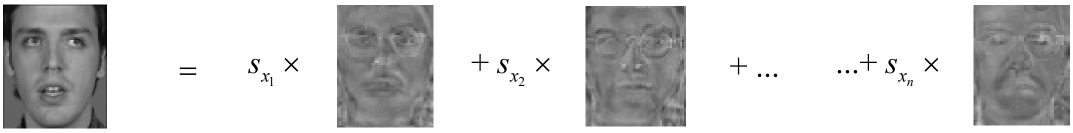


Figure 4.12. Factorial code representation attained by second architecture of ICA, s_x's are statistically independent coefficients and basis images are columns of mixing matrix A_x.

As given in eq. (2.14) the de-mixing matrix can be written as

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

where E_{100}^{T} reduces the dimensionality of the input data to 100 by projecting it onto the PC 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.

4.3.2 Estimating Statistics of Noise and Feature Vectors of PC Subspaces:

In order to estimate a priori information that we need to have for application of subspace superresolution methods described in sections 3.5 and 3.6, we have divided both databases into two equal groups as training and testing. We have used the training group to estimate the statistics of noise and feature vectors of principal component analysis. From high resolution face images $I_1,...,I_K$ (K=16 for VPA K=18 for M2VTS databases) in the training set, we estimated the statistics of principal component feature vectors a and together with low resolution face video frames of the corresponding high resolution images we estimated statistics of noise given in eq.(3.28). The unbiased estimate for mean and covariance of a can be obtained by using sample mean and variances:

$$\mu_a \simeq \frac{1}{K}\sum_{j=1}^{K}(\phi^T I_j) \tag{4.3}$$

and

$$\Lambda_a \simeq \frac{1}{K}\sum_{j=1}^{K}(\phi^T I_j - \mu_a)(\phi^T I_j - \mu_a)^T . \tag{4.4}$$

The off-diagonal elements of the matrix Λ_a are set to zero, in order to make more reliable estimate of the covariance by using limited number of training images.

Similarly, we estimated the mean and covariance matrices of noise in pixel domain for each face as follows;

$$\mu_v^a \approx \frac{1}{KM}\sum_{j=1}^{K}\sum_{i=1}^{M}(y_j^{(i)} - H^{(i)}\phi\phi^T I_j), \tag{4.5}$$

and

$$Z_v^a \simeq \frac{1}{KM}\sum_{j=1}^{K}\sum_{i=1}^{M}(y_j^{(i)} - H^{(i)}\phi\phi^T I_j - \mu_v)(y_j^{(i)} - H^{(i)}\phi\phi^T I_j - \mu_v)^T . \tag{4.6}$$

where $y_j^{(i)}$ is the i'th observation of the j'th face obtained from low resolution face video. M is the number of low resolution observations used in estimating the statistics of the noise. As mentioned before for our case M is five. Again, diagonal elements of Z are set to zero. The mean, η, and the covariance matrix, Q, for the noise in principal component subspace (given in eq. (3.28)) are found using the following formulations,

$$\eta = \psi^T \mu_v^a \tag{4.7}$$

and,

$$Q = \psi^T Z_v^a \psi . \tag{4.8}$$

4.3.3 Estimating Statistics of Noise and Feature Vectors of IC Subspace:

Once more, in order to capture a priori information for IC subspace reconstructions, we have used the same training scheme described for principal component analysis. However, in this case independent component feature vectors are denoted by s_x and W_x (de-mixing) and A_x (mixing) subspace matrices are used. The mean and covariance matrices of independent component feature vectors are

$$\mu_x \simeq \frac{1}{K}\sum_{j=1}^{K}(W_x I_j) \tag{4.9}$$

and

$$\Lambda_x \simeq \frac{1}{K}\sum_{j=1}^{K}(W_x I_j - \mu_x)(W_x I_j - \mu_x)^T. \tag{4.10}$$

Similarly, the off-diagonal elements of the matrix Λ_x are set to zero.

Again, we estimated the mean and covariance matrices of noise in pixel domain for each face as follows;

$$\mu_v^x \approx \frac{1}{KM}\sum_{j=1}^{K}\sum_{i=1}^{M}(y_j^{(i)} - H^{(i)}A_x W_x I_j), \tag{4.11}$$

and

$$Z_v^x \simeq \frac{1}{KM}\sum_{j=1}^{K}\sum_{i=1}^{M}(y_j^{(i)} - H^{(i)}A_x W_x I_j - \mu_v^x)(y_j^{(i)} - H^{(i)}A_x W_x I_j - \mu_v^x)^T. \tag{4.12}$$

The mean, η, and the covariance matrix, Q, for noise given in eq. (3.41) are found using the following formulations,

$$\eta = W_y \mu_v^x \tag{4.13}$$

and,

$$Q = W_y Z_v^x W_y^T \tag{4.14}$$

where W_y is the de-mixing matrix of the subspace of the low resolution face video frames.

4.3.4 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 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. 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, as we have mentioned before. 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 section 3.5 and 3.6 with ten iterations for each sequence and a priori information contribution coefficient for Bayesian estimation, λ , is set to 0.5.

4.4 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 face and the camera and 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).

$$\text{L1: } d = \sum_{j=1}^{100} \left| a_{training,j} - a_{test,j} \right| \tag{4.15}$$

$$\text{L2: } d = \left(\sum_{j=1}^{100} \left(a_{training,j} - a_{test,j} \right)^{2} \right)^{1/2} \tag{4.16}$$

$$\text{CC:}\ d = \frac{a_{training} \bullet a_{test}}{\left\| a_{training} \right\| * \left\| a_{test} \right\|} \tag{4.17}$$

As we have mentioned before, we used two face video databases to test the proposed methods and compare them with those already available in the literature. 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 in Figures 4.13 to Fig 4.15, in case of without superresolution (dashed lines, bilinearly interpolated frames are used), Fig. 4.13 supports the claims given in [9] and [16] that ICA offers a better representation than PCA for face recognition in L1 norm but they provide same recognition performance in L2 norm and CC distance metrics . If we compare the proposed independent component based superresolution algorithm (ICA-SR) with the principal component based one (PCA-SR), in L1 and L2 norm as noise level increases, the proposed method outperforms PCA-SR with 100% recognition rate in all noise levels. Note that we have mentioned that blur in pixel-domain can be thought as noise in canonical representation, in other words in subspace representation.

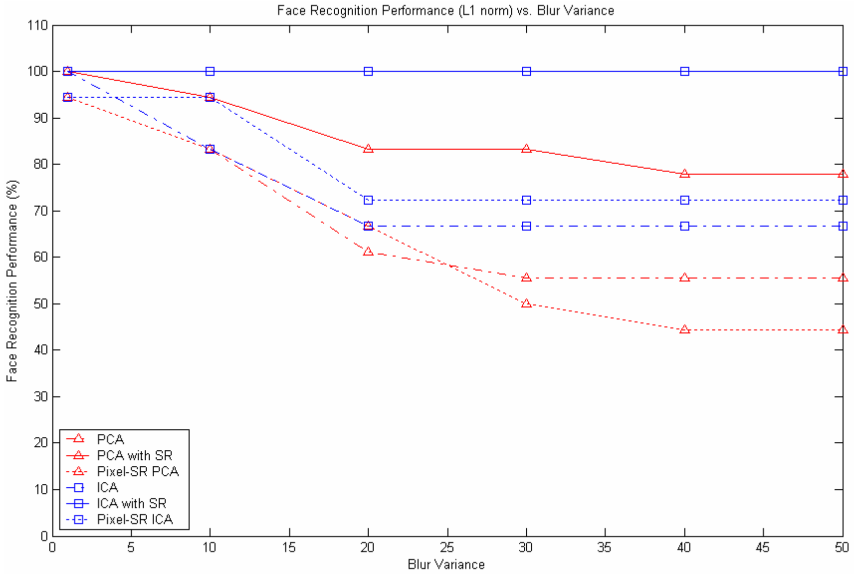


Figure 4.13. Comparison of face recognition performances with L1 distance metric

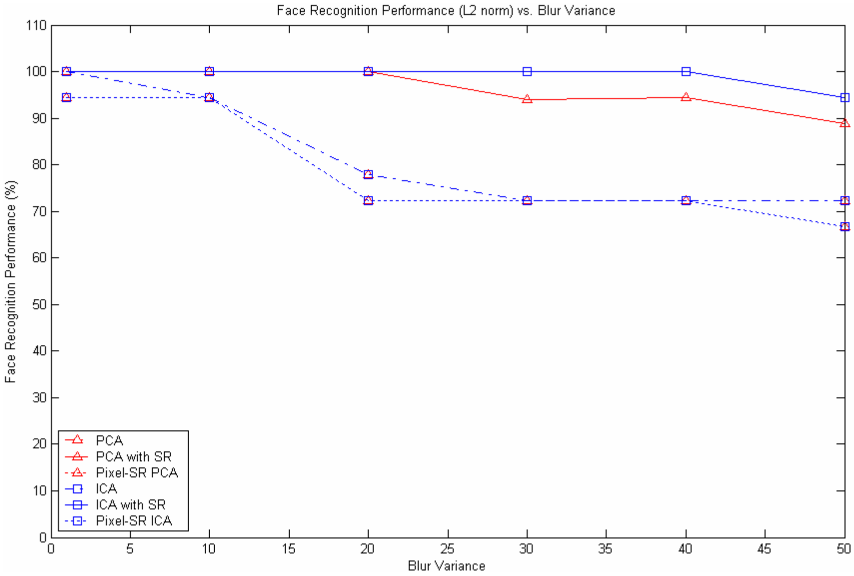


Figure 4.14. Comparison of face recognition performances with L2 distance metric

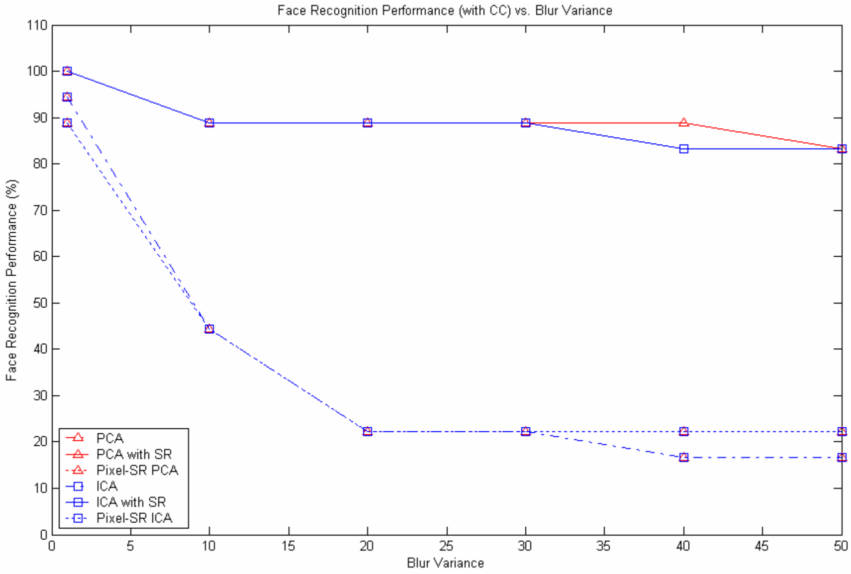


Figure 4.15. Comparison of face recognition performances with CC distance metric

The face recognition scheme for M2VTS database is formed by selecting a high-resolution frame as training face. 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 Figures 4.13 to 4.15, face recognition is performed by comparing distances between features of testing high resolution face and features of training low-resolution center frame whose size is incremented by the application of subspace superresolution methods (PCA-SR and ICA-SR legends in the figures), pixel domain superresolution methods (Pixel-SR) and just by bilinear interpolation (ICA and PCA legends in the figures). Observe that results of the pixel domain superresolution are close to the results of the interpolated ones.

The use of 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. The pixel domain imaging model given in eq. (3.4), we observe that the low-resolution image, *y*, is formed by blurring and downsampling of 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 section 4.1.2 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 in Table 1 that pixel domain superresolution 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 superresolution, 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*

5. CONCLUSIONS

In this thesis, we aim 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 superresolution techniques on face recognition problem is given for the first time in the literature.

One of the outcomes of this thesis is to implement and demonstrate that superresolution 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 superresolution techniques in a real setup. For this a new database, called VPA SR Face Database, is created.

This thesis presents how pixel-domain superresolution can be achieved from face videos where there are non-rigid motions together with rigid ones. In preceding works, pixel-domain superresolution is just applied in case of single or multiple objects having rigid motions. By extending studies in pixel-domain superresolution into non-rigid motion circumstances, this thesis is the first effort. Besides, POCS based signal reconstruction algorithm in subspace domain is formulated for superresolution and it is found to improve face recognition performance by 6.25 % in VPA SR Face Database compared to pixel-domain superresolution.

Application of eigenface-based superresolution method proposed by Gunturk et al [14] 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 in this thesis independent component based superresolution technique using Bayesian estimation is revealed to be superior to eigenface-based one as noise level increases. Moreover, an increase in recognition rate by 18.75% is achieved by using Bayesian estimation based subspace methods compared to non-SR applications.

In conclusion, we can say that subspace based superresolution methods together with computational advantages, provides robustness against noise compared to pixel-domain superresolution algorithms. Hence, enables better face recognition performances.

REFERENCES

[1] W. Zhao, R. Chellappa, P.J. Philips, and A. Rosenfeld, “Face Recognition: A Literature Survey,” *Technical Report CAR-TR-948*, University of Maryland, 2000.

[2] T. Kanade, “Computer Recongition of Human Faces,” Basel and Stuttgart: Birkhauser, 1973.

[3] T. Kanade, “Picture processing by computer complex and recognition of human faces,” *Technical Report*, Department of Information Science, Kyoto University, Japan, 1973.

[4] M.D. Kelly, “Visual Identification of People by Computer,” *Technical Report AI-130*, Stanford AI Project, Stanford, CA, 1970.

[5] R. Brunelli and T. Poggio, “Face Recognition: Features versus templates,” IEEE Trans. on Pattern Analysis and Machine Intelligence, vol. 15, no. 10, October 1993, pp. 1042-1052.

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

[7] 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.

[8] Belhumeur, P. N., J. P. Hespanha and D. J. Kriegman, “Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection”, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 19, No. 7, 1997, pp. 711-720.

[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] 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.

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

[12] 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.

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

[14] 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

[15] 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.

[16] 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.

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

[18] 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.

[19] W.K. Pratt, “Digital Image Processing,” *John Wiley and Sons*, NY, 1978.

[20] H.C. Andrews and B.R. Hunt, “Digital Image Restoration,” *Prentice-Hall*, NJ, 1977.

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

[22] 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.

[23] 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.

[24] M.Z. Nashed, *IEEE Trans. Antennas Propag.* Vol 29, 1981, p:220.

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

[26] C.R. Vogel, “Computational Methods for Inverse Problems,” *SIAM Frontiers in Applied Mathematics*, P.A., 2002.

[27] H.J. Trussell, *IEEE Trans. Acoust. Speech Signal Process.* ASSP-28, 1980, p:114.

[28] 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

[29] D.C. Youla and H. Webb, “Image restoration by the method of convex peojections: Part 1-Theory,” *IEEE Trans. Med. Imaging,* vol. MI-1, Oct. 1981, pp:81-94.

[30] 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.

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

[32] Y. Nakaya and H. Harashima, “Motion Compensation Based on Spatial transformations,” *IEEE Trans. On Circuits and Systems for Video Tech.*, vol.4, no.3, June 1994, pp:339-367.

[33] Y. Altunbasak and A. M. Tekalp, “Occlusion-Adaptive, Content-Based Mesh Design and Forward Tracking,” *IEEE Trans. on Image Processing*, vol. 6, no. 9, September 1997, pp:1270-1280.

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

[35] 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.

[36] 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.

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

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

APPENDIX

We have mentioned for the convergences of the projections, P_V's, the necessary condition is to have a positive λ value. To demonstrate, we will use proof of the essential features for the convergence of POCS defined in the text. Hence, rather than calculating λ for each constraint set, it is also possible to assign a fixed λ value which can be determined by trial-and-error to achieve the convergence.

Proof for non-expansiveness

We need to show that the projection operator is non-expansive, in other word;

$$\|Ox_1 - Ox_2\| \le \|x_1 - x_2\| \tag{A.1}$$

where x_1 and x_2 are two vectors in the Hilbert space and O is the projection operator for 'Variance of Residual', i.e., P_V.

In order to prove that our projection operator is non-expansive randomly select two vectors x_1 and x_2 from the Hilbert space defined by subspace methods. Then,

$$Ox_1 = x_1 + (M^T M + \frac{1}{\lambda} I)^{-1} M^T (g - Mx_1) \tag{A.2}$$

$$Ox_2 = x_2 + (M^T M + \frac{1}{\lambda} I)^{-1} M^T (g - Mx_2) \tag{A.3}$$

where residual is defined as $r = g - Mx_2$ (for our case $g=s_y^{(i)}$, $x_2=s_x$, $M=W_yH^{(i)}A_x$ and λ is the Lagrange multiplier)

Subtraction of eq. (A.2) from eq. (A.3) will give us the following;

$$\begin{aligned} Ox_1 - Ox_2 &= (x_1 - x_2) - \left[(M^T M + \frac{1}{\lambda} I)^{-1} M^T M x_1 - (M^T M + \frac{1}{\lambda} I)^{-1} M^T M x_2 \right] \\ &= (x_1 - x_2) - \left[(M^T M + \frac{1}{\lambda} I)^{-1} M^T M (x_1 - x_2) \right] \\ &= \left[I - (M^T M + \frac{1}{\lambda} I)^{-1} M^T M \right] (x_1 - x_2) \end{aligned} \tag{A.4}$$

The matrix M has real values and from the linear algebra we know M^TM is a symmetric matrix and can be written as;

Let $\Gamma=M^TM$

$$\Gamma = Q\Lambda Q^T \tag{A.5}$$

where Q represents the orthogonal eigenvectors and Λ contains eigenvalues for the corresponding eigenvectors.

$$\Lambda = \begin{bmatrix} \lambda_1 & & . & 0 \\ 0 & \lambda_2 & . & 0 \\ . & . & . & 0 \\ 0 & 0 & 0 & \lambda_N \end{bmatrix} \tag{A.6}$$

Hence eq. (A.4) evolves to

$$\begin{aligned} Ox_1 - Ox_2 &= \left[I - \left(Q(\Lambda + \frac{1}{\lambda} I)Q^T \right)^{-1} Q\Lambda Q^T \right](x_1 - x_2) \\ &= \left[I - QZQ^T \right](x_1 - x_2) = \left[QYQ^T \right](x_1 - x_2) \end{aligned} \tag{A.7}$$

where Y is

$$Y = \begin{bmatrix} 1 - \dfrac{\lambda_1}{\lambda_1 + \dfrac{1}{\lambda}} & & . & 0 \\ 0 & 1 - \dfrac{\lambda_2}{\lambda_2 + \dfrac{1}{\lambda}} & . & 0 \\ . & . & . & 0 \\ 0 & 0 & 0 & 1 - \dfrac{\lambda_N}{\lambda_N + \dfrac{1}{\lambda}} \end{bmatrix} \tag{A.8}$$

Hence finally we have following equality

$$Ox_1 - Ox_2 = \left[QYQ^T \right](x_1 - x_2) \tag{A.9}$$

Using Schwartz inequality norm of difference of the projected vectors can be written as;

$$\| Ox_1 - Ox_2 \| \leq \| QYQ^T \| . \| x_1 - x_2 \| \tag{A.10}$$

Schwartz inequality can be extended as follows;

$$\| QYQ^T \| \leq \| Q \| . \| YQ^T \| \leq \| Q \| . \| Y \| . \| Q^T \| \tag{A.11}$$

Since norm of orthonormal vectors is one, eq.(A.10) turns into;

$$\| Ox_1 - Ox_2 \| \leq \| Y \| . \| x_1 - x_2 \| \tag{A.12}$$

Norm of Y is equal to the largest singular value of Y. Since Y is diagonal its largest singular value will be one of its diagonal elements. Let l be the largest singular value of Y then in order to have non-expansive projection operator l should be greater than zero which implies;

$$l = 1 + \frac{\lambda_m}{\lambda_m + \frac{1}{\lambda}} \geq 0 \Rightarrow \frac{1}{\lambda} \geq 0 \Rightarrow \lambda > 0 \qquad \text{(A.13)}$$

Thus, if positive Lagrange multipliers are obtained from optimization which we built to minimize the difference between original and the projected signal with the variance constraint, non-expansiveness of the projection operator will be satisfied. Hence if λ is positive, non-expansiveness is satisfied.

Proof of asymptotically regular projection

In order to prove the convergence of the algorithm, we need to show that our projection operator is asymptotically regular. In other words following equation should hold;

$$\lim_{n \to \infty} \left\| O^n x - O^{n+1} x \right\| = 0 \qquad \text{(A.14)}$$

for $x \in \mathcal{H}$ and O: $\mathcal{H} \to \mathcal{H}$

Now let x1 be a vector in the Hilbert space defined by subspace methods. The we can write following difference equation for the projection operator *O;*

$$\begin{aligned} Ox_1 - x_1 &= x_1 + (M^T M + \frac{1}{\lambda} I)^{-1} M^T g - (M^T M + \frac{1}{\lambda} I)^{-1} M^T M x_1 - x_1 \\ &= (M^T M + \frac{1}{\lambda} I)^{-1} M^T (g - Mx_1) \end{aligned} \qquad \text{(A.15)}$$

If x_1 is the projection of another vector x_0 in the Hilbert space then,

$$x_1 = x_0 + (M^T M + \frac{1}{\lambda} I)^{-1} M^T (g - Mx_0) \qquad \text{(A.16)}$$

Substitute eq.(A.16) into eq.(A.15)

$$\begin{aligned} O^2 x_0 - Ox_0 &= (M^T M + \frac{1}{\lambda} I)^{-1} M^T g - (M^T M + \frac{1}{\lambda} I)^{-1} M^T M x_0 - \\ &(M^T M + \frac{1}{\lambda} I)^{-1} M^T M (M^T M + \frac{1}{\lambda} I)^{-1} M^T g + \\ &(M^T M + \frac{1}{\lambda} I)^{-1} M^T M (M^T M + \frac{1}{\lambda} I)^{-1} M^T M x_0 \end{aligned} \qquad \text{(A.17)}$$

Let's define two symmetric matrices *K* and *L* as

$$\begin{aligned} K &\equiv (M^T M + \frac{1}{\lambda} I)^{-1} \\ L &\equiv M^T M \end{aligned} \qquad \text{(A.18)}$$

Then eq. (A.17) turns into

$$\begin{aligned} O^2x_0 - Ox_0 &= KM^Tg - KLx_0 - KLKM^Tg + KLKLx_0 \\ &= (KM^T - KLKM^T)g - (KL - KLKL)x_0 \\ &= \left[(I-KL)KM^T\right]g - \left[(I-KL)KL\right]x_0 \\ &= (I-KL)\left[KM^Tg - KLx_0\right] \end{aligned} \tag{A.19}$$

From these derivations following equality is found

$$O^{n+1}x_0 - O^nx_0 = (I-KL)^n\left[KM^Tg - KLx_0\right] \tag{A.20}$$

Since *KL* and *I-KL* are symmetric matrices, we can write the following

$$(I\text{ - }KL) = QYQ^T \tag{A.21}$$

where *Y* is derived and defined in the previous proof. For the non-expansive projection operator we need to have

$$\lambda > 0 \Leftrightarrow \frac{1}{\lambda} \geq 0 \tag{A.22}$$

which implies that every diagonal element of *Y* is between zero and one;

$$0 < 1 - \frac{\lambda_m}{\lambda_m + \frac{1}{\lambda}} \leq 1 \tag{A.23}$$

From Schwartz inequality we can write norms as;

$$\left\|O^{n+1}x_0 - O^nx_0\right\| \leq \left\|(I-KL)^n\right\| . \left\|(KM^Tg - KLx_0)\right\| \tag{A.24}$$

Norm of *(I-KL)^n* is the largest singular value of the matrix say *l*. Since every diagonal element, including *l*, is between zero and one we can write

$$\lim_{n\to\infty} l = 0 \tag{A.25}$$

Thus,

$$\lim_{n\to\infty}\left\|O^{n+1}x_0 - O^nx_0\right\| \leq 0 \tag{A.26}$$

Since norm of a matrix is always non-negative. We have;

$$\lim_{n\to\infty}\left\|O^{n+1}x_0 - O^nx_0\right\| = 0 \tag{A.27}$$

Therefore, we have proved that the projection operator is asymptotically regular if λ is positive. Moreover, positive λ ensures the projection operator to be asymptotically regular and non-expansive, which in return guarantees the convergence of the projection operator.