A 3-D assisted generative model for facial texture super-resolution

Pouria Mortazavian*, Josef Kittler*, and William Christmas*

Abstract— **This paper describes an example-based Bayesian method for 3D-assisted pose-independent facial texture super-resolution. The method utilizes a 3D morphable model to map facial texture from a 2D face image to a pose- and shape-normalized texture map and vice versa. The center piece of this method is a generative model to describe the process of forming an image from a pose- and shape-normalized texture map. The goal is to reconstruct a high-resolution texture map given an low-resolution face image. The prior knowledge about the sought high-resolution texture is incorporated into the Bayesian framework by using a recognition-based prior that encourages the gradient values of the texture map to be close to some predicted values.**

We develop the generative model and formulate the problem as MAP estimation. The results show that this framework is capable of performing pose-independent face recognition even when the sample set only contains exemplar face images with frontal pose. We present results in frontal and non-frontal poses. We also demonstrate that the technique can be utilized to improve face recognition results when the probe images have a lower resolution compared to the gallery images.

I. INTRODUCTION

Image super-resolution is the process of constructing a high-resolution image given one or more lower-resolution images of the same scene or object. Many methods have been proposed in the literature which can be largely classified as reconstruction-based (e.g. [1], [2], [3], [4]) or example-based (e.g. [5], [6], [7], [8], [9], [10], [11], [12]) methods. The reconstruction-based methods try to fuse information from multiple low-resolution (LR) observations of the scene to build a higher-resolution image. The main source of information for such methods in order to build the sought high-resolution (HR) image is low-resolution observations and prior knowledge about the image formation process. The information about the image formation process is generally incorporated into a generative model which models the process of generating the low-resolution observations given the sought HR image. These methods usually try to build an HR image that can satisfy the super-resolution *reconstruction constraint* that the HR image, if appropriately warped and down-sampled to simulate the image formation process, should yield the low-resolution observations. This constraint was first used by Peleg *et al.*[1]. They assumed that the imaging process is known and that the low-resolution images have been registered with sub-pixel accuracy. Starting from an initial guess for the unknown HR image, the imaging process is simulated to generate a set of simulated LR images resembling the LR inputs. An error was then defined between the original and simulated LR images. By minimizing this error with respect to HR pixels iteratively, the final HR image is obtained. Irani and Peleg [2] presented an iterative back-projection (IBP) super-resolution approach in which the error between the observed LR images and the LR images estimated by the generative model is back-projected to update the estimate for the HR image in each iteration.

Tom and Katsaggelos [13] used a *maximum Likelihood* (ML) estimation approach for simultaneous registration and super-resolution. Although the ML method improves the input images, for large magnification factors the estimation becomes highly ill-conditioned and the solution is highly sensitive to noise in the input observations, the parameters of the generative model, and to registration error. Attempts to overcome this problem include placing hard constraints on the individual pixel intensities or the use of a Bayesian prior model of the super-resolved image by modelling the image as a first-order stationary Markov Random Field (MRF) and including a spatial prior to model the spatial dependencies of the neighboring pixels. The latter would result in a *maximum a posteriori* (MAP) estimation. Such priors are typically either chosen to impose some kind of smoothness on the image (e.g. Gaussian MRF) or have some edge preserving characteristics (e.g. Huber-MRF). Schultz and Stevenson [4] proposed a discontinuity preserving MAP reconstruction method using the Huber-Markov Gibbs prior for super-resolving low-resolution video sequences. A joint MAP estimation for simultaneous registration and super-resolution was proposed by Hardie *et al.* [14] which uses a Gaussian prior.

Baker and Kanade [15] showed that reconstruction constraints provide far less useful information as the magnification factor increases. They showed that for large enough magnification factors, any smoothness prior leads to overly smooth results with very little amount of high-frequency information regardless of the number of LR observations used. This means that there is effectively a theoretical limit to how well reconstruction-based methods can perform. As a remedy, example-based methods use additional sources of information. In such methods, a set of exemplar images are used to model or learn the relationship between high- and low-resolution image pairs.

Freeman *et al.* [5], [6] proposed a patch-based approach that used a set of training images to learn the relationship between a sharp image and its low-resolution counterpart. They used a Markov network to probabilistically model the relationship between the HR and LR patches. In their network, each HR patch is connected to an LR patch and its neighboring HR patches. By blurring and downsampling a

*The authors are with the Centre for Vision, Speech and Signal Processing, University of Surrey, United Kingdom. ({p.mortazavian, j.kittler, w.christmas} @surrey.ac.uk)

978-1-4244-XXXX-X/09/$25.00 ©2009 Crown

set of training HR images, the authors generated a training set of sharp and blurred image pairs which is used to learn the compatibility functions of this network.

In applications where the problem can be limited to a specific object of interest (e.g. human face) building the sample set with images of the same object can result in obtaining much more plausible information for example-based super-resolution. Many methods have been specifically tailored for face super-resolution ([7], [8], [9], [10], [11], [12]).

Wang and Tang [10] proposed an eigentransformation approach in which the input image is represented as a linear combination of the training LR images using a PCA transform. The HR image is generated by replacing the LR images with their HR counterparts while keeping the mixture coefficients. Liu *et al.*[7] proposed a two-step method for face super-resolution. This method is based on the fact that the HR image is a combination of common (global) face properties and individual (local) characteristics. The global and individual properties are captured by a global parametric model (H^g) and a local non-parametric model (H^l) respectively. The final HR image (H) is assumed to be a combination of these two models: $H = H^g + H^l$. Li and Lin [9] have further improved this approach by applying PCA to both LR and HR images in the first step in order to incorporate the estimation of noise model and also using a MAP framework instead of the MRF model for finding the optimal local face.

In [16], Capel and Zisserman presented another example-based, domain-specific method for faces. Each face image is broken down into 6 regions and each region is represented with PCA. The PCA representation is then used to constrain the results in an ML estimation framework and also to define appropriate prior terms in a MAP estimation framework.

A very well-known approach for object-specific super-resolution is *Face Hallucination* (also named as the *recogstruction* method) proposed by Baker and Kanade [17], [18], [8][1]. This approach is applicable to one or more input images. In the case of multiple input images, the geometric registration transformation is assumed to be only translational and the exact translations are assumed to be known *a priori*. This approach uses a MAP framework with a **recognition-based prior** which is defined on the gradients of the HR image using "recognition" results of some generic local "features". Recognition - in this context - means finding the most "similar" pixel from the training set for each pixel of the LR input image(s). The prior term is then defined such that it encourages the gradient of the super-resolution image to be close to the closest matching training samples. The Name *recogstruction* for this method is a combination of the terms *recognition* and *reconstruction* which suggest the two main sources of information for the method: the reconstruction constraints and the recognition-based prior term.

[1]Although the term *Face Hallucination* was first coined for this specific method, it is also used in a more general sense in the literature to refer to (example-based) face super-resolution

The above face super-resolution methods generate an HR image of a single facial modality (i.e. at a fixed expression, pose and illumination). In [12] Jia and Gong have presented a *generalized* face super-resolution method which aims at multi-modal super-resolution of faces. In other words, given an LR image at a specific modality, it aims at generating HR images of the same and other different modalities. A tensor structure is used to incorporate information and interactions of (training) images of multiple modalities at different resolutions. The task is then formulated as MAP estimation of the high-resolution images given the LR observation(s) with the inter-modal interactions inferred using multi-linear analysis. This technique is able to generate HR images in a novel pose. However, it requires sample training images of multiple modalities in order to perform multi-modal super-resolution thus it is limited to poses available in the sample set. In [19] the authors have presented a 3-D assisted facial texture super-resolution framework in which a 3D morphable model is used in order to perform pose-independent super-resolution in a single frame. The problem is posed as MAP estimation of the morphable model parameters and the HR texture given the LR observation. Using two assumptions about the model parameters and the HR texture, the problem is then broken down into two steps: 1) fitting the 3D model on the LR observation, 2) super-resolving the facial texture in a texture super-resolution framework similar to the *recogstruction* ([17], [18], [8]) framework. In the framework proposed by [19], it is assumed that the HR texture can be fully recovered from LR texture extracted from the LR observation and a generative model is utilized that generates the LR *texture* given the HR *texture*. The likelihood term in the MAP framework is defined using this generative model while the prior term is defined in a process similar to that of [17], [18], [8]. Using this framework, once the HR texture is constructed, it can be used to generate HR images of the face by utilizing the parameters of the fitted 3D morphable model.

In this work we follow the general approach of [19] but instead of making the above assumption, we propose a more suitable generative model to be used in this framework. Our proposed generative model describes the formation process of the LR *image* given the HR *texture*. In section II we will describe the generative model used in our framework, and in section III we formulate the 3D-assisted super-resolution in a MAP framework. Some experimental results are presented in section IV and section V concludes the paper.

II. Generative Model

We describe the image formation process with a generative model. The model will describe the process through which an LR image is formed using a 3D morphable model to map HR texture from a *texture* plane to the *image* plane. A 3D morphable model is a vector space representation of 3D faces. We used a morphable model built by Tena [20]. The model is built using 69 3D face scans. Each raw scan comprises a 3D mesh and a 2D texture map. The (x, y, z) coordinates of the vertices of the 3D mesh of each face scan and the corresponding texture values are concatenated into a *shape* (S) and a *texture* (R) vector respectively. Assuming these vectors are in full dense correspondence for all face scans, PCA is used to compress the data:

$$S_{mod} = \bar{S} + \sum_i \alpha_i s_i \quad , \quad R_{mod} = \bar{R} + \sum_i \beta_i r_i \tag{1}$$

where $\bar{S}$ and $\bar{R}$ are the mean shape and texture vectors respectively and s_i and r_i are the i^{th} eigenvectors of the shape and texture covariance matrices respectively. Also, α_i and β_i are the mixture coefficients known as the model shape and texture parameters respectively.

Given a single 2D face image as input and a set of landmarks, the parameters of the 3D morphable model can be estimated such that they represent the 3D shape and texture of the input face. This process is called model fitting which estimates the model parameters together with a set of rendering parameters -including lighting parameters, 3D translation, pose angles, camera focal length etc.- such that rendering the model with the estimated parameters will produce an image which resembles the input face image. The model obtained by this process represents the 3D shape and the texture of the input 2D face image. Model fitting can be formulated as a MAP estimation of the model and rendering parameters given the input face image and the landmarks. Assuming independence between some parameters:

$$\begin{aligned} \alpha^*, \beta^*, \rho^* &= \underset{\alpha,\beta,\rho}{\operatorname{argmax}}\, P(\alpha,\beta,\rho|f,L) \\ &= \underset{\alpha,\beta,\rho}{\operatorname{argmax}}\, P(f|\alpha,\beta,\rho)P(L|\alpha,\beta,\rho)P(\alpha)P(\beta)P(\rho) \end{aligned} \tag{2}$$

where f is the input face image, L is a set of landmarks marked on f, α and β are the model shape and texture parameters respectively, and ρ is the set of rendering parameters.

In order to use the real facial texture from the input image (as opposed to the texture *estimated* by model fitting), the shape and rendering parameters estimated during the model fitting can be used to extract the facial texture from the input image (where available) and map it to a predefined texture coordinate frame. The coordinate frame for such a texture map is defined by a 2D representation of the 3D surface of the face model which preserves the geodesic distance between the vertices of the 3D surface. Such a representation is obtained using the *isomap* algorithm proposed by Tenenbaum *et al.*[21]. The texture value for any texel[2] on this texture map is then extracted from the image by projecting each texel center to the image plane and finding the texture value of the projected point. For texels that are not visible in the input image, the texture value estimated by model fitting will be used.

In order to project any point from the texture map to the image plane, the equivalent point on the 3D surface is found and projected to the image though a pinhole camera model.

[2]We use the term *texel* to refer to the pixels of a 2D *texture map* in order to avoid confusion with pixels of an *image*

The texture map extracted from a 2D face image in such a manner represents the shape- and pose-normalized texture of the face. Hence, in order to generate the image from such a texture map, the image formation process includes projecting all texels to the image plane according to the shape and pose data, and integrating the texture values within each pixel of the image plane to obtain the final pixel value. In order to project a texel to the image plane, we project its 4 corners and obtain a quadrangle on this plane. We define each pixel (x,y) of the image as "*affected* by texel (p,q)" if the area of the overlap between texel (p,q) when projected to the image plane, and pixel (x,y) is not zero. Each pixel value is defined in our generative model as the weighted sum of the texel values which affect it. The weights define the contribution of each texel to the pixel value:

$$f(x,y) = \sum_{p,q} \frac{W(x,y,p,q,\alpha,\rho)}{\sum_{p,q} W(x,y,p,q,\alpha,\rho)} T(p,q) + \eta \tag{3}$$

where (p,q) indexes the texels of the texture map T, (x,y) indexes the pixels of the image f, W is the weight that determines the contribution of texel $T(p,q)$ to the value of pixel $f(x,y)$, and η is an additive pixel noise term. The weight $W(x,y,p,q,\alpha,\rho)$ is defined as the fraction of the area of the projected texel that lies within pixel (x,y) on the image plane.

Note that projecting any point of the texture map to the image plane requires information about the 3D shape of the model in order to find the equivalent point on the 3D surface as well as pose and rendering parameters in order to project the point from the 3D surface to the image plane. Hence, W in (3) is also a function of α and ρ.

III. A Bayesian Formulation

In this section we formulate the 3D-assisted texture super-resolution as a MAP estimation problem. Given an LR image f, the sought HR texture map is:

$$\begin{aligned} T^* &= \underset{T}{\operatorname{argmax}} \sum_{\mu,\rho} P(T,\mu,\rho|f) \\ &= \underset{T}{\operatorname{argmax}} \sum_{\mu,\rho} \{P(T|\mu,\rho,f)P(\mu,\rho|f)\} \end{aligned} \tag{4}$$

where T is an HR texture map, and μ is the set of model parameters ($\mu = \{\alpha, \beta\}$). As in [19], we assume that $P(\mu,\rho|f)$ peaks at the optimum values of μ and ρ, and has a dense distribution around these values:

$$P(\mu,\rho|f) \backsimeq \delta(\mu-\mu^*)\delta(\rho-\rho^*) \tag{5}$$

where δ is the dirac function and μ^* and ρ^* are the optimum model and rendering parameters respectively. The assumption implies that the fitting process is ideal. These values can be found by the model fitting process described in II. Instead of fitting the model on the LR face image, we use bilinear interpolation to enlarge the input image prior to fitting. Given the above assumption and the optimized parameters from the fitting process, 4 simplifies to:

$$T^* = \underset{T}{\operatorname{argmax}}\, P(T|\mu^*, \rho^*, f) = \underset{T}{\operatorname{argmax}}\, P(f|T, \mu^*, \rho^*)P(T)$$
$$= \underset{T}{\operatorname{argmin}}\, \{-\log P(f|T, \mu^*, \rho^*) - \log P(T)\} \quad (6)$$

Using the generative model in (3) and assuming that the pixel noise η has an *i.i.d.* Gaussian distribution with covariance σ_η^2, the negative log-likelihood term in 6 can be defined as:

$$-\log P(f|\mu^*, \rho^*, T) =$$
$$\frac{1}{2\sigma_\eta^2} \sum_{x,y} \left(f(x,y) - \sum_{p,q} \frac{W(x,y,p,q,\mu^*,\rho^*)}{\sum_{p,q} W(x,y,p,q,\mu^*,\rho^*)} T(p,q) \right)^2 \quad (7)$$

A. The recognition-based prior

In order to define the prior term, $P(T)$, in (6) we follow an approach similar to that of [17]. First we need to build a texture sample set. We do this by fitting the model to a set of exemplar HR face images and extracting texture from them. The extracted texture maps are considered as level zero of a *texture pyramid*. Now let f be an LR face image which is 2^l times smaller than the HR images used for the sample set, and t be the LR texture map extracted from f. The amount of information available in t is equivalent to the l^{th} level of the texture pyramid. Hence, for each texel (p,q) of t we find the most similar texel in the same location from the sample set by comparing the l^{th} level parent structures ([22]) of the sample texels with the l^{th} level parent structure of $t(p,q)$. The horizontal and vertical gradient values of this "most similar" sample texel are then taken as the predicted values for the horizontal and vertical gradient values of $t(p,q)$ respectively.

Having defined predictions for the gradient values of all texels of t, the prior term is defined by assuming *i.i.d.* Gaussian error with covariance σ_∇^2 for the gradient prediction process:

$$-\log P(T) = \frac{1}{2\sigma_\nabla^2} \sum_{p,q} (H_{pr}(p,q) - H(p,q))^2 +$$
$$\frac{1}{2\sigma_\nabla^2} \sum_{p,q} (V_{pr}(p,q) - V(p,q))^2 \quad (8)$$

where $H(p,q)$ and $V(p,q)$ are the horizontal and vertical gradient values at texel (p,q), and $H_{pr}(p,q)$ and $V_{pr}(p,q)$ are their predicted values respectively. The gradient values in (8) can be written as linear expressions in the unknowns $T(p,q)$. By replacing (7) and (8) in (6), the final cost function is derived. We minimize this cost function with respect to the HR texel values, $T(p,q)$, to obtain the sought HR texture map (T^*).

IV. Experimental Results

In this section we will present some experimental results of our proposed approach and compare the results with Baker and Kanade's benchmark face hallucination algorithm [8]. These experiments were performed using the XM2VTS database [23]. The sample set for the super-resolution algorithm was built by extracting texture from 8 frontal shots of 95 subjects which were chosen randomly as the training subjects. The test subjects were chosen from the remaining subjects in the database. The LR data was generated by fitting the 3D morphable model on HR shots of the test subjects and rendering the model with the optimized parameters and extracted texture while setting the focal length of the virtual camera to 8 times smaller than what was estimated for the HR samples. In all cases the magnification ratio for the super-resolution algorithm is 8.

A. Frontal face super-resolution

Figure IV-A shows the super-resolution results for some experiments with a frontal face image as the input. From left to right, the first column (a) is the LR image, the second column (b) is the result of using bilinear interpolation to increase the size of the image 8 times. The third column (c) shows the results obtained with Baker and Kanade's face hallucination. The fourth column (d) of figure IV-A shows the results obtained with our algorithm while the fifth column (e) is the HR face. As seen in the figure IV-A the results of our approach are comparable, and in some cases visually more pleasing than the other approach and in most cases contain less artifacts.

The two right-most columns of figure IV-A illustrate two novel views of the same face in high resolution. Using the super-resolved texture map we can render novel views of the HR face. Note that in some areas which were not visible in the input frontal image, the texture estimated in the fitting process is used which can be rather inaccurate in some cases.

B. Non-frontal face super-resolution

A major advantage of our approach is that it does not assume any pose for the input face. It can be applied to LR face images of any pose without the need to extend the sample set. Figure IV-B illustrates the results of super-resolving facial texture when the input is a side view LR image. The super-resolved texture is rendered to produce the same pose as the input and super-imposed on a bilinearly interpolated version of the input image.

Note that the results in Figure IV-B were generated using a sample set that only consisted of frontal face images. This is to show that even without using other poses our method manages to do an acceptable job in super-resolving different poses. However, the performance can be further enhanced by adding samples of other poses to the sample set.

C. Face Recognition

An important challenge in face recognition is the resolution of the input face. It has been shown (e.g. in [24]) that the performance of a face recognition algorithm can drop dramatically if the resolution of the probe images drops below a certain level. In such cases, super-resolution techniques can be utilized to improve the resolution of probe images as a pre-processing step in order to improve the recognition results. In [25], Yeomans *et al.*proposed a framework for simultaneous super-resolution and recognition which generally yields better results than performing super-resolution as pre-prosessing. However, in this paper we resort

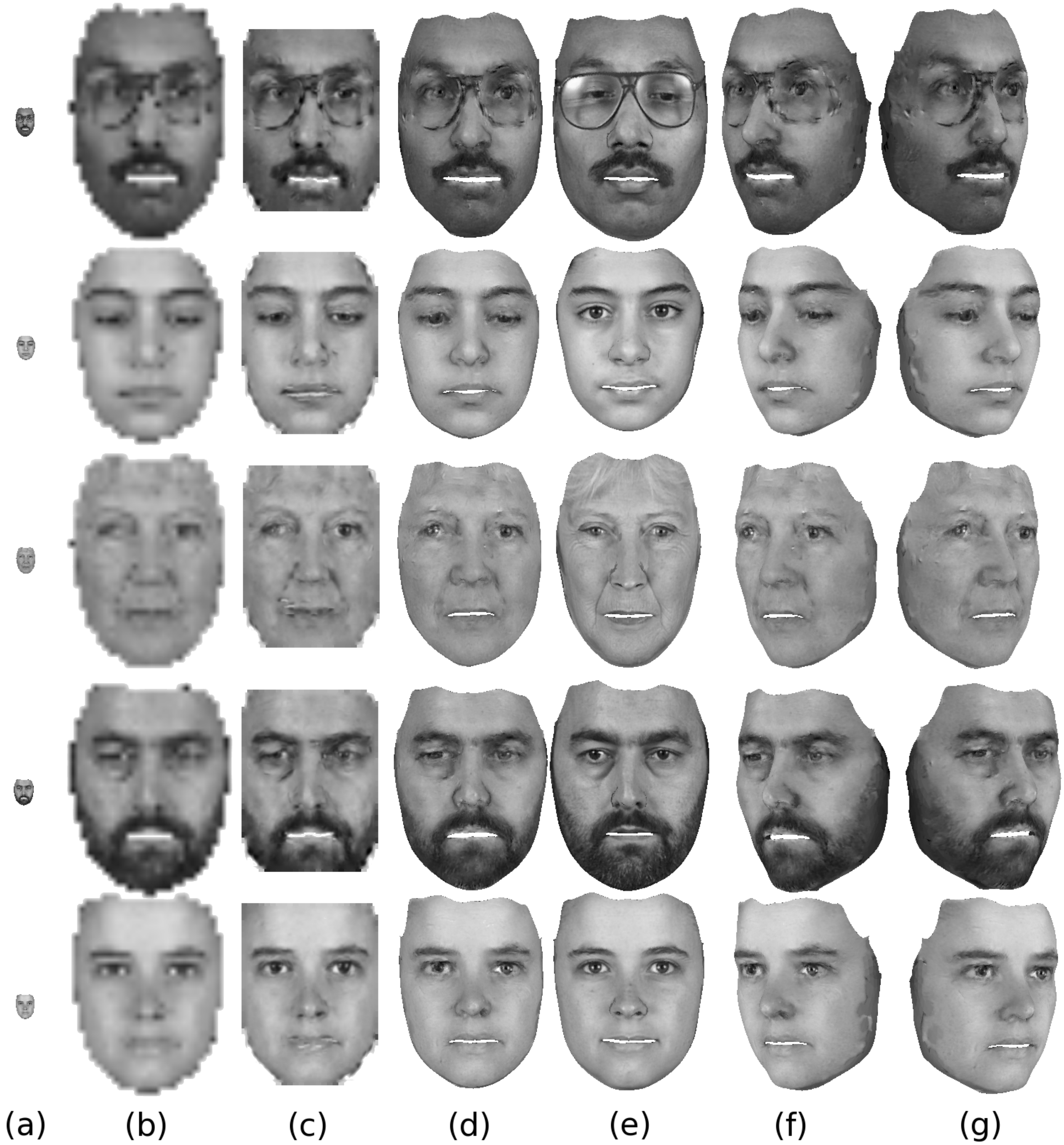

Fig. 1. (a) Examples of the low-resolution images. (b) Bilinear interpolation. (c) Frontal face super-resolution of the faces in 'a', using Baker and Kanade's framework. (d) Frontal face super-resolution using our method. (e)Original high-resolution images. (f,g) Generated super-resolved imaged of novel poses using our method.

Fig. 2. Examples of non-frontal face super-resolution. Top row: (bilinear interpolation of) low-resolution images. Bottom row: Super-resolution results using our method.

to the traditional approach since our goal is to evaluate the performance of our super-resolution framework as opposed to focusing on improvement of the recognition rate.

We use a face identification algorithm similar to the one proposed by Shan *et al.* [26]. An ensemble of classifiers is built by dividing the image into 100 regions. Spatial histograms of Local Binary Patterns (LBP) are used in each region as feature vectors which are then projected into LDA space. A sum rule is then used to combine the similarity scores (normalized correlation) from the 100 classifiers. Nearest Neighbor(NN) is then employed for final classification. 140 subjects were chosen from the XM2VTS database (not including those already used as super-resolution sample set). The HR data for training the LDA transform was generated by fitting the morphable model to 3 training shots of each subject and rendering the model with focal length value of 1500. The test data was generated by fitting the model to another shot of the same subject and rendering it with focal length 1500 for HR test data and focal length 187.5 for LR test data.

Table I shows the identification rates. It is obvious from table I that the high recognition rate of HR probe images drops dramatically when the resolution of probe images is decreased by a factor of 8 and the images are re-sized to the original size by only bilinear interpolation (LR+Bilinear). This rate can be improved significantly by using super-resolution as a pre-processing step. The identification rate obtained with our approach (LR+Our Method) is marginally less than that obtained with Baker and Kanade's face hallucination technique (LR+Baker_Kande). However, our approach has the potential capacity to super-resolve faces of arbitrary pose. In any case, we believe that the lower performance is caused by errors in fitting the model to LR images which our current work is attempting to rectify. Nevertheless, the fact that the identification rate is significantly increased with both super-resolution methods shows that the additional information injected is beneficial for recognition.

TABLE I

FACE IDENTIFICATION RESULTS

Method	Identification Rate
HR	99.28
LR+Bilinear	78.57
LR+Baker_Kande	96.43
LR+Our Method	95

V. CONCLUSIONS

In this paper we proposed a 3D-assisted generative model for super-resolving pose- and shape-normalized facial texture. A 3D morphable model is utilized to map facial texture from a 2D image to a pose- and shape-normalized texture map which is then super-resolved in an example-based framework using our generative model. The super-resolution problem is formulated in a maximum aposteriori estimation framework and the prior knowledge about the process of image formation from the texture map is incorporated into the framework using our generative model. A recognition-based prior term is also used to incorporate prior information about the HR texture map. The results show that our proposed framework can be utilized for pose-independent facial texture super-resolution with visually acceptable results and that it can also be used to improve face recognition results when the probe image has a low resolution.

REFERENCES

[1] S. Peleg, D. Keren, and L. Schweitzer. Improving image resolution using subpixel motion. *Pattern Recognition Letters*, 5(3):223–226, 1987.

[2] Michal Irani and Shmuel Peleg. Improving resolution by image registration. *CVGIP: Graph. Models Image Process.*, 53(3):231–239, 1991.

[3] Hanoch Ur and Daniel Gross. Improved resolution from subpixel shifted pictures. *CVGIP: Graph. Models Image Process.*, 54(2):181–186, 1992.

[4] Richard R. Schultz and Robert L. Stevenson. Extraction of high-resolution frames from video sequences. *IEEE Transactions on Image Processing*, 5:996–1011, 1996.

[5] William T. Freeman, Thouis R. Jones, and Egon C Pasztor. Example-based super-resolution. *IEEE Computer Graphics and Applications*, 22(2):56–65, 2002.

[6] William T. Freeman, Egon C. Pasztor, and Owen T. Carmichael Y. Learning low-level vision. *International Journal of Computer Vision*, 40:2000, 2000.

[7] Ce Liu, Heung-Yeung Shum, and Chang-Shui Zhang. A two-step approach to hallucinating faces: Global parametric model and local nonparametric model. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 1:192, 2001.

[8] Simon Baker and Takeo Kanade. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(1):1167 – 1183, September 2002.

[9] Yang Li and Xueyin Lin. An improved two-step approach to hallucinating faces. *Image and Graphics, International Conference on*, 0:298–301, 2004.

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

[11] Jiangang Yu, Bir Bhanu, Yilei Xu, and Amit K. Roy Chowdhury. Super-resolved facial texture under changing pose and illumination. In *ICIP (3)*, pages 553–556. IEEE, 2007.

[12] Kui Jia and Shaogang Gong. Generalized face super-resolution. *IEEE Transactions on Image Processing*, 17(6):873–886, 2008.

[13] B. C. Tom and A. K. Katsaggelos. Reconstruction of a high-resolution image by simultaneous registration, restoration, and interpolation of low-resolution images. In *ICIP '95: Proceedings of the 1995 International Conference on Image Processing (Vol.2)-Volume 2*, page 2539, Washington, DC, USA, 1995. IEEE Computer Society.

[14] Russell C. Hardie, Kenneth J. Barnard, and Ernest E. Armstrong. Joint map registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Transactions on Image Processing*, 6:1621–1633, 1997.

[15] Simon Baker and Takeo Kanade. Super-resolution: Limits and beyond. In S. Chaudhuri, editor, *Super-Resolution Imaging*. Kluwer Academic Press, 2001.

[16] David Capel and Andrew Zisserman. Super-resolution from multiple views using learnt image models. *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, 2:627, 2001.

[17] Simon Baker and Takeo Kanade. Hallucinating faces. Technical Report CMU-RI-TR-99-32, Robotics Institute, Pittsburgh, PA, September 1999.

[18] Simon Baker and Takeo Kanade. Limits on super-resolution and how to break them. In *Proceedings of the 2000 IEEE Conference on Computer Vision and Pattern Recognition*, June 2000.

[19] Josef Kittler Pouria Mortazavian and William Christmas. 3d-assisted facial texture super-resolution. *To appear in The British Machine Vision Conference proceedings*, 2009.

[20] J.R. Tena Rodriguez. *3D Face Modelling for 2D+3D Face Recognition*. PhD thesis, Centre for Vision, Speech, and Signal Processing, University of Surrey, November 2007.

[21] J. B. Tenenbaum, V. de Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, December 2000.

[22] Jeremy S. De Bonet and Paul Viola. A non-parametric multi-scale statistical model for natural images. In *NIPS '97: Proceedings of the 1997 conference on Advances in neural information processing systems 10*, pages 773–779, Cambridge, MA, USA, 1998. MIT Press.

[23] K Messer, J Matas, J Kittler, J Luettin, and G Maitre. Xm2vtsdb: The extended m2vts database. In *Second International Conference on Audio and Video-based Biometric Person Authentication*, March 1999.

[24] Jingdong Wang, Changshui Zhang, and Heung yeung Shum. Face image resolution versus face recognition performance based on two global methods. In *Proceedings of Asia Conference on Computer Vision (ACCV04*, 2004.

[25] Pablo H. Hennings-Yeomans, Simon Baker, and B. V. K. Vijaya Kumar. Simultaneous super-resolution and feature extraction for recognition of low-resolution faces. In *CVPR*, 2008.

[26] Shiguang Shan, Wenchao Zhang, Yu Su, Xilin Chen, and Wen Gao. Ensemble of piecewise fda based on spatial histograms of local (gabor) binary patterns for face recognition. *Pattern Recognition, International Conference on*, 4:606–609, 2006.