ASSESSMENT OF SUPER-RESOLUTION FOR FACE RECOGNITION FROM VERY-LOW RESOUTION IMAGES

JAMES ROGER ROEDER

Department of Electrical & Computer Engineering, ECE

APPROVED:

Sergio D. Cabrera, Ph.D., Chair

Jose Gerardo Rosiles, Ph.D.

Christian Meissner, Ph.D.

Patricia D. Witherspoon, Ph.D.
Dean of the Graduate School

Dedication

To my family,

for always believing in me, encouraging me and supporting me throughout this endeavor.

ASSESSMENT OF SUPER-RESOLUTION FOR FACE RECOGNITION FROM VERY-LOW RESOUTION IMAGES

by

JAMES ROGER ROEDER, BSEE

THESIS

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

MASTER OF SCIENCE

Department of Electrical & Computer Engineering, ECE

THE UNIVERSITY OF TEXAS AT EL PASO

May 2009

Acknowledgements

I would like to express my gratitude and appreciation to the following people whose help during this research I found invaluable. I would like to thank two professors in the Wireless Sensor Networks Group. I need to thank Dr. Sergio Cabrera for introducing me to digital signal processing, image processing, and computer vision and inspiring me to pursue work in this area and for his guidance during this thesis. I also want to thank Dr. Cabrera for providing funding for my degree through the Texas Instruments Foundation Endowed Scholarship to help with the tuition. I would like to thank Dr. Gerardo Rosiles for accepting me as into the Wireless Sensor Networks Group.

Many of my colleagues were also instrumental to the completion of this thesis. Joel Quintana reminded me to work hard, but always take a break to not become overworked, and Ricardo Zapeda kept me relaxed even during the most stressful times.

This material is based upon work supported by the National Science Foundation under Grant No. CNS-0709438. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation. Network switches were provided by Cisco Systems, Inc.

Abstract

Super-Resolution (SR) involves the registration of multiple images/frames and reconstruction of a single higher resolution image. The goal of this research is to use multiple, very-low resolution images, such as those produced from a video sequence in a wireless sensor network system, as input to the super-resolution process in a face recognition system. The algorithm used for face recognition is the Fisherfaces method with a nearest neighbor classifier used for the recognition decision. Super-resolution consists of two stages, a registration stage and a reconstruction stage.

The testing images were segmented using a simple skin color detection approach. After the cropping, the images were combined into groups of four from a sliding window that would take the current image and the following three images repeating this process by moving to the next image in the sequence and the subsequent three images until the end of the current class, or person, is reached. This same sliding window was used for the super-resolution algorithm using faces from the three people or classes. Each group of four images was used as an input to the Keren registration algorithm where the rotational and translation information was saved that was then entered into the robust super-resolution reconstruction algorithm to create a single high quality image, which was processed by the face recognition algorithm. The methods tested to compare were the average of the same groups of four, the centroid shifted average and the minimum of the four faces in the group. The comparison was based on nearest neighbor classifier and on classification rates. The results were not in favor of the super-resolution method but instead, the centroid shifted average was the best in this study.

Table of Contents

List of Tables

List of Figures

Chapter 1: Introduction

This section covers topics that introduce the general information regarding cameras, face recognition, and super-resolution. A literature review of face recognition, super-resolution, and the combination of face recognition and super-resolution is given to describe what has been done and what is currently being done, and the scope of the project is covered in the section as well.

1.1 CAMERA

A camera captures the certain range of electromagnetic waves that are reflected from an object and allows those wavelengths to be seen by a human eye. Electromagnetic waves are classified into different ranges depending on their wavelength such as visible light and infrared to name a few. The visible wavelength, which is what most cameras operate at, is broken down into the colors violet, blue, green, yellow, orange, and red and the wavelengths are between 400 nm to 700 nm.

Most camera systems in use are CCD, charge-coupled device, or CMOS, complementary metal oxide semiconductor, devices. The CCD sensor has every pixel's charge transferred through a very limited number of output nodes that are then converted to voltage, buffered, and sent off-chip as an analog signal as shown in Figure 1.1. With the CMOS sensor each pixel has its own charge-to-voltage conversion, and the sensor often also includes amplifiers and other processing circuits resulting in the chip outputs being digital bits as shown in Figure 1.2 [1].

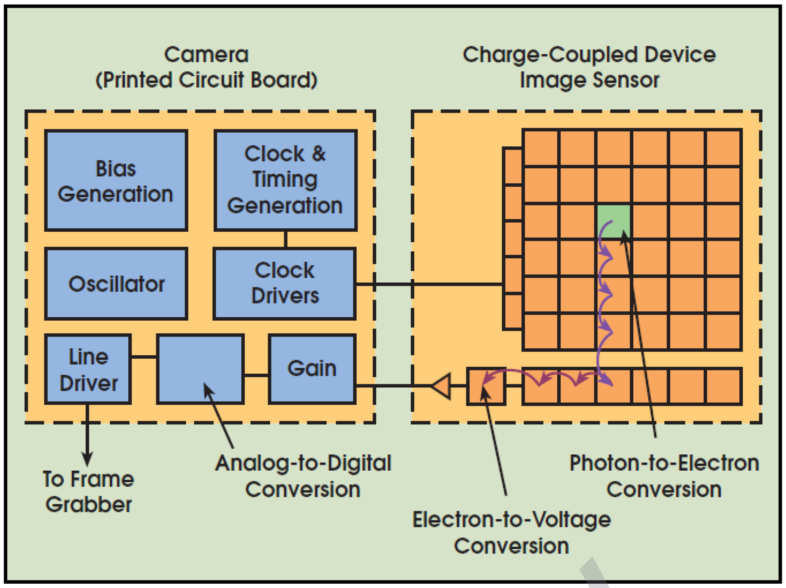

Figure 1.1: CCD Camera [1]

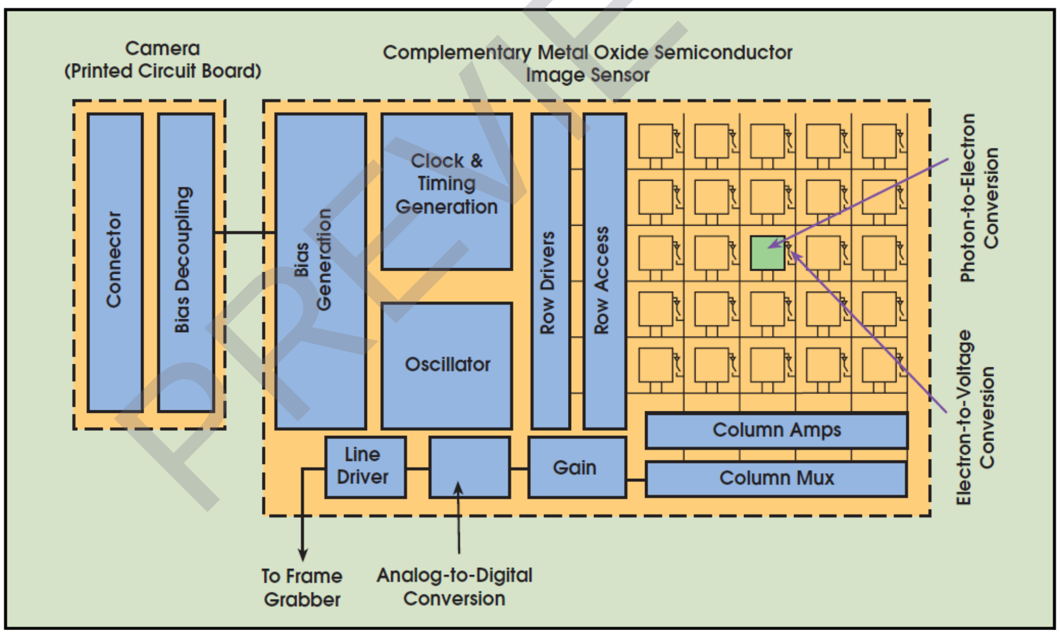

Figure 1.2: CMOS Camera [1]

1.2 WHAT IS FACE RECOGNITION?

Automated face recognition is clearly an important application of computer imaging though it remains a difficult problem and is not as reliable as would be desirable. Most advances have been made that include only small variations in lighting and in pose [2]. Some faces with medium to large

variations due to lighting are difficult for humans to identify and even harder for a computer to identify. The two most common methods for face recognition are Eigenfaces and Fisher Linear Discriminant (FLD) or Fisherfaces. Both methods take the images from a higher dimensionality to a lower dimensionality for classification. This lower dimensionality is useful in that fewer features are needed and therefore it is an ideal simplification for being sent through a low-capacity data channel like those encountered in wireless sensor networks. Each method reduces the dimensionality by using a projection matrix that is created using a small amount of images referred to as the training set of images. The training sets allows for the algorithm to know the different faces for each class.

1.3 WHAT IS SUPER-RESOLUTION?

When images are taken from a camera or a video source, the highest resolution is not always possible. Two main constraints in a wireless sensor network system are the low available data transmission bandwidth, and the need to use low-cost hardware. Some trade-offs must be made and the quality of the images is what will be sacrificed. However, the ability to take multiple images or a video sequence in a short period of time may allow for low quality images to be sufficient. High-resolution images can be reconstructed out of a series of low-resolution images. High-resolution or Super-Resolution (SR) involves registration of multiple images/frames and reconstruction of a single higher resolution image. This SR resulting image, called the high-resolution (HR) image, contains information from all of the images used and can be created to have multiple times as many pixels as the original images. The low-quality images used must have slight differences or the reconstructed image will not contain any new information and will be an interpolated version of any one of the identical input images.

Super-Resolution consists of two stages, a registration stage and a reconstruction stage. The registration stage is where the images are compared to each other to find any differences between the images in terms of rotations and linear shifts as shown in Figure 1.3. These types of differences can be from camera and/or object motion as shown in Figure 1.4. If the motion between the low-resolution images is a shift by an integer number of units, the resulting SR of HR image will not contain any new information. A good SR image will contain new information if the shifts have sub-pixel accuracy and are sampled at the same rate, that is, they are ideally subsampled and aliased versions of a higher

resolution ideal image which is what the SR method would attempt to recover [3]. Registration can be done either in the spatial domain or in the frequency domain. The spatial domain approach estimates the relative motion between the low resolution images, using nonuniform interpolation to produce a high quality image, and removes any blur or additional noise that was introduced. The frequency domain approach is based on the shifting property of the Fourier Transform, the aliasing relationship between the continuous Fourier transform of the Super-Resolution image and the discrete Fourier transform of the low resolution images, and the assumption that the signal is bandlimited.

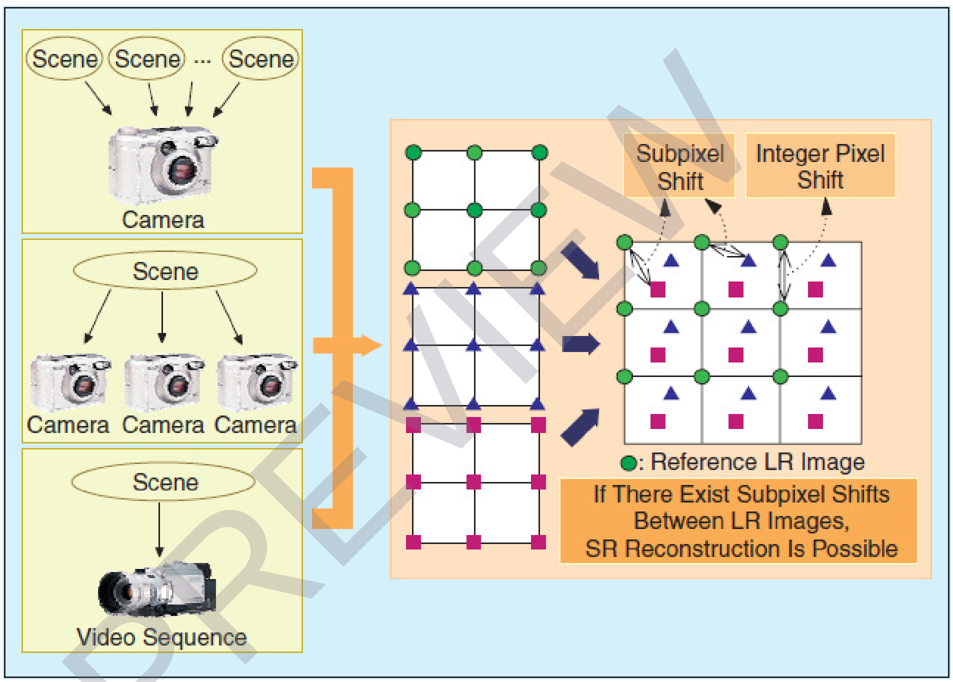

Figure 1.3: Sub-pixel Shifts Between Input Images [3]

The reconstruction stage is where the estimated registration parameters are taken and used to create a new, higher-resolution image. The higher resolution images are then mapped to an image that has 4 times as many pixels using interpolation as shown in Figure 1.5.

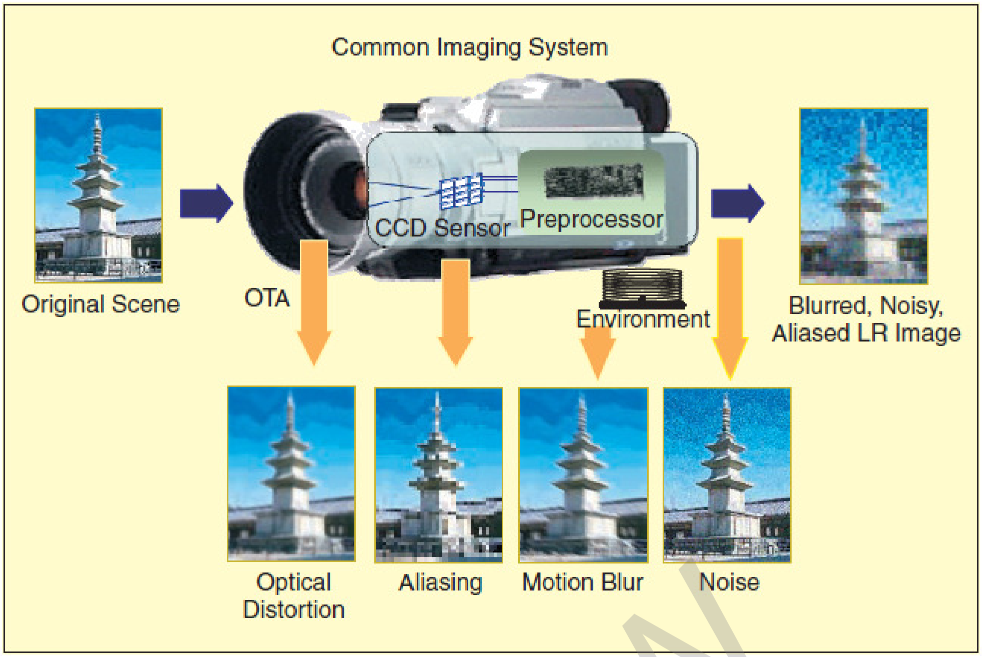

Figure 1.4: Common Sources of Sub-pixel Shifts [3]

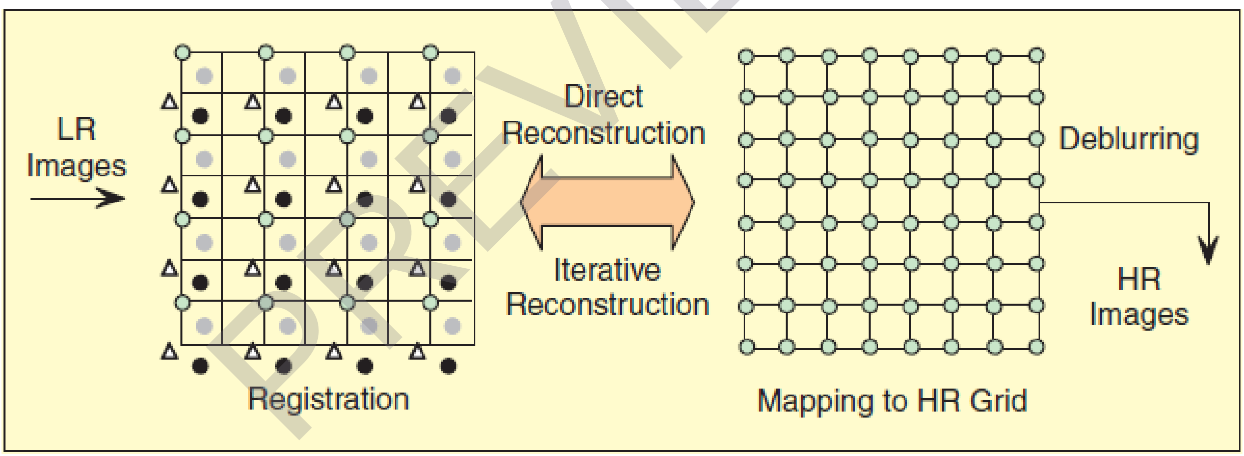

Figure 1.5: High-Resolution Interpolation [3]

1.4 LITERATURE REVIEW

1.4.1 Face Recognition

Face recognition has had many different methods proposed to solve the problem of having a computer classify a complex feature, a human face. A typical face recognition algorithm consists of three steps: face detection, feature extraction, and face recognition. In most cases the face detection and the feature extraction stages are done simultaneously, however, this is not always the case.

The face detection stage detects a face and performs a rough normalization of the face. Until the middle 1990's most segmentation was focused on segmenting a single face from a simple or complex background. The approaches used included a whole face template, a deformable feature-based template, skin color segmentation, and a neural network [4]. Recently, appearance or image based methods [5], [6] train machine systems using a large number of samples have achieved the best results. Some applications of the face detection stage are face tracking and pose estimation.

The feature extraction stage takes the face that has been detected and obtains features that are then inputted into the face classification system. The features that are extracted can be local features such as lines or points of reference or facial features such as eyes, nose, and mouth. There are three types of feature extraction methods currently: generic methods that are based on edges, lines and curves, feature template based methods detect facial features, and structural matching methods that consider geometrical constraints on the features [4]. Some applications of face detection alone are face tracking and pose estimation where feature extraction applications include face feature tracking, emotion recognition, and gaze estimation.

For face recognition there are three different types of approaches: holistic templates, feature geometry, and a hybrid between holistic templates and feature geometry. These approaches are used with intensity images and give a high-level categorization similar to how humans use holistic and local features. A guideline suggested by the psychological study of how humans use holistic and local features is shown in Figure 1.6.

Approach
Holistic Methods
Principal-component analysis (PCA)
Eigenfaces
Probabilistic Eigenfaces
Fisherfaces/subspace LDA
SVM
Evolution pursuit
Feature Lines
ICA
Other representations
LDA/FLD
PDBNN
Feature Based Methods
Pure geometry methods
Dynamic link architecture
Hidden Markov Model
Convolution Neural Network
Hybrid Methods
Modular Eigenfaces
Hybrid LFA
Shape-normalized
Component-based

Figure 1.6: Further Classification of Face Recognition Methods [4]

The holistic methods use a whole face region while the feature matching methods use local features like the eyes, nose, and mouth which are extracted first while their locations and geometries are inputted subsequently into a structural classifier. The hybrid methods use a combination of the two methods similar to a human's perception.

1.4.2 Super-Resolution

Super-resolution is a signal processing technique to create a high resolution image from a sequence of low resolution images. Super-resolution consists of two separate and important stages: the registration stage and the reconstruction stage. Large amounts of work have been done in each stage and brief overviews of the most common methods are described.

The registration process for super-resolution can be done in either the spatial domain or the frequency domain. The advantage to using the spatial domain is that more general motion models are allowed [3]. Some spatial domain registration methods are by [7], which is an iterative planar motion estimation algorithm that is based on Taylor expansion using a pyramid scheme to increase the precision for large motion parameters. The frequency domain approaches are limited by the properties of the Fourier transform resulting in global motion models [3]. Only planar shifts and planar rotation and scale are considered in the frequency domain approaches. The advantage to these approaches is that it is much easier to describe and handle aliasing in the frequency domain than in the spatial domain. Most of the frequency domain registration approaches are based on the fact two shifted images differ by a phase shift only in the frequency domain, which can be found from their correlation [8]. A log-polar transform of the magnitude of the frequency spectrum, image rotation and scale can be converted into horizontal and vertical shifts. [8] apply a phase correlation technique to estimate the planar shifts. To minimize the errors due to aliasing, [8] uses the low-frequency part of the images since that part of the frequency domain is almost alias free.

Once the images have been properly registered, they can be reconstructed into a single high resolution image. Spatial domain approaches usually use interpolation based approaches [Keren] and others use an estimate of the image using an imaging model that is then updated during each iteration until an acceptable amount of error is reached or the number of preset iterations has been reached [9]. [9] uses the median of the errors improving the results and making the algorithm more robust when outliers are present. Another reconstruction method is to use a stochastic approach. Stochastic super-resolution usually involves a Bayesian approach that is flexible and a convenient way to model a priori knowledge [3]. A maximum a priori (MAP) estimator is used to recreate the high resolution image that will be stable because of the priori constraints [3], [8], [10], [11]. The most common priori assumptions are Gaussian Markov random fields (GMRF) or Huber Markov random fields (HMRF) [10]. HMRF's are used because the GMRF priors are too smooth and sharp edges are lost. The HRMF's use a Gibb's prior

$$p(x) = \begin{cases} x^2 & if\ |x| \leq \alpha \\ 2\alpha|x| - \alpha^2 & otherwise \end{cases}$$

where x is the first derivative of the image and α is a scaling factor shown in Figure 1.7.

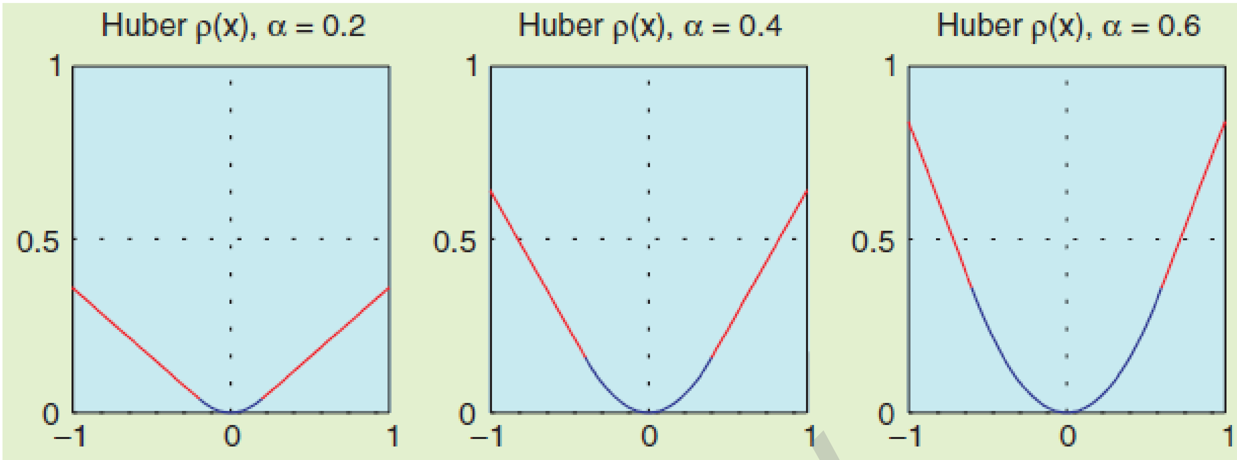

Figure 1.7: Different Values of $\boldsymbol{\alpha}$ [10]

1.4.3 Combining Face Recognition and Super-Resolution

Combining super-resolution and face recognition is usually a simple and straight forward procedure as shown in Figure 1.8. This combination takes the low resolution images and applies super-resolution to the images creating a single high-resolution image which is then fed into the face recognition algorithm. This method is the most intuitive and most common approach when using super-resolution.

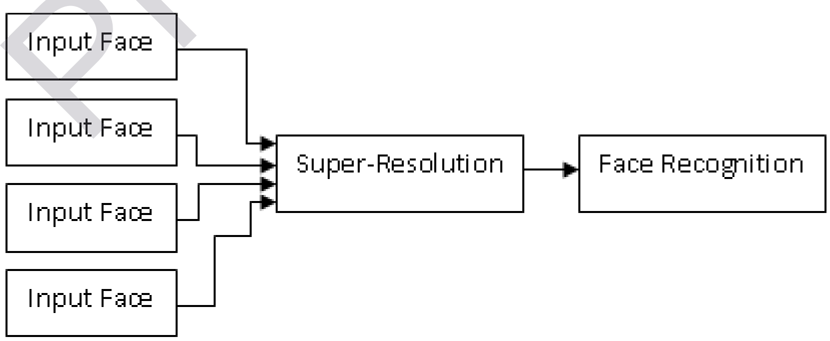

Figure 1.8: Normal Combination for Face Recognition and Super-Resolution

An alternative approach to combining face recognition and super-resolution is to take the low resolution input face images and transform those images into the face space and perform the face recognition in that domain instead of the spatial or pixel domain before super-resolution is applied as