Robust Multi-Image Based Blind Face Hallucination

Yonggang Jin
University of Bristol

Christos-Savvas Bouganis
Imperial College London

Abstract

This paper proposes a robust multi-image based blind face hallucination framework to super-resolve LR faces. The proposed framework first estimates both blurring kernel and transformations of multiple LR faces by robust deblurring and registration in PCA subspace. A patch-wise mixture of probabilistic PCA prior is then incorporated for face super-resolution. Previous work on face SR using PCA prior can be viewed as special cases of the framework. Experimental results in both simulated and real LR sequences demonstrate very promising performance of the proposed method.

1. Introduction

The paper focuses on multi-image based blind face hallucination to super-resolve low-resolution (LR) faces in image sequences. There has been much previous work carried out on face hallucination [1, 3, 15, 29, 7, 4, 27, 13, 17, 10] and face subspace based methods are often employed for face super-resolution (SR). PCA prior based face SR was presented in [3] with two MAP formulations, FS-MAP and IS-MAP. PCA prior is used in [15] as global constraints in combination with a patch-based local constraint from [6] for face SR, where the soft-constraint is essentially the same as FS-MAP in [3]. Nonnegative Matrix Factorization (NMF) was adopted in [29] as the global constraint, which is then combined with sparse coding for face hallucination. Face SR in eigen-domain was presented in [7]. Tensor-based approaches were introduced in [13, 17] to hallucinate facial expression using multi-modal faces. The asymmetry of registration of LR faces was pointed out in [4], where a Resolution-Aware Fitting (RAF) based on AAM [5] was proposed for joint registration and super-resolution to avoid interpolating LR images. Face SR based on image alignment was recently proposed where exemplars of HR faces are matched to an LR face by either dense SIFT flow [24, 11] or facial feature matching [28].

Most previous work on face SR utilizes only a single LR face, which may limit SR performance due to insufficient measurements available. Multiple LR images are often exploited for generic image SR [22, 8, 19], in combination with generic image priors, such as Tikhonov regularization, Total Variation (TV) or Huber MRF based sparsity prior. However for multi-image based SR, often a lot of LR images are used and registration errors will arise for moving objects or multi-view images. Recent work [26, 16, 23, 9] proposed to combine registration and/or deblurring with multi-image SR.

This paper presents a robust multi-image based blind face hallucination framework to super-resolve LR faces in image sequences. The proposed method first estimate both blurring kernel and transformations of multiple LR faces by robust deblurring and registration in PCA subspace using PCA prior. A patch-wise mixture of probabilistic PCA (MPPCA) prior [25] is then incorporated for face SR. Previous work on face SR using PCA prior can be viewed as special cases and it can be shown that FS-MAP[3]/Soft-Constraint[15] reduces to one iteration step in the framework. Previous work [15, 4, 27, 13] on face SR using single LR face manually or automatically aligned a face to a normalized face. For face SR using multiple LR faces [3], additional transformations between multiple LR faces need to be estimated by feature matching or tracking. However, due to low visual quality of LR faces, there are often significant errors in the initial alignment, which actually may degrade performance of face SR. In addition, fixed and small blurring is often assumed in previous work, be it Gaussian or square, which may not be valid in practice. The proposed method is in the similar spirit of [26, 16], but explores face PCA subspace, rather than original high resolution (HR) image space, for robust deblurring and registration. In addition, a new patch-wise MPPCA prior, rather than weak generic image priors, is applied to improve SR performance. The proposed method improves face SR performance, by combining robust blurring kernel and transformation estimation with the patch-wise MPPCA prior, with very promising results in both simulated and real LR sequences.

The remainder of the paper is organized as follows. Face hallucination is formulated in section 2. Section 3 describes the multi-image based blind face hallucination framework. Results in both simulated and real LR sequences are pre-

sented in section 4 and the paper is concluded in section 5.

2. Face Hallucination Formulation

2.1. Observation model

The process of image acquisition can be modeled by an observation model,which relates an original HR image $\mathbf{x}$ to an observed LR image $\mathbf{z}_t$, both vectors in lexicographical order,

$$\mathbf{z}_t = \mathcal{S}(s)\mathcal{K}(\mathbf{k})\mathcal{W}(\mathbf{w}_t)\mathbf{x} + \mathbf{n}_t \tag{1}$$

where $\mathcal{W}(\mathbf{w}_t) = \mathcal{W}(\mathbf{H}(\mathbf{w_t}))$ warps an HR image via transformation $\mathbf{H}(\mathbf{w_t})$ parameterized by $\mathbf{w}_t$, $\mathcal{K}(\mathbf{k})$ represents blurring effect caused by blurring kernel $\mathbf{k}$, $\mathcal{S}(s)$ is the sub-sampling operator with fixed factor s, and $\mathbf{n}_t$ is additive sensor noise. In the process of image acquisition, loss of image spatial resolution is mainly caused by blurring (optical blurring, motion blurring and sensor Point Spread Function (PSF)), sub-sampling and additive sensor noise. (1) can be rewritten as $\mathbf{z}_t = \mathcal{H}_t(\mathbf{k}, \mathbf{w}_t)\mathbf{x} + \mathbf{n}_t$, where $\mathcal{H}_t(\mathbf{k}, \mathbf{w}_t) = \mathcal{S}(s)\mathcal{K}(\mathbf{k})\mathcal{W}(\mathbf{w}_t)$ combines warping, blurring and sub-sampling operations. The observation model provides data constraint [15] for an HR image to be estimated.

Given multiple LR images, $\mathcal{Z} = \left[\ \mathbf{z}_{t-T}; \cdots ; \mathbf{z}_{t+T}\ \right]$, generated by multiple transformations but the same blurring kernel $\mathbf{k}$, the augmented observation model provides more constraints for estimation,

$$\mathcal{Z} = \mathcal{H}(\mathbf{k}, \mathbf{w})\mathbf{x} + \mathbf{n} \tag{2}$$

where $\mathcal{H}(\mathbf{k}, \mathbf{w}) = \left[\ \mathcal{H}_{t-T}(\mathbf{k}, \mathbf{w}_{t-T}); \cdots ; \mathcal{H}_{t+T}(\mathbf{k}, \mathbf{w}_{t+T})\ \right]$ is an augmented operator using multiple transformations $\mathbf{w} = \{\mathbf{w}_t\}$, and $\mathbf{n} = \left[\ \mathbf{n}_{t-T}; \cdots ; \mathbf{n}_{t+T}\ \right]$ is the augmented noise which is often assumed to be zero mean Gaussian with covariance $\mathbf{R} = \eta^2\mathbf{I}$. Note that except sub-sampling factor s, both blurring kernel $\mathbf{k}$ and transformation $\mathbf{w}$ are unknown and need to be estimated in practice. The assumption of constant blurring kernel is often valid for multiple LR images within short temporal sliding windows, and it also significantly reduces the number of unknowns thus make estimation more stable given insufficient number of LR images.

2.2. Priors

For face subspace based method using PCA prior [3, 15], it is assumed that latent PCA coefficients $\boldsymbol{\alpha}$ are in a zero-mean Gaussian distribution with diagonal covariance $\boldsymbol{\Lambda}$. A HR face $\mathbf{x}$ is then generated using eigenfaces $\mathbf{D}$ and mean face $\boldsymbol{\mu}$ via a generative model,

$$p(\boldsymbol{\alpha}) = \mathcal{N}(\boldsymbol{\alpha}; \mathbf{0}, \boldsymbol{\Lambda}) \text{ and } p(\mathbf{x}|\boldsymbol{\alpha}) = \mathcal{N}(\mathbf{x}; \boldsymbol{\mu} + \mathbf{D}\boldsymbol{\alpha}, \boldsymbol{\Sigma}) \tag{3}$$

where diagonal covariance $\boldsymbol{\Lambda}$ is composed of eigenvalues of corresponding eigenfaces and covariance $\boldsymbol{\Sigma} = \varepsilon^2\mathbf{I}$. In addition to the PCA prior, priors of transformations $p(\mathbf{w})$ and blurring kernel $p(\mathbf{k})$ can also be exploited. Combining likelihood (2) and priors, the posterior is

$$p(\mathbf{x}, \mathbf{k}, \mathbf{w}, \boldsymbol{\alpha}|\mathcal{Z}) \propto p(\mathcal{Z}|\mathbf{x}, \mathbf{k}, \mathbf{w})p(\mathbf{x}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})p(\mathbf{k})p(\mathbf{w}) \tag{4}$$

and if both blurring kernel $\mathbf{k}^*$ and transformation $\mathbf{w}^*$ are known beforehand as in [3, 15], then the posterior can be simplified to

$$p(\mathbf{x}, \boldsymbol{\alpha}|\mathcal{Z}, \mathbf{k}^*, \mathbf{w}^*) \propto p(\mathcal{Z}|\mathbf{x}, \mathbf{k}^*, \mathbf{w}^*)p(\mathbf{x}|\boldsymbol{\alpha})p(\boldsymbol{\alpha}) \tag{5}$$

Although PCA prior assumes a Gaussian distribution for latent coefficients in PCA subspace, it can be shown that for MAP estimation it is equivalent to a prior in image space,

$$\begin{aligned}
\hat{\mathbf{x}} &= \arg\max_{\mathbf{x},\boldsymbol{\alpha}} p(\mathcal{Z}|\mathbf{x}, \mathbf{k}^*, \mathbf{w}^*)p(\mathbf{x}|\boldsymbol{\alpha})p(\boldsymbol{\alpha}) \\
&= \arg\max_{\mathbf{x}} p(\mathcal{Z}|\mathbf{x}, \mathbf{k}^*, \mathbf{w}^*) \left\{ \max_{\boldsymbol{\alpha}} p(\boldsymbol{\alpha}|\mathbf{x})p(\mathbf{x}) \right\} \\
&= \arg\max_{\mathbf{x}} p(\mathcal{Z}|\mathbf{x}, \mathbf{k}^*, \mathbf{w}^*)p(\mathbf{x}) \left\{ \max_{\boldsymbol{\alpha}} p(\boldsymbol{\alpha}|\mathbf{x}) \right\} \\
&= \arg\max_{\mathbf{x}} p(\mathcal{Z}|\mathbf{x}, \mathbf{k}^*, \mathbf{w}^*)p(\mathbf{x})
\end{aligned}$$

where the last step is due to the fact that $\max_{\boldsymbol{\alpha}} p(\boldsymbol{\alpha}|\mathbf{x}) = \max_{\boldsymbol{\alpha}} \mathcal{N}(\boldsymbol{\alpha}; \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Lambda}}) = \left|2\pi\hat{\boldsymbol{\Lambda}}\right|^{-\frac{1}{2}}$ is a constant independent of $\mathbf{x}$, and the equivalent prior in image space is

$$p(\mathbf{x}) = \int p(\mathbf{x}|\boldsymbol{\alpha})p(\boldsymbol{\alpha})d\boldsymbol{\alpha} = \mathcal{N}(\mathbf{x}; \tilde{\mathbf{x}}, \tilde{\mathbf{P}}) \tag{6}$$

where $\tilde{\mathbf{x}} = \boldsymbol{\mu}$ and $\tilde{\mathbf{P}} = \mathbf{D}\boldsymbol{\Lambda}\mathbf{D}^T + \boldsymbol{\Sigma}$ are predicted HR face and its covariance respectively. Note that this image prior can also be interpreted in two terms [21]: Euclidean distance to the PCA subspace and Mahalanobis distance within the PCA subspace. Now the MAP estimation can be simplified to

$$\begin{aligned}
\hat{\mathbf{x}} &= \arg\max_{\mathbf{x}} p(\mathbf{x}|\mathcal{Z}, \mathbf{k}^*, \mathbf{w}^*) \\
&= \boldsymbol{\mu} + \mathbf{K}\left[\mathcal{Z} - \mathcal{H}\boldsymbol{\mu}\right]
\end{aligned} \tag{7}$$

where $p(\mathbf{x}|\mathcal{Z}, \mathbf{k}^*, \mathbf{w}^*) = \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}, \hat{\mathbf{P}})$ is the posterior given blurring kernel and transformation with mean $\hat{\mathbf{x}} = \boldsymbol{\mu} + \mathbf{K}\left[\mathcal{Z} - \mathcal{H}\boldsymbol{\mu}\right]$ and covariance $\hat{\mathbf{P}} = (\mathbf{I} - \mathbf{K}\mathcal{H})\tilde{\mathbf{P}}$ by using Kalman gain $\mathbf{K} = \tilde{\mathbf{P}}\mathcal{H}^T\left[\mathbf{R} + \mathcal{H}\tilde{\mathbf{P}}\mathcal{H}^T\right]^{-1}$. It generalizes previous work of both FS-MAP/IS-MAP [3] and soft-constraint/hard-constraint [15], which can be shown to become special cases.

2.3. Patch-wise MPPCA prior

Limitations of PCA prior It can be noted that (7) actually is a linear estimator to estimate an HR image, where $\mathbf{K}$ is a dense linear coefficient matrix, so each row in $\mathbf{K}$ linearly combines all LR image pixels to estimate an HR pixel. This holistic PCA approaches comes with an over-fitting issue, as all pixels in an LR image are involved to estimate an HR pixel, no matter how far they are spatially far away from the HR pixel. It is well-known that correlation between pixels decreases dramatically as the distance between pixels increases. The over-fitting can be shown that although fitting error of a training set is the lowest of all linear estimators by using all LR pixels, fitting error of a testing set is much higher in comparison with patch-based approaches.

The second limitation is that very roughly predicted mean and covariance are used in the equivalent image prior (6). The predicted mean is actually training sample mean $\boldsymbol{\mu} = \frac{1}{N}\sum_n \mathbf{x}_n$, while the predicted covariance can be further rewritten as $\tilde{\mathbf{P}} = \mathbf{C} - d\mathbf{C} + \boldsymbol{\Sigma}$, where $\mathbf{C} = \frac{1}{N}\sum_n [\mathbf{x}_n - \boldsymbol{\mu}][\mathbf{x}_n - \boldsymbol{\mu}]^T$ is the covariance of training samples, $d\mathbf{C} = \mathbf{D}'\boldsymbol{\Lambda}'\mathbf{D}'^T$ consists of discarded eigenvectors and eigenvalues of $\mathbf{C}$, so the predicted covariance becomes the covariance of training samples if all eigenvectors are used.

Patch-wise MPPCA We improve the simple PCA prior model by a patch-wise MPPCA model [25], where an HR image is spatially partitioned into local overlapping patches as in [6, 29], patch-wise local MPPCA dictionaries are learnt for each patch. The local patch-wise approach is to address the first limitation of the holistic approach, while MPPCA aims to improve both mean and covariance prediction, which is the second limitation of the simple PCA prior.

For MPPCA, the generative model of an HR patch $\mathbf{y}$ is

$$p(\boldsymbol{\alpha}) = \mathcal{N}(\boldsymbol{\alpha}; \mathbf{0}, \mathbf{I}) \text{ and } p(\mathbf{y}|\boldsymbol{\alpha}) = \sum_j \pi_j \mathcal{N}(\mathbf{y}; \boldsymbol{\mu}_j + \mathbf{D}_j\boldsymbol{\alpha}, \boldsymbol{\Sigma}_j) \tag{8}$$

where π_j, $\mathbf{D}_j$, $\boldsymbol{\mu}_j$ and $\boldsymbol{\Sigma}_j$ are the mixture weight, dictionary, mean and covariance of jth component respectively, normalized latent coefficients $\boldsymbol{\alpha}$ is in zero-mean Gaussian distribution with identity covariance. Given a HR training set, patch-wise MPPCA can be learned for each patch, where mixture weights $\{\pi_{i,j}\}$, dictionaries $\{\mathbf{D}_{i,j}\}$, means $\{\boldsymbol{\mu}_{i,j}\}$, covariances $\{\boldsymbol{\Sigma}_{i,j}\}$ of the ith patch can be estimated iteratively as shown in [25]. Assuming independent patches, they can be combined to

$$p(\boldsymbol{\alpha}) = \prod_i p(\boldsymbol{\alpha}_i) \text{ and } p(\{\mathbf{y}_i\}|\boldsymbol{\alpha}) = \prod_i p(\mathbf{y}_i|\boldsymbol{\alpha}_i) \tag{9}$$

where $p(\mathbf{y}_i|\boldsymbol{\alpha}_i) = \sum_j \pi_{i,j}\mathcal{N}(\mathbf{F}_i\mathbf{x}; \boldsymbol{\mu}_{i,j} + \mathbf{D}_{i,j}\boldsymbol{\alpha}_i, \boldsymbol{\Sigma}_{i,j})$, $\mathbf{F}_i$ denotes an operator to extract the ith patch from an HR image, and $\boldsymbol{\alpha} = \{\boldsymbol{\alpha}_i\}$. The probability that image patches are from learnt patch-wise MPPCA priors becomes

$$p(\{\mathbf{y}_i\}) = \prod_i p(\mathbf{y}_i) \tag{10}$$

where $p(\mathbf{y}_i) \propto \sum_j \pi_{i,j}\mathcal{N}(\mathbf{y}_i; \boldsymbol{\mu}_{i,j}, \tilde{\mathbf{P}}_{i,j})$, and $\tilde{\mathbf{P}}_{i,j} = \mathbf{D}_{i,j}\mathbf{D}_{i,j}^T + \mathbf{I}$ are patch-wisely predicted covariance. As overlapping patches are not independent, logarithm of (10) can be viewed as "Expected Patch Log Likelihood"(EPLL) [30].

3. Multi-image based Blind Face Hallucination

In the previous section, both blurring kernel and transformation are assumed to be known. However in practice, they are unknown and need to be estimated, and it will be shown in this section that both blurring kernel and multiple transformations can be estimated efficiently in PCA subspace using PCA prior. Given estimated blurring kernel and transformations, an HR face is then estimated using patch-wise MPPCA prior.

3.1. Blurring kernel and transformation estimation

Instead of directly optimizing the posterior (4) to estimate all the unknowns, $\arg\min_{\mathbf{x},\mathbf{k},\mathbf{w},\boldsymbol{\alpha}} p(\mathbf{x}, \mathbf{k}, \mathbf{w}, \boldsymbol{\alpha}|\mathcal{Z})$, we optimize marginalized posterior,

$$p(\mathbf{k}, \mathbf{w}|\mathcal{Z}) = \int\int p(\mathbf{x}, \mathbf{k}, \mathbf{w}, \boldsymbol{\alpha}|\mathcal{Z}) d\boldsymbol{\alpha} d\mathbf{x} \tag{11}$$

where high dimensional HR image $\mathbf{x}$ and PCA coefficients $\boldsymbol{\alpha}$ are regarded as latent variables. Marginalized posterior $p(\mathbf{k}, \mathbf{w}|\mathcal{Z})$ can be approximated by its variational lower bound,

$$\begin{aligned} &\log p(\mathbf{k}, \mathbf{w}|\mathcal{Z}) \\ &\geq \int\int q(\mathbf{x}, \boldsymbol{\alpha}) \log\left[p(\mathcal{Z}|\mathbf{x}, \mathbf{k}, \mathbf{w}) p(\mathbf{x}|\boldsymbol{\alpha}) p(\boldsymbol{\alpha}) p(\mathbf{k}) p(\mathbf{w})\right] d\mathbf{x} d\boldsymbol{\alpha} \\ &\quad - \int\int q(\mathbf{x}, \boldsymbol{\alpha}) \log q(\mathbf{x}, \boldsymbol{\alpha}) d\mathbf{x} d\boldsymbol{\alpha} + \text{const} \\ &= \mathcal{L}(\log p(\mathbf{k}, \mathbf{w}|\mathcal{Z})) \end{aligned} \tag{12}$$

where $\mathcal{L}(\log p(\mathbf{k}, \mathbf{w}|\mathcal{Z}))$ is the lower bound of $\log p(\mathbf{k}, \mathbf{w}|\mathcal{Z})$. Let $q(\mathbf{x}, \boldsymbol{\alpha}) = \delta(\mathbf{x} - \boldsymbol{\mu} - \mathbf{D}\boldsymbol{\alpha}) q(\boldsymbol{\alpha})$, then the variational lower bound becomes

$$\begin{aligned} &\mathcal{L}(\log p(\mathbf{k}, \mathbf{w}|\mathcal{Z})) \\ &= \int q(\boldsymbol{\alpha}) \log\left[p(\mathcal{Z}|\mathbf{k}, \mathbf{w}, \boldsymbol{\alpha}) p(\boldsymbol{\alpha})\right] d\boldsymbol{\alpha} \\ &\quad - \int q(\boldsymbol{\alpha}) \log q(\boldsymbol{\alpha}) d\boldsymbol{\alpha} + \log p(\mathbf{k}) + \log p(\mathbf{w}) + \text{const} \end{aligned} \tag{13}$$

where $q(\boldsymbol{\alpha})$ is the approximated posterior of latent PCA coefficients. Note that we utilize reduced dimensionalities

of PCA subspace, rather than the high dimensional image space or filter space in [14], for blurring kernel and transformation estimation, and full covariance of $q(\boldsymbol{\alpha})$ is explored, rather than a diagonalized covariance in [14]. Lower bound (13) can be optimized by updating $q(\boldsymbol{\alpha})$, $\mathbf{k}$ and $\mathbf{w}$ iteratively as follows:

Updating coefficients distribution $q(\boldsymbol{\alpha})$. Given $\mathbf{k}$ and $\mathbf{w}$, it can be shown that

$$q(\boldsymbol{\alpha}) = p(\boldsymbol{\alpha}|\mathcal{Z}, \mathbf{k}, \mathbf{w}) = \mathcal{N}(\boldsymbol{\alpha}; \hat{\boldsymbol{\alpha}}, \hat{\boldsymbol{\Lambda}}) \tag{14}$$

which is the posterior given blurring kernel and transformation with mean $\hat{\boldsymbol{\alpha}} = \mathbf{K}[\mathcal{Z} - \mathcal{H}\boldsymbol{\mu}]$ and covariance $\hat{\boldsymbol{\Lambda}} = (\mathbf{I} - \mathbf{K}\mathcal{H}\mathbf{D})\boldsymbol{\Lambda}$ by using Kalman gain $\mathbf{K} = \boldsymbol{\Lambda}\mathbf{D}^T\mathcal{H}^T\left[\mathbf{R} + \mathcal{H}\mathbf{D}\boldsymbol{\Lambda}\mathbf{D}^T\mathcal{H}^T\right]^{-1}$. It is essentially FS-MAP[3]/Soft-Constraint[15], however only mean $\hat{\boldsymbol{\alpha}}$ is used as the MAP estimation in FS-MAP[3]/Soft-Constraint[15]. It can be shown that in addition to the mean, covariance $\hat{\boldsymbol{\Lambda}}$, which represents uncertainties of the MAP estimation, is important and will be used to estimate the blurring kernel and transformation.

Updating blurring kernel $\mathbf{k}$. Given $q(\boldsymbol{\alpha})$ and $\mathbf{w}$, it can be shown that the lower bound can be rewritten in quadratic form

$$\mathcal{L}(\mathbf{k}) \propto \mathbf{k}^T\mathbf{A}(\hat{\boldsymbol{\Lambda}})\mathbf{k} + 2\mathbf{b}^T(\hat{\boldsymbol{\alpha}})\mathbf{k} - 2\log p(\mathbf{k}) + \text{const} \tag{15}$$

where $\mathbf{A}(\hat{\boldsymbol{\Lambda}})$ and $\mathbf{b}(\hat{\boldsymbol{\alpha}})$ involve the previously estimated mean $\hat{\boldsymbol{\alpha}}$ and covariance $\hat{\boldsymbol{\Lambda}}$. Prior $p(\mathbf{k})$ includes L1 norm constraint $\|\mathbf{k}\|_1 = 1$, non-negativity constraint $\mathbf{k} \geq \mathbf{0}$. For Gaussian blurring kernel, additional symmetric constraint can be applied so that the dimensionality of the blurring kernel can be further reduced. The blurring kernel is initialized with an estimated blurring kernel at previous frame and for the first frame it is initialized to be the Dirac delta function.

Updating multiple transformations $\mathbf{w}$. Given $q(\boldsymbol{\alpha})$ and $\mathbf{k}$, it can be shown that the lower bound reduces to

$$\begin{aligned}\mathcal{L}(\mathbf{w}) \propto \left\|\mathcal{Z} - \mathcal{H}(\mathbf{w})(\boldsymbol{\mu} + \mathbf{D}\hat{\boldsymbol{\alpha}})\right\|^2 - 2\log p(\mathbf{w}) \\ + Tr\left[\mathbf{D}^T\mathcal{H}^T(\mathbf{w})\mathcal{H}(\mathbf{w})\mathbf{D}\hat{\boldsymbol{\Lambda}}\right] + \text{const}\end{aligned} \tag{16}$$

Note that covariance $\hat{\boldsymbol{\Lambda}}$ can be eigen-decomposed by $\hat{\boldsymbol{\Lambda}} = \mathbf{V}\mathbf{L}\mathbf{V}^T$, where $\mathbf{V} = [\mathbf{v}_1, \cdots, \mathbf{v}_L]$ and $\mathbf{L} = \text{diag}([\lambda_1, \cdots, \lambda_L])$, $\{\mathbf{v}_l\}_{l=1}^L$ and $\{\lambda_l\}_{l=1}^L$ are eigenvectors and eigenvalues of $\hat{\boldsymbol{\Lambda}}$ respectively. Now the second term becomes $\sum_{l=1}^{L}\lambda_l\left\|\mathcal{H}(\mathbf{w})\mathbf{D}\mathbf{v}_l\right\|^2$, so

$$\begin{aligned}\mathcal{L}(\mathbf{w}) \propto \left\|\mathcal{Z} - \mathcal{H}(\mathbf{w})(\boldsymbol{\mu} + \mathbf{D}\hat{\boldsymbol{\alpha}})\right\|^2 - 2\log p(\mathbf{w}) \\ + \sum_{l=1}^{L}\lambda_l\left\|\mathcal{H}(\mathbf{w})\mathbf{D}\mathbf{v}_l\right\|^2 + \text{const}\end{aligned} \tag{17}$$

It can be observed that optimal transformation needs to not only reduce the first two terms which are due to the observation model and prior respectively, but also reduce the third term which involves covariance $\hat{\boldsymbol{\Lambda}}$, which represents uncertainties of $\hat{\boldsymbol{\alpha}}$.

Multiple transformations are initialized by tracking forwardly and backwardly in neighboring frames, and is compositional updated as in [2] but in a backward direction as in [4] to avoid interpolating LR images. Given an initialized transformation $\tilde{\mathbf{H}}_t$, the transformation is compositional updated by $\mathbf{H}_t = \tilde{\mathbf{H}}_t\Delta\mathbf{H}(\mathbf{w}_t)$, where $\mathbf{w}_t$ are coefficients of Lie algebra generators $\left\{\mathbf{G}^{(i)}\right\}$ and $\Delta\mathbf{H}(\mathbf{w}_t) = \exp\left(\sum_i \mathbf{w}_t^{(i)}\mathbf{G}^{(i)}\right)$ is the incremental transformation. The prior of $\mathbf{w}$ is $p(\mathbf{w}) = \prod_t p(\mathbf{w}_t)$, where $p(\mathbf{w}_t) = \mathcal{N}(\mathbf{w}_t; \mathbf{0}, \mathbf{Q})$ is a zero mean Gaussian distribution with covariance $\mathbf{Q}$.

3.2. Super-resolution with patch-wise MPPCA prior

Given estimated blurring kernel $\mathbf{k}^*$ and multiple transformations $\mathbf{w}^*$, it can be shown that the variational lower bound of $\log p(\mathcal{Z})$ is

$$\begin{aligned}\mathcal{L}\left(\log p(\mathcal{Z})\right) = \int q(\mathbf{x})\log p(\mathcal{Z}|\mathbf{x}, \mathbf{k}^*, \mathbf{w}^*)d\mathbf{x} \\ + \sum_{i,j} q_{i,j}\int q(\mathbf{x})\log\left[\mathcal{N}(\mathbf{F}_i\mathbf{x}; \boldsymbol{\mu}_{i,j}, \tilde{\mathbf{P}}_{i,j})\right]d\mathbf{x} \\ + \sum_{i,j} q_{i,j}\log\frac{\pi_{i,j}}{q_{i,j}} - \int q(\mathbf{x})\log q(\mathbf{x})d\mathbf{x} + \text{const}\end{aligned} \tag{18}$$

where $q(\mathbf{x})$ is an approximated posterior of an HR image, latent association weights $\mathcal{Q} = \{q_{i,j}\}$ indicates probabilities that patch i is from mixture component j. The lower bound can also be optimized by updating $q(\mathbf{x})$ and $\mathcal{Q}$ alternatively:

Updating HR image distribution $q(\mathbf{x})$. Given $\mathcal{Q}$, it can be shown that

$$q(\mathbf{x}) \propto \mathcal{N}(\mathbf{x}; \hat{\mathbf{x}}, \hat{\mathbf{P}}) \tag{19}$$

with mean $\hat{\mathbf{x}} = \hat{\mathbf{P}}\left[\mathcal{H}^T\mathbf{R}^{-1}\mathcal{Z} + \sum_i\left[\sum_j q_{i,j}\boldsymbol{\mu}_{i,j}^T\tilde{\mathbf{P}}_{i,j}^{-1}\right]\mathbf{F}_i\right]$

Algorithm 1 Multi-image Based Blind Face Hallucination

1. Initialize $\mathbf{k}^{(0)} = \mathbf{k}^*$, $\mathbf{w}^{(0)} = \mathbf{0}$ and $\left\{\tilde{\mathbf{H}}_t\right\}$ by tracking
2. Blurring kernel/transformation estimation: set $k = 1$

   (i) Updating coefficients distribution $\hat{\boldsymbol{\alpha}}^{(k)}$ and $\hat{\boldsymbol{\Lambda}}^{(k)}$ by (14)
   (ii) Updating blurring kernel $\mathbf{k}^{(k)}$ by optimizing (15)
   (iii) Updating multiple transformations $\mathbf{w}^{(k)}$ by optimizing (17)
   if $\left\|\mathbf{w}^{(k)} - \mathbf{w}^{(k-1)}\right\| < \varepsilon$ and $\left\|\mathbf{k}^{(k)} - \mathbf{k}^{(k-1)}\right\| < \varepsilon$ **then**
   $\mathbf{k}^* = \mathbf{k}^{(k)}$, $\mathbf{w}^* = \mathbf{w}^{(k)}$, $\left\{\mathbf{H}_t^* = \tilde{\mathbf{H}}_t \Delta\mathbf{H}(\mathbf{w}_t^*)\right\}$, and go to 3
   else
   $k = k + 1$ go to (i)
   end if
3. Super-resolution: set $k = 1$

   (i) Updating HR image distribution $\hat{\mathbf{x}}^{(k)}$ and $\hat{\mathbf{P}}^{(k)}$ by (19)
   (ii) Updating association weights $\mathcal{Q}$ by (20)
   if $\left\|\hat{\mathbf{x}}^{(k)} - \hat{\mathbf{x}}^{(k-1)}\right\| < \varepsilon$ **then**
   $\mathbf{x}^* = \hat{\mathbf{x}}^{(k)}$ and stop
   else
   $k = k + 1$ go to (i)
   end if

and covariance $\hat{\mathbf{P}} = \left[\mathcal{H}^T\mathbf{R}^{-1}\mathcal{H} + \sum_i \mathbf{F}_i^T\left[\sum_j q_{i,j}\tilde{\mathbf{P}}_{i,j}^{-1}\right]\mathbf{F}_i\right]^{-1}$

Updating association weights $\mathcal{Q}$. Given $q(\mathbf{x})$, association weights $\mathcal{Q}$ is updated by

$$q_{i,j} \propto \pi_{i,j}\mathcal{N}(\mathbf{F}_i\hat{\mathbf{x}}; \boldsymbol{\mu}_{i,j}, \tilde{\mathbf{P}}_{i,j}) \exp\left(-\tfrac{1}{2}Tr\left[\mathbf{F}_i^T\tilde{\mathbf{P}}_{i,j}^{-1}\mathbf{F}_i\hat{\mathbf{P}}\right]\right)$$
$$\sum_j q_{i,j} = 1, \forall i \tag{20}$$

where it can be noted that covariance $\hat{\mathbf{P}}$, which represents uncertainties of an estimated HR image $\hat{\mathbf{x}}$, is also involved to estimate association weights.

Main stages of the proposed method using patch-wise MPPCA combined with blurring kernel and transformation estimation are summarized in algorithm 1.

4. Results

Experiments were carried out to test SR performance of the proposed method in both simulated and real LR image sequences. Frontal HR faces from FERET database [18] are used as the training set, where all faces are normalized to size 60×60 by affine transformation using eyes and mouth centres. For patch-wise MPPCA prior, overlapping patches with size 6×6 are used to train mixtures with ten components. Testing set includes BioID database [12], PubFig83 database [20] and real LR image sequences. Homography is used in all experiments, and is initialized in the first frame to be identity transformation for normalized BioID database whereas affine transformations for PubFig83 database and real LR sequences by using centers of eyes and mouth.

4.1. Simulated LR sequences

100 random HR faces from BioID database [12] are used to simulate multiple LR faces, as shown in Fig. 1(a). An HR face $\mathbf{x}$ is first transformed by multiple random homographies, and then blurred with a Gaussian kernel with $\sigma = 0.6s = 2.4$ and sub-sampled by $s = 4$ to 15×15. Finally Gaussian noise with $\eta = 2$ are added to obtain a sequence of multiple (3) LR faces $\mathcal{Z}$, as shown in Fig. 1(b). The Ground-Truth (GT) blurring kernel is shown in Fig. 4(a) whereas an example of estimated blurring kernels in Fig. 4(b). Our unoptimized MATLAB implementation takes about 15-20 minutes on 2.6GHz Xeon when 3 LR faces are used, where registration and blurring kernel estimation is the most time-consuming part. Speed can be much improved as both registration of multiple LR faces and patch-wise MPPCA prior can be implemented in parallel.

SR results using PCA prior, essentially FS-MAP/Soft-Constraint [3, 15], in combination with the GT blurring kernel and transformations initialized by tracking, are shown in Fig. 1 (d). It serves as an indicator of SR performance of previous work [3, 15] in practice. It can be noted that even given the GT blurring kernel, inaccurate transformation estimation degrades SR performance significantly.

Results of generic HMRF prior [19], PCA prior [3, 15] and patch-wise MPPCA prior, in combination with the GT blurring kernel and the GT transformations, are shown in Fig. 1(c), (f) and (h) respectively, which can be regarded as, in general, upper bounds of SR performance that can be achieved for these methods. It should be noted that results of patch-wise MPPCA are often sub-optimal due to the iterative optimization.

Results of the proposed method using PCA prior and patch-wise MPPCA prior, in combination with an estimated blurring kernel and estimated transformations, are shown in Fig. 1(e) and (g) respectively, which represent SR performance of PCA prior and patch-wise MPPCA prior achieved in practice so far. The results are visually close to the corresponding results using the GT blurring kernel and transformations in Fig. 1(f) and (h). It is also interesting to observe in the last row how details of eyes and mouth were kept along with a white dot in the corner of background using the patch-wise MPPCA prior, in comparison with PCA prior.

Performance measures of SR using both PSNR and

Figure 1. Results of face hallucination using multiple simulated LR faces. (a) HR faces $\mathbf{x}$. (b) LR faces $\mathcal{Z}$ obtained by warping HR faces in (a) with random homographies, then blurring with Gaussian blurring kernel with $\sigma = 0.6s$ and sub-sampling by $s = 4$ to 15×15, and finally adding Gaussian noise with $\eta = 2$. (c) Results of generic HMRF prior in combination with the GT blurring kernel and transformations. (d) Results of PCA prior in combination with the GT blurring kernel and transformations initialized by tracking. (e) Results of PCA prior in combination with an estimated blurring kernel and transformations. (f) Results of PCA prior in combination with the GT blurring kernel and transformations. (g) Results of the proposed patch-wise MPPCA prior in combination with an estimated blurring kernel and transformations. (h) Results of the patch-wise MPPCA prior in combination with the GT blurring kernel and transformations. Number below images is PSNR/SSIM. **The images are better viewed by zooming.**

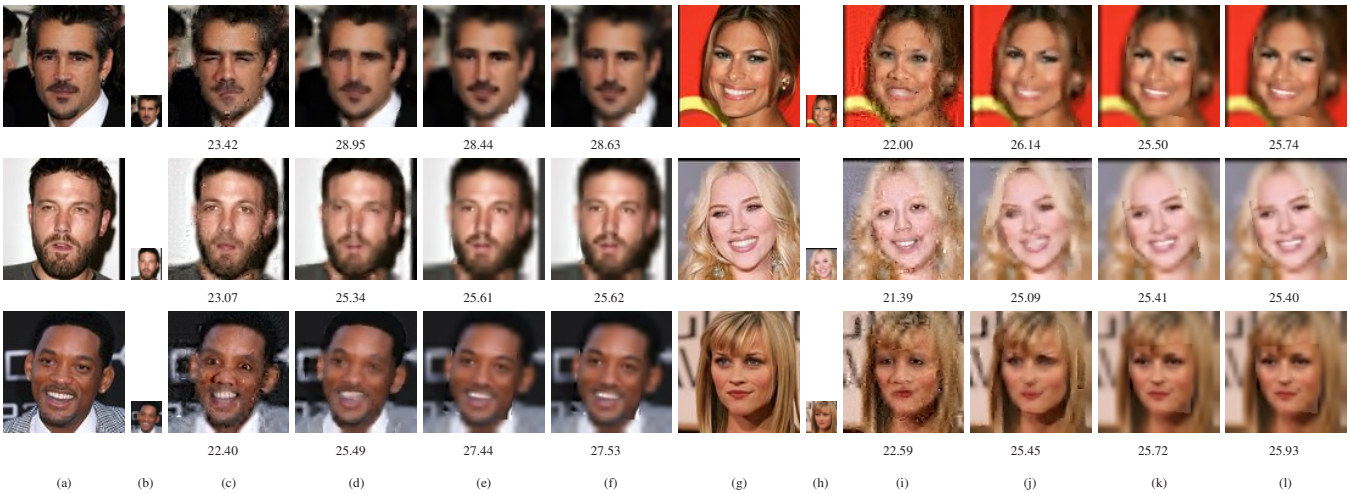

Figure 2. Results of face hallucination using single simulated LR face. (a)(g) HR faces $\mathbf{x}$. (b)(h) LR faces $\mathcal{Z}$ obtained by blurring HR faces in (a)(g) with $\sigma = 0.4s$ and sub-sampling by $s = 4$. (c)(i) Results of [28] with the GT blurring kernel and GT HR faces for facial feature detection. (d)(j) Results of [11, 24] with the GT blurring kernel and GT HR faces of same individuals. (e)(k) Results of patch-wise MPPCA prior in combination with estimated blurring kernel and transformation. (f)(l) Results of patch-wise MPPCA prior in combination with GT blurring kernel and estimated transformation. Number below images is PSNR. **The images are better viewed by zooming.**

Figure 3. Examples of masked faces used to measure PSNR.

Figure 4. GT blurring kernels and examples of estimated blurring kernels. (a) GT Gaussian blurring kernel with $\sigma = 0.6s = 2.4$. (b) An example of estimated blurring kernel of (a). (c) GT Gaussian blurring kernel with $\sigma = 0.4s = 1.6$. (d) An example of estimated blurring kernel of (c).

SSIM are summarized in Table 1, where the effect of accurate blurring kernel and transformation estimation becomes evident. The importance of prior can be observed by comparing results of weak generic prior in Fig. 1 (c) and Table 1(c), with results of specific face subspace prior in Fig. 1 (f)(h) and Table 1(f)(h). In general, large number of LR images is essential for generic image prior based SR to improve SR performance [8, 26, 19, 16].

The effect of improving face prior from PCA to patch-wise MPPCA can be noted by comparing results in Fig. 1 (f) and Table 1(f), with Fig.1(h) and Table 1(h). It can be seen that the upper bound of SR performance using patch-wise MPPCA prior is much improved in comparison with that of holistic PCA prior.

The combined effect of prior and blurring kernel and transformation estimation can also be observed by comparing results in Fig. 1 (g) and Table 1(g), with Fig.1(e)(f) and Table 1(e)(f). The proposed method, using patch-wise MPPCA prior in combination with estimated blurring kernel and transformations, not only is much better than that of PCA prior using the same estimated blurring kernel and transformations, but also surpass the upper bound of holistic PCA prior, where GT blurring kernel and transformations are utilized.

The proposed method is also compared with state of the art face SR methods using single LR face [28, 11, 24][1], which are all based on image-alignment and authors' implementations are adopted. 100 HR faces in PubFig83 database [20], which includes 83 individuals, are used for testing. There is at least one face for each individual, including some faces used in experiments of [11, 24]. Those faces, which are from individuals of different ages, genders, poses, expressions and illumination conditions, are blurred with Gaussian blurring kernel with $\sigma = 0.4s = 1.6$ and then sub-sampling by $s = 4$ without adding any noise. It follows similar test setting in their implementations in [28, 11], however there is no blurring applied in [11, 24] whereas LR faces are quite large with size about 30×30 in [28], which are much less challenging in comparison with LR faces used here. The GT blurring kernel is shown in Fig. 4(c) whereas an example of estimated blurring kernels in Fig. 4(d).

[1]Note that the performance of [11] was shown to be as good as that of [24], and it can serve as an indicator of the performance of [24].

Methods	HMRF [19]	PCA [3, 15]			MPPCA	
	(c)	(d)	(e)	(f)	(g)	(h)
Blurring	GT	GT	Est	GT	Est	GT
Trans.	GT	Initial	Est	GT	Est	GT
PSNR	19.67	19.52	21.27	21.51	21.78	22.32
SSIM	0.67	0.72	0.76	0.77	0.79	0.81

Table 1. Performance comparison of face SR using multiple simulated LR faces.

Methods	Bicubic	[28]	[11]	MPPCA	
		(c)(i)	(d)(j)	(e)(k)	(f)(l)
Blurring		GT	GT	Est	GT
PSNR	24.78	22.29	26.12	25.85	26.17

Table 2. Performance comparison of face SR using single simulated LR face.

As blurring kernel needs to be known beforehand in [28, 11, 24], GT blurring kernel are provided. In addition, GT HR faces are supplied to the method in [28] for more accurate facial feature detection, whereas other GT HR faces of same individuals are made available to the method in [11, 24] for SIFT flow estimation. Results are shown in Fig. 2. As the proposed method only super-resolves a facial region rather than a full image, performance is measured only in a pre-processed GT facial mask by excluding background pixels, so only PSNR is reported in Table. 2 due to irregular shape of facial masks as shown in Fig. 3. It is more accurate than unmasked measures in [28, 11, 24], which actually is biased due to a large portion of background pixels and similar background elsewhere in the dataset. Note that for all methods, only the luminance channel is super-resolved and then combined with bi-cubic interpolated color channels. It can be noted that results of patch-wise MPPCA using GT blurring kernel even outperforms those of [11], which have other GT HR faces of same individuals available, whereas results of patch-wise MPPCA using estimated blurring kernel in Fig. 2 (e)(k) are visually very close to those using GT blurring kernel in Fig. 2 (f)(l).

The proposed method works well for faces with poses from -30 to +30 degree, though it should be noted that errors in registration and blurring kernel may increase for faces with poses from ± 60 to ± 90 degree, as after all only frontal faces are used in training and homography becomes not sufficient. The performance of methods in [28, 11, 24]

Figure 5. Results of face hallucination in real LR sequences. (a) Multiple LR faces obtained by tracking forwardly and backwardly in sequences. (b) Results of bi-cubic interpolation. (c) Results of generic HMRF prior in combination with a fixed Gaussian blurring kernel with $\sigma = 0.4s$ and transformations initialized by tracking. (d) Results of PCA prior in combination with a fixed Gaussian blurring kernel with $\sigma = 0.4s$ and transformations initialized by tracking. (e) Results of the proposed method using patch-wise MPPCA in combination with an estimated blurring kernel and estimated transformations. **The images are better viewed by zooming.**

also relies on the accuracy of face alignment, however a whole dataset needs to be aligned to an LR face and usually efficient early pruning by PatchMatch is essential. In practice, their performance may further degrade due to unknown blurring kernel and unavailable GT HR faces of same individuals as shown in [24].

4.2. Real LR sequences

The proposed method was also applied to super-resolve LR faces in challenging real sequences, where LR faces in those sequences are usually not frontal and heavily blurred. GT of both blurring kernel and transformations are not available. LR faces with size range $12-30$ in the sequences are super-resolved by $s = 3$, where 5 consecutive fields are used for SR due to interlaced video format. Note that only LR faces, rather than whole images, in sequences are super-resolved, and only the luminance channel is super-resolved.

Examples of results using bi-cubic interpolation, generic HMRF prior [19] and PCA prior [3, 15], in combination with a fixed Gaussian blurring kernel with $\sigma = 0.4s$ and transformations initialized by tracking, are shown in Fig. 5(b-d), where multiple LR faces are obtained by tracking forwardly and backwardly in sequences. Results of the proposed method, using patch-wise MPPCA prior combined with estimated blurring kernel and transformations, are shown in Fig. 5 (e). It can be observed that although the non-frontal faces are heavily blurred, results are much improved to become clearer and sharper by the proposed method.

5. Conclusions

The paper presented a robust multi-image based blind face hallucination framework to super-resolve LR faces in image sequences. The proposed method first explored face PCA subspace, rather than original HR image space, for robust deblurring and registration. A new patch-wise MPPCA prior, rather than weak generic image priors, was then incorporated for face SR. Previous work on face SR using PCA prior becomes special cases of the framework. The proposed method improves face SR performance by combining robust blurring kernel and transformation estimation with the patch-wise MPPCA prior, and the experimental results in challenging simulated and real LR sequences shown the very promising performance of the proposed method.

References

[1] S. Baker and T. Kanade. Limits on super-resolution and how to break them. *IEEE TPAMI*, 24(9):1167–1183, 2002.

[2] S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework. *IJCV*, 56(3):221 – 255, 2004.

[3] D. Capel and A. Zisserman. Super-resolution from multiple views using learnt image models. In *Proc. IEEE CVPR*, pages 627–634, 2001.

[4] G. Dedeoglu, T. Kanade, and S. Baker. The asymmetry of image registration and its application to face tracking. *IEEE TPAMI*, 29(5):807–823, 2007.

[5] T. F.Cootes, G. J. Edwards, and C. J.Taylor. Active appearance models. *IEEE TPAMI*, 23(6):681–685, 2001.

[6] W. T. Freeman, E. C. Pasztor, and O. T. Carmichael. Learning low-level vision. *IJCV*, 20(1):25–47, 2000.

[7] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes, and R. M. Mersereau. Eigenface-domain super-resolution for face recognition. *IEEE TIP*, 12(5):597–606, 2003.

[8] R. Hardie, K. Barnard, and E. Armstrong. Joint map registration and high-resolution image estimation using a sequence of undersampled images. *IEEE TIP*, 6(12):1621–1633, 1997.

[9] S. Harmeling, S. Sra, M. Hirsch, and B. Scholkopf. Multiframe blind deconvolution, super-resolution, and saturation correction via incremental EM. In *Proc. IEEE ICIP*, 2010.

[10] Y. Hu, K.-M. Lam, G. Qiu, and T. Shen. From local pixel structure to global image super-resolution: A new face hallucination framework. *IEEE TIP*, 20(2):433–445, Feb. 2011.

[11] P. Innerhofer and T. Pock. A convex approach for image hallucination. In *Proc. DAGM AAPR Workshop*, 2013.

[12] O. Jesorsky, K. Kirchberg, and R. Frischholz. Robust face detection using the Hausdorff distance. In *Proc. IAPR International Conference on Audio- and Video-Based Personal Authentication*, pages 90–95, 2003.

[13] K. Jia and S. Gong. Generalized face super-resolution. *IEEE TIP*, 17(6):873–886, 2008.

[14] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman. Efficient marginal likelihood optimization in blind deconvolution. In *Proc. IEEE CVPR*, pages 2657–2664, 2011.

[15] C. Liu, H. Y. Shum, and W. T. Freeman. Face hallucination: Theory and practice. *IJCV*, 75(1):115–134, 2007.

[16] C. Liu and D. Sun. A Bayesian approach to adaptive video super resolution. In *Proc. IEEE CVPR*, pages 209–216, 2011.

[17] W. Liu, D. Lin, and X. Tang. Hallucinating faces: TensorPatch super-resolution and coupled residue compensation. In *Proc. IEEE CVPR*, pages 478–484, 2005.

[18] P. J. Philips, H. Moon, P. Rauss, and S. A. Rizvi. The FERET evaluation methodology for face-recognition algorithms. *IEEE TPAMI*, 22(10):1090–1104, 2000.

[19] L. C. Pickup, S. J. Roberts, and A. Zisserman. Optimizing and learning for super-resolution. In *Proc. BMVC*, 2006.

[20] N. Pinto, Z. Stone, T. Zickler, and D. Cox. Scaling up biologically-inspired computer vision: A case study in unconstrained face recognition on facebook. In *Proc. IEEE CVPR Workshop*, page 3542, 2011.

[21] D. A. Ross, J. Lim, R. S. Lin, and M. H. Yang. Incremental learning for robust visual tracking. *IJCV*, 77:125–141, 2000.

[22] R. Schultz and R. Stevenson. Extraction of high-resolution frames from video sequences. *IEEE TIP*, 5(6):996–1011, 1996.

[23] F. Sroubek, G. Cristobal, and J. Flusser. Simultaneous super-resolution and blind deconvolution. *Journal of Physics: Conference Series*, 2008.

[24] M. F. Tappen and C. Liu. A Bayesian approach to alignment-based image hallucination. In *Proc. ECCV*, 2012.

[25] M. E. Tipping and C. M. Bishop. Mixtures of probabilistic principal component analysers. *Neural Computation*, 11(2):443–482, 1999.

[26] M. E. Tipping and C. M. Bishop. Bayesian imge super-resolution. In *NIPS*, pages 1279–1286, 2003.

[27] X. Wang and X. Tang. Hallucinating face by eigentransformation. *IEEE TSMC(C)*, 35(3):425–434, 2005.

[28] C.-Y. Yang, S. Liu, and M.-H. Yang. Structured face hallucination. In *Proc. IEEE CVPR*, 2013.

[29] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE TIP*, 19(11):2861–2873, 2010.

[30] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. In *Proc. IEEE ICCV*, pages 479–486, 2011.