[3] C. L. Yang, L. M. Po, and W. H. Lam, "A fast H.264 intra prediction algorithm using macroblock properties," in *Proc. IEEE ICIP'04*, Oct. 2004, vol. 1, pp. 461–464.

[4] F. Pan, X. Lin, S. Rahardja, and K. P. Lim *et al.*, "Fast mode decision algorithm for intraprediction in H.264/AVC video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 7, pp. 813–822, Jul. 2005.

[5] P. Yin, H. C. Tourapis, A. M. Tourapis, and J. Boyce, "Fast mode decision and motion estimation for JVT/H.264," in *Proc. IEEE ICIP'03*, Sept. 2003, vol. 3, pp. 853–856.

[6] X. Jing and L. P. Chau, "Fast approach for H.264 inter mode decision," *Electron. Lett.*, vol. 40, no. 17, pp. 1051–1052, Aug. 2004.

[7] D. Wu, F. Pan, K. P. Lim, and S. Wu *et al.*, "Fast intermode decision in H.264/AVC video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 7, pp. 953–958, Jul. 2005.

[8] L. Yang, K. Yu, J. Li, and S. Li, "An effective variable block-size early termination algorithm for H.264 video coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 6, pp. 784–788, Jun. 2005.

[9] H.264/AVC Reference Software JM9.5 [Online]. Available: http://iphome.hhi.de/suehring/tml/

[10] Y. H. Moon, G. Y. Kim, and J. H. Kim, "An improved early detection algorithm for all-zero blocks in H.264 video encoding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 15, no. 8, pp. 1053–1057, Aug. 2005.

[11] H. S. Malvar, A. Hallapuro, M. Karczewicz, and L. Kerofsky, "Low-complexity transform and quantization in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 598–603, Jul. 2003.

Super-Resolution of Face Images Using Kernel PCA-Based Prior

Ayan Chakrabarti, A. N. Rajagopalan, and
Rama Chellappa, *Fellow, IEEE*

Abstract—We present a learning-based method to super-resolve face images using a kernel principal component analysis-based prior model. A prior probability is formulated based on the energy lying outside the span of principal components identified in a higher-dimensional feature space. This is used to regularize the reconstruction of the high-resolution image. We demonstrate with experiments that including higher-order correlations results in significant improvements.

Index Terms—Face, higher-order statistics, kernel PCA, principal component analysis (PCA), super-resolution.

I. INTRODUCTION

Super-resolution is becoming increasingly important for many multimedia applications [1]. It refers to the process of reconstructing a high-resolution image from low-resolution frames. Most methods [2]–[6] assume knowledge of the geometric warp of each observation and the nature of the blur. However, the effectiveness of such reconstruction-based super-resolution algorithms, that do not incorporate any specific prior information about the image being super-resolved,

has been shown to be inherently limited [7], [8]. A learning-based method has been suggested in [7] to super-resolve face images. It uses a prior based on the error between the gradient values of the corresponding high-resolution pixel in the training image and in the estimated image. But this renders it sensitive to image alignment, scale, and noise. Gunturk *et al.* [9] perform super-resolution in the eigenface space. Since their goal is face recognition, they reconstruct only the weights along the principal components instead of attempting to produce a high-resolution estimate that is visually superior. In [10], a method is presented which super-resolves faces by first finding the best fit to the observations in the eigenface domain. A patch-based Markov network is then used to add residual high-frequency content. Some more learning-based approaches are discussed in [11]–[13].

In this paper, we introduce a learning-based method for super-resolution of faces that uses kernel principal component analysis (PCA) for deriving prior knowledge about the face class. Kernel PCA is a nonlinear extension of traditional PCA for capturing higher-order correlations in a data set. Using kernel PCA, we extract valuable prior information in a computationally efficient manner and show how it can be used within a maximum *a posteriori* (MAP) framework along with the observation model for improving the quality of the super-resolved face image. Experimental results on still images and real video sequences are given to demonstrate the effectiveness of using higher-order statistics for super-resolution. The proposed method outperforms traditional reconstruction-based algorithms and is more robust to errors in motion and blur parameters. We also show comparisons with the learning-based method described in [7] and demonstrate improved performance when the magnification is high.

It may be mentioned that in [14], an application of the *Hebbian* algorithm is described where kernel PCA is used for image zooming by projecting an interpolated version of the low-resolution image on to the high-resolution principal subspace. The method is, however, limited to using a single image and does not incorporate any knowledge of the imaging process.

The remainder of the paper is organized as follows: In Section II, we discuss PCA-based super-resolution within the MAP framework. Construction of a prior based on kernel PCA is discussed in Section III. Experimental results are given in Section IV while Section V concludes the paper.

II. MAP FRAMEWORK

Each low-resolution image is obtained by applying a geometric warp, blurring and down-sampling to the original high-resolution image, followed by addition of noise. Since all of these are linear operations, for each of the r low-resolution images, we have in matrix-vector form

$$Y_i = DB_iG_iX + N_i, \qquad i = 1, 2, \ldots, r \qquad (1)$$

where X is the high-resolution image, Y_i is the ith low-resolution image, N_i is the noise in the ith observation, D is the decimation matrix, while B_i and G_i are the blur and geometric warp matrices, respectively, for the ith observation. Combining the equations in (1), we can write

$$Y = HX + Z \qquad (2)$$

where $Y = [Y_1^{\mathsf{T}} \ldots Y_r^{\mathsf{T}}]^{\mathsf{T}}$, $H = [(DB_1G_1)^{\mathsf{T}} \ldots (DB_rG_r)^{\mathsf{T}}]^{\mathsf{T}}$ and $Z = [N_1^{\mathsf{T}} \ldots N_r^{\mathsf{T}}]^{\mathsf{T}}$. Assuming that Z is white Gaussian with variance

σ_Z^2, the expression for the MAP estimate $\hat{X}$ of X can be derived using Bayes' rule as

$$\hat{X} = \arg\min_X \left\{ \frac{\|Y - HX\|^2}{2\sigma_Z^2} - \log P(X) \right\}. \qquad (3)$$

Markov Random Fields (MRFs) have been used to derive the prior term $P(X)$ [4], [5]. Such models are based on smoothness constraints and have to compromise between blurring out edges and letting in noise. We develop an image model that uses the prior knowledge that the image to be super-resolved belongs to the face class.

A. Super-Resolution Using Linear PCA

Given an N-dimensional zero-mean random vector X, PCA finds orthonormal matrix P such that

$$\begin{aligned} \operatorname{Cov}(P^\mathsf{T} X) &= E[P^\mathsf{T} X X^\mathsf{T} P] \\ &= P^\mathsf{T} \operatorname{Cov}(X) P \\ &= \operatorname{diag}([\lambda_1, \ldots, \lambda_N]), \ \lambda_1 > \ldots > \lambda_N. \end{aligned} \qquad (4)$$

The column vectors $\{V_i\}$ of P are the eigen-vectors of $\operatorname{Cov}(X)$, with the eigen-values $\{\lambda_i\}$ giving the corresponding variance. In our problem, X is a random super-resolved (zero-mean) face image, arranged lexicographically as an $N \times 1$ vector. The covariance matrix of the data set is calculated empirically from a set of training vectors (high-resolution face images) $\{T_i\}_{i=1\ldots M}$ as $\operatorname{Cov}(X) = (1/M)\sum_{i=1}^M T_i T_i^\mathsf{T} = AA^\mathsf{T}$ where $A = (1/\sqrt{M})[T_1 T_2 \ldots T_M]$ is an $N \times M$ matrix. Since the eigen-vectors $\{V_i\}$ lie in the span of $\{T_i\}$, it is sufficient to compute the eigen-vectors of $A^\mathsf{T} A$ (which is an $M \times M$ matrix, $M \ll N$). The eigen-vectors $\{e_i\}$ of $A^\mathsf{T} A$ can be related to $\{V_i\}$ as $V_i = Ae_i$ where $\lambda_i e_i^\mathsf{T} e_i = 1$. The first m principal eigen-vectors are chosen such that they account for more than 90% of the total variance.

Since the image to be super-resolved is a face, we expect most of its energy to lie within the subspace spanned by the principal components and very little energy outside of it. We use a Gibb's prior to derive the term $P(X)$ in (3) as

$$P(X) = \frac{1}{C} \exp \left\{ -\frac{1}{\zeta} U(X) \right\} \qquad (5)$$

where $U(X)$ is the energy of X outside the principal subspace, and C and ζ are normalization constants. The energy along the principal components is considered legitimate facial variation and is not penalized. We define the projection of X on the principal subspace as $\Psi_X = \sum_{i=1}^m (X^\mathsf{T} V_i) V_i$. Due to the orthogonality of V_i's, we have

$$U(X) = \|X\|^2 - \sum_{i=1}^m (X^\mathsf{T} V_i)^2 = \|\Psi_X - X\|^2. \qquad (6)$$

Consequently, $U(X)$ is the square of the distance between X and Ψ_X. Therefore,

$$P(X) = \frac{1}{C} \exp \left\{ -\frac{\|\Psi_X - X\|^2}{\zeta} \right\} \qquad (7)$$

Note that $P(X)$ increases as X converges to Ψ_X. Combining (3) and (7), we obtain

$$\begin{aligned} \hat{X} &= \arg\min_X \left\{ \frac{\|Y - HX\|^2}{2\sigma_Z^2} + \frac{\|\Psi_X - X\|^2}{\zeta} \right\} \\ &= \arg\min_X \left\{ \|Y - HX\|^2 + \Gamma_Z \|\Psi_X - X\|^2 \right\} \end{aligned} \qquad (8)$$
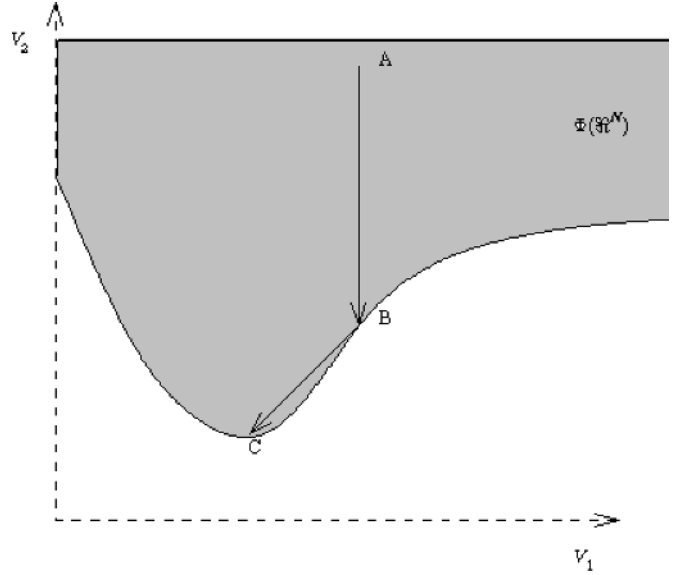

Fig. 1. Shaded portion indicates the image of $\Re^N$ in $\Re^F$. If we try to minimize the distance of $\tilde{\Phi}(X)$ from the face space starting from A, the estimate will move from A to B but then drift to C.

where $\Gamma_Z \propto \sigma_Z^2$. Equation (8) yields a regularized solution that can be computed by gradient-descent.

III. KERNEL PCA-BASED PRIOR

In the previous section, we developed a prior image model for the super-resolution problem using linear PCA, but this is limited to capturing second-order correlations. We now extend PCA to take higher-order correlations among the face pixels into account for achieving better super-resolution capability.

Based on the ideas in [15], [16], we define a nonlinear function Φ on the *image space* $\Re^N$ as $\Phi : \Re^N \rightarrow \Re^F$ where typically $F \gg N$. We call $\Re^F$ the *feature space*. The function Φ is usually chosen such that $\Re^F$ contains higher-order product terms from $\Re^N$. Carrying out PCA on $\Re^F$ will then capture higher-order correlations from $\Re^N$. However, explicitly calculating the nonlinear map Φ is computationally expensive, especially when the space is high-dimensional. We define a kernel function k such that $k(x, y) = \Phi(x)^\mathsf{T} \Phi(y)$. Given vectors in the image space, the kernel function calculates their dot product in the feature space. We can choose k appropriately to induce a desired Φ.

Since we need to carry out PCA in $\Re^F$, the map of our data set (face images) in $\Re^F$ needs to be mean-adjusted. We modify Φ and k to get $\tilde{\Phi}$ and $\tilde{k}$ in the following manner:

$$\tilde{\Phi}(x) = \Phi(x) - \frac{1}{M} \sum_{i=1}^M \Phi(T_i) \qquad (9)$$

$$\begin{aligned} \tilde{k}(x, y) &= \tilde{\Phi}(x)^\mathsf{T} \tilde{\Phi}(y) \\ &= k(x, y) - \frac{1}{M} \sum_{i=1}^M k(x, T_i) - \frac{1}{M} \sum_{i=1}^M k(T_i, y) \\ &\quad + \frac{1}{M^2} \sum_{i=1}^M \sum_{j=1}^M k(T_i, T_j). \end{aligned} \qquad (10)$$

We need to first compute $A^\mathsf{T} A$ where $A = (1/\sqrt{M})[\tilde{\Phi}(T_1) \ldots \tilde{\Phi}(T_M)]$. Each element of $A^\mathsf{T} A$ is given by

$$(A^\mathsf{T} A)_{i,j} = \frac{1}{M} \tilde{\Phi}(T_i)^\mathsf{T} \tilde{\Phi}(T_j) = \frac{1}{M} \tilde{k}(T_i, T_j). \qquad (11)$$

TABLE I
ALGORITHMS FOR TRAINING AND SUPER-RESOLUTION

Training	Super-resolution
Require: Set of M training face images $\{T_i\}_{i=1,2,\dots,M}$.	**Require:** Observations $\{Y_i\}$, blur and warp matrices $\{B_i, G_i\}$, and training data.
1: Compute matrix $A^\mathsf{T} A$ using (11).	1: Combine $\{Y_i\}$ to form Y, and $\{B_i, G_i\}$ to form H as in (2).
2: Eigen-decompose $\frac{1}{M} A^\mathsf{T} A$ to get $\{\lambda_i\}$ and $\{e_i\}$.	2: Calculate X_0 using (20).
3: Choose m according to $\{\lambda_i\}$ so as to contain $> 90\%$ of the total variance.	3: Calculate projections of X_0 on $\{V_i\}_{i=1\dots m}$ using (12).
4: Store $\{T_i\}_{i=1\dots M}$, m, $\{e_i\}_{i=1\dots m}$.	4: Calculate $\hat{X}$ from (15) using gradient descent.

Next, we eigen-decompose $A^\mathsf{T} A$ to get the eigen-values $\{\lambda_i\}$, and normalized eigen-vectors $\{e_i\}$. Based on the values of $\{\lambda_i\}$, we choose the first m eigen-vectors as the principal components. We do not explicitly calculate $\{V_i\}$ since finding A is computationally prohibitive. Given a new image z, we calculate the projection of $\tilde{\Phi}(z)$ on eigen-vector V_i as

$$\tilde{\Phi}(z)^\mathsf{T} V_i = \tilde{\Phi}(z)^\mathsf{T} A e_i = \sum_{j=1}^M e_{i_j} \tilde{k}(z, T_j). \quad (12)$$

To build an appropriate prior image model, we choose k as

$$k(x, y) = (1 + x^\mathsf{T} y)^n \quad (13)$$

This choice of k induces a mapping Φ that will contain all product terms of order 0 to n. Thus, the corresponding prior would incorporate all correlations from order 0 to $2n$.

It may be noted that unlike in the linear PCA case, using $U(X)$ as defined in (6) may not yield good estimates. This is because the estimate $\tilde{\Phi}(X)$ is constrained to move only in the volume $\tilde{\Phi}(\Re^N)$, which will not be at a uniform distance from the face space. Thus, maximizing a prior having the same form as in (7) can cause the estimate to move away from its own projection as illustrated in Fig. 1. We derive $P(X)$ using a modified formulation for the energy $U(X)$, which uses the distance of X from the projection of an initial estimate $\tilde{\Phi}(X_0)$, rather than from that of $\tilde{\Phi}(X)$ itself. The initial estimate will act as an anchor and prevent the estimate from drifting away. The projection Ψ_{X_0} is given by

$$\Psi_{X_0} = \sum_{i=1}^m \left(\tilde{\Phi}(X_0)^\mathsf{T} V_i \right) V_i. \quad (14)$$

Modifying (7) and (8) appropriately, we get $P(X) = (1/C)\exp\{-(\|\Psi_{X_0} - \tilde{\Phi}(X)\|^2/\zeta)\}$. Hence,

$$\hat{X} = \arg\min_X \left\{ \|Y - HX\|^2 + \Gamma_Z \left\| \Psi_{X_0} - \tilde{\Phi}(X) \right\|^2 \right\}. \quad (15)$$

In the above equation

$$\left\| \Psi_{X_0} - \tilde{\Phi}(X) \right\|^2 = \sum_{i=1}^m \left(\tilde{\Phi}(X_0)^\mathsf{T} V_i \right)^2 + \tilde{k}(X, X)$$
$$- 2 \sum_{i=1}^m \left(\tilde{\Phi}(X_0)^\mathsf{T} V_i \right) \left(\tilde{\Phi}(X)^\mathsf{T} V_i \right). \quad (16)$$

The terms $\tilde{\Phi}(X_0)^\mathsf{T} V_i$ and $\tilde{\Phi}(X)^\mathsf{T} V_i$ can be calculated using (12). From (16), the expression for the gradient of $\|\Psi_{X_0} - \tilde{\Phi}(X)\|^2$ with respect to X can be calculated as

$$\nabla_X \left\| \Psi_{X_0} - \tilde{\Phi}(X) \right\|^2 = \nabla_X \tilde{k}(X, X)$$
$$- 2 \sum_{i=1}^m \left(\tilde{\Phi}(X_0)^\mathsf{T} V_i \right) \sum_{j=1}^M e_{i_j} \nabla_X \tilde{k}(X, T_j) \quad (17)$$

where

$$\nabla_X \tilde{k}(X, X) = 2n(1 + X^\mathsf{T} X)^{n-1} X$$
$$- \frac{2}{M} \sum_{i=1}^M n(1 + X^\mathsf{T} T_i)^{n-1} T_i \quad (18)$$

and

$$\nabla_X \tilde{k}(X, T_j) = n(1 + X^\mathsf{T} T_j)^{n-1} T_j$$
$$- \frac{1}{M} \sum_{i=1}^M n(1 + X^\mathsf{T} T_i)^{n-1} T_i. \quad (19)$$

As we can determine all these terms with only the kernel, $\hat{X}$ can be calculated using gradient-descent in a computationally efficient manner. For the initial X_0, we use a finite-energy constrained least-squares estimate, i.e.,

$$X_0 = \arg\min_X \|Y - HX\|^2 + \gamma \|X\|^2. \quad (20)$$

The algorithms for training and super-resolution using kernel PCA are summarized in Table I.

IV. EXPERIMENTAL RESULTS

The training data is common to all the experiments. We used 450 images of 350 different individuals taken from the well-known FERET data set [17] for training. Actual eye positions were used to crop, scale and align the images. The final images had a resolution of 84×108 with 256 gray levels (normalized to have a common intensity).

In the first experiment, we consider super-resolution of face images from the [17] and YALE [18] databases (top and bottom row, respectively, in Fig. 2). The FERET image is of an individual not represented in the training set. The parameters of the experiment are as follows: the magnification factor is 5, the uniform blur is of size 5×5, observation noise variance is 100, and nine low-resolution observations are used. The low-resolution images were generated by blurring, shifting and down-sampling the high-resolution image and adding noise. Thus, the shifts between the low-resolution frames were known exactly and these were fed to the algorithms.

In Fig. 2, we have given results for different methods—linear PCA-based prior, Gaussian MRF (GMRF) prior [4], Huber MRF (HMRF) prior [5] and kernel PCA-based prior with $n = 5$ and 9. For the PCA and kernel PCA-based priors, the number of principal components m is chosen as 250. The values of Γ_Z were found empirically for each method to yield best results. From the figure, we note that while the GMRF-based method reduces noise, it leads to excessive blurring of the facial features. The HMRF-based method does a comparatively better job but lets in some noise and also blurs out some features (around the eyes). The kernel PCA-based prior not only suppresses noise but also brings out the facial features sharply. It yields good estimates even for the YALE image, although the imaging conditions (camera properties, etc.) were different from the training set which was FERET. Note that there is a significant improvement as we go from linear PCA to kernel
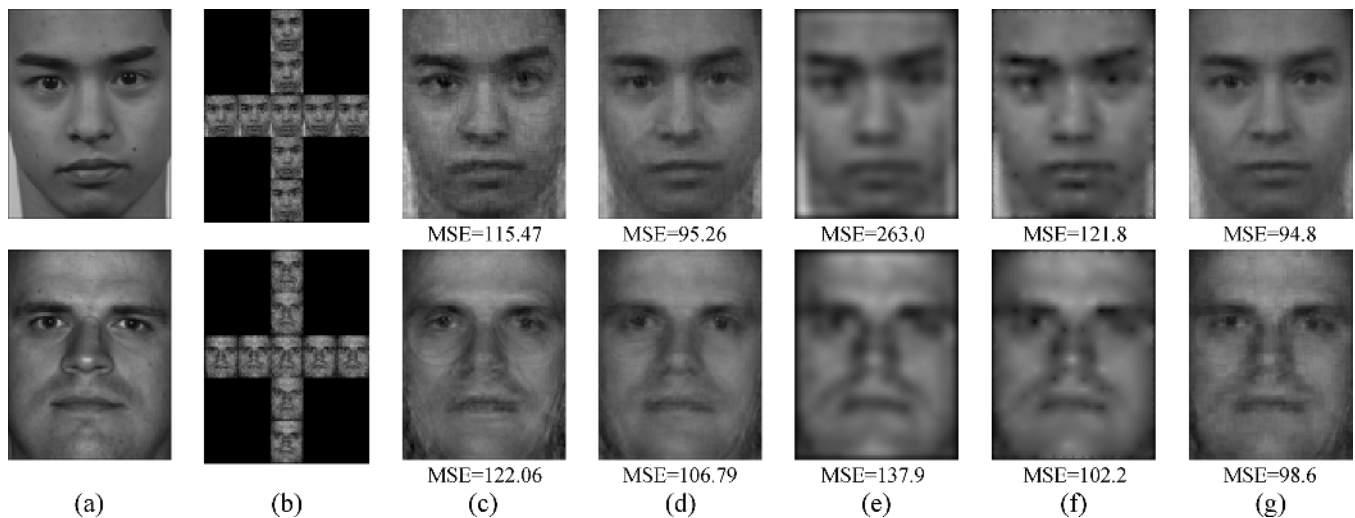
Fig. 2. (a) Original high-resolution images. (b) Low-resolution observations. Super-resolved image and corresponding mean square error (MSE) value using (c) linear PCA-based, (d) kernel PCA-based ($n = 5$), (e) GMRF-based, (f) HMRF-based, and (g) kernel PCA-based ($n = 9$) priors.
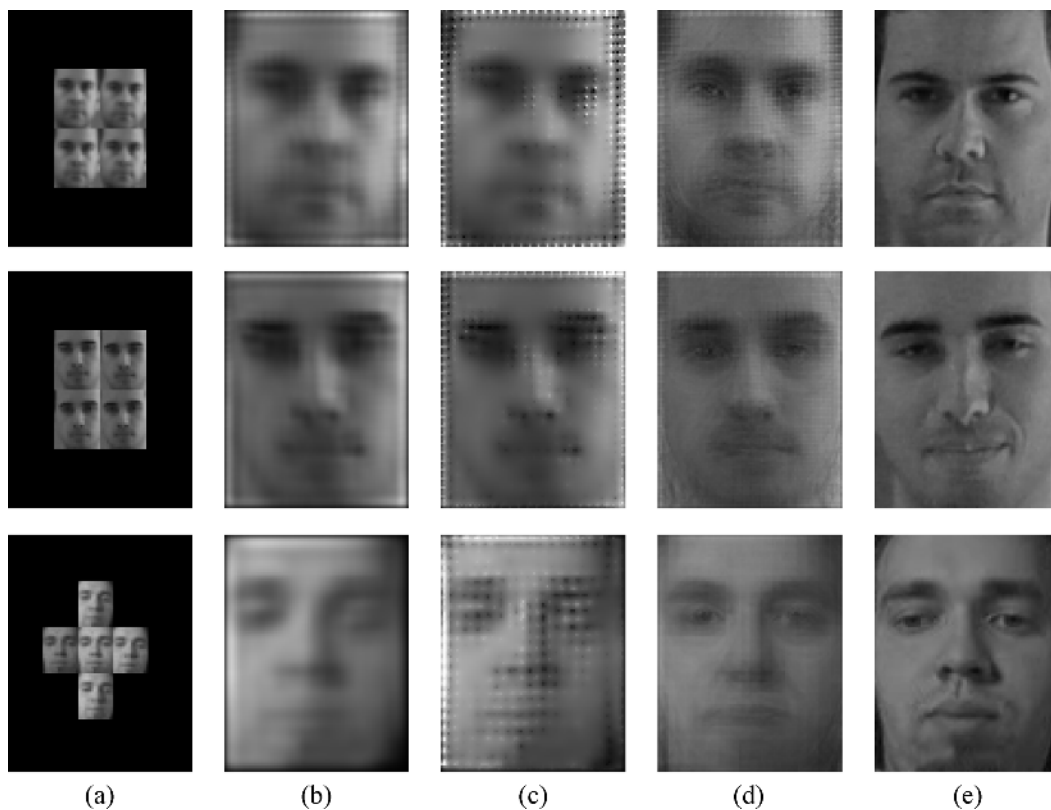

Fig. 3. Results on real video sequences (top two rows: frames from VPA database videos. Bottom row: frames from UO database video). (a) Low-resolution frames. Super-resolved images using (b) GMRF-based method, (c) HMRF-based method, and (d) the proposed method. (e) High-resolution images of individuals.

PCA-based prior. Since best results for kernel PCA were obtained with $n = 9$, this was used in the remaining experiments.

We next carried out super-resolution of face videos from the VPA database [19] and the UO database [20]. These sequences were re-encoded at lower-resolutions and then super-resolved with magnification factors of 4 and 5, using four and five low-resolution frames, respectively. For each case, the defocus blur PSF was approximated by a Gaussian, the eye positions were marked manually for one frame, and rudimentary correlation was used to estimate the translation shifts (and hence the matrices G_i) for the remaining frames. Fig. 3 shows the results for different methods along with high-resolution

images of the individuals (though not corresponding exactly to any of the video frames) for comparison. The overall quality of results for the real experiments is lower because of the combined errors in estimating the motion parameters, the blur PSF, and the eye positions. The GMRF-based method causes blurring of features while the HMRF-based method yields estimates with many false edges. However, the proposed method does a better job of bringing out the true facial features while suppressing artifacts.

We also compared our method with the well-known learning-based *hallucination* algorithm [7]. An implementation of the algorithm, provided by the authors themselves for super-resolving single frames, was
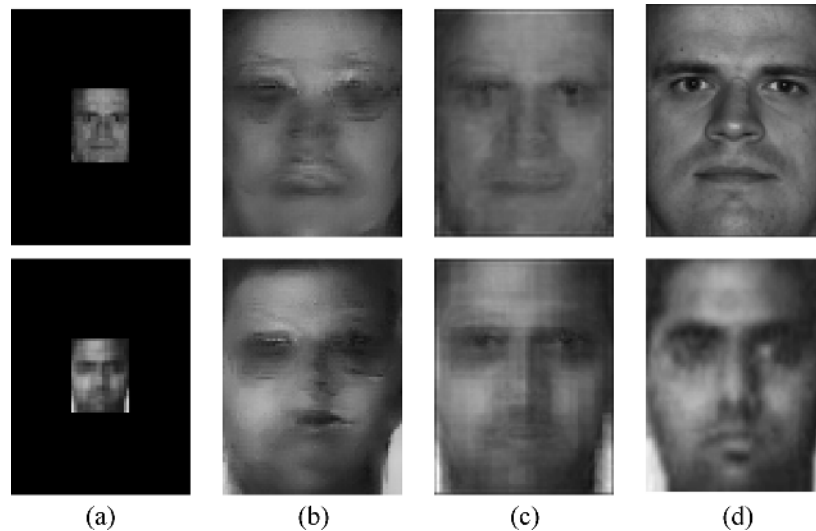
Fig. 4. Comparison of the proposed method with the hallucination algorithm. (a) Low-resolution frame. Super-resolved images using (b) the hallucination algorithm, and (c) the proposed method. (d) High-resolution images.

used. Both algorithms were given only single low-resolution images and we compared results using the YALE database image and a frame from a real low-resolution video taken in our lab. Both images were super-resolved by a factor of 4 and the results are given in Fig. 4. Since the method in [7] works by matching image gradient values, we note that for a high magnification factor and in the presence of noise (synthetically added in the YALE image, and naturally present in the real case) it produces several artifacts, particularly around the eyes and the mouth. On the other hand, the proposed method yields improved estimates in both the situations. This is despite the fact that the low-resolution frame for the real case is of low quality.

V. CONCLUSIONS

We proposed a learning-based method for super-resolution of face images that utilizes kernel PCA to build a prior model for frontal face images. This model is used to regularize the reconstruction of high-resolution face images from blurred and noisy low-resolution observations. By nonlinearly mapping the face images to a higher-dimensional feature space and performing PCA on the feature space, we capture higher-order correlations present in face images. The effectiveness of the proposed method was demonstrated with several experiments.

ACKNOWLEDGMENT

The authors thank the reviewers for their useful suggestions. They also thank S. Baker for providing the implementation of the hallucination algorithm and O. Sezer for granting access to the VPA database.

REFERENCES

[1] Y. Altunbasak, A. J. Patti, and R. M. Mersereau, "Super-resolution still and video reconstruction from MPEG-coded video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, pp. 217–226, 2002.

[2] S. P. Kim, N. K. Bose, and H. M. Valenzuela, "Recursive reconstruction of high resolution image from noisy undersampled multiframes," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1013–1027, 1990.

[3] N. Nguyen, M. Milanfar, and G. Golub, "A computationally efficient superresolution image reconstruction algorithm," *IEEE Trans. Image Processing*, vol. 10, pp. 1187–1193, 2001.

[4] R. C. Hardie, K. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Trans. Image Processing*, vol. 6, pp. 1621–1632, 1997.

[5] R. R. Schultz and R. L. Stevenson, "Extraction of high-resolution frames from video sequences," *IEEE Trans. Image Processing*, vol. 5, pp. 996–1011, 1996.

[6] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multi-frame super-resolution," *IEEE Trans. Image Processing*, vol. 13, pp. 1327–1344, 2004.

[7] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 24, pp. 1167–1183, 2002.

[8] Z. Lin and H. Y. Shum, "Fundamental limits of reconstruction-based super-resolution algorithms under local translation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, pp. 83–97, 2004.

[9] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes, and R. M. Mersereau, "Eigenface-domain super-resolution for face recognition," *IEEE Trans. Image Processing*, vol. 12, pp. 597–606, 2003.

[10] C. Liu, H. Shum, and Z. Zhang, "A two-step approach to hallucinating faces: Global parametric model and local non-parametric model," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recog. (CVPR)*, 2001, vol. 1, pp. 192–198.

[11] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Trans. Comput. Graphics Applicat.*, vol. 2, pp. 56–65, 2002.

[12] J. Sun, N.-N. Zheng, H. Tao, and H.-Y. Shum, "Image hallucination with primal sketch priors," in *Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recog. (CVPR)*, 2003, vol. 2, pp. 729–736.

[13] L. C. Pickup, S. J. Roberts, and A. Zisserman, "A sample texture prior for image super-resolution," in *Advances in Neural Info. Proc. Systems (NIPS)*, S. Thrun, L. Saul, and B. Schölkopf, Eds. Cambridge, MA: MIT Press, 2003.

[14] K. I. Kim, M. O. Franz, and B. Schölkopf, "Iterative kernel principal component analysis for image modeling," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 27, pp. 1351–1366, 2005.

[15] B. Schölkopf, A. Smola, and K.-R. Müller, "Nonlinear principal component analysis as a kernel eigen value problem," *Neural Comput.*, vol. 10, pp. 1299–1319, 1998.

[16] C. Liu, "Gabor-based kernel PCA with fractional power polynomial models for face recognition," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, pp. 572–581, 2004.

[17] P. J. Philips, H. Wechsler, J. Huang, and P. Rauss, "The FERET database and evaluation procedure for face recognition algorithms," *Image Vis. Comput. J.*, vol. 16, pp. 295–306, 1998.

[18] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 23, pp. 643–660, 2001.

[19] O. G. Sezer, Y. Altunbasak, and A. Ercil, "Face recognition with independent component based super-resolution," in *Proc. SPIE*, 2006, pp. 52–66.

[20] B. Martinkauppi, M. Soriano, S. Huovinen, and M. Laaksonen, "Face video database," in *Proc. 1st Eur. Conf. Color in Graphics, Imaging, Vision*, 2002, pp. 380–383.