CHAPTER I

INTRODUCTION

1.1 Background and Signification of the Research Problems

The term super-resolution (SRR) refers to the process of obtaining higher-resolution (HR) images from several lower-resolution (LR) ones, i.e. resolution enhancement. The quality improvement is caused by fractional-pixel displacements between images. Super-resolution allows to overcome the limitations of the imaging system (resolving limit of the sensors) without the need for additional hardware. The reconstruction attempts to take advantage of the additional spatio-temporal data available in the sequence of images portraying the same scene. The fundamental problem addressed in super-resolution is a typical example of an inverse problem, wherein multiple low-resolution (LR) images are used to solve for the original high-resolution (HR) image.

Hallucination or recogstruction is a super-resolution algorithm that uses a different kind of constraint, in addition to the reconstruction constraint. This algorithm attempts to recognize local features in the low-resolution images and then enhances their resolution in an appropriate manner. Moreover, face hallucination is still a very active field of research and challenging because people are so familiar with the face. A small error, e.g. an asymmetry of the eyes, might be significant to human perception, whereas for super resolution of generic images the errors in textured regions, e.g. leaves, are often overlooked. It can be widely applied in many fields ranging from image compression to face identification. Especially in video surveillance, a higher resolution face image with detailed facial features will be obviously significant to raise the systems performance.

Face hallucination with the reconstruction-based methods which try to model the process of image formulation to build the relationship between LR and HR based on reconstruction constraints and smoothness constraints, is quite limited by the number of input LR and usually cannot work well in single-image super-resolution problem. Then, the face hallucination with learning-based methods becomes very popular. These methods use some training set directly or indirectly to reconstruct unknown HR images but a major problem of these methods is the high computation requirement due to complex learning process.

In our frameworks, we concentrate in the color image processing which differs from grayscale image processing because of the redundancy and the complementary information within the color bands. The processing is much more complicated due to the increased dimensionality of the problem and exchanges information from and among all bands. Almost all super-resolution methods to date have been designed to increase the resolution of a single channel (monochromatic) image. A related problem such as color super-resolution (SR), addresses fusing a set of previously demosaiced color low-resolution frames to enhance their spatial resolution. To date, there is very little work addressing the problem of color SR . The typical solution involves applying monochromatic SR algorithms to each of the color channels independently, while using the color information to improve the accuracy of the motion estimation. Another approach is transforming the problem to a different color space, where chrominance layers are separated from luminance, and SR is applied only to the luminance channel. Both of these methods are suboptimal as they do not fully exploit the correlation across the color bands.

Real data of natural and social sciences is often very high-dimensional especially the color face images. They can be naturally described as tensors or multilinear arrays. In the most of previous works on face representation, the face is represented as a vector in high-dimensional space. However, an image is intrinsically a matrix, or the second order tensor. In vector representation, the face image has to be converted to a vector. A typical way to do this so-called matrix-to-vector alignment is to concatenate all the rows in the matrix together to get a single vector. To acquire such linear transformation, traditional subspace learning methods, Principal Component Analysis (PCA) need to eigen- decomposition of some matrices. Moreover, the learning parameters in PCA is very large. These methods might not acquire good performance when the number of training samples is small. Recently, multilinear algebra, the algebra of higher-order tensors, is applied for analyzing the multifactor structure of image ensembles. Tensorface which is a novel face representation algorithm represents the set of face images by a higher-order tensor and extends traditional PCA to higher-order tensor decomposition [1, 2]. Then, we can apply multilinear principal component analysis (MPCA) to face hallucination. The MPCA performs feature extraction by determining a multilinear projection that captures most of the original tensorial input variation. In this way, the multiple factors related to expression, illumination and pose can be separated from different dimensions of the tensor. In addition, when the MPCA is implemented in the color space RGB, YCbCr, HSV and CIELAB, it can be investigated that there is a correlation between each color channel.

1.2 Literature Review

The super-resolution restoration idea was first presented by Huang et al. [3] in 1984. It is the process of combining multiple low-resolution images to form a higher resolution one. Numerous super-resolution algorithms have been proposed in the literature [4–8]. Most try to produce a super-resolution image from a sequence of low-resolution images [9, 10]. Based on the definition of SRR , the relevant research papers, published in the conferences and journals are comprehensively reviewed and are broadly categorized into two classes [11]. Specifically, the categorization is into the classes of reconstruction-based SRR algorithm and recognition-based SRR algorithm (or hallucination).

This reconstruction-based SRR algorithm does not require images for training therefore this algorithm does not depend on observed images but reconstruction-based approach inherits limitations when magnification factor increases. The frequency domain approach is a part of the reconstruction-based method. It makes explicit use of the aliasing that exists in each LR image to reconstruct an HR image [12, 13]. Although the frequency domain methods are intuitively simple and computationally cheap, the observation model is restricted to only global translational motion and blur. Due to the lack of data correlation in the frequency domain, it is also difficult to apply the spatial domain a priori knowledge for regularization. Next, the projection onto convex sets (POCS) approach is proposed to describe an alternative iterative approach and it can also incorporate prior knowledge about the solution into the reconstruction process. With the estimates of registration parameters, this algorithm simultaneously solves the restoration and interpolation problem to estimate the SR image [14]. Since the SRR algorithm is an ill-posed problem from an insufficient number of LR images and ill-conditioned blur operators, the regularized ML approach, called regularization, is proposed to stabilize the inversion of ill-posed problem. The last approach of reconstruction based is a nonuniform interpolation approach. In addition, it is the most intuitive method for SR image reconstruction and a fast super-resolution reconstruction based on a non-uniform interpolation using a frequency domain registration is proposed by Vandewalle et al [15–17].

In recognition-based SRR algorithm (or hallucination), this algorithm require images for training therefore this algorithm depend on observed images but this algorithm have high performance when magnification factor increases [18]. With statistical approach, Baker and Kanade [19] proposed another super-resolution algorithm (hallucination or recognition-based super-resolution) that attempts to recognize local features in the low-resolution image and then enhances their resolution in an appropriate manner. Due to the training database, therefore, this algorithm performance depends on the image type (such as face or character) and this algorithm is not robust enough to be sued in typical surveillance video.

For face identification, especially by human, it is desirable to render a high-resolution face image from the low-resolution one. This technique is called face hallucination or face super-resolution [19]. They infer the high frequency components from a parent structure by recognizing the local features from the training set, but there exists some noise in certain area. The simplest way to increase image resolution is a direct interpolation of input images with such algorithms as nearest neighbor or cubic spline. However, the performance of direct interpolation is usually poor since no new information is added in the process. Some other approaches [6, 20–26] are based on learning from the training set containing high and low-resolution image pairs, with the assumption that high-resolution images are Markov random fields (MRFs) [20, 21, 27]. These methods are more suitable for synthesizing local texture, and are usually applied to generic images without special consideration of the property of face images. Baker and Kanade [19, 28] developed a hallucination method based on the property of face images. Abandoning the MRFs assumption, it infers the high-frequency components from a parent structure by recognizing the local features from the training set. Liu et al. [29] developed a two-step statistical modeling approach integrating global and local parameter models. Both methods rely on explicit resolution-reduction-function, which is sometimes unavailable in practice. Wang and Tang develop an efficient face hallucination algorithm using an eigentransformation algorithm [30, 31]. However, the method only utilizes global information without paying attention to local details. Inspired by a well-known manifold learning method, locally linear embedding (LLE), Chang et al. [32] develop the Neighbor Embedding algorithm based on the assumption that the local distribution structure in sample space is preserved in the down-sampling process, where the structure is encoded by patch-reconstruction weights.

To go beyond the current super-resolution techniques which only consider face images under fixed imaging conditions in terms of pose, expression and illumination, these factors are crucial to face analysis and synthesis. Recently, Vasilescu et al. introduce multilinear analysis to face modelling [1, 2] and demonstrate its promising application in computer vision. In the method, equipped with tensor algebra, the multiple factors are unified in the same framework with the coordination between factors expressed in an elegant tensor product form. Wu et al. propose a novel regrssion model to use tensor principal component analysis (PCA) subspace as the face representation [33], which is a special case of the concurrent subspace analysis and Ayan et al. introduce a learning-based method for super-resolution of face that uses kernal principal component analysis (PCA) for deriving prior knowledge about the face class [34]. In addition, Takahiro et al. present a kernel PCA-based adaptive resolution enhancement method of still images. The proposed method introduces two novel approaches into the kernel PCA-based reconstruction of high frequency components missed from a high-resolution (HR) image [35]. Jia et al. propose a multimodal tensor model for face super-resolution with nonlinear deformations and choose a global image-based tensor to perform synthesis across different facial modalities, and a local patch-based multiresolution tensor for hallucination [36–38]. Ma et al. present a simple and efficient multiview face hallucination (MFH) method to generate high-resolution (HR) multiview faces from a single given low-resolution (LR) one [39].

The problem of decomposing tensors (also called n-way arrays or multidimensional arrays) is approached in a variety of ways by extending the Singular Value Decomposition (SVD), principal components analysis (PCA), and other methods to higher orders; see, e.g., [40–46]. Multilinear analysis is a general extension of traditional linear methods such as PCA or matrix SVD and Lathauwer et al. propose a multilinear generalization of the symmetric eigenvalue decomposition for pair-wise symmetric tensors and investigate how tensor symmetries affect the decomposition [47]. Kolda et al. explore the orthogonal decomposition of tensors (also known as multidimensional arrays or n-way arrays) using two different definitions of orthogonality and present numerous examples to illustrate the difficulties in understanding such decompositions. For example, color images are often stored as a sequence of RGB triplets, i.e.,as separate red, green and blue overlays [48].

Note that almost all super-resolution methods to date have been designed to increase the resolution of a single channel (monochromatic) image. A related problem, color SR, addresses fusing a set of previously demosaiced color LR frames to enhance their spatial resolution. To date, there is very little work addressing the problem of color SR. The typical solution involves applying monochromatic SR algorithms to each of the color channels independently [49, 50], while using the color information to improve the accuracy of motion estimation. Another approach is transforming the problem to a different color space, where chrominance layers are separated from luminance, and SR is applied only to the luminance channel [51]. Both of these methods are suboptimal as they do not fully exploit the correlation across the color bands.

In color image super-resolution, Patil et al. propose efficient registration and wavelet based interpolation technique to yield a color super resolved image from four low resolution color images [52]. Therefore, this technique is efficient and computationally fast having clear perspective of real time implementation. The new algorithm in adaptive color super-resolution reconstruction, robust M-estimation is proposed [53]. Using a robust error norm in the objective function, and adapting the estimation process to each of the low-resolution frames, the proposed method effectively suppresses the outliers due to violations of the assumed observation model, and results in color super-resolution estimates with crisp details and no color artifacts, without the use of regularization.

Because abstractdigital color cameras sample the continuous color spectrum using three or more filters therefore, each pixel represents a sample of only one of the color bands. This arrangement is called a mosaic. To produce a full-resolution color image, the recorded image must be processed to estimate the values of the pixels for all the other color bands. This restoration process is often called demosaicing. Ron et al. proposes method involves two successive steps for color super-resolution with CCD sensors [54]. His technique is to let the edges support the color information, and the color channels support the edges, and thereby achieve better perceptual results than those that are bounded by the sampling theoretical limit. Next, Trussell et al. uses stacked notation to represent the mosaiced image capture and derives the minimum mean square error (MMSE) estimator for the demosaiced image [55]. Farsui et al. propose a fast and robust hybrid method of super-resolution and demosaicing, based on a maximum a posteriori estimation technique by minimizing a multiterm cost function [56]. They used L_1 norm for measuring the difference between the projected estimate of the high-resolution image and each low-resolution image, removing outliers in the data and errors due to possibly inaccurate motion estimation. Bilateral regularization is used for spatially regularizing the luminance component, resulting in sharp edges and forcing interpolation along the edges and not across them. Moreover, an additional regularization term is used to force similar edge location and orientation in different color channels.

In super-resolution of color video Sequences, Nimish et al. propose a new multiframe algorithm to enhance the spatial resolution of frames in video sequences and this technique specifically accounts for the possibility that motion estimation will be inaccurate and compensates for these inaccuracies [57]. In [58], an iterative algorithm for enhancing the resolution of monochrome and color image sequences is proposed and two sets of experiments are presented. First, several different experiments using the same motion estimator but three different data fusion approaches to merge the individual motion fields were performed. Second, estimated high-resolution images using the block matching estimator were compared to those obtained by employing a recursive scheme.

In this dissertation, two frameworks are applied to improve the performance of color face hallucination. In the first framework, we employ multilinear principal component analysis (MPCA) in linear regression model for face reconstruction. In the training set, we compute the MPCA subspace projections for both the HR images and the LR images. Next, the color testing image is hallucinated by back-projection in subspace process.

The second frameworks are the combination LR and HR images in a unified tensor which can be reduced to two parts: a global image-based tensor and a local patch-based multiresolution tensor for incorporating high-resolution image details.

1.3 Objectives

Propose a novel face super-resolution (hallucination) with higher-order tensors for color image. The higher-order tensor can be suitable for color face images. For this reason, it can be overcome curse of dimensionality and it also preserves the significant information when the images are in a feature subspace. The two frameworks of color face super-resolution reconstruction with MPCA are proposed to increase the resolution of hallucination performance. In addition, the complexity in hallucination process can be reduced from our proposed method.

1.4 Scope

1. Develop the face hallucination technique with a linear regression model in MPCA that can reconstruct the reasonable color facial image.
   (a) Only the color face image of full frontal view faces will be presented to the PCA in each color channel.
   (b) Only the color face image of full frontal view faces will be presented to the linear regression model with MPCA.
   (c) The color face image of full frontal view faces with partially occluded will be presented to the linear regression model with MPCA.

1.5 Expected Prospects

1. Acquire a basic knowledge of principal component analysis (PCA) for applying to face hallucination.

2. Obtain the new PCA analysis techniques.
3. Obtain new color face hallucination systems.
4. Publish the international journal or conference papers.
5. Know the advantages and disadvantages of using the proposed MPCA techniques in color face hallucination.
6. Understand the necessity of the MPCA techniques for color face hallucination.

1.6 Research Procedure

1. Study previous research papers relevant to the research works of the dissertation.
2. Develop the novel hallucination techniques.
3. Develop simulation programs.
4. Test the proposed algorithms by using standard face databases such as FERET.
5. Perform the proposed algorithm on a color facial database.
6. Collect and analyze computational results obtained from simulation programs.
7. Summarize the major findings as we found in step 6 and conclude the performance of the proposed framework in all concerned aspects.
8. Publish the international journal or conference papers.
9. Check whether the conclusions meet all the objectives of the research work of the dissertation.
10. Write the dissertation.