IEEE Fifth International Conference on Advanced Video and Signal Based Surveillance

Super-resolution of Facial Images in Video with Expression Changes

Jiangang Yu and Bir Bhanu
Center for Research in Intelligent Systems
University of California, Riverside, CA 92521

Abstract

Super-resolution (SR) of facial images from video suffers from facial expression changes. Most of the existing SR algorithms for facial images make an unrealistic assumption that the "perfect" registration has been done prior to the SR process. However, the registration is a challenging task for SR with expression changes. This paper proposes a new method for enhancing the resolution of low-resolution (LR) facial image by handling the facial image in a nonrigid manner. It consists of global tracking, local alignment for precise registration and SR algorithms. A B-splin based Resolution Aware Incremental Free Form Deformation (RAIFFD) model is used to recover a dense local non-rigid flow field. In this scheme, low-resolution image model is explicitly embedded in the optimization function formulation to simulate the formation of low resolution image. The results achieved by the proposed approach are significantly better as compared to the SR approaches applied on the whole face image without considering local deformations. The results are also compared with two state-of-the-art SR algorithms to show the effectiveness of the approach in super-resolving facial images with local expression changes.

1. Introduction

Video-based applications such as surveillance, monitoring, security and access control have received significant attention in the past decade. In many of the above scenarios, the distance between the objects and the cameras may be large, which makes the quality of video usually low and face images small. To overcome this problem, enhancing low-resolution (LR) images from the video sequence has been studied by many researchers in the past decades[4][1][7][14][10].

Super-resolution reconstruction is one of the most difficult and ill-posed problems due to the demand of accurate alignments between multiple images and multiple solutions for a given set of images. In particular, human face is much more complex compared to other objects which have been used in the majority of the super-resolution literature. Super-resolution from facial video may suffer from subtle facial expression variation, non-rigid complex motion model, occlusion, illumination and reflectance variations. Figure 1 shows six low-resolution facial frames from one video sequence with corresponding high-resolution frames. It is clear that the face undergoes non-rigid motions because of the expression changes.

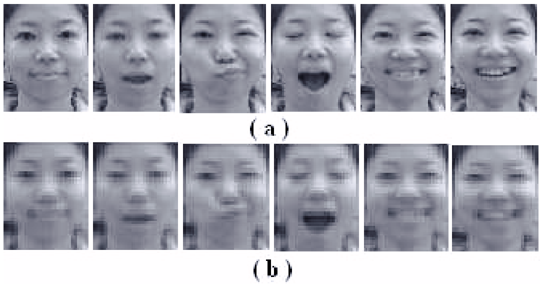

Figure 1. An example of facial images with expression changes. (a) High-resolution (92x73) images. (b) Low-resolution (30x24)images.

In order to tackle the problems brought by the complexity of facial images, in this paper, we propose a novel global-to-local approach to locally align the images before applying SR algorithms. The approach consists of three steps: global tracking, local alignment and SR algorithm. In the global tracking step, a global transformation is used to track the face through the video sequence. Following the global registration, a space warping technique – Free Form Deformations (FFD) is used for modeling the local deformation of faces. The globally aligned image is parameterized using a set of B-spline functions and a set of control points overlaid on its volumetric embedding space. Correspondence is obtained through evolving a control lattice overlaid on the source image. We explicitly embed the LR image formation into the formulation of the Free Form Deformation to simulate the process of LR imaging. We use

978-0-7695-3341-4/08 $25.00 © 2008 IEEE
DOI 10.1109/AVSS.2008.72

Table 1. Sample work for face SR (L-learning-based method, R-reconstruction-based method)

Authors	Approach	Comments
Baker and Kanade[1]	Find the closest matching pixels in the training data using Gaussian, Laplacian and feature pyramids (L).	Manual affine registration by 3 hand marked points on the high-resolution data.
Liu et al.[10]	Combine a global parametric linear model for the whole face with a local nonparametric model which learns local texture from training faces (L).	Manually align the data using 5 hand-selected points on the high-resolution data.
Liu et al.[11]	Fuse high-resolution patches to form a high-resolution image integrating TensorPatch model and coupled residue compensation (L).	Assume a preamble step for alignment using locations of eyes and mouth on the high-resolution data.
Capel and Zisserman[2]	Divide face into six regions and recover SR face image from a high-resolution eigenface space (L).	SR images have visible artifacts on the regions which have expression changes.
Dedeoglu et al.[3]	Find the best-matched patch in training set for the probe image by encoding spatial-temporal consistency of the LR images using a graphical model (L).	Consider translation only global motion on the high-resolution data.
Wang and Tang[14]	Render SR facial image from high-resolution training set using eigen-transformation algorithm (L).	Manual registration using the locations of eyes on high-resolution data.
Jia and Gong[8]	Super-resolve facial image given multiple partially occluded LR images using a Bayesian framework (L).	Manual registration using 3 hand-selected points on the high-resolution data.
Pan et al.[12]	Super-resolve 3D human face by maximizing $a\ posteriori$ probability using progressive resolution chain (PRC) model (L).	Correspondence between 3D model and an image is achieved by mesh parameterizations and there is no local deformation of the mesh.
Lin et al.[9]	Reconstruct SR face image based on a layered predictor network by integrating the local predictors and learning-based fusion strategy (L).	Manual registration by using locations of eyes and mouth on high-resolution data.
This paper	Reconstruct SR facial image with expression changes by registering them with global transformation and local deformation (R).	Automatic registration is done using global-to-local approach to handle the expression changes.

three SR algorithms [7][15][4] in the last step and compare performance results on many real video sequences.

1.1. Related Work

In the past decades, a number of SR techniques have been proposed [4][1][3][11]. Based on whether a training step is employed in SR restoration, they are categorized as: reconstruction-based methods [4][7] and learning-based methods [1][3][11][2]. Table 1 presents a summary of the recent work on super-resolution of facial images and compares it with the work in this paper. All the methods in Table 1 are learning-based SR approaches and need a certain amount of training data of faces. They do not handle local deformations. They assume that alignment has been performed before applying SR methods. However, accurate alignment is the most critical step for SR of facial videos. Our proposed approach integrates the alignment and super-resolution steps.

1.2. Contribution of this paper

• **Super-resolve facial images by handling them in a non-rigid way:** Considering the facial expression changes on a human face, we propose a hierarchical registration scheme that combines a global parametric transformation with a local free-form deformation. In addition to the global transformation which tracks the face through the video sequence using a global motion model, a B-spline based Free-form Deformation is used to locally warp the input LR images to register with the reference LR image.

• **Resolution aware incremental FFD:** The performance of tracking and registration algorithms, which are not designed for LR data, degrades as the quality of input images become poor and the size of images become small. In order to relieve this difficulty brought by LR data, we explicitly embed low resolution imaging model in the formulation of FFD to simulate the formation of LR images.

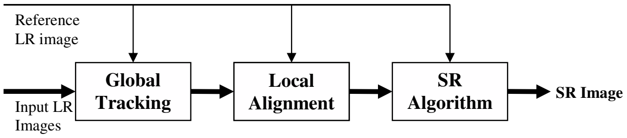

Figure 2. Block diagram of our approach.

• **Part-based super-resolution of facial images:** We design a matching statistic to measure the alignment and then super-resolve the images.

2. Technical Approach

The overall approach is shown in Figure 2. Given a sequence of facial images, we first track the facial region using a global motion model [5]. After this global tracking step, we find the optimal transformation $\mathbf{T}: (x, y) \longrightarrow (x_0, y_0)$ which maps any point of the facial region in the dynamic image sequence $\mathbf{I}(x, y, t)$ at time t into its corresponding point in the reference image $\mathbf{I}(x_0, y_0, t_0)$. Reference image is the first image in the video sequence. We extract the facial region from the video sequence after the global tracking. In order to handle the warping errors and outliers, we partition the facial regions into six regions as left/right eyes, left/right eyebrows, mouth and the rest of the face. Following this step, we deform the globally aligned facial regions (rectified image) to locally register with the regions in the reference image using our specially designed FFD algorithm that accounts the nature of LR data. The last step is to super-resolve the facial image on these registered images to acquire the SR image. Therefore, we design a combined transformation $\mathbf{T}$ consisting of a global transformation and a local deformation as follows

$$\mathbf{T}(x, y, t) = \mathbf{T}_{global}(x, y, t) + \mathbf{T}_{local}(x, y, t) \quad (1)$$

In Fig. 2, the input LR images are tracked to extract the facial region and globally aligned with the reference LR image. Following this step, a Resolution Aware Incremental

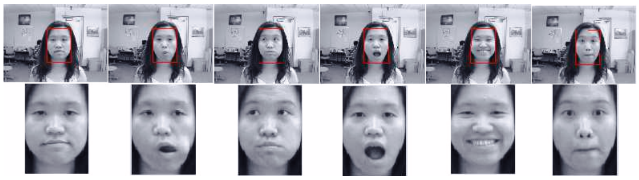

Figure 3. Tracking results for one sequence. The first row shows the video frames. The second row shows the tracked facial regions.

Free-Form Deformation (RAIFFD) is used to locally deform the LR facial images to align with the reference facial image. The details are given below.

2.1. Tracking of Facial Regions

A plane tracking algorithm based on minimizing the sum of squared difference between stored image of the reference facial region and the current image of it [5] is used for tacking the facial region. The motion parameters are obtained by minimizing the sum of squared difference between the template and the rectified image. For efficiency we compute the Jacobian matrix only once on the template image as $\mathbf{M}_0$ instead of recomputing it for each input image [5].

We use a similarity transformation as the motion model to track the facial regions through the video sequence. We locate the facial region interactively on the first image. We recover the motion parameters of translation, rotation and scale for the input images and register them with the reference image by applying these parameters. Fig. 3 shows some frames of tracking results for one video sequence. The first row shows the original video frames and the second one demonstrates the tracked facial regions.

2.2. Local Deformation

The global motion model only captures the global motion of the facial images. As shown in Fig. 1, the local deformations vary significantly across the video sequence. In order to handle the local deformations, we choose a B-splines based FFD model [13] [6], which is a popular approach in graphics, animation and rendering. The basic idea of FFD is to deform an object by space warping the control points which are overlaid on the object. Compared to the pixel-wise optical flow techniques, FFD methods support smoothness constraints and exhibit robustness to noise. Moreover, FFD produces one-to-one correspondences after deformation, which is important for carrying out the SR process.

We consider an Incremental Free Form Deformation (IFFD) [6] formulation to model the local deformation and integrate it into our framework. In order to cope with the large local motion such as open mouth in Figure 1, we adopt a multi-level IFFD from coarse to fine control points levels. Given an incoming frame which is globally aligned, a control lattice P is overlaid on the image space. By evolving the control points in the lattice, the displacements of all the control points are acquired. Subsequently, B-spline basis functions are used as interpolation functions to get the dense deformation field for each pixel in the image space.

2.2.1 Free Form Deformation Formulation

Let us denote the domain of image space as $\Omega = \{\mathbf{x}\} = \{(x,y)|1 \leq x \leq X, 1 \leq y \leq Y\}$ and a lattice of control points overlaid to the image space

$$P = \{(p^x_{m,n}, p^y_{m,n})\};\ m = 1,...,M,\ n = 1,...,N$$

Let us denote the initial configuration of the control lattice as P^0, and the deforming one as $P = P^0 + \delta P$. The parameters of FFD are the deformations of the control points of the lattice in both directions (x,y):

$$\Delta\mathbf{P} = \{(\delta P^x_{m,n}, \delta P^y_{m,n})\};\ (m,n) \in [1,M] \times [1,N]$$

The deformed location of pixel given the deformation of the control lattice from P^0 to P, is defined as the tensor product of cubic B-splines

$$\begin{aligned} & \mathbf{T}_{local}(\mathbf{x};\Delta\mathbf{P}) \\ = & \sum_{k=0}^{3}\sum_{l=0}^{3} B_k(u)B_l(v)(P^0_{i+k,j+l} + \delta P_{i+k,j+l}) \\ = & \sum_{k=0}^{3}\sum_{l=0}^{3} B_k(u)B_l(v)P^0_{i+k,j+l} \\ + & \sum_{k=0}^{3}\sum_{l=0}^{3} B_k(u)B_l(v)\delta P_{i+k,j+l} \end{aligned} \tag{2}$$

where $\mathbf{P}_{i+k,j+l}(k,l) \in [0,3] \times [0,3]$ are pixel (x,y)'s sixteen adjacent control points, $k = \lfloor \frac{x}{X} \cdot (M-1) \rfloor + 1, j = \lfloor \frac{y}{Y} \cdot (M-1) \rfloor + 1$ and $B_k(u)$ represents the $k\text{-}th$ basis function of cubic B-splines

$$B_0(u) = (1-u)^3/6,\ B_1(u) = (3u^3 - 6u^2 + 4)/6$$
$$B_2(u) = (-3u^3 + 3u^2 + 3u + 1)/6,\ B_3(u) = u^3/6$$

where $u = \frac{x}{X} \cdot M - \lfloor \frac{x}{X} \cdot M \rfloor, v = \frac{y}{Y} \cdot M - \lfloor \frac{y}{Y} \cdot M \rfloor$.

2.2.2 Cost Function

Given the local deformation formulations, we need to find the deformation parameters of the control lattice $\Delta\mathbf{P}$. Then we can warp the input frame $\mathbf{I}(x,y)$ to register with the reference frame $\mathbf{I}(x_0,y_0)$ using the cubic B-spline functions. Similar to [6], we use Sum of Squared Differences (SSD) as the data-driven term for our optimization energy function

$$E_{data}(\Delta\mathbf{P}) = \iint_{\Omega} (\mathbf{I}(x,y) - g(\mathbf{T}(x,y;\Delta\mathbf{P})))^2 dxdy \tag{3}$$

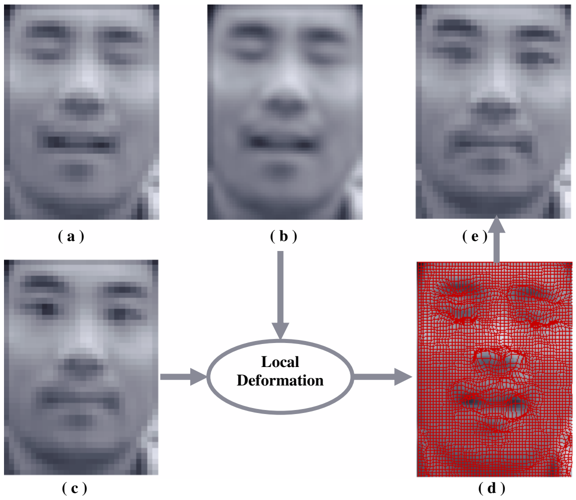

Figure 4. The Resolution Aware Incremental Free Form Deformation. (a) The input LR image. (b) Bicubic interpolated image of (a). (c) The reference LR image. (d) The interpolated image (b) overlaid with control lattice. (e) The deformed LR image .

In order to account for outliers and noise, we consider an additional smoothness term on the deformation field δP as

$$E_{smoothness}(\Delta\mathbf{P}) = \iint_{\Omega} \left(\left\| \frac{\partial \delta \mathbf{T}}{\partial x} \right\|^2 + \left\| \frac{\partial \delta \mathbf{T}}{\partial y} \right\|^2 \right) dxdy \quad (4)$$

Combining Equation (3) with Equation (4), we can write the engergy term as

$$E(\Delta\mathbf{P}) = E_{data}(\Delta\mathbf{P}) + \lambda E_{smoothness}(\Delta\mathbf{P}) \quad (5)$$

where λ is a constant which defines the tradeoff between the displacements and the smoothness of the transformation. We choose λ as a value between 2 and 20. The calculus of variations and a gradient decent method can be used to optimize the energy function. We can take the derivative of $E(\Delta\mathbf{P})$ with respect to the deformation parameters $\Delta\mathbf{P}$ as $\frac{\alpha E(\Delta\mathbf{P})}{\alpha \Delta\mathbf{P}^{[x]}_{(m,n)}}$ and $\frac{\alpha E(\Delta\mathbf{P})}{\alpha \Delta\mathbf{P}^{[y]}_{(m,n)}}$ to find the deformation $\Delta\mathbf{P}$ by minimizing the energy function.

2.2.3 Resolution Aware Local Deformation

In order to account for the complexity brought by LR data, we integrate LR image formation model into the FFD formulation. Considering the LR imaging model of a digital camera, LR images are blurred and subsampled (aliased) from the high resolution data with additive noise. While FFD works well with high resolution data, its accracy of deformation degrades quickly at low resolution. We integrate LR imaging model into the FFD formulation. Fig. 4 shows the process of our proposed Resolution Aware Incremental Free Form Deformation (RAIFFD). In Fig. 4, (a) shows the input LR image which is need to be deformed with reference to the LR image in (d). (b) is the interpolated image of (a). (e) shows the control lattice overlaid on (b) to show the deformation. (c) is the LR image deformed from (a). We perform local deformations on (b) based on the following considerations:

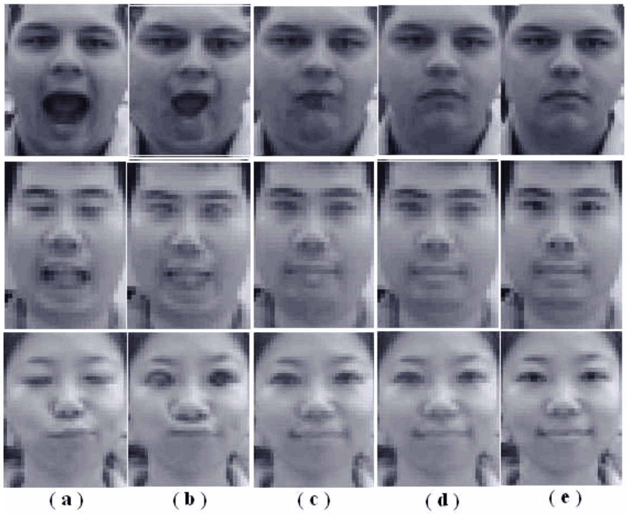

Figure 5. Examples of our Resolution Aware IFFD local registration. (a) Source frames which need to be deformed. (b)-(c) illustrate the deformation process from coarse-to-fine control lattice. (d) represents the final deformed LR images warped to the reference frames. (e) represents the reference frames.

- **Deform local motion on high resolution data**. Without the loss of generality, we assume the camera is fixed and the relative motion between the object and the camera is due to the motion of the object. Local motion occurs on the object while the acquired LR image is the process of digital camera. To better model the local motion, instead of deforming the control lattice on LR image we perform FFD on the high resolution data. Then we can simulate the process of LR imaging from the deformed high resolution image to get the motion compensated LR image. Plugging the LR imaging model into the data driven term in Equation (3), we can rewrite Equation (3) as

$$\begin{aligned} & E_{data}(\Delta\mathbf{P}) \\ & \quad = \iint_{\Omega} \left(\mathbf{I}_{LR}(x,y) - f(X,Y;\Delta\mathbf{P})\right)^2 dxdy \end{aligned} \quad (6)$$

where

$$\begin{aligned} & f(X,Y;\Delta\mathbf{P}) \\ & \quad = g(\mathbf{T}(X,Y;\Delta\mathbf{P})) * \mathbf{h}) \downarrow_s \\ & \quad = \iint_{(X',Y') \in bin(X,Y)} \varphi(\cdot) dX' dY' \end{aligned} \quad (7)$$

In Equation (7), $\varphi(\cdot)$ is defined as

$$\varphi(\cdot) = g(\mathbf{T}(X',Y';\Delta\mathbf{P}))\mathbf{h}(X-X',Y-Y')dX'dY' \quad (8)$$

where (X,Y) is the pixel coordinates on high resolution image, $bin(X,Y)$ is the sensing area of the discrete pixel (X,Y) and $\mathbf{h}$ is the blurring function (Point Spread Function). The continuous integral in Equation (7) is defined over $bin(X,Y)$ to simulate formation of LR image. The smoothness term in Equation (4) is rewritten as

$$E_{smoothness}(\Delta\mathbf{P}) = \iint\limits_{\Omega} \phi(\cdot)dXdY \quad (9)$$

where

$$\phi(\cdot) \quad = \left\| \frac{\partial\delta\mathbf{T}(X,Y;\Delta\mathbf{P})}{\partial X} \right\|^2 + \left\| \frac{\partial\delta\mathbf{T}(X,Y;\Delta\mathbf{P})}{\partial Y} \right\|^2$$

The derivatives of $E_{data}(\Delta\mathbf{P})$ along $\Delta P^{[x]}_{(m,n)}$ can be calculated as

$$\frac{\partial E_{data}(\Delta\mathbf{P})}{\partial\Delta P^{[x]}_{(m,n)}} = -\iint\limits_{\Omega} 2r\frac{\partial r}{\partial\Delta P^{[x]}_{(m,n)}}dxdy \quad (10)$$

where

$$\frac{\partial r}{\partial\Delta P^{[x]}_{(m,n)}}$$

$$= \quad \iint_{(X',Y')\in bin(X,Y)} \varphi(\cdot)\frac{\partial\mathbf{T}(X',Y';\Delta\mathbf{P})}{\partial\Delta P^{[x]}_{(m,n)}}dX'dY' \quad (11)$$

The derivatives of $E_{smoothness}(\Delta\mathbf{P})$ along $\Delta P^{[x]}_{(m,n)}$ is calculated as

$$\frac{\partial E_{smoothness}(\Delta\mathbf{P})}{\partial\Delta P^{[x]}_{(m,n)}}$$

$$\begin{aligned} = \quad & 2\iint \frac{\partial\delta\mathbf{T}(X,Y;\Delta\mathbf{P})}{\partial X}\cdot\frac{\frac{\partial\delta\mathbf{T}(X,Y;\Delta\mathbf{P})}{\partial X}}{\partial\Delta\mathbf{P}^{[x]}_{(m,n)}}dXdY \\ & +2\iint \frac{\partial\delta\mathbf{T}(X,Y;\Delta\mathbf{P})}{\partial Y}\cdot\frac{\frac{\partial\delta\mathbf{T}(X,Y;\Delta\mathbf{P})}{\partial Y}}{\partial\Delta\mathbf{P}^{[x]}_{(m,n)}}dXdY \quad (12) \end{aligned}$$

where r in data term is defined as $\mathbf{I}_{LR}(x,y)-f(X,Y;\Delta\mathbf{P})$. The derivation for $\Delta P^{[y]}_{(m,n)}$ can be similarly obtained.

- **Super-resolution methodology requires sub-pixel registration**. SR reconstruction can be seen as "combining" new information from LR images to obtain a SR image. If the LR images have sub-pixel shifts from each other, it is possible to reconstruct a SR image. Otherwise, if the LR images are shifted by integer units, then each image contains the same information, and thus there is no new information that can be used to reconstruct a SR image. High frequency areas in human face usually correspond to facial features such as eyes, eyebrows and mouth. During facial expression changes, these facial features deform the most among the face. In LR facial images, the fact is that these facial features have a few pixels, which are blurred from high resolution image and noisy. If we deform the LR image, the deformation can not capture the subtle movements. Moreover, interpolation in LR image during the process of deformation smooths out the high frequency features since they only occupy a few of pixels. This leads to the loss of new information which can be used to reconstruct SR image.

2.3. Super-resolution Algorithm

We work with three algorithms [7][15][4]. The first super-resolution algorithm is based on IBP [7]. We also perform experiments on using methods in [15] and [4]. In [15], the authors proposed a robust approach which combines a median estimator and the iterative framework to reconstruct SR images. The authors demonstrate that their method is robust to the outliers due to motion errors, inaccurate motion models, noise, moving objects and motion blurs [15]. The authors in [4] proposed a general hybrid SR approach that combines the benefits of the stochastic approaches ML (or MAP) and the POCS approach.

2.4. A Match Measure for Warping Errors

Even though a global transformation and a local deformation method are used to handle the non-rigidity of the facial image, there may exist warping errors due to the violation of basic assumptions such as Lambertian surface, particularly for low-resolution images. In order to detect anomalies in flow based warping, we partition the facial images into six regions based on facial features as left/right eyebrows, left/right eyes, mouth and the other part of the face. We design a match statistics to measure how well the warped patches align with the target patches. If the match score is below a certain threshold, the corresponding part will be ignored during super-resolving the texture. We define our match measure as follows,

$$\mathbf{E}_j = \sum_{x=1}^{M}\sum_{y=1}^{N}\frac{\Big((\mathbf{Y}^j_k(x,y)-\mu_1)([\mathbf{X}^j_n(x,y)]^{\mathbf{B}_n}-\mu_2)\Big)}{M{*}N{*}\sigma_1{*}\sigma_2} \quad (13)$$

where M and N are the image size, μ_1 and μ_2 are respective means of image region, σ_1 and σ_2 are respective image variances within the region. j is from 1 to 6 representing the j-th part of the face, $\mathbf{Y}^j_k$ represents the j-th local region of the k-th input LR frame and $\mathbf{X}^k_n$ denotes the intensity value of the SR image at n-th iteration. The absolute value of $\mathbf{E}_j$ is between 0 and 1. We choose 0.9 as the matching threshold in this paper.

3. Experimental Results

- **Data and Parameters:** We record 10 video sequences for 10 people with each of them lasting about 3 minutes.

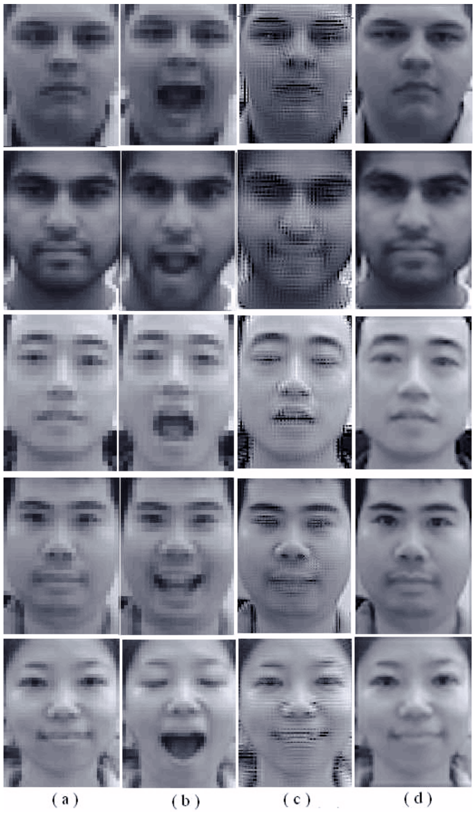

Figure 6. Super-resolution Results. (a)-(b) Low-resolution images. (c)Reconstructed SR images using global registration. (d)Reconstucted SR images using global and our RAIFFD local deformation approach.

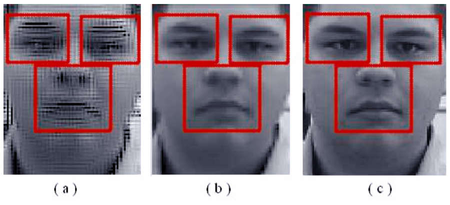

Figure 8. Marked regions for calculating peak signal-to-noise ratio (PSNR). (a) Reconstructed SR image using global registration. (b) Reconstructed SR image using global + local deformation approach. (c) Original High-resolution image.

During the recording, each person was asked to look at the camera at the distance of about 20 feet from the camera and make continuous expression changes. The average size of the face is about 75 × 70 with maximum size at 115 × 82 and minimum size at 74 × 58. We then blur these videos with a 5 × 5 Gaussian kernel and down-sample them into image sequences with average size 35 × 27. We then use our proposed approach to estimate the flow fields and super-resolve these LR images to acquire SR images. We found that 35 frames were enough to cover most of the expression changes the people made during recording. We then use 35 frames for each person for super-resolving.

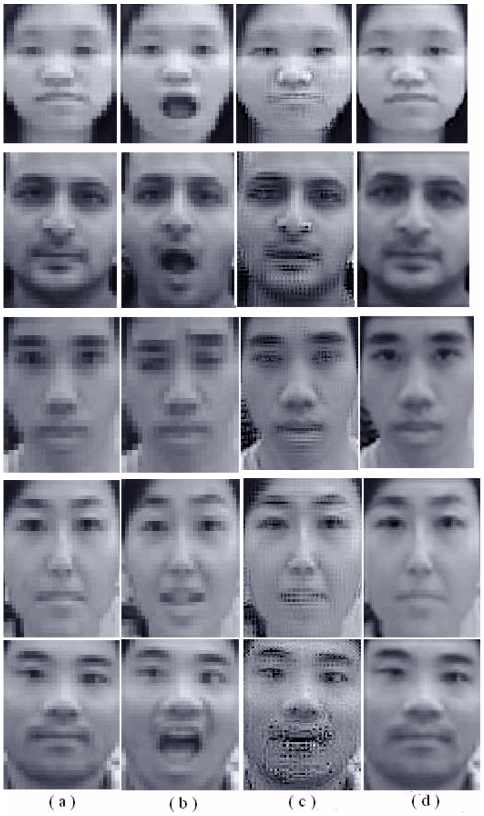

Figure 7. Super-resolution Results. (a)-(b) Low-resolution images. (c)Reconstructed SR images using global registration. (d)Reconstucted SR images using global and our RAIFFD local deformation approach.

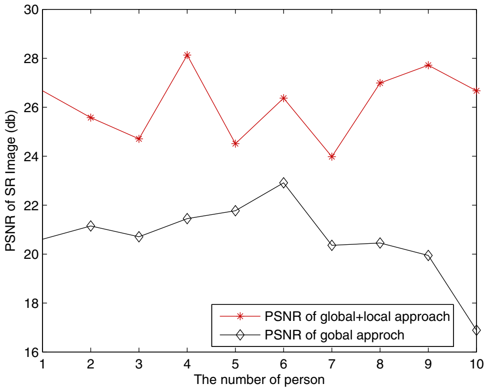

Figure 9. Comparison of PSNR values between global reconstructed SR images and our global+local reconstructed SR images.

• **Results of Resolution Aware FFD:** The examples of the local deformation process are shown in Figure 5. In this figure, the input frames which need to be locally deformed are shown in (a). (b)-(c) demonstrate the process of local deformation from coarse-to-fine control lattice configurations. The reference frames are shown in (e). This fig-

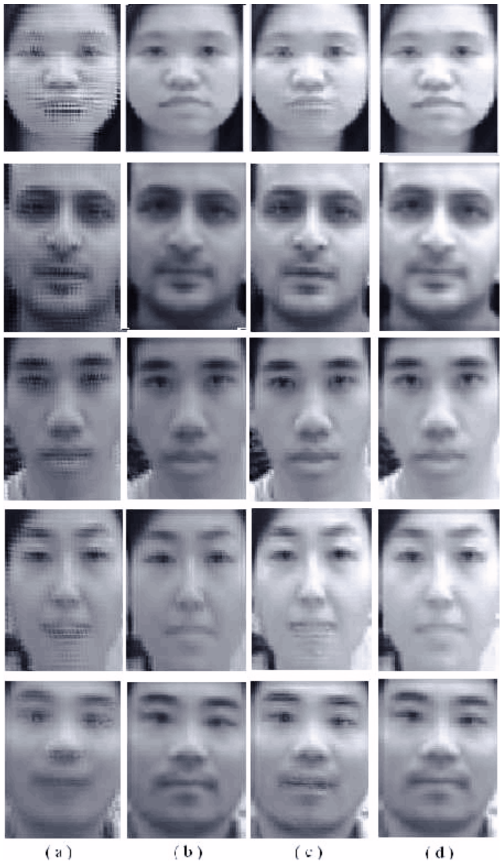

Figure 10. Super-resolution results using SR algorithms in [15] and [4] (1). (a) SR results of globally aligned data using method in [4]. (b) SR results of globally+locally aligned data using method in [4]. (c) SR results of globally aligned data using method in [15]. (d) SR results of globally+locally aligned data using method in [15].

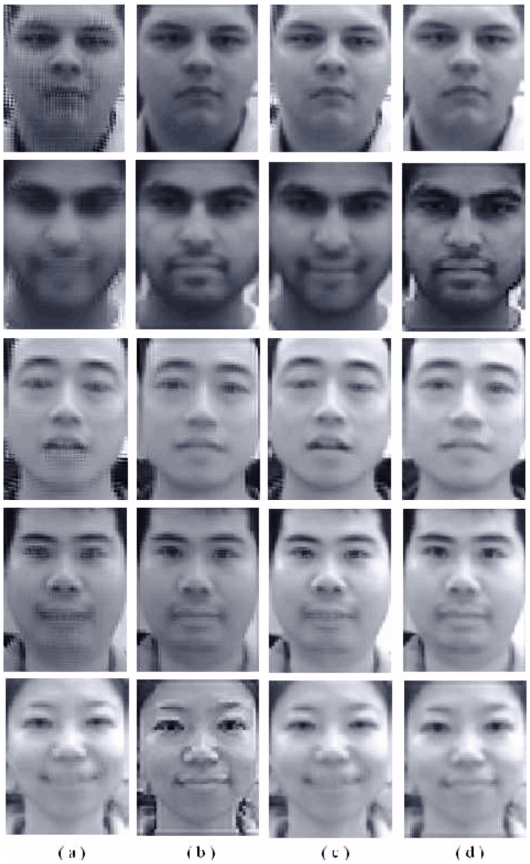

Figure 11. Super-resolution results using SR algorithms in [15] and [4] (2). (a) SR results of globally aligned data using method in [4]. (b) SR results of globally+locally aligned data using method in [4]. (c) SR results of globally aligned data using method in [15]. (d) SR results of globally+locally aligned data using method in [15].

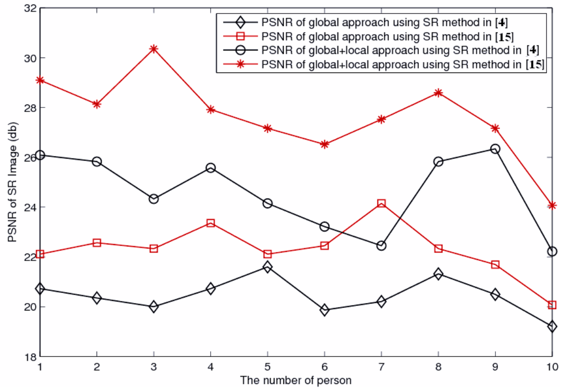

Figure 12. Comparison of PSNR values between global reconstructed SR images and our global+local reconstructed SR images using the SR methods in [15] and [4].

ure clearly shows our proposed Resolution Aware FFD approach can successfully deforms the facial images with expression changes to register with the reference images.

- **Super-resolution Results on global registration vs. global + RAIFFD local deformation:** We super-resolve the input LR images to acquire SR images using the SR algorithms discussed in section 2.3. We perform SR reconstruction on two kind of aligned data: globally aligned LR images using the registration method in section 2.1 and LR images aligned using the proposed global and local alignment methods in section 2.1 and 2.2. We show the results for the complete data in Fig. 6 and Fig. 7. In Fig. 6 and Fig. 7, two of the input LR frames are shown in (a) and (b). The reconstructed SR images on globally aligned LR data are shown in (c). SR images reconstructed using our global+local method are shown in (d). The SR results in (d) are much better than the results in (c), especially in the high-frequency areas such as mouth, eyes and eyebrows. This is due to the reason that the global only alignment method can not capture the local deformations on these local parts when face have expression changes. Our global+local approach not only finds the global transformation for the global motion, but also captures the local deformations.

Quantification of performance: In order to measure the performance of our algorithm, we compute peak signal-to-noise ratio (PSNR) as the measurement between original high-resolution and reconstructed SR images. Considering the fact that local motions mostly occur in the high-frequency areas such as eyes and mouth (see Fig. 8, we calculate the PSNR values only on the marked regions between the reconstructed SR image ((a) and (b)) and the original high-resolution image. SR image in (a) is reconstructed using global registration and SR image in (b) is the SR image using our proposed global+local approach. We calculate the PSNR values for our data (10 people) and show this plot in Fig. 9. We find that PSNR of the areas for our global+local approach is much better than that for global only approach. The average PSNR value is 20.6261 for the global approach and 26.1327 for our global+local approach.

• **Proposed approach with two different SR algorithms:** Similar to the previous experiments, we implement two methods [4][15]both on the globally aligned data and the data aligned using our global+local model. The results are shown in Fig. 10 and 11. We find both SR images in (b) and (d) are better than in (a) and (c), which are the results on the globally aligned data. We also compute the PSNR values using these two methods and show them in Figure 12. This verifies again that global only alignment is not good enough to register facial images with expression changes for SR purpose. After comparing the results between (b) and (d) or (a) and (c), we find the quality of SR images using [15] outperforms the method in [4]. This does not surprise us because the proposed method in [15] uses median estimator to replace the sum in the iterative minimization in method [4], which reduces the influence of the large projection errors due to inaccurate motion estimation. We also find SR algorithm in [15] achieves better results than [7] from the PSNR values in Fig. 9 and Fig. 12 while [7] has similar results compared to method in [4].

4. Conclusions

Reconstruction of SR facial images from multiple LR images suffers from the special characteristics of human face. Among these difficulties, non-rigidity of human face is a critical issue since accurate registration is an important step for SR. In this paper, we proposed a Resolution Aware Incremental Free Form Deformation (RAIFFD) approach which embedded the low resolution imaging model explicitly in the formulation to handle the non-rigidity and low-resolution issues. Our experimental results showed that the proposed SR framework can effectively handle the complicated local deformation of human face and produced better SR image than the global approach. Since most of the SR approaches presented in Table 1 are learning-based methods which do not handle local deformations in real data, we did not compare them with our approach.

References

[1] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 24(9):1167–1183, September 2002. 1, 2

[2] D. Capel and A. Zisserman. Super-resolution from multiple views using learnt image models. *CVPR'01*, 2:627–634, 2001. 2

[3] G. Dedeoglu, T. Kanade, and J. August. High-zoom video hallucination by exploiting spatio-temporal regularities. *CVPR'04*, 2:151–158, Jun. 2004. 2

[4] M. Elad and A. Feuer. Restoration of a single superresolution image from several blurred, noisy and undersampled measured images. *IEEE Trans. Image Processing*, 6:1646–1658, December 1997. 1, 2, 5, 7, 8

[5] G. D. Hager and P. N. Belhumeur. Efficient region tracking with parametric models of geometry and illumination. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(10):1025–1039, 1998. 2, 3

[6] X. Huang, N. Paragios, and D. Metaxas. Shape registration in implicit spaces using information theory and free form deformations. *IEEE Trans. on Medical Imaging*, 28(8):1303–1318, August 2006. 3

[7] M. Irani and S. Peleg. Motion analysis for image enhancement: Resolution, occlusion, and transparency. *Journal Visual Communication Image Represent*, 4:324–335, December 1993. 1, 2, 5, 8

[8] K. Jia and S. Gong. Hallucinating multiple occluded face images of different resolutions. *Pattern Recogn. Lett.*, 27(15):1768–1775, 2006. 2

[9] D. Lin, W.Liu, and X. Tang. Layered local prediction network with dynamic learning for face super-resolution. *Proc. IEEE Int. Conf. Image Processing*, a:885–888, 2005. 2

[10] C. Liu, H. Shum, and C. Zhang. A two-step aprroach to hallucinating faces: Global parametric model and local nonparametric model. *CVPR'01*, 1:192–198, 2001. 1, 2

[11] W. Liu, D. Lin, and X. Tang. Hallucinating faces: Tensor patch super-resolution and coupled residue compensation. *CVPR'05*, 2:478–484, 2005. 2

[12] G. Pan, S. Han, Z. Wu, and Y. Wang. Super-resolution of 3d face. *The 9th European Conference on Computer Vision (ECCV'06),*, 3952:389–401, 2006. 2

[13] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. Nonrigid registration using free-form deformations: Application to breast mr images. *IEEE Trans. on Medical Imaging*, 8:712–721, 1999. 3

[14] X. Wang and X. Tang. Hallucinating face by eigentransformation. *IEEE Transactions on Systems, Man, and Cybernetics, Part C*, 35(3):425–434, 2005. 1, 2

[15] A. Zomet, A. Rav-Acha, and S. Peleg. Robust super resolution. *Proc. of IEEE International Conf. on Computer Vision and Patern Recognition (CVPR)*, 1:645–650, 2001. 2, 5, 7, 8