

Probabilistic Super-resolution Reconstruction for Face Images Using Sequential Monte-Carlo Method

Hua HUANG, Xin FAN, Chun QI and Shi H. ZHU

Abstract—Most previous super-resolution (SR) approaches are implemented with two individual cascade steps, image registration and image fusion. They face the dilemma that the fusion demands accurate motion estimation, while the lack of image information of high quality, which is the output of the fusion step, leads to inaccurate motion fields estimation. In this paper, we pose SR reconstruction as Bayesian state estimation which results in that image alignment and image fusion are combined into one unified framework. We build a part-based face model that encodes the structural information of human faces. The prior information from the face model is incorporated into both registration and fusion process of super-resolution, and high resolution images are reconstructed by using Sequential Monte Carlo based algorithm. Experiments performed on synthesized frontal face sequences show that the proposed approach gains superior performance in registration as well as reconstruction.

Index Terms—Super-resolution, human face, Sequential Monte-Carlo Method, motion estimation

1. INTRODUCTION

Super-resolution (SR) refers to reconstructing a high resolution (HR) image or image sequence from a series of low resolution (LR) images. Due to its wide applications in video surveillance, video communication and digital publishing, SR has gained great attentions and numerous SR algorithms have been proposed during the past decade [1, 2]. In most traditional approaches, SR reconstruction is implemented with two individual cascade steps, i.e., image registration and image fusion [1]. Image registration is performed prior to the fusion step in order to compute the relative motion fields between pixels of consecutive image frames. The LR image frames are aligned by using the estimated motion fields and then fused into a HR image by some sophisticated techniques, in which additional constraints on the desired HR image are usually imposed so as to resolve the inherent illness of the inverse process of reconstruction [1, 3].

Image registration is such a critical step to SR reconstruction that the accuracy of the estimated motion fields determines the quality of reconstructed HR images [4]. Various methods have been devised to estimate the motion fields represented as simple parametric forms or dense optical flows. Dating back to Lucas and Kanade's pioneer work [5], the Lucas-Kanade (LK) algorithm and its

variants are widely used for motion estimation in SR reconstruction due to their efficiency [6]. Though the LK based algorithms with pyramid searching strategy can handle a wide range of translation, it is not robust enough to obtain accurate motion fields in many critical cases, e.g., non-linear motion, cluttered background, low quality images and partial occlusion. This incapability can be attributed to the following reasons: 1) the gradient-based search strategy used in LK algorithm is sensitive to the initial point, and is prone to find a local minimum; 2) the local intensity information used can hardly recover accurate motion fields when part of image information is obscured; 3) motion estimation is performed prior to data fusion so that no information from HR images can be employed to resolve the ambiguity at this stage. In the field of object tracking, various approaches have been developed to deal with occlusion, cluttered background or non-linear motions. For example, Sequential Monte-Carlo (SMC) based methods are used to accommodate non-linear motions and to combat cluttered background [7], and the statistical models representing the structural information of the object of interest are employed to handle occlusion [8]. And the performance is improved further by combining the statistical models with SMC [3, 9]. These techniques may be helpful for SR reconstruction.

The second step, data fusion, essentially needs to make use of the information about the image formation process that yields the LR image sequences. Besides, the prior knowledge pertaining to the original HR image is required to regularize the ill-posed inverse process of reconstruction. Under a Bayesian framework, this type of information can be formulated as a prior probability model, and then the SR reconstruction is posed as maximum a posterior (MAP) estimation [10]. Great efforts have been made to develop advanced statistical models, e.g., Markov Random Fields [11, 12], which characterize local spatial properties in the vicinity of image pixels. These priors can be regarded as smoothness priors that encourage smooth solutions, and they are less effective on recovering image details as the magnification factor increases dramatically [1]. Recently, recognition-based priors have been incorporated into the fusion step to substitute the smoothness priors [1, 13, 14, 15]. These approaches take into account the structures of what is expected in the LR images and the desired HR image, and use the recognized features to "hallucinate" or synthesize the image details lost at the process of down-sampling. Significant improvements on the recovery of local details are reported. However, it is assumed (supposed) that accurate motion fields have been estimated prior to fusion, that is, the recognition-based priors are not incorporated into both steps of the SR reconstruction.

Manuscript received December 18, 2005; revised April 23, 2006. This work was supported by the High Technology Research and Development Program of China under Grant No. 2004AA1Z2312.

Hua Huang, Chun Qi and Shi H. Zhu are with the School of Electronics & Information Engineering, Xi'an Jiaotong University, Xi'an, Shaan'xi, PR. China, (e-mail: huanghua@xjtu.edu.cn, qichun@xjtu.edu.cn and szhu@mail.xjtu.edu.cn); Xin Fan is with the Institute of Signal and Image Processing, Dalian Maritime University, Dalian, Liaoning, PR. China, (e-mail: fanxin@dlnu.edu.cn).

We observe that there exists a dilemma in the previous SR methods that the fusion of LR image frames demands accurate motion estimation. However, the lack of HR image information, which is the output of the fusion step, leads to the difficulties in obtaining accurate motion fields. The cause of this dilemma lies in that the process of SR reconstruction is divided into two steps connected in an open-loop way. On the other hand, the previous work on motion tracking and image fusion shows that it is beneficial for both steps to take into account the prior information of the object. In this paper, we propose a part-based face model composed of several conditional dependent statistical models to represent facial components with geometrical constraints. The textures of the facial components are generated from their appearance models, and then warped to appropriate positions to obtain the desired HR face image. Both the appearance parameters and location parameters are simultaneously estimated by a novel SMC based method, which allows both Markov transition along the temporal chain and Markov Random Filed coupling between facial components in one framework. In the proposed approach, the high level prior knowledge in terms of statistical models is incorporated into the both steps of SR reconstruction, and the image information provided by the LR images is accumulated along the temporal axis. This means that the proposed approach combines motion estimation and image fusion into one framework.

2. PART-BASED FACE MODEL

As known, a face is such an object that presents highly structured characteristics. These characteristics can be summarized as follows: 1) a face can be composed of several components in a semantic sense, e.g., eyes, nose and mouth; 2) each component exhibits structural appearance; and 3) the locations and appearance of the components are highly constrained. We use a graphical structure to represent the appearance of the facial components as well as the relationships between the components, since graphical models provide a powerful tool for characterizing conditional dependence relationships and for performing probabilistic inference [16]. Figure 1 gives the graphical structure of the proposed model, which is described in details below.

The graph is composed of four nodes and exhibits such a tree structure that three of the nodes are connected to one root node. We only consider four facial components, namely two eyes, nose and mouth, because these components are the key features to determine the identity of a face and facial expressions. The nodes \mathbf{X}_i represent the locations and appearance parameters of the facial components, denoted as $\mathbf{X}_i \doteq (\mathbf{l}_i, \mathbf{a}_i)$. Locations are given by affine parameters, i.e. translations in the image plane and scale $\mathbf{l} = (t_x, t_y, s)$, while appearance parameters for each node are determined via principal

component analysis (PCA) on example images of each facial component, which is similar to modular eigenface [17]. The appearance of each facial component is generated by:

$$\mathbf{T}_i = \mathbf{T}_{i0} + \Phi_i \mathbf{a}_i \quad (1)$$

where \mathbf{T}_{i0} denotes the mean feature of the i th component and Φ_i is a matrix composed of the principal eigen vectors of the covariance of the i th component.

The edges \mathbf{E} in the graph indicate the conditional dependences between facial components. These dependences are related by pair wise interaction potentials:

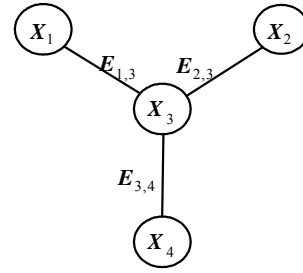


Fig. 1. The graphical structure

$$\psi(\mathbf{X}_i, \mathbf{X}_j) = \psi_l(\mathbf{l}_i, \mathbf{l}_j) \psi_a(\mathbf{a}_i, \mathbf{a}_j) \quad (2)$$

where i and j are the indices of graph nodes. The tree structure in this graphical model is specified in advance for the simplicity of the learning process. We notice that some algorithms allowing for automatic learning of graphical structure give the same tree structure as the graph shown in Figure 1 [18]. In addition, this type of structure makes it viable to achieve the inference via a SMC based algorithm along the temporal chain in which pair wise Markov potentials between facial components are coupled.

The appearance model for each facial component and the potentials in Eq. (2) are learnt from a number of example images with annotated positions and masks of facial components. Performing standard PCA on the training images of facial components can yield the appearance models. We model the potential between the positions of facial components ψ_l as a Gaussian density. The mean and variance of the Gaussian density are estimated from the annotated positions. And the potential functions ψ_a relating coefficients of appearance models are approximated by kernel density estimates [19] from the training images.

Once the trained models are available, images of any facial components can be generated by specifying appearance parameters \mathbf{a}_i and then warp these images to their corresponding locations \mathbf{l}_i to produce a face image. It means that a plausible face image can be obtained with appropriate appearance parameters and locations estimated from the observations with low quality. As a result, SR

reconstruction of a face image becomes probabilistic estimation of the values of four nodes \mathbf{X}_i from the available LR images. We elaborate the probabilistic formulation for SR reconstruction in the next section.

3. BAYESIAN FORMULATION OF SR RECONSTRUCTION

We cast the SR reconstruction of a face image as the probabilistic state estimation of N nodes $\mathbf{X}_t = (\mathbf{X}_{1t}, \dots, \mathbf{X}_{it}, \dots, \mathbf{X}_{Nt})$ given LR image frames $\mathbf{Y}^t = (\mathbf{Y}_1, \dots, \mathbf{Y}_t)$ up to time t . It is worth noting that the state of each node is composed of both appearance and location parameters. These estimated parameters give a plausible high quality image by accumulating observed information from the LR image sequence and incorporating prior knowledge from the defined face model. The posterior probability density can be recursively updated as [20]:

$$p(\mathbf{X}_t | \mathbf{Y}^t) \propto p(\mathbf{Y}_t | \mathbf{X}_t) \int_{\mathbf{X}_{t-1}} p(\mathbf{X}_t | \mathbf{X}_{t-1}) p(\mathbf{X}_{t-1} | \mathbf{Y}^{t-1}) \quad (3)$$

where the likelihood $p(\mathbf{Y}_t | \mathbf{X}_t)$ expresses how the current state \mathbf{X}_t fits the observations available at time t . We assume that a face performs the movement with slight deviation from frontal view so that facial components do not occlude each other. The likelihood can be represented as the product of the likelihood densities of facial components, that is

$$p(\mathbf{Y}_t | \mathbf{X}_t) = \prod_i p(\mathbf{Y}_{it} | \mathbf{X}_{it}). \quad (4)$$

The transition density $p(\mathbf{X}_t | \mathbf{X}_{t-1})$ gives the relationship between the face states of two consecutive time steps, t and $t-1$. Motivated by the idea of using MRF to model the constraints between components within a time step [21, 22], we factorize the transition density as

$$p(\mathbf{X}_t | \mathbf{X}_{t-1}) \propto \prod_i p(\mathbf{X}_{it} | \mathbf{X}_{i(t-1)}) \prod_{ij \in E} \psi(\mathbf{X}_{it}, \mathbf{X}_{jt}) \quad (5)$$

This factorization extracts the interactions between nodes out of the integral over \mathbf{X}_{t-1} in Eq. (3).

4. SMC BASED INFERENCE ALGORITHM

In a SMC method [20], the posterior density $p(\mathbf{X}_{t-1} | \mathbf{Y}^{t-1})$ is approximated by a set of samples \mathbf{X}_{t-1}^k associated with corresponding weights w_{t-1}^k , i.e., $\{\mathbf{X}_{t-1}^k, w_{t-1}^k\}_{k=1}^{N_s}$, where N_s is the number of samples. Then the integral in Eq. (3) is approximated as the summation of the weighted samples:

$$p(\mathbf{X}_t | \mathbf{Y}^t) \approx p(\mathbf{Y}_t | \mathbf{X}_t) \sum_k w_{t-1}^k p(\mathbf{X}_t | \mathbf{X}_{t-1}^k) \quad (6)$$

If we draw samples \mathbf{X}_t^k from the prior transition density $p(\mathbf{X}_t | \mathbf{X}_{t-1}^k)$, that is

$$\mathbf{X}_t^k \sim p(\mathbf{X}_t | \mathbf{X}_{t-1}^k) \quad (7)$$

and then the associated weights w_t^k are updated as

$$w_t^k = p(\mathbf{Y}_t | \mathbf{X}_t^k) w_{t-1}^k \quad (8)$$

With the samples \mathbf{X}_t^k properly weighted by w_t^k , the state up to a time step t can be inferred by a MAP estimate or a mean estimate. Eqs. (7) and (8) give one iteration step of a standard SMC method. We need to devise a novel sample propagating and weight updating algorithm to accommodate the proposed graphical structure.

4.1 Sample propagation

Substituting Eq. (5) into Eq. (6), we obtain:

$$p(\mathbf{X}_t | \mathbf{Y}^t) \approx [p(\mathbf{Y}_t | \mathbf{X}_t) \bullet \prod_{ij \in E} \psi(\mathbf{X}_{it}, \mathbf{X}_{jt})] \bullet [\sum_k \prod_i p(\mathbf{X}_{it} | \mathbf{X}_{i(t-1)}^k) w_{t-1}^k] \quad (9)$$

This means that we can use the samples of $\prod_i p(\mathbf{X}_{it} | \mathbf{X}_{i(t-1)}^k)$ instead of directly sampling from $p(\mathbf{X}_t | \mathbf{X}_{t-1}^k)$ as Eq. (7) to approximate the integral over \mathbf{X}_{t-1} in Eq. (3). We independently sample the transition density of each component, $p(\mathbf{X}_{it} | \mathbf{X}_{i(t-1)}^k)$, in order to form the samples $\{\mathbf{X}_t^k\}_{k=1}^{N_s}$:

$$\mathbf{X}_t^k \sim \prod_i p(\mathbf{X}_{it} | \mathbf{X}_{i(t-1)}^k) \quad (10)$$

For any nodes in the graph shown in Figure 1, the state vector is composed of location and appearance parameters, that is,

$$p(\mathbf{X}_{it} | \mathbf{X}_{i(t-1)}^k) = p(\mathbf{l}_{it}, \mathbf{a}_{it} | \mathbf{l}_{i(t-1)}, \mathbf{a}_{i(t-1)}) \quad (11)$$

Derived from multiplication rule, we rewrite it as

$$p(\mathbf{l}_{it}, \mathbf{a}_{it} | \mathbf{l}_{i(t-1)}, \mathbf{a}_{i(t-1)}) = p(\mathbf{a}_{it} | \mathbf{l}_{i(t-1)}, \mathbf{a}_{i(t-1)}) \bullet p(\mathbf{l}_{it} | \mathbf{l}_{i(t-1)}, \mathbf{a}_{i(t-1)}) \quad (12)$$

We assume that location \mathbf{l}_{it} is determined only by the given location at the previous time step (t-1), and is independent from the appearance at (t-1), $\mathbf{a}_{i(t-1)}$, i.e.,

$$p(\mathbf{l}_{it} | \mathbf{l}_{i(t-1)}, \mathbf{a}_{i(t-1)}) = p(\mathbf{l}_{it} | \mathbf{l}_{i(t-1)}) \quad (13)$$

Thus, we can sample a new particle for location \mathbf{l}_{it}^k from the above dynamics model as typical SMC methods

do. The probability of the appearance at t given $\mathbf{l}_{it}, \mathbf{l}_{i(t-1)}$ and $\mathbf{a}_{i(t-1)}$ is assumed as a Gaussian:

$$p(\mathbf{a}_{it} | \mathbf{l}_{it}, \mathbf{l}_{i(t-1)}, \mathbf{a}_{i(t-1)}) = N(\mathbf{a}_{i(t-1)}, \mathbf{Q}_{i(t-1)}) \quad (14)$$

where $\mathbf{Q}_{i(t-1)}$ is a diagonal matrix with the elements depending on the difference between \mathbf{l}_{it} and $\mathbf{l}_{i(t-1)}$. Large movements are likely to yield great appearance variations, and thus the elements in $\mathbf{Q}_{i(t-1)}$ are rewarded with larger values. Otherwise, smaller variances will be specified. Noticing (14), we use Kalman-like updating equations to propagate appearance coefficients \mathbf{a}_{it} similar to those in [23]. Starting from an initial density $\mathbf{a}_{i0} \sim N(\bar{\mathbf{a}}_{i0}, \mathbf{P}_{i0})$, the probability density of appearance is updated as:

$$\mathbf{P}_{it} = (\Phi_{it}^T \Phi_{it} + (\mathbf{Q}_{i(t-1)} + \mathbf{P}_{i(t-1)})^{-1})^{-1} \quad (15)$$

$$\bar{\mathbf{a}}_{it} = \mathbf{P}_{it} (\Phi_{it}^T \mathbf{Y}_{it} + (\mathbf{Q}_{i(t-1)} + \mathbf{P}_{i(t-1)})^{-1} \bar{\mathbf{a}}_{i(t-1)}) \quad (16)$$

These updating equations circumvent sampling in the higher dimensional appearance subspace (with the dimension larger than 10) so that the proposed method can be implemented in an efficient way. It should be noted that in (9) a state vector is decomposed into location and appearance and the co-inference between location and appearance is proposed in [24]. However, the appearance is integrated out for tracking applications. We employ both location and appearance information to generate HR images.

4.2 Weight updating

Noticing Eq. (9), we can treat the constraint between nodes as an additional term to update the weights,

$$\begin{aligned} w_{it}^k &= (p(\mathbf{Y}_{it} | \mathbf{X}_{it}^k) \prod_{ij \in E} \psi(\mathbf{X}_{it}^k, \mathbf{X}_{jt}^k)) w_{it-1}^k \\ &= \left(\prod_i p(\mathbf{Y}_{it} | \mathbf{X}_{it}^k) \prod_{ij \in E} \psi_l(\mathbf{l}_{it}^k, \mathbf{l}_{jt}^k) \psi_a(\mathbf{a}_{it}^k, \mathbf{a}_{jt}^k) \right) w_{it-1}^k \end{aligned} \quad (17)$$

where $\psi_l(\mathbf{l}_{it}^k, \mathbf{l}_{jt}^k)$ and $\psi_a(\mathbf{a}_{it}^k, \mathbf{a}_{jt}^k)$ are obtained during the training stage as described above. The likelihood $p(\mathbf{Y}_{it} | \mathbf{X}_{it}^k)$ is calculated as:

$$p(\mathbf{Y}_{it} | \mathbf{X}_{it}^k) \propto \exp\left(-\frac{1}{2} \left\| \mathbf{T}_{it}^k - \mathbf{T}_{i0} - \Phi_{it} \mathbf{a}_{it}^k \right\|_{\Sigma}^2\right) \quad (18)$$

where \mathbf{T}_{it}^k is the image obtained by warping the observed LR image with the affine transformation f^k , which is determined by the location samples \mathbf{l}_{it}^k . The distance in Eq. (18) is defined as

$$\left\| \mathbf{x} \right\|_{\Sigma}^2 = \mathbf{x}^T \Sigma^{-1} \mathbf{x} \quad (19)$$

where Σ is a diagonal matrix with the elements as the eigen values of the covariance of each component.

We perform the modified SMC iteration as Eqs. (13), (15), (16) and (17) to obtain the appearance and location parameters via given the LR images \mathbf{Y}^t up to any time step t . We generate patches of the facial components with the estimated appearance parameters and fuse them by using the estimated location parameters to yield HR face image.

5. EXPERIMENTAL RESULTS

We selected 143 frontal neutral face images from AR data set [25] and flipped them horizontally to double the amount of images. All the images were manually annotated 4 points: the centers of the eyeballs, the tip of the nose, and the center of the mouth. Then PCA models of the four components were built from the extracted images of the corresponding components. We used 10 PCA coefficients to represent the appearance of each component. The interaction potentials between the coefficients were obtained by kernel density estimation from the training data. For simplicity, we used three parameters to represent locations, i.e. translations along X and Y axis and a scale parameter, $\mathbf{l} = (t_x, t_y, s)$. These location parameters determine a simplified affine transformation f that maps any point in the image plane $\mathbf{p} = [l_x, l_y]^T$ to a new one $\mathbf{p}' = [l'_x, l'_y]^T$ as:

$$\mathbf{p}' = f(\mathbf{p}) = s \cdot [l_x, l_y]^T + [t_x, t_x]^T \quad (20)$$

In our experiments, the initial location parameter \mathbf{l}_0 was manually specified and the initial location samples were generated by a uniform distribution around the specified \mathbf{l}_0 ; the mean $\bar{\mathbf{a}}_{i0}$ of the initial normal density for appearance updating was obtained as follows: firstly interpolating the first observed LR frame into the desired dimension (resolution), and then projecting the interpolated image into the PCA subspace to initial density mean. The covariance matrix \mathbf{P}_{i0} was specified as the matrix Σ , which was the byproduct of performing PCA on the training examples of each component.

The performance of the proposed approach is examined on synthetic face image sequences. We firstly use the transformations determined by the specified location parameters to warp a HR face image into several LR images, and then add Gaussian white noise with various power dependent on the location parameters to the LR images so as to simulate the fact that large spatial movements are likely to bring great appearance variances.

In order to investigate the performance of the proposed approach on alignment parameter estimation, we compare the estimated location parameters of the proposed approach with those of the widely used LK algorithm. Figure 2 shows the mean estimation errors obtained by the proposed approach and LK algorithm when performing various translations with the scale fixed to 0.5. It can be seen from

Figure 2 that both LK algorithm and the proposed approach estimate translations with high accuracy. However, the estimation error of LK algorithm is relatively higher than that of the proposed approach in the cases of large translation, although the pyramid scheme is adopted in our implementation of LK algorithm. It is reported that LK algorithm is able to gain high estimation accuracy even when large translations take place, but the additive noise varying with movement parameters in the experiments breaks the brightness consistence assumption supposed in LK algorithm. In contrast, our approach can accommodate the appearance variations (see Eqs. (15) and (16)) so that it gains superior performance. The effects of scale variations on estimation accuracy are demonstrated in Figure 3. Since the pyramid scheme effective to cope with the problem of local minima is disabled by large scale variations, LK algorithm does not work well in these cases. However, in our approach, the sampling based inference algorithm that maintains multiple hypotheses is not sensitive to non-maximal modes. Figure 3 shows that the proposed approach gives accurate estimation even when the relative scale variation is as large as 1.8.

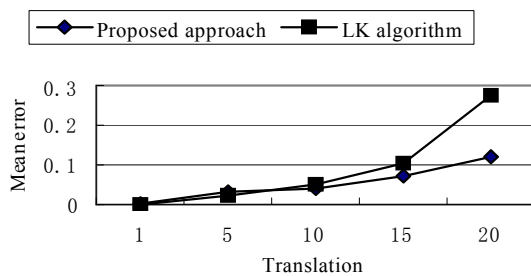


Fig. 2. The mean estimation errors obtained by the proposed approach and LK algorithm

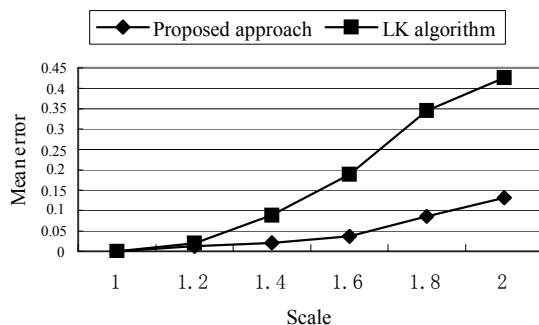


Fig. 3. The effects of scale variations on estimation accuracy

Figure 4 shows the results of reconstructing LR face images down-sampled with the factor 4 and corrupted by the noise varying with location parameters. It can be shown from Figure 4(b) that the HR image obtained by bi-cubic interpolation smears the noise as well as image details. Figure 4(c) shows the reconstructed HR images with the

appearance parameters estimated by the proposed approach. It should be noted that we only concentrate on the four facial components (i.e. eyes, nose, and mouth) that are crucial to determine face expressions and face identities. Thus, we only estimate the appearance parameters of these facial components and reconstruct the regions where these components locate. Though some trivial artifacts present and some facial details with very high frequency (e.g., eyelash) are smeared compared with the original images shown in Figure 4(a), the proposed approach provides reconstruction HR images with superior visual quality. Because we use the PCA based models derived from the statistical characteristics of face images, the resultant images are robust to additive noise. In the experiments, we find that the proposed approach in the case of large translation can work well. It means that the reconstructed appearance is not sensitive to location parameters, especially translations, in contrast to the results of traditional two-step SR algorithms [4]. Figure 4(d) shows that there exist distinct boundaries between facial components and inconsistent overall brightness. This deficiency can be amended by imposing compatible constraints among the overlapped regions of the facial components as what Freeman *et al.* [26] did.

6. CONCLUSION

In this paper, we propose a Bayesian super-resolution approach that integrates image alignment and image fusion into one unified framework. In addition, we build a graphical face model that represents the appearance of individual facial components as well as the interactions between the components. Under the proposed Bayesian framework, the prior information from the constructed face model is incorporated into both alignment and fusion processes of super-resolution, and the higher resolution images are reconstructed via an SMC based inference algorithm. Experimental results show that the proposed approach gains superior performance in the alignment as well as reconstruction. In addition, the proposed approach provides such a new perspective for super-resolution that SR is posed as a problem of parameter estimation. Different face models can be introduced in order to improve the performance further. And it is likely that the proposed approach can be extended to other highly structured objects different from face. The proposed scheme is a primary trying for SR via a perspective different from existing ones. The deficiency, such as existing distinct boundaries, needs to be investigated in the future.

ACKNOWLEDGEMENT

The Matlab code of Lucas-Kanade alignment algorithm is provided by Wang Xinnian, Institute of Signal and Image Processing, Dalian Maritime University.



Fig. 4. The results of reconstructing LR face images.

REFERENCES

- [1] Baker, S. and Kanade T., "Limits on Super-Resolution and How to Break them," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol 29, No 9, Sep. 2002, pp. 1167-1183.
- [2] Sung, C. P., Min K. P., et al., "Super-resolution: A Technical Overview," *IEEE Signal Processing Magazine*, Vol 20, No 5, May 2003, pp. 21~36.
- [3] Torre D. I., F., Yacoob Y., et al. "A probabilistic framework for rigid and non-rigid appearance based tracking and recognition," in *Proc. of International Conference on Face Recognition and Gesture Analysis*, Grenoble, France, March 28-30, 2000, pp. 491-498
- [4] Robinson, D. and Milanfar P., "Fundamental Performance Limits in Image Registration," *IEEE Trans. on Image Processing*, Vol 13, No 6, June, 2004, pp. 1185-1199.
- [5] Lucas, B. D. and Kanade T., "An Iterative Image Registration Technique with an Application to Stereo Vision?" in *Proc. of International Joint Conference on Artificial Intelligence*, Vancouver, BC, Canada, August 1981.

- [6] Baker, S. and Matthews I., "Lucas-Kanade 20 Years On: A Unifying Framework," *International Journal of Computer Vision*, Vol 56, No 3, Mar. 2004, pp. 221-255.
- [7] Isard, M. and Blake A., "Contour tracking by stochastic propagation of conditional density," in *Proc. of European Conference on Computer Vision*, Cambridge, UK, April 15-18, 1996, pp. 343-356
- [8] Black, M. J. and Jepson A. D., "EigenTracking: Robust matching and tracking of articulated objects using a view-based representation," *International Journal of Computer Vision*, Vol 26, No 1, Jan. 1998, pp. 63-84.
- [9] Zhou, S. K., Chellappa R., et al., "Visual Tracking and Recognition Using Appearance-Adaptive Models in Particle Filters," *IEEE Trans. on Image Processing*, Vol 13, No 11, Nov. 2004, pp. 1491-1506.
- [10] Elad, M. and Feuer A., "Restoration of a Single Superresolution Image from Several Blurred, Noisy, and Undersampled Measured Images," *IEEE Trans. on Image Processing*, Vol 6, No 12, Nov. 1997, pp. 1646-1658.
- [11] Schultz, R. R. and Stevenson R. L., "Extraction of high-resolution frames from video sequences," *IEEE Trans. on Image Processing*, Vol 5, No 6, June 1996, pp. 996-1011
- [12] Smelyanskiy, V. N., Cheeseman P., et al., "Bayesian super-resolved surface reconstruction from images," in *Proc. of CVPR'00*, South Carolina, USA, June 13-15, 2000, Vol I, pp. 375-382
- [13] Capel, D. P. and Zisserman A., "Super-resolution from multiple views using learnt image models," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, 2001, vol 1, pp. 627-634
- [14] Liu, C., Shum H. Y., et al., "A two-step approach to hallucinating faces: global parametric model and local nonparametric model," in *Proc. of CVPR'01*, Hawaii, USA, December 8-16, 2001, Vol I, pp. 192-198
- [15] Wang, X. and Tang X., "Hallucinating face by eigentransformation," *IEEE Trans. on Systems, Man and Cybernetics, Part C*, Vol 35, No 3, Aug. 2005, pp. 425-434
- [16] Jordan, M. I., "Graphical models," *Statistical Science*, Vol 19, No 1, Feb. 2004, pp. 140-155
- [17] Moghaddam, B. and Pentland A., "Probabilistic Visual Learning for Object Representation," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol 19, No 7, July 1997, pp. 696-710
- [18] Felzenszwalb, P. F. and Huttenlocher D. P., "Pictorial Structures for Object Recognition," *International Journal of Computer Vision*, Vol 61, No 1, Jan. 2005, pp. 55-79
- [19] Sudderth, E. B., Ihler A. T., et al., "Nonparametric Belief Propagation," in *Proc. of CVPR'03*, Madison, WI, USA, June 16-22 2003, Vol 1, pp. 605-612
- [20] Arulampalam, S., Maskell S., et al., "A Tutorial on Particle Filters for On-line Non-linear/Non-Gaussian Bayesian Tracking," *IEEE Trans on Signal Processing*, Vol 50, No 2, Feb. 2002, pp. 174-188
- [21] Chang, C., Ansari R., et al., "Cyclic articulated human motion tracking by sequential ancestral simulation," in *Proc. of CVPR2004*, Washington, DC, USA, June 27- July 2, 2004, pp. 45-52
- [22] Khan, Z., Balch T., et al., "MCMC-Based Particle Filtering for Tracking a Variable Number of Interacting Targets," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, Vol 27, No 11, Nov. 2005, pp. 1805-1918.
- [23] Khan, Z., Balch T., et al., "A Rao-Blackwellized Particle Filter for Eigen Tracking," in *Proc. of CVPR2004*, Washington, DC, USA, June 27- July 2, 2004, pp. 980-986
- [24] Wu, Y. and Huang T. S., "Robust Visual Tracking by Integrating Multiple Cues Based on Co-Inference Learning," *International Journal of Computer Vision*, Vol 58, No 1, Jan. 2004, pp. 55-71
- [25] Martinez, A. M. and Benavente R., "The AR face database," *CVC Technical Report #24*. 1998
- [26] Freeman W.T., Pasztor E.C., and Carmichael O.T., "Learning Low-Level Vision," *International Journal of Computer Vision*, Vol 20, No 1, Jan. 2000, pp. :25-47



Hua Huang received the B.S. and M.S. degree in Information & Communication Engineering in 1996 and 2001 from Xi'an Jiaotong University, China.

He has been a Ph.D. candidate and a lecturer of Xi'an Jiaotong University since 2002. His research interests are image processing and reconstruction.

Xin Fan received B.S. and Ph.D. degree in information and communication engineering in 1998 and 2005 from Xi'an Jiaotong University, China.

He is a lecturer of the Institute of Signal and Image Processing, Dalian Maritime University, China. His research interests include computer vision, face image processing and recognition.

Chun Qi received B.S. degree in electronic engineering in 1982 from Xidian University, China, and Ph.D. degree in information and communication engineering in 2000 from Xi'an Jiaotong University, China.

He is a Professor of Xi'an Jiaotong University. His main research interests include image processing, pattern recognition and signal processing.

Shihua Zhu received B.S. degree in radio engineering in 1982 from Xi'an Jiaotong University, China, and Ph.D. degrees in electronic systems engineering in 1987 from the University of Essex, United Kingdom.

He is a Professor of Xi'an Jiaotong University. His main research interests are mobile communication.