

Super-resolution from multiple views using learnt image models

David Capel and Andrew Zisserman
Robotics Research Group
Department of Engineering Science
University of Oxford
Oxford, OX1 3PJ

Abstract

The objective of this work is the super-resolution restoration of a set of images, and we investigate the use of learnt image models within a generative Bayesian framework.

It is demonstrated that restoration of far higher quality than that determined by classical maximum likelihood estimation can be achieved by either constraining the solution to lie on a restricted sub-space, or by using the sub-space to define a spatially varying prior. This sub-space can be learnt from image examples.

The methods are applied to both real and synthetic images of text and faces, and results are compared to Schultz and Stevenson’s MAP estimator [15]. We consider in particular images of scenes for which the point-to-point mapping is a plane projective transformation which has 8 degrees of freedom. In the real image examples, registration is obtained from the images using automatic methods.

1. Introduction

Super-resolution restoration aims to solve the following problem: given a set of observed images, estimate an image at a higher-resolution than is present in any of the individual images. The observed images are regarded as degraded observations of a real, high-resolution ‘texture’. These degradations typically include geometric warping, optical blur, spatial sampling and noise. Given several such observations, a *maximum likelihood* (ML) estimate of the super-resolution image may be obtained such that, when reprojected back into the images via a generative imaging model, it minimizes the difference between the actual and “predicted” observations. Successful ML algorithms have been engineered for geometric transformations between the observed images ranging from pure translation to an eight parameter projective transformation [10, 11, 12, 19].

Unfortunately, this estimation procedure becomes highly ill-conditioned as the extent of the super-resolution enlargement is increased. Consequently the estimate is extremely sensitive to noise in the observed images and errors in the parameters of the generative model, such as in the registration transformation. Attempts to regularize the solution are

typically cast as *maximum a posterior* (MAP) estimation, which allow the use of a Bayesian prior model of the super-resolution image. The traditional approach is to model the image as a first-order, stationary Markov Random Field (MRF) [2, 4, 6, 7, 15, 16] and include a *spatial* prior which models the spatial correlation between neighbouring pixels. Such priors are typically very simple, chosen to encourage a smooth solution (e.g. Gaussian MRF), possibly with some edge-preserving characteristics (e.g. Huber MRF, Generalized Gaussian MRF). They are also deliberately chosen to be convex functions of the image gradient/curvature in order to allow a deterministic optimization procedure to be used.

Although not properly explored in the super-resolution literature, it is widely recognized in the single-image restoration literature that placing *hard constraints* on the individual pixel intensities, which restrict the solution to some sub-space of the full image space, can give excellent results without the need for a spatial prior. Examples of such constraints are *non-negativity* and upper/lower bound-constraints on pixel values.

As a motivating example, figure 1 compares super-resolution estimates using both unconstrained and bound-constrained ML estimators. The 20 synthetic images are generated by perspective warping, blurring with a Gaussian ($\sigma = 0.7$ pixels), and down-sampling by a factor of 3. Three different levels of Gaussian noise are added to the synthetic images. Even for these tiny levels of noise, the unconstrained ML estimate is completely corrupted by reconstruction error. The bound-constrained estimate on the other hand is unaffected and of good quality.

Figure 2 compares the performance of the bound-constrained estimator to results obtained using Schultz and Stevenson’s edge-preserving MAP estimator which uses a Huber MRF prior (HMRF) (defined in section 2.3). In this case the noise on the synthetic images is a much more realistic $\sigma = 5$ grey-levels. The constrained result is clearly superior to both MAP results. The lack of any imposed spatial correlation means that the constrained estimator is able to generate sharper results than the MAP estimators using spatial MRF priors.

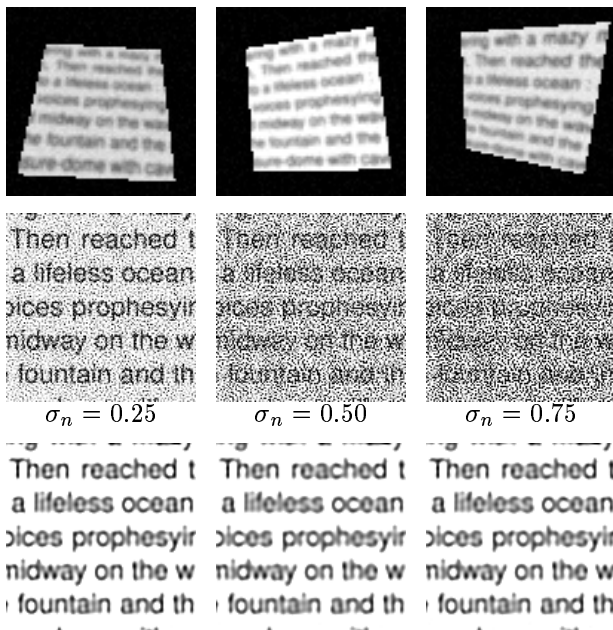


Figure 1: (Top) 3 of 30 synthetic perspective images generated as described in the text. (Middle) ML reconstructions at $3\times$ zoom for three different levels of Gaussian noise (in grey-levels) added to the low-resolution images. (Bottom) The same ML reconstructions with a 0/255 lower/upper bound on the pixel intensities. The reconstruction error has been almost entirely eliminated in the constrained cases.

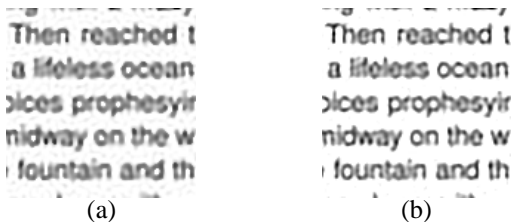


Figure 2: Reconstructions computed from the same 30 images used in figure 1, but with $\sigma = 5$ grey-levels of Gaussian noise added. (a) HMRF prior with $\lambda = 0.005$, $\alpha = 0.1$. (b) Bound-constrained. The constrained estimator out-performs the MAP estimator on this problem.

Figure 3 compares ML, HMRF and bound-constrained reconstructions of a sequence of 10 real images. The images were captured using a CCD camera which was translating parallel to the planar scene. Registration was obtained automatically using the feature-based method of [4]. The reconstruction is performed at $1.75\times$ pixel zoom. The ML estimate is badly corrupted, whilst the bound-constrained estimator is comparable in quality to the MAP estimate using the HMRF spatial prior. The bound-constrained quadratic minimization is performed using More & Toraldo’s “Gradient-Projection Conjugate Gradient” algorithm [13], which is efficient for the solution of very large, sparse systems with a large number of constraints, as here.

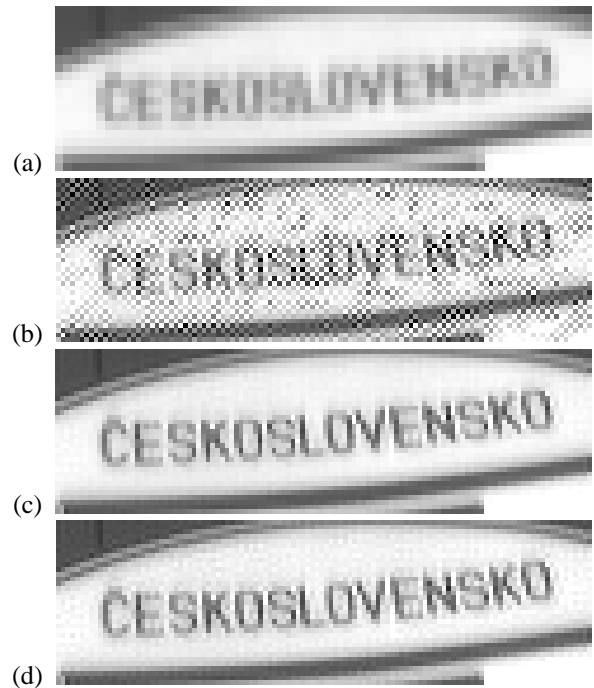


Figure 3: Super-resolution reconstructions from 10 real images taken by a CCD camera which was translated parallel to the planar scene. The images were acquired by Tomas Pajdla and Daniel Martinec at CMP Prague. The reconstruction are at $1.75\times$ pixel-zoom. (a) The region of interest in one of the low-resolution image. (b) The ML estimate. (c) The HMRF estimate with $\lambda = 5 \times 10^{-3}$, $\alpha = 5 \times 10^{-2}$ (d) The bound-constrained ML estimate. The ML estimate is badly corrupted, whilst the bound-constrained estimator produces a clean, clear result close to that of the HMRF estimator.

In this paper we explore learning such constrained models and priors from training images [18], and applying these to super-resolution restoration. We concentrate here on classes of scenes where intensity neighbourhood distributions are *not* spatially homogeneous. The target application is faces where the intensity distribution on the cheek, for example, is very different from that around the eye, as opposed to brick texture where a homogeneous representation is more appropriate. Here, the constrained model is a sub-space which is computed using principal component analysis of a set of registered face images. This sub-space is also used to define a spatially varying prior within a MAP estimator.

The methods of this paper are similar in spirit to the approach of Baker and Kanade [1] who also examined the use of spatially varying priors for face images. Their input consists of images which are all at the same scale and are related by translations. The novelty of our method is two-fold: first, the low resolution images need not be at the same resolution and indeed the resolution can vary across the image.

This generalization is essential in the case that the low resolution images are related by a transformation more general than 2D pure translation. In the work of this paper the images are related by an 8 degrees of freedom plane projective transformation; second, a generative model is used throughout – both in the ML case with a likelihood function constrained to a sub-space, or in the MAP case with the prior measuring distance from a sub-space.

The use of PCA sub-space priors provides an interesting middle ground between the generic local priors previously used, and the type of priors based on sampling exemplars from training images (e.g. for homogeneous texture generation [8]) which have been applied to super-resolution by Candocia and Principe [3], and Freeman and Pasztor [9]

2. Super-resolution background

This section summarizes the standard super-resolution methods and defines the notation used throughout the paper.

2.1. Generative imaging model

It is assumed that the set of observed low-resolution images were produced by a single high-resolution image under the following generative model

$$g_n = s\downarrow(h * \mathcal{T}_n \bar{f}) + \eta \quad (1)$$

- \bar{f} - ground truth, high-resolution image
- g_n - n^{th} observed low-resolution image
- \mathcal{T}_n - geometric transformation of n^{th} image
- h - point spread function
- $s\downarrow$ - down-sampling operator by a factor S
- η - noise term

More explicitly

$$g_n(x, y) = s\downarrow(h(u, v) * \bar{f}(\mathcal{T}_n(x, y))) + \eta(x, y) \quad (2)$$

Transformation \mathcal{T} is assumed to be projective. The point spread function h is assumed to be linear and spatially invariant. The noise η is assumed to be Gaussian with mean zero.

After discretization, the model can be expressed in matrix form as

$$\mathbf{g}_n = \mathbf{M}_n \bar{\mathbf{f}} + \boldsymbol{\eta} \quad (3)$$

in which the operators \mathcal{T}_n , h and $s\downarrow$ have been combined into a single matrix \mathbf{M}_n .

Finally, the generative models of all N images are stacked vertically to form an over-determined linear system

$$\begin{bmatrix} \mathbf{g}_0 \\ \mathbf{g}_1 \\ \vdots \\ \mathbf{g}_{N-1} \end{bmatrix} = \begin{bmatrix} \mathbf{M}_0 \\ \mathbf{M}_1 \\ \vdots \\ \mathbf{M}_{N-1} \end{bmatrix} \bar{\mathbf{f}} + \boldsymbol{\eta} \quad (4)$$

$$\mathbf{g} = \mathbf{M} \bar{\mathbf{f}} + \boldsymbol{\eta} \quad (5)$$

Knowledge of the PSF for any given image sequence is usually unavailable, so here it is modelled as a Gaussian. Comparisons with the measured PSF of several CCD based imaging systems show this approximation to be quite reasonable and this is further verified by good super-resolution results obtained on real images. A procedure for estimating the PSF is described in [14].

2.2. Maximum likelihood estimation

Assuming the image noise to be Gaussian with mean zero, variance σ^2 , the total probability of the observed image g_n given an estimate of the super-resolution image $\hat{\mathbf{f}}$ is

$$\Pr(g_n | \hat{\mathbf{f}}) = \prod_{\forall x, y} \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(\hat{g}_n(x, y) - g_n(x, y))^2}{2\sigma^2}\right) \quad (6)$$

and hence the associated negative log-likelihood function is

$$\mathcal{L}(g_n) = -\sum_{\forall x, y} (\hat{g}_n(x, y) - g_n(x, y))^2 \quad (7)$$

The maximum likelihood estimate \mathbf{f}_{ML} is obtained by maximizing this function over all observed images.

$$\mathbf{f}_{ML} = \arg \max_{\mathbf{f}} \sum_n \mathcal{L}(g_n) \quad (8)$$

The ML formulation given above is simply a very large, sparse system of linear equations. Analysis of the eigenvalues of this system shows that in general it is extremely poorly conditioned. However, in principle the solution of equation (8) in the matrix notation of equation (5) is given by

$$\hat{\mathbf{f}} = (\mathbf{M}^T \mathbf{M})^{-1} \mathbf{M}^T \mathbf{g} = \mathbf{M}^+ \mathbf{g} \quad (9)$$

where \mathbf{M}^+ is the pseudo-inverse of \mathbf{M} . In practice the matrices are too large to compute the pseudo-inverse directly, and a solution is obtained by iterative methods.

2.3. Maximum a posterior estimation

As noted above the basic ML estimator is highly ill-conditioned. However, if a prior probability distribution on the super-resolution image is available then this information may be used to “regularize” the estimation. The MAP estimator has the form:

$$\mathbf{f}_{MAP} = \arg \max_{\mathbf{f}} \left(\sum_n \mathcal{L}(g_n) + \lambda \mathcal{L}(\mathbf{f}) \right) \quad (10)$$

where $\mathcal{L}(\mathbf{f})$ provides a measure of the likelihood of a particular estimate \mathbf{f} .

For example, Schultz and Stevenson [15] propose a prior which applies a penalty to the image curvature. The MAP estimator has the form

$$f_{\text{MAP}} = \arg \max_f \left(\sum_n \mathcal{L}(g_n) + \lambda \sum_{\forall x,y} \rho(K(f, x, y)) \right) \quad (11)$$

where the prior is defined by the spatial-activity function $K(f, x, y)$, defined as a sum of second-order finite differences in the horizontal, vertical and diagonal directions, and the penalty function $\rho(x)$ which is the Huber function,

$$\begin{aligned} \rho(x) &= x^2, & \text{if } x \leq \alpha \\ &= 2\alpha |x| - \alpha^2, & \text{otherwise} \end{aligned}$$

This prior encourages local smoothness, whilst being more lenient toward step edges than $\rho(x) = x^2$.

3 Super-resolution using a sub-space

In this section the image is modelled using a PCA basis computed from training images at the target resolution. Given a set of images the variation in the signals is optimally modelled for a given number of principal components. This representation is applicable to signals of a particular class where the aim is to model the within-class variation.

The image will be modelled by its PCA components as $\hat{\mathbf{f}} = \mathbf{V}\mathbf{y} + \boldsymbol{\mu}$, where \mathbf{V} represents the set of PC basis vectors and $\boldsymbol{\mu}$ is the average of the training images. The aim is a low dimensional representation where the dimension of the parameter vector \mathbf{y} is far less than the number of pixels in the image (the dimension of \mathbf{f}). This representation is applied here to registered face images. It is known from extensive use of PCA since at least [17] that a compact representation can be achieved in this case.

Rather than learn the PC for the entire image, the face is divided into four key facial regions: the eyes, nose, mouth and cheek areas; and a separate PCA basis is learnt for each feature. The intuition is that these regions are relatively uncorrelated, and that by considering small regions, better models can be learnt than would be by performing PCA on the whole face. The resulting model is analogous to the ‘‘IdentiKit’’ system often used to compose images of police suspects.

The region segmentation is shown in figure 4. To learn the model, PCA analysis is applied to 160 face images which have been geometrically registered using a similarity transformation. Both male and female faces are used in the training set. To increase the number of training samples, each face is flipped symmetrically around the y-axis, providing a total of 320 faces. The average features and the first few ‘‘eigen-features’’ are shown in figure 5. The composed average regions are shown in figure 4(b).

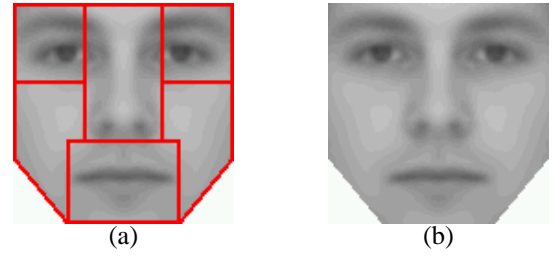


Figure 4: (a) The regions of the face over which independent PCA bases are computed. (b) The average image computed over each of the regions independently using the 320 training images (160 \times 2 using symmetry.)

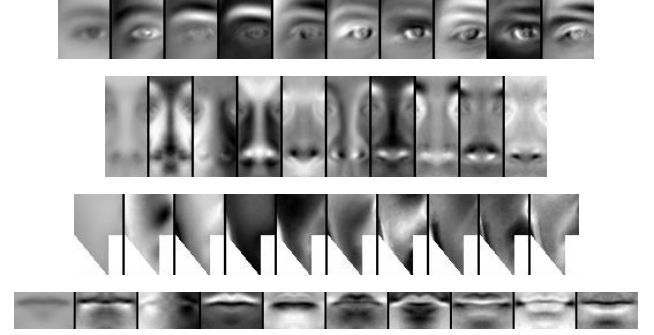


Figure 5: The first 10 eigenimages of each of the four facial features.

A face image is thus represented here by 6 sets of principal component coefficients (one for each of the left and right eyes, nose, mouth, left and right cheeks),

$$\begin{aligned} \mathbf{f} &= [\mathbf{V}_{lefteye} \mathbf{V}_{righteye} \dots \mathbf{V}_{rightcheek}] \mathbf{y} + \boldsymbol{\mu} \\ &= \mathbf{V}\mathbf{y} + \boldsymbol{\mu} \end{aligned} \quad (12)$$

where the matrices \mathbf{V} are the projection matrices whose columns form the PCA basis for each feature, \mathbf{y} is the stacked vector of coefficients associated with each basis, and $\boldsymbol{\mu}$ is a vector containing the average image features for each region. Using the PCA model, the number of parameters required to represent a face image is reduced from (in this case) 120×120 pixels, down to at most $6 \times 319 = 1914$ PC coefficients.

When using the model, we often do not want to use *all* 319 basis vectors per feature. For instance, we would intuitively expect that the cheek, which is a smooth, low-detail feature, will required fewer principal components to accurately represent it than (say) the eye. Also, by reducing the number of components used to represent each feature to the lowest acceptable number, we are reducing the dimensionality of the model still further, and we would expect this to improve the condition of the super-resolution estimators. The criterion we will use to choose an appropriate dimensionality for each feature is to keep the minimum number of

ν	0.95	0.98	0.99	0.995	1.0
Eye	27	64	103	149	319
Nose	39	92	142	190	319
Mouth	22	59	105	157	319
Cheek	8	29	69	126	319

Table 1: The number of principal components required to represent a particular fraction ν of the total variance within each feature.

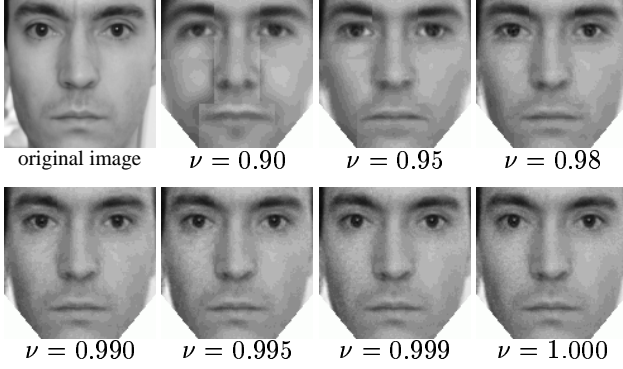


Figure 6: A face image **not** in the training set is projected onto the PCA face model. The number of components used to represent each feature is chosen to span some fraction (shown underneath) of the total variance.

components that span some fraction ν of the total variance. This calculation is extremely simple, since the variances of the principal component are obtained as a by-product of the PCA. Table 1 shows the required dimensionality of each feature in order to span various fractions of the total variance.

To demonstrate the ability of the model to represent a new face figure 6 shows a face which was *not in the training set* projected onto the PCA model, using feature dimensions chosen to preserve increasing fractions of the total variance. Reconstructions for which $\nu \geq 0.99$ are very close to the ground-truth.

3.1 An ML estimator

The simplest way to use the PCA model is to constrain the super-resolution reconstruction to lie on the sub-space defined by the column span of V . The ML solution using the sub-space parameterization follows directly from equation (8):

$$\mathbf{f}_{mle} = \arg \min_{\mathbf{y}} \|\mathbf{M}(\mathbf{V}\mathbf{y} - \boldsymbol{\mu}) - \mathbf{g}\|^2 \quad (13)$$

for which the minimizer is given by

$$\mathbf{V}^T \mathbf{M}^T \mathbf{M} \mathbf{V} \mathbf{y} = \mathbf{V}^T \mathbf{M}^T (\mathbf{g} - \mathbf{M}\boldsymbol{\mu}) \quad (14)$$

3.2 MAP estimators

There are two straightforward ways to develop priors based on the face model, and these are now described.

A prior over face-space (FS-MAP) A by-product of the PCA is an estimate of the variance of each principal component. This immediately gives us a very simple prior defined on the coefficients of the principal components. We use this to augment the ML estimator above, producing a MAP estimator :

$$\mathbf{f}_{map} = \arg \min_{\mathbf{y}} \|\mathbf{M}(\mathbf{V}\mathbf{y} - \boldsymbol{\mu}) - \mathbf{g}\|^2 + \lambda \mathbf{y}^T \boldsymbol{\Sigma}^{-1} \mathbf{y}$$

for which the minimizer is given by

$$(\mathbf{V}^T \mathbf{M}^T \mathbf{M} \mathbf{V} + \lambda \boldsymbol{\Sigma}^{-1}) \mathbf{y} = \mathbf{V}^T \mathbf{M}^T (\mathbf{g} - \mathbf{M}\boldsymbol{\mu}) \quad (15)$$

where $\boldsymbol{\Sigma}$ is the diagonal matrix of component variances obtained from the PCA. We refer to this as a prior over face-space since, as with the ML estimator, the solution is constrained to lie on the sub-space defined by the PCA model, and the prior is defined in the sub-space. This estimator imposes a penalty proportional to the Mahalanobis distance of the features in \mathbf{y} to the average features in $\boldsymbol{\mu}$.

A prior over image-space (IS-MAP) A different way to use the learnt model is as part of a prior which encourages the estimated image to lie *near* to the PCA sub-space. We assume that the probability of obtaining a super-resolution image $\hat{\mathbf{f}}$ is Gaussian in the distance of $\hat{\mathbf{f}}$ from V . The resulting MAP estimator has the form

$$\mathbf{f}_{MAP} = \arg \min_{\mathbf{f}} \|\mathbf{M}\mathbf{f} - \mathbf{g}\|^2 + \lambda \|(\mathbf{I} - \mathbf{V}\mathbf{V}^T)(\mathbf{f} - \boldsymbol{\mu})\|^2$$

The minimizer is given by

$$(\mathbf{M}^T \mathbf{M} + \lambda(\mathbf{I} - \mathbf{V}\mathbf{V}^T)) \mathbf{f} = \mathbf{M}^T \mathbf{g} + \lambda(\mathbf{I} - \mathbf{V}\mathbf{V}^T) \boldsymbol{\mu} \quad (16)$$

We refer to this as a prior over image-space, since the Gaussian distribution attached to the face-space defines a prior over all images \mathbf{f} . The solution is not constrained to lie on the face-space.

For all three proposed estimators, the optimization is efficiently performed using the conjugate gradient descent algorithm. Note however that the ML and FS-MAP estimators require optimization over the compact face-model parameters \mathbf{y} , and can therefore be computed much more rapidly than the IS-MAP, which is parameterized in terms of the actual super-resolution pixels \mathbf{f} .

4 Examples using real images

The behaviour of the estimators is examined in detail on synthetic data in [5]. Here we only include results on real



Figure 7: Five frames from a sequence of 25 showing a moving face. The face occupies 40×40 pixels in these images.



Figure 8: For the sequence shown in figure 7 (a) the region of interest in one of the input images, (b) the average image, (c) the super-resolution HMRF-MAP estimate at $3\times$ zoom with $\lambda = 0.025$, $\alpha = 0.05$.

image sequences. Two cases are considered : a moving face and fixed camera, and vice versa. In both cases, registration is obtained automatically using the feature-based method of [4]. The only manual input is the requirement to select the eyes and mouth in one of the input images in order to register the faces in the sequence with the PCA model.

Fixed camera, moving face Figure 7 shows 5 frames from a sequence of 25 of a moving face captured using a monochrome Cohu CCD camera. The face occupies a 40×40 pixel region. The super-resolution reconstructions are 120×120 pixels, i.e. $3\times$ pixel zoom. All 25 images are used in forming the reconstructions.

For the purpose of later comparison, figure 8 shows the region of interest in one of the input images; the average image; and the super-resolution reconstruction using the HMRF-MAP estimator with $\lambda = 0.025$, $\alpha = 0.05$ (tuned by trial-and-error).

Figure 9 shows reconstructions using the face-space constrained ML estimator as the number of components per features varies. The constrained MLE performs quite well on this data set. As ν increases, the reconstruction improves, moving away from the average face μ . But for $\nu > 0.995$, the reconstruction quality decreases rapidly as the more spurious and unstructured components allow noise to be introduced into the solution. The reconstruction for which $\nu = 0.995$ shows good detail with relatively little noise.

Figure 10 shows reconstructions using the FS-MAP estimator as λ varies. The quality of reconstruction is consistently quite high, even though λ varies over large range. This is probably due to the fact that the variance of the principal components falls off very rapidly. For instance, in

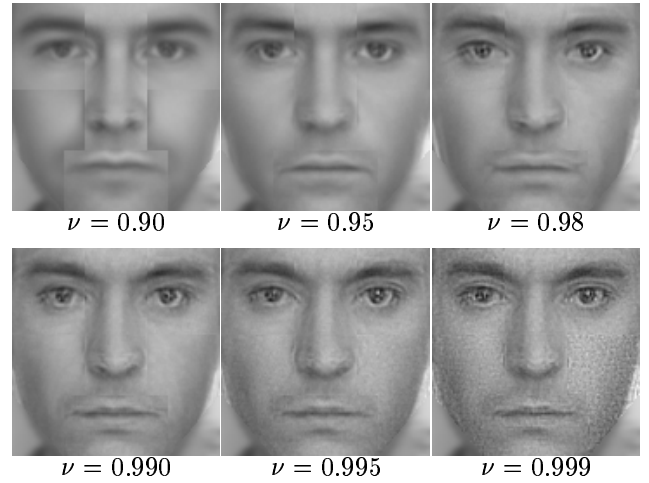


Figure 9: Face-space constrained ML estimates at $3\times$ zoom as the number of components per feature increases. As ν increases the estimate moves away from the average face μ .



Figure 10: FS-MAP reconstructions as λ varies. The quality of the reconstruction is consistently high over a large range of values of λ . As λ increases, the solution tends toward the average face μ .

the case of the mouth components, the variance of the 40^{th} component is 1000 times smaller than that of the 1^{st} component. Consequently, even a small contribution from this prior is enough to severely dampen all the but the first few tens of principal components. As λ increases, \mathbf{y} is forced towards 0, and hence the solution tends toward μ . When λ is zero, the solution is equivalent to the face-space constrained ML.

Figure 11 shows the reconstructions using the IS-MAP estimator as λ varies, ν being fixed as 0.99. When $\lambda = 0.005$, the estimate tends toward the unconstrained MLE of equation 8, and reconstruction error is evident. But when $\lambda = 0.05$, the reconstruction is very good.

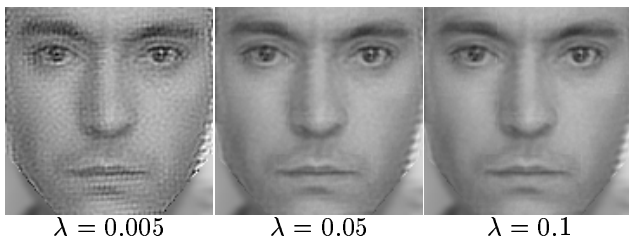


Figure 11: IS-MAP reconstructions as λ varies, ν being fixed as 0.99. As λ tends to zero the reconstruction tends towards the constrained MLE of figure 9

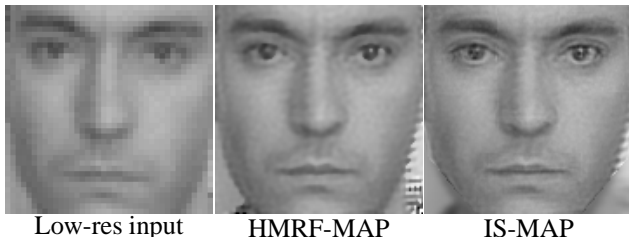


Figure 12: A comparison of the quality of the input imagery with the HMRF-MAP estimate, and the face-model IS-MAP.

Finally, figure 12 compares the quality of the input images, the HMRF-MAP estimate and the IS-MAP estimate with $\lambda = 0.05$, $\nu = 0.99$.

Fixed face, moving camera Figure 13 shows a mosaic featuring a face. The mosaic was created from a sequence of 30 PAL size, JPEG compressed images captured using a Cohu CCD camera which was rotated on a tripod. Nine of the original frames are also shown. The face occupies only 30×30 pixels in the low-resolution images. The mosaic was created using the method described in [4].

Figure 14 shows reconstructions using the FS-MAP and IS-MAP face-space estimators with various parameters settings. For comparison, one of input images, the average image, and an reconstruction using the HMRF-MAP estimator are also shown. The size of the reconstructed image is 120×120 pixels, 16 times as many pixels as in the 30×30 pixel low-resolution region of interest. The $4\times$ zoom ratio and the fairly poor quality of the input imagery means that the face-space ML estimator does not produce good results and is omitted here. In this example, the quality of the IS-MAP estimates is arguably equal or superior to the HMRF estimate, although some artifacts are visible on the boundaries between the different facial features.

5 Summary and future work

A new method of performing super-resolution restoration of low-resolution image sequence has been described which uses learnt image models either to directly constrain the ML



Figure 13: (Top) 9 of the 30 PAL size, JPEG compressed images captured using a rotating CCD camera. (Bottom) A mosaic created from the images sequence. The mosaic features a face to which we apply super-resolution restoration.

estimate or as a prior on the ML estimate. In both cases it has been demonstrated that the instability present in ‘pure’ ML estimation is removed, and that the learnt model gives superior results to traditional MAP estimators.

The quality of the results based on the face-space model may be improved by using a larger set of training images, or by employing a more sophisticated method of registering the face images, such as one based on a deformable mesh. The artifacts which are sometimes visible on the boundaries between the different features may be ameliorated by including a prior term which penalizes spatial incoherence across the boundaries.

We are currently looking at a more sophisticated technique for learning constrained image models which is based on learning a lexicon of exemplars which accurately and compactly represent the high-frequency characteristics of the particular image class. This method will be applicable to a wide range of image classes.

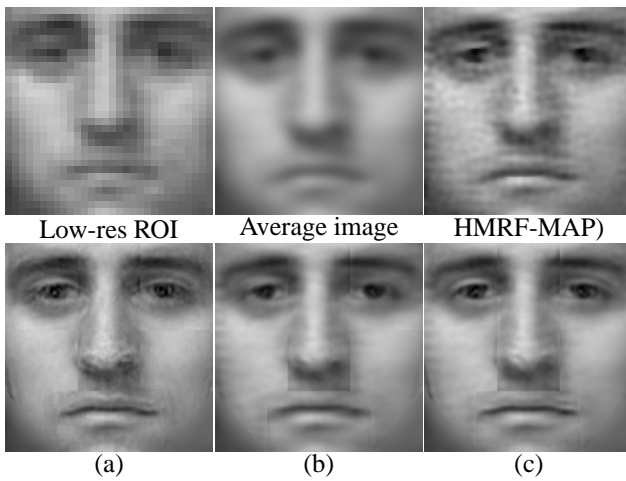


Figure 14: Super-resolution estimation from 30 images. (Top) The region of interest in one of the low-resolution input images; the average image; and the HMRf-MAP estimate with $\lambda = 0.025$, $\alpha = 0.075$. (Bottom) Super-resolution reconstructions using (a) FS-MAP with $\lambda = 0.1$, (b) IS-MAP with $\lambda = 0.1$, $\nu = 0.90$, and (c) IS-MAP with $\lambda = 0.1$, $\nu = 0.95$. The size of the reconstructed image is 120×120 pixels, 16 times as many pixels as in the 30×30 pixel low-resolution region of interest. The quality of the IS-MAP estimates is arguably equal or superior to the HMRf estimate, although some artifacts are visible on the boundaries between the different facial features.

References

- [1] S. Baker and T. Kanade. Limits on super-resolution and how to break them. In *Proc. CVPR*, 2000.
- [2] S. Borman and R.L. Stevenson. Simultaneous multi-frame MAP super-resolution video enhancement using spatio-temporal priors. In *Proc. ICIP*, 1999.
- [3] F. M. Candocia and J. C. Principe. Super-resolution of images based on local correlations. *IEEE-NN*, 10(2):372, 1999.
- [4] D. Capel and A. Zisserman. Automated mosaicing with super-resolution zoom. In *Proc. CVPR*, pages 885–891, Jun 1998.
- [5] D.P. Capel. *Image Mosaicing and Super-resolution*. PhD thesis, University of Oxford, 2001.
- [6] P. Cheeseman, B. Kanefsky, R. Kraft, and J. Stutz. Super-resolved surface reconstruction from multiple images. Technical report, NASA, 1994.
- [7] F. Dellaert, S. Thrun, and C. Thorpe. Mosaicing a large number of widely dispersed, noisy, and distorted images: A bayesian approach. Technical report, School of Computer Science, Carnegie Mellon University, 1999.
- [8] A. Efros and T. Leung. Texture synthesis by non-parametric sampling. In *Proc. ICCV*, pages 1039–1046, Sep 1999.
- [9] W. Freeman and E. Pasztor. Learning low-level vision. In *Proc. ICCV*, pages 1182–1189, 1999.
- [10] M. Irani and S. Peleg. Improving resolution by image registration. *GMIP*, 53:231–239, 1991.
- [11] M. Irani and S. Peleg. Motion analysis for image enhancement: resolution, occlusion, and transparency. *J. Visual Communication and Image Representation*, 4:324–335, 1993.
- [12] S. Mann and R. W. Picard. Virtual bellows: Constructing high quality stills from video. In *Proc. ICIP*, 1994.
- [13] J. More and G. Toraldo. On the solution of large quadratic programming problems with bound constraints. *SIAM J. Optimization*, 1(1):93–113, 1991.
- [14] S. E. Reichenbach, S. K. Park, and R. Narayanswamy. Characterizing digital image aquisition devices. *Optical Engineering*, 30(2):170–177, 1991.
- [15] R. R. Schultz and R. L. Stevenson. Extraction of high-resolution frames from video sequences. *IEEE Transactions on Image Processing*, 5(6):996–1011, Jun 1996.
- [16] V.N. Smelyanskiy, P. Cheeseman, D. Maluf, and R. Morris. Bayesian super-resolved surface reconstruction from images. In *Proc. CVPR*, pages I:375–382, 2000.
- [17] M. Turk and A.P. Pentland. Face recognition using eigenfaces. In *Proc. CVPR*, pages 586–591, 1991.
- [18] S.C. Zhu and D. Mumford. Learning generic prior models for visual computation. In *Proc. CVPR*, pages 463–469, 1997.
- [19] A. Zomet and S. Peleg. Applying super-resolution to panoramic mosaics. In *Proc. WACV*, 1998.