

Bayesian Methods for Image Super-Resolution

LYNDSEY C. PICKUP^{1*}, DAVID P. CAPEL², STEPHEN J. ROBERTS¹ AND ANDREW ZISSERMAN¹

¹Information Engineering Building, Department of Engineering Science, Parks Road, Oxford OX1 3PJ, UK
²2D3 Ltd

*Corresponding author: elle@robots.ox.ac.uk

We present a novel method of Bayesian image super-resolution in which marginalization is carried out over latent parameters such as geometric and photometric registration and the image point-spread function. Related Bayesian super-resolution approaches marginalize over the high-resolution image, necessitating the use of an unfavourable image prior, whereas our method allows for more realistic image prior distributions, and reduces the dimension of the integral considerably, removing the main computational bottleneck of algorithms such as Tipping and Bishop's *Bayesian image super-resolution*. We show results on real and synthetic datasets to illustrate the efficacy of our method.

Keywords: super-resolution; Bayesian; image prior; image registration

Received 22 January 2007; revised 11 September 2007

1. INTRODUCTION

Problem of fusing information from several low-resolution images into a single clearer high-resolution image is referred to as multi-frame super-resolution. The output image has a higher spatial resolution, or includes more visible detail in the high spatial frequencies [1]. Super-resolution can be used to improve the aesthetic quality of images for media publication [2], or to feed into higher-level vision tasks such as object recognition or localization [3].

The low-resolution images taken as inputs can come from a wide variety of sources, from television broadcasts and holiday snaps to surveillance footage and satellite terrain imagery. These sets of low-resolution images are related to the high-resolution version of the scene by imaging parameters, including the point-spread function (PSF), lighting variations and geometric warpings. In order to perform super-resolution by exploiting the relative sub-pixel motion between the scene and the imaging planes, we need to estimate these imaging parameters, but no matter how accurate the estimate, there will be some residual uncertainty associated with the values [4].

This paper develops a scheme to deal with this problem by performing Bayesian marginalization over the uncertainty in the set of estimated imaging parameters, leading to an iterative algorithm to estimate a high-resolution image. These parameters are also known as *nuisance parameters* because they are not directly part of the desired

output of the algorithm, which in this case is the high-resolution image.

We compare our results to standard techniques, and show improved results on synthetic and real digital image datasets. We also examine an alternative Bayesian approach [5] and show that our method has several advantages in terms of high-resolution image quality and computation time.

1.1. Background

Parameter estimation for the super-resolution problem and the computation of the high-resolution image are usually treated as distinct processes, to be considered sequentially [1, 6–12]. Hardie *et al.* [3] demonstrated that the low-resolution image registration can be updated using the super-resolution image estimate, and that this improves a *maximum a posteriori* (MAP) super-resolution image estimate. More recently, Pickup *et al.* [13] used a similar joint MAP approach to learn more general geometric and photometric registrations, the super-resolution image and the values for the prior's parameters simultaneously.

Tipping and Bishop's *Bayesian Image Super-Resolution* work [5] uses a maximum likelihood (ML) point estimate of the registration parameters and the camera imaging blur, found by integrating the high-resolution image out of the registration problem and optimizing the marginal probability of the observed low-resolution images directly. This gives

an improvement in the accuracy of the recovered registration (measured against known truth on synthetic data) compared to the MAP approach.

Tipping and Bishop's method has two limitations: (i) it is restricted to a Gaussian image prior in order for the marginalization to remain tractable, but others have shown that the improved image super-resolution results are produced using distributions with heavier tails [7, 14], and (ii) it is computationally expensive due to the very large matrices required by the algorithm. In contrast, our approach allows for a more realistic image prior and operates with smaller matrix sizes.

In the following, we describe both these methods (Tipping and Bishop's and ours) and compare their performance. We give an overview of the generative model that forms the basis for both techniques in Section 2. In Section 3, we outline the derivation of both marginalization approaches: our marginalization over the registration parameters in Section 3.1, and the marginalization of the image pixel values of [5] in Section 3.2. Implementation details for our method are given in Section 3.3. Section 4 evaluates the results on several synthetically generated sequences—which provide the ground truth high-resolution images for comparison and performance evaluation—and on a real data example, with comparisons being drawn with Tipping and Bishop's method and the standard MAP algorithm. A discussion of the registration-marginalizing Bayesian approach is given in Section 5, with some intuition as to why it works. Concluding remarks are in Section 6.

Preliminary findings for registration-only problems were published in [15], but in this paper we outline an extended model capable of handling the PSF, and include further experiments and comparisons with alternative methods.

2. GENERATIVE MODEL

Generative model for multi-frame super-resolution assumes that a known high-resolution image generates each low-resolution image according to pre-determined geometric and photometric transforms, and a given PSF kernel.

Formally, the high-resolution image \mathbf{x} (vectorized, size $N \times 1$) generates K low-resolution images, each of which is $\mathbf{y}^{(k)}$ (vectorized, size $M \times 1$). The set of K such images $\{\mathbf{y}^{(k)}\}$ is generated from geometric registration parameters $\{\boldsymbol{\theta}^{(k)}\}$ and photometric registration parameters $\{\boldsymbol{\lambda}^{(k)}\}$. A common blur parameter γ , corresponding to the standard deviation of a Gaussian blur kernel, is used for all the images in the sequence. Gaussian *i.i.d.* noise with precision (inverse variance) β is then added to each $\mathbf{y}^{(k)}$,

$$\mathbf{y}^{(k)} = \lambda_{\alpha}^{(k)} \mathbf{W}(\boldsymbol{\theta}^{(k)}, \gamma) \mathbf{x} + \boldsymbol{\lambda}_{\beta}^{(k)} + \mathcal{N}(\mathbf{0}, \beta^{-1} \mathbf{I}). \quad (1)$$

Photometric parameters λ_{α} and λ_{β} provide a global linear correction for the scene illumination, and $\boldsymbol{\lambda}_{\beta}$ is simply an $M \times 1$ vector filled out with the value of λ_{β} . Each row of $\mathbf{W}^{(k)}$ constructs a single pixel in $\mathbf{y}^{(k)}$; the row's entries are formed from the vectorized PSF response for each low-resolution pixel, mapped into the frame of the super-resolution image [5, 7, 16]. The PSF is usually assumed to be an isotropic Gaussian on the imaging plane, though a disk may also be used, in particular to model out-of-focus blurring.

For a set of low-resolution images, given their registrations, the PSF, and the high-resolution image \mathbf{x} , the data likelihood is

$$p\left(\left\{\mathbf{y}^{(k)}\right\} \middle| \mathbf{x}, \left\{\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\right\}, \gamma\right) = \prod_{k=1}^K p(\mathbf{y}^{(k)} | \mathbf{x}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}, \gamma) \quad (2)$$

$$= \prod_{k=1}^K \left(\frac{\beta}{2\pi}\right)^{M/2} \exp\left\{-\frac{\beta}{2} \left\|\mathbf{r}^{(k)}\right\|_2^2\right\}, \quad (3)$$

where the residual

$$\mathbf{r}^{(k)} = \mathbf{y}^{(k)} - \lambda_{\alpha}^{(k)} \mathbf{W}(\boldsymbol{\theta}^{(k)}, \gamma) \mathbf{x} - \boldsymbol{\lambda}_{\beta}^{(k)} \quad (4)$$

and the distribution over $\mathbf{r}^{(k)}$ comes from the Gaussian noise assumption. Optimizing this likelihood with respect to the pixels in \mathbf{x} gives the ML solution. This has a closed form, though it is usual to solve it iteratively for all but the smallest of super-resolution images. The system is ill-conditioned, however, so in general a prior distribution $p(\mathbf{x})$ over the high-resolution image \mathbf{x} is also required.

The most popular approach to super-resolution is to obtain a MAP estimate, typically using an iterative scheme to maximize $p(\mathbf{x} | \{\mathbf{y}^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\}, \gamma)$ with respect to \mathbf{x} , where

$$p\left(\mathbf{x} \middle| \left\{\mathbf{y}^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\right\}, \gamma\right) = \frac{p(\mathbf{x}) \prod_{k=1}^K p(\mathbf{y}^{(k)} | \mathbf{x}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}, \gamma)}{p(\{\mathbf{y}^{(k)}\} | \{\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\}, \gamma)} \quad (5)$$

and the denominator is an unknown scaling factor. Gradient descent methods on the negative log of the numerator can be used to compute \mathbf{x} easily for any of a wide range of priors $p(\mathbf{x})$ that are convex in \mathbf{x} . This leads to an objective function \mathcal{L} of the form

$$\mathcal{L} = -\log(p(\mathbf{x})) - \frac{\beta}{2} \sum_{k=1}^K \left\|\mathbf{r}^{(k)}\right\|_2^2. \quad (6)$$

We will now briefly consider a few useful forms for $p(\mathbf{x})$ before proceeding with the Bayesian marginalization.

2.1. Forms of the image prior

The most common failing of an ML super-resolution image is that it contains a lot of excess high-frequency noise [6, 7]. To combat this problem, image priors are generally chosen to promote a certain degree of smoothness in the super-resolution image, often by penalizing gradients or higher-order spatial derivatives in the image.

In this work, a Huber prior is assumed over directional image gradients $\mathbf{D}\mathbf{x}$ in the super-resolution image \mathbf{x} ; in this case, the directions are horizontal, vertical and two diagonals, so that

$$p(\mathbf{x}) = \frac{1}{\mathcal{Z}_x} \exp\left\{-\frac{\nu}{2} \rho(\mathbf{D}\mathbf{x}, \alpha)\right\}, \quad (7)$$

$$\rho(\mathbf{z}, \alpha) = \sum_i \rho(z_i, \alpha), \quad (8)$$

$$\rho(z, \alpha) = \begin{cases} z^2, & \text{if } |z| < \alpha, \\ 2\alpha|z| - \alpha^2, & \text{otherwise,} \end{cases} \quad (9)$$

where α is a parameter of the Huber potential function, ν the prior strength parameter and \mathcal{Z}_x is required to make the distribution normalize. This belongs to a family of functions often favoured over Gaussians for super-resolution image priors [7, 4, 6] because the Huber distribution's heavy tails mean image edges are penalized less severely. The difficulty in computing the partition function \mathcal{Z}_x is a consideration when marginalizing over \mathbf{x} as in [5], though for the MAP image estimate, a value for this scale factor is not required.

At the limit as $\alpha \rightarrow 0$, the Huber prior is equivalent to the total variation (TV) prior, or L_1 norm of the image gradients $\mathbf{D}\mathbf{x}$. An even closer match to the statistics of natural images can be obtained using a mixture of Gaussians or other such distributions [17]. However, the resulting prior over the image is no longer convex, and we find empirically that the local minima onto which the system converges tend to give poorer solutions than the simple Huber-type prior employed here.

2.2. Possibilities for marginalization

Tipping and Bishop's approach obtains an ML estimate of the geometric registration by marginalizing over \mathbf{x} , then calculates the super-resolution estimate as in (5). Although Tipping and Bishop did not include a photometric model, the equivalent expression to be maximized with respect to parameters θ, λ

and γ is the marginal data likelihood

$$p\left(\left\{\mathbf{y}^{(k)}\right\} \middle| \left\{\theta^{(k)}, \lambda^{(k)}\right\}, \gamma\right) \\ = \int p(\mathbf{x}) p\left(\left\{\mathbf{y}^{(k)}\right\} \middle| \mathbf{x}, \left\{\theta^{(k)}, \lambda^{(k)}\right\}, \gamma\right) d\mathbf{x} \quad (10)$$

$$= \int p(\mathbf{x}) \prod_{k=1}^K p\left(\mathbf{y}^{(k)} \middle| \mathbf{x}, \theta^{(k)}, \lambda^{(k)}, \gamma\right) d\mathbf{x}, \quad (11)$$

where the data part of the integral is the same expression shown in (2).

This integral is generally only tractable if a Gaussian form is chosen for the image prior $p(\mathbf{x})$. This is in contrast to a more suitable image prior, such as the Huber distribution over image gradients, as discussed above.

Another approach that integrates over the super-resolution image in the same way is proposed by Molina *et al.* [18]. Again, the prior takes a Gaussian form, though in this instance the resulting marginal data likelihood is optimized with respect to parameters representing the strength of the prior and of the noise terms on each individual low-resolution image.

Our approach is to integrate over the registration and PSF parameters, rather than over the high-resolution image. This gives us a probability distribution directly over the high-resolution image we seek to reconstruct, and allows us to use a non-Gaussian image prior. We chose a Gaussian prior over the registration parameters, which is a preferable approach for two reasons: first after an initial image registration, it is not unreasonable to assume the residual registration errors might be approximately Gaussian-distributed, and second, natural image statistics tend not to be well modelled by Gaussian distributions, so we benefit from being able to retain a non-Gaussian form for the image prior.

3. BAYESIAN MARGINALIZATION

Both of the Bayesian marginalizations proceed along similar lines mathematically, though the effect of choosing a different variable of integration makes a large difference to the way in which the super-resolution algorithms proceed.

3.1. Marginalizing over registration parameters

The goal of super-resolution is to obtain a high-resolution image, so our proposed approach treats the other parameters — geometric and photometric registrations and the PSF — as 'nuisance variables' which might be marginalized out of the

problem. This leads to an integral of the form

$$p(\mathbf{x}|\{\mathbf{y}^{(k)}\}) = \int p(\{\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\}, \gamma) \times p(\mathbf{x}|\{\mathbf{y}^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}, \gamma\}) d\boldsymbol{\theta} d\boldsymbol{\lambda} d\gamma \quad (12)$$

$$= \frac{p(\mathbf{x})}{p(\{\mathbf{y}^{(k)}\})} \int \left[\prod_{k=1}^K p(\{\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\}, \gamma) \times p(\mathbf{y}^{(k)}|\mathbf{x}, \{\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\}, \gamma) \right] d\boldsymbol{\theta} d\boldsymbol{\lambda} d\gamma, \quad (13)$$

where in going from (12) to (13), we have made use of the conditional distribution in (5), taken the image prior term outside the integral, and simplified the denominator by assuming the marginal probability of the the observed images, $p(\{\mathbf{y}^{(k)}\}|\{\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\}, \gamma)$, is independent of the registration and imaging parameter variables, $\{\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\}$ and γ .¹

We require a prior distribution over the registration and PSF parameters, which appears within the integral. Notice that $p(\mathbf{x})$ can be taken outside the integral, leaving us free to chose any suitable super-resolution image prior, rather than being constrained to picking a Gaussian merely to make the integral tractable, as in the image-marginalizing case.

In the following, it is assumed that a preliminary image registration (geometric and photometric) and an estimate of the PSF are available. The registration estimate can be obtained using classical registration methods, either intensity-based [20] or estimation from image points [21]. There is a rich literature of *Blind Image Deconvolution* concerned with estimating an unknown blur on an image [22]. It is then assumed that compared to the ground truth values, this set of registration and PSF parameter values have some unknown Gaussian-distributed additive noise.

We introduce a vector $\boldsymbol{\delta}$ to represent the perturbations from ground truth in our initial parameter estimate. For the parameters of a single image, we have

$$\begin{bmatrix} \boldsymbol{\theta}^{(k)} \\ \lambda_{\alpha}^{(k)} \\ \lambda_{\beta}^{(k)} \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{\theta}}^{(k)} \\ \bar{\lambda}_{\alpha}^{(k)} \\ \bar{\lambda}_{\beta}^{(k)} \end{bmatrix} + \boldsymbol{\delta}^{(k)}, \quad (14)$$

where $\bar{\boldsymbol{\theta}}^{(k)}$ and $\bar{\boldsymbol{\lambda}}^{(k)}$ are the estimated registration, and $\boldsymbol{\theta}^{(k)}$ and $\boldsymbol{\lambda}^{(k)}$ are the true registration. The stacked vector $\boldsymbol{\delta}$ is

¹This approximation is justified by experiments which show that the gradient and Hessian of $p(\{\mathbf{y}^{(k)}\}|\{\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\}, \gamma)$ with respect to the registration parameters are much smaller than those we introduce in (19)[19].

then composed as

$$\boldsymbol{\delta}^T = [\boldsymbol{\delta}^{(1)T}, \boldsymbol{\delta}^{(2)T}, \dots, \boldsymbol{\delta}^{(K)T}, \boldsymbol{\delta}_{\gamma}], \quad (15)$$

where the final entry is for the PSF parameter so that $\gamma = \bar{\gamma} + \delta_{\gamma}$ and $\bar{\gamma}$ is the initial estimate. For our experiments, a Gaussian kernel is used for the PSF, and the log of its standard deviation is used as the unknown parameter, since the standard deviation itself must be a positive quantity.

The vector $\boldsymbol{\delta}$ is assumed to be distributed according to a zero-mean Gaussian and the diagonal matrix \mathbf{V} is constructed to reflect the confidence in each parameter estimate. This might mean a standard deviation of one-tenth of a low-resolution pixel on image translation parameters, or a few grey levels' shift on the illumination model, for instance, giving

$$\boldsymbol{\delta} \sim \mathcal{N}(\mathbf{0}, \mathbf{V}). \quad (16)$$

Defining the data error term as the sum of norms of $\{\mathbf{r}^{(k)}\}$,

$$\mathbf{e}(\boldsymbol{\delta}) = \sum_{k=1}^K \left\| \mathbf{y}^{(k)} - \lambda_{\alpha}^{(k)} \mathbf{W}(\boldsymbol{\theta}^{(k)}, \gamma) \mathbf{x} - \lambda_{\beta}^{(k)} \right\|_2^2, \quad (17)$$

where $\boldsymbol{\theta}$, $\boldsymbol{\lambda}$ and γ are functions of $\boldsymbol{\delta}$ and the initial registration values, we can then expand the integral in (13) to an integral over $\boldsymbol{\delta}$,

$$p(\mathbf{x}|\{\mathbf{y}^{(k)}\}) = \frac{|\mathbf{V}^{-1}|^{1/2} \beta^{KM/2}}{p(\{\mathbf{y}^{(k)}\}) (2\pi)^{(K(M+n)+1)/2}} \times \frac{1}{Z_x} \exp\left\{-\frac{\nu}{2} \rho(\mathbf{D}\mathbf{x}, \alpha)\right\} \times \int \exp\left\{-\frac{\beta}{2} \mathbf{e}(\boldsymbol{\delta}) - \frac{1}{2} \boldsymbol{\delta}^T \mathbf{V}^{-1} \boldsymbol{\delta}\right\} d\boldsymbol{\delta}. \quad (18)$$

The exponent of the data error part of the integral can be written as function of the parameter estimates $\{\bar{\boldsymbol{\theta}}^{(k)}, \bar{\boldsymbol{\lambda}}^{(k)}\}$ and $\bar{\gamma}$ and the vector $\boldsymbol{\delta}$, and a second-order Taylor series can be taken about these estimated parameters giving:

$$\mathbf{e}(\boldsymbol{\delta}) = f + \mathbf{g}^T \boldsymbol{\delta} + \frac{1}{2} \boldsymbol{\delta}^T \mathbf{H} \boldsymbol{\delta}. \quad (19)$$

Values for f , \mathbf{g} and \mathbf{H} can be found numerically (for geometric registration parameters) or analytically (for the photometric parameters) from \mathbf{x} and $\{\mathbf{y}^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\}$. Thus,

the exponent in the integral of (18), becomes

$$a = -\frac{\beta}{2} \mathbf{e}(\boldsymbol{\delta}) - \frac{1}{2} \boldsymbol{\delta}^T \mathbf{V}^{-1} \boldsymbol{\delta} \quad (20)$$

$$= -\frac{\beta}{2} f - \frac{\beta}{2} \mathbf{g}^T \boldsymbol{\delta} - \frac{1}{2} \boldsymbol{\delta}^T \left[\frac{\beta}{2} \mathbf{H} + \mathbf{V}^{-1} \right] \boldsymbol{\delta}. \quad (21)$$

We are now in a position to evaluate the integral in (18) using the identity

$$\int \exp \left\{ -\mathbf{b}\mathbf{x} - \frac{1}{2} \mathbf{x}^T \mathbf{A} \mathbf{x} \right\} d\mathbf{x} = 2\pi^{d/2} |\mathbf{A}|^{-1/2} \exp \{ \mathbf{b}^T \mathbf{A}^{-1} \mathbf{b} \}, \quad (22)$$

where d is the dimension of the vector \mathbf{b} .

Considering only the integral itself, this gives us

$$\begin{aligned} \int \exp \{a\} d\boldsymbol{\delta} &= \exp \left\{ -\frac{\beta}{2} f \right\} \\ &\times \int \exp \left\{ -\frac{\beta}{2} \mathbf{g}^T \boldsymbol{\delta} - \frac{1}{2} \boldsymbol{\delta}^T \mathbf{S} \boldsymbol{\delta} \right\} d\boldsymbol{\delta} \end{aligned} \quad (23)$$

$$= \exp \left\{ -\frac{\beta}{2} f \right\} (2\pi)^{(nK+1)/2} |\mathbf{S}|^{-1/2} \exp \left\{ \frac{\beta^2}{8} \mathbf{g}^T \mathbf{S}^{-1} \mathbf{g} \right\}, \quad (24)$$

where $\mathbf{S} = \left[\frac{\beta}{2} \mathbf{H} + \mathbf{V}^{-1} \right]$ and n the number of registration parameters (geometric and photometric) per image.

The objective function, \mathcal{L} , to be minimized with respect to \mathbf{x} is obtained by taking the negative log of (18), and neglecting the constant terms:

$$\mathcal{L} = \frac{\nu}{2} \rho(\mathbf{D}\mathbf{x}, \alpha) + \frac{\beta}{2} f + \frac{1}{2} \log |\mathbf{S}| - \frac{\beta^2}{8} \mathbf{g}^T \mathbf{S}^{-1} \mathbf{g}. \quad (25)$$

This is the function we optimize with respect to \mathbf{x} to compute the super-resolution image.

3.2. Marginalizing over the super-resolution image

We will outline the marginalization method used in [5] here, since it is useful for comparison with our method, and also because the model used here extends theirs, by adding photometric parameters, which introduce a few extra terms.

The prior used in [5] takes the form of a zero-mean Gaussian over the pixels in \mathbf{x} with covariance \mathbf{Z}_x ,

$$p(\mathbf{x}) = (2\pi)^{-N/2} |\mathbf{Z}_x|^{-1/2} \exp \left\{ -\frac{1}{2} \mathbf{x}^T \mathbf{Z}_x^{-1} \mathbf{x} \right\}. \quad (26)$$

We can substitute this expression for the image prior, and the Gaussian observation density of (3), into (11), which marginalizes over the super-resolution image. Using the identity in (22), this evaluates to a Gaussian distribution over $\{\mathbf{y}^{(k)}\}$. Taking \mathbf{y} to be a stacked vector of all the input images, and $\boldsymbol{\lambda}_\beta$ to be a stack of the $\boldsymbol{\lambda}_\beta^{(k)}$ vectors, this distribution is

$$\mathbf{y} \sim \mathcal{N}(\boldsymbol{\lambda}_\beta, \mathbf{Z}_y), \quad (27)$$

where

$$\mathbf{Z}_y = \beta^{-1} \mathbf{I} + \boldsymbol{\Lambda}_\alpha \mathbf{W} \mathbf{Z}_x \mathbf{W}^T \boldsymbol{\Lambda}_\alpha^T \quad (28)$$

where $\boldsymbol{\Lambda}_\alpha$ are matrices whose diagonals are given by the λ_α value of the corresponding low-resolution image, and where \mathbf{W} is the stack of individual $\mathbf{W}^{(k)}$ matrices.

The objective function, which does not depend on \mathbf{x} , is optimized with respect to $\{\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\}$ and γ , and is given by

$$\begin{aligned} \mathcal{L} &= \frac{1}{2} \left[\beta \sum_{k=1}^K \left\| \mathbf{y}^{(k)} - \lambda_\alpha^{(k)} \mathbf{W}^{(k)} \boldsymbol{\mu} - \boldsymbol{\lambda}_\beta^{(k)} \right\|_2^2 \right. \\ &\quad \left. + \boldsymbol{\mu}^T \mathbf{Z}_x^{-1} \boldsymbol{\mu} - \log |\boldsymbol{\Sigma}| \right], \end{aligned} \quad (29)$$

where

$$\boldsymbol{\Sigma} = \left[\mathbf{Z}_x^{-1} + \beta \sum_{k=1}^K \lambda_\alpha^{(k)2} \mathbf{W}^{(k)T} \mathbf{W}^{(k)} \right]^{-1}, \quad (30)$$

$$\boldsymbol{\mu} = \beta \boldsymbol{\Sigma} \left(\sum_{k=1}^K \lambda_\alpha^{(k)} \mathbf{W}^{(k)T} (\mathbf{y}^{(k)} - \boldsymbol{\lambda}_\beta^{(k)}) \right), \quad (31)$$

which can be derived by taking the log of (27), using the same standard matrix manipulations as in [5], and dropping terms which are independent of the variables of interest.

The expression for the posterior mean $\boldsymbol{\mu}$ is the closed form of the overall MAP solution for the super-resolution image. However, in [5], the optimization over registration and blur parameters is carried out with low-resolution image patches of just 9×9 pixels, rather than the full low-resolution images, because of the computational cost involved in computing the terms in (29) — even for a tiny 50×50 -pixel high-resolution image, the \mathbf{Z}_x and $\boldsymbol{\Sigma}$ matrices are 2500×2500 . The full-sized super-resolution image can then be computed by fixing the optimal registration and PSF values and taking the usual MAP super-resolution approach of (6), with $p(\mathbf{x})$ defined as in (26), which is exactly equivalent to find $\boldsymbol{\mu}$ using the full low-resolution images.

In comparison, the dimensionality of the matrices in the terms comprising the registration–marginalizing objective function (25) is in most cases much lower than those in (29). This means the terms arising from the marginalization are far less costly to compute, so our algorithm can be run on entire low-resolution images, rather than just patches.

3.3. Implementation notes for the registration–marginalizing approach

The objective function (25) can be optimized using *Scaled Conjugate Gradients* (SCG) [23], noting that the gradient can be expressed

$$\begin{aligned} \frac{d\mathcal{L}}{d\mathbf{x}} = & \frac{\nu}{2} \mathbf{D}^T \frac{d}{d\mathbf{x}} \rho(\mathbf{D}\mathbf{x}, \alpha) + \frac{\beta}{2} \frac{df}{d\mathbf{x}} - \frac{\beta^2}{4} \boldsymbol{\xi}^T \frac{d\mathbf{g}}{d\mathbf{x}} \\ & + \left[\frac{\beta}{4} \text{vec} \left(\mathbf{S}^{-1} + \frac{\beta^2}{8} \boldsymbol{\xi} \boldsymbol{\xi}^T \right)^T \right] \frac{\text{dvec}(\mathbf{H})}{d\mathbf{x}}, \end{aligned} \quad (32)$$

where

$$\boldsymbol{\xi} = \mathbf{S}^{-1} \mathbf{g} \quad (33)$$

and where *vec* is the matrix vectorization operator. Further details can be found in the technical note [19]. Derivatives of f , \mathbf{g} and \mathbf{H} with respect to \mathbf{x} can be found analytically for photometric parameters, and numerically (using the analytic gradient of $e^{(k)}(\boldsymbol{\delta}^{(k)})$ with respect to \mathbf{x}) with respect to the geometric parameters. The upper part of \mathbf{H} is block-diagonal $nK \times nK$ sparse matrix, and the final $(nK + 1)$ th row and column are non-sparse, assuming that the blur parameter is shared between the images, as it might be in a short video sequence, for instance, and that the images registration errors for two different images are independent.

Notice that the value f from (21) is simply the reprojection error of the current estimate of \mathbf{x} at the mean registration parameter values, i.e. the value of (17) evaluated at $\bar{\boldsymbol{\theta}}^{(k)}$, $\bar{\boldsymbol{\lambda}}^{(k)}$ and $\bar{\gamma}$. Gradients of this expression with respect to the $\boldsymbol{\lambda}$ parameters, and with respect to \mathbf{x} can both be found analytically. To find the gradient with respect to a geometric registration parameter $\boldsymbol{\theta}_i^{(k)}$, and elements of the Hessian involving it, a central difference scheme involving only the k th image is used.

Mean values for the registration are computed by standard registration techniques, and \mathbf{x} is initialized at these parameters. A useful concept here is the *average image*, \mathbf{a} , which is a stable though excessively smooth approximation to \mathbf{x} [7]. Each pixel in \mathbf{a} is a weighted combination of pixels in \mathbf{y} , such that a_i depends strongly on y_j if y_j depends strongly on x_i , according to the weights in \mathbf{W} . Lighting changes must also be taken into consideration, so

$$\mathbf{a} = \boldsymbol{\Phi}^{-1} \mathbf{W}^T \boldsymbol{\Lambda}_\alpha^{-1} (\mathbf{y} - \boldsymbol{\Lambda}_\beta), \quad (34)$$

where $\boldsymbol{\Phi}$ is a diagonal matrix whose elements are the column sums of \mathbf{W} . Notice that both inverted matrices are diagonal, so \mathbf{a} is simple to compute. In addition to using this form to obtain \mathbf{a} as a starting estimate for the super-resolution image \mathbf{x} , this is also useful for an iterative intensity-based starting estimate for the registration parameters, since minimizing the reprojection error from the average image (the quantity $\sum_{k=1}^K \|\mathbf{y}^{(k)} - \lambda_\alpha^{(k)} \mathbf{W}^{(k)} \mathbf{x} - \lambda_\beta^{(k)}\|_2^2$) with respect to the registration parameters gives values for the parameters which lie reliably close to ground truth, with an error typically on the order of a few hundredths of a pixel for shift parameters, for example.

To initialize \mathbf{x} , we start from the average image, then use SCG to minimize the reprojection residuals (the ML solution), but terminate it after around $K/4$ steps, before the instabilities dominate. This gives a sharper result than initializing with \mathbf{a} alone, as in [7]. When only a few images are available, a more stable ML solution can be found by using a constrained optimization to bound the pixel intensity values so they must lie in the permitted image intensity range. For all our experiments, pixel values are scaled to lie between $-1/2$ and $1/2$, in accordance with [5].

In our implementation, the parameters representing the λ values are scaled so that they share the same standard deviations as the $\boldsymbol{\theta}$ parameters, which represent the sub-pixel geometric registration shifts, which makes the matrix \mathbf{V} a multiple of the identity. The scale factors are chosen so that one standard deviation in λ_β gives a 10-grey-level shift, and one standard deviation in λ_α varies pixel values by around 10 grey levels at mean image intensity.

4. RESULTS

The first three experiments use synthetic datasets in order to allow a quantitative measure of performance to be taken with respect to known ground truth high-resolution images. In the first two experiments, only the marginalization over the geometric and photometric registrations is tested, then the PSF parameter is also introduced to the list over which the system integrates. The final experiment shows the full system working on real data, and compares the results to standard Huber–MAP and to the approach of [5].

4.1. Butterfly sequence

The first experiment in this section demonstrates the improvement this registration marginalization can attain on a simple synthetic dataset. The generative model is used to generate 16-image datasets at a zoom factor of four, with a fixed PSF (standard deviation 0.2 low-resolution pixels), and low noise (30 dB, corresponding to additive noise with standard deviations of approximately 1.55 grey levels for this image). Each 30×30 -pixel low-resolution image is related to the

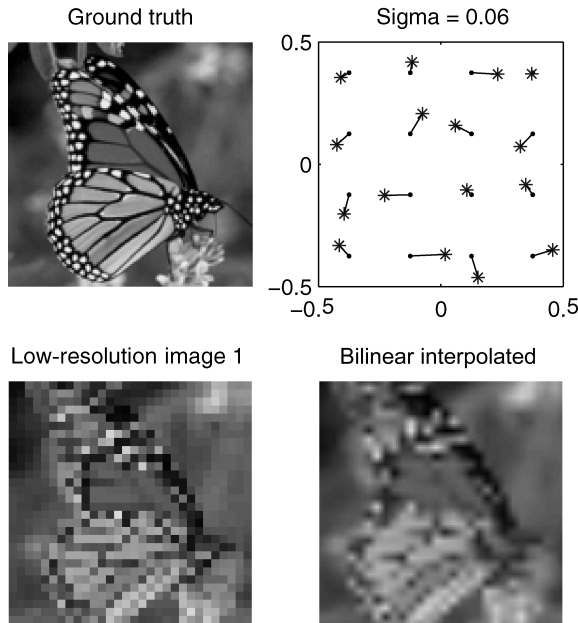


FIGURE 1. Top left: ground truth high-resolution image. Top right: example of sub-pixel image registration and artificially introduced geometric registration errors, plotted on a single low-resolution image pixel. Dots on the grid denote the true shift sizes (measured in low-resolution pixels), and stars mark the values as passed to the super-resolution reconstruction algorithms. Bottom left: an example of low-resolution image. Bottom right: a second image warped into the high-resolution image frame using bilinear interpolation, increasing the number of pixels by a factor of 16. See Figure 2 and Table 1 for the superresolution results.

ground truth image by a 2D translation-only motion model (in addition to the zoom factor), and the two-parameter global linear illumination model, giving a total of four registration parameters per low-resolution image. The sub-pixel perturbations are evenly spaced over a grid up to plus or minus one half of a low-resolution pixel, giving a similar setup to that described in [11], but with additional lighting variation.

Before solving to find super-resolution images, the registration data are corrupted by the addition of *i.i.d.* Gaussian noise. The noise corrupting the dataset had standard deviations of 0.02–0.14 low-resolution pixels, with corresponding intensity shift standard deviations of around 1.53–10.71 grey-levels.

Figure 1 shows the ground-truth butterfly image, a plot of the image registration errors within a single low-resolution pixel for the case when the standard deviation on the registration error is 0.06 low-resolution pixels, and two of the 16 low-resolution images, one of which has been interpolated up to the original image resolution size using bilinear interpolation.

Three super-resolution algorithms are used to construct super-resolution images from these datasets: ML, plain Huber–MAP and registration–marginalizing (which also

uses the same Huber prior). In this first experiment, we will consider integrating only over the geometric and photometric registrations, so the true value of the PSF parameter is supplied to all three algorithms. In both cases where the Huber prior is involved, the Huber parameter is set to $\alpha = 0.04$. For the registration–marginalizing approach, the error distribution over the registration parameters is assumed to have a standard deviation of 0.005 pixels, which is found to perform well in spite of being smaller than the true registration perturbation distributions.

Figure 2 shows the results of the three super-resolution algorithms. The ML case quickly becomes unrecognizable as a butterfly as the registration error increases. The Huber–MAP approach (using prior strength $\log_{10}(\nu/\beta) = -1.6$) performs better, but errors and speckling in the vicinity of the black-white edges become apparent as the image registration noise level increases. The most plausible-looking high-resolution image for each of the four datasets comes from the registration–marginalizing approach, which succeeds in maintaining much cleaner transitions at the edge boundaries. The strength of the Huber prior used in the registration–marginalizing approach does not need to be as high as the regular Huber–MAP approach, since some of the regularization power comes from the new terms in the objective function (25), so in this case we choose $\log_{10}(\nu/\beta) = -2.2$. Considering only the images in the final column, the registration–marginalizing approach reduces the RMSE by over 32%.

Table 1 gives the RMSE for each of the images of Fig. 2 (ML, Huber and Integr. ‘A’). Also given for comparison are the RMSE values for the registration–marginalizing algorithm when the stronger image prior ($\log_{10}(\nu/\beta) = -1.6$) is used (Integr. ‘B’).

4.2. Eyechart sequence

The second experiment examines the behaviour of the registration–marginalizing method over a larger range or parameter values, again without marginalizing over the PSF. To show that the approach can handle multiple classes of image, an image of an eyechart is used. The 16-image synthetic dataset is created, where each 30×30 -pixel low-resolution image is generated at a zoom factor of 4, again using the model with two translational degrees of freedom and two photometric degrees of freedom. A Gaussian PSF with a standard deviation of 0.4 low-resolution pixels is used, and Gaussian noise (30 dB; standard deviation equivalent to approximately 3.4 grey levels) is added to the intensity of each low-resolution pixel independently.

The observed images are shown in Fig. 3. The mean intensity is clearly different, and the horizontal shift is most easily observed in the ‘D’ characters to the right of each image.

Geometric and photometric registration parameters were initialized to the identity, and the images were registered using an

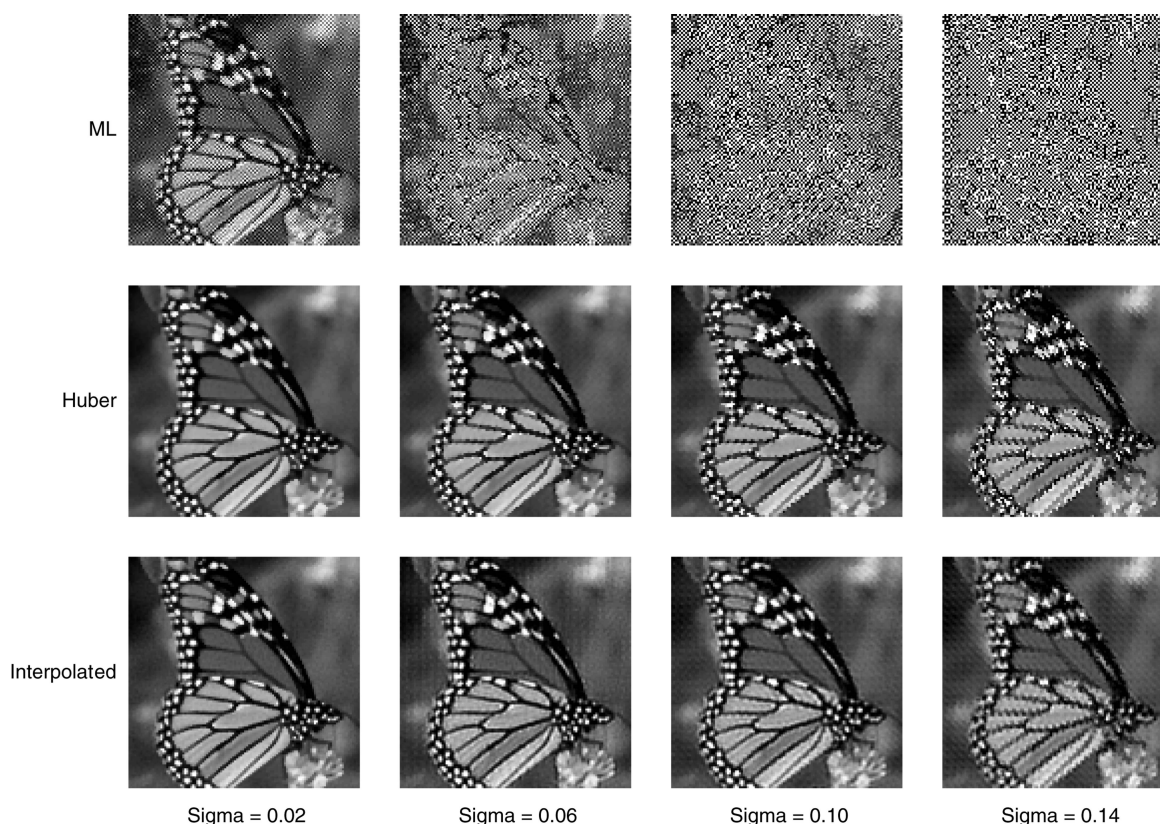


FIGURE 2. Super-resolution images found using ML (top), Huber–MAP (middle) and registration–marginalizing recovery methods. The mis-registration error gets progressively worse from left to right, with the standard deviations being 0.02, 0.06, 0.10 and finally 0.14 low-resolution pixels. The bottom right corner (registration–marginalizing) image is visibly better than the equivalent Huber image directly above. The RMS errors for each image compared to the ground truth high-resolution image are given in Table 1.

TABLE 1. RMS errors (compared to ground truth image) for the images in the butterfly experiment, as shown in Fig. 2.

RMSE method	Standard deviation of registration error			
	0.02	0.06	0.10	0.14
ML	50.56	328.07	904.91	1771.06
Huber	10.61	15.59	20.12	31.53
Integr. A	7.33	11.89	13.45	21.32
Integr. B	9.27	12.96	13.10	22.62

The two rows for the registration–marginalizing approach indicate the error scores for different ν values in the Huber prior. The ‘A’ row uses a slightly weaker prior (see text) and corresponds to the ‘Int’ images of Fig. 2, whereas the ‘B’ row uses a stronger prior equal to that used in the Huber–MAP super-resolution algorithm, and is given here for comparison only.

iterative intensity-based scheme. The resulting parameter values were used to recover two sets of super-resolution images: one using the standard Huber–MAP algorithm, and the other one using our extension integrating over the geometric and photometric registration uncertainties. The Huber parameter α was

fixed at 0.01 for all runs, and ν was varied over a range of possible values representing ratios between ν and the image noise precision β .

The images giving lowest RMS error from each set are displayed in Fig. 4. Visually, the differences between the images are subtle, though the bottom row of letters is better defined in the output from our algorithm. Plotting the RMSE as a function of ν in Fig. 5, we see that the registration–marginalizing approach achieves a lower error, compared to the ground truth high-resolution image, than the standard Huber–MAP algorithm for any choice of prior strength, $\log_{10}(\nu/\beta)$. Because ν and β are free parameters in the algorithm, it is an advantage that the marginalizing approach is less sensitive to variation in their values.

4.3. Face sequence

The face experiment uses a 16-image synthetic dataset generated from a face image in the same manner as the eyechart dataset above. The image size is smaller— 15×15 pixels—and the noise level is set to 25 dB (approximately 3.94 grey levels). It is difficult to identify the individual in any one of the low-

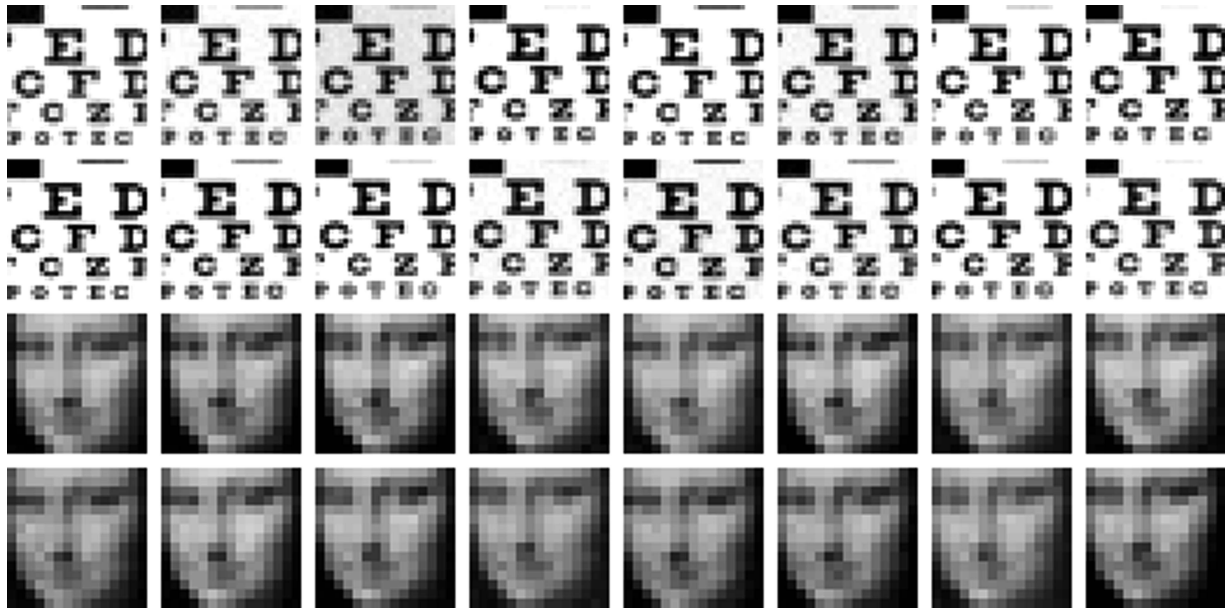


FIGURE 3. The eyechart and face datasets. Each contains 16 low-resolution images generated from a common high-resolution version at a zoom factor of 4. The variation in intensity is clearly visible, and the sub-pixel displacements necessary for multi-frame image super-resolution are most apparent on the 'D' characters to the right of each image in the eyechart sequence, and can also be observed in the face sequence. Eyechart images are 30×30 pixels, and face images are 15×15 pixels in size.

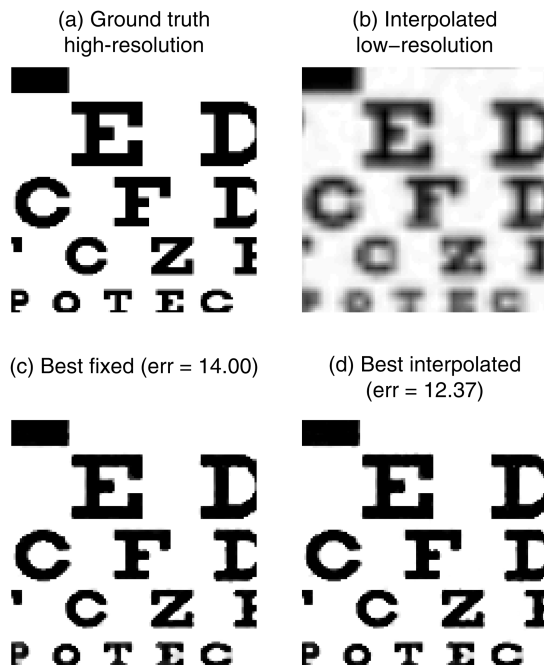


FIGURE 4. Super-resolving the eyechart dataset. (a) Ground truth image (only the central recoverable part is shown). (b) Interpolated low-resolution image. (c) Best (minimum MSE) image from the regular Huber-MAP algorithm, having super-resolved the dataset multiple times with different prior strength settings. (d) The best result using our approach of integrating over θ and λ . As well as having a lower RMSE, note the improvement in black-white edge detail on some of the letters on the bottom line.

resolutions images; the full set of 16 is shown in the lower section of Fig. 3.

In addition to the error in registration parameters arising from registering these small input images, we also introduce uncertainty in the value of the PSF parameter, γ , by choosing a value that is 0.05 low-resolution pixels bigger than the true PSF standard deviation. PSF kernel estimation, particularly in such small images with such a high proportion of pixels

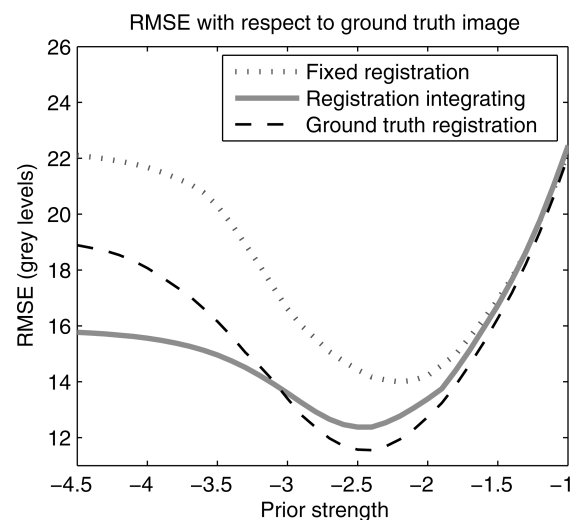


FIGURE 5. Eyechart dataset. Plot showing the variation of RMSE with prior strength for the standard Huber-prior MAP super-resolution method and our approach integrating over θ and λ .

lying near the image borders, is a hard problem, so this level of error is not unexpected in a general super-resolution problem with unknown parameters.

The full registration–marginalizing approach was applied (including marginalizing over the PSF, γ), so δ was a 65-element vector comprised of four registration parameters per image, plus the single log-PSF parameter. This was compared to the plain Huber–MAP super-resolution algorithm with no integration, over a range of possible image prior strengths. The registration–marginalizing without the PSF-integration was also used, providing a third image estimate for each prior strength value. For the results shown, we take the distribution over the registration to have standard deviation 0.01 low-resolution pixels, and use the same standard deviation for the PSF parameter.

The errors for each approach are plotted against prior strength in Fig. 6. As anticipated, the plain Huber–MAP approach gives the greatest error overall, and the full registration–marginalizing approach which takes the PSF uncertainty into account gives the lowest error overall, though the version which handles only the 64 image registration parameters and also handles the problem better than the plain Huber–MAP method.

The best output images from the plain Huber–MAP and the full registration–marginalizing approach are shown in Fig. 7, along with one of the low-resolution images interpolated up in to the high-resolution frame, to give an idea of the improvement made by the super-resolution.

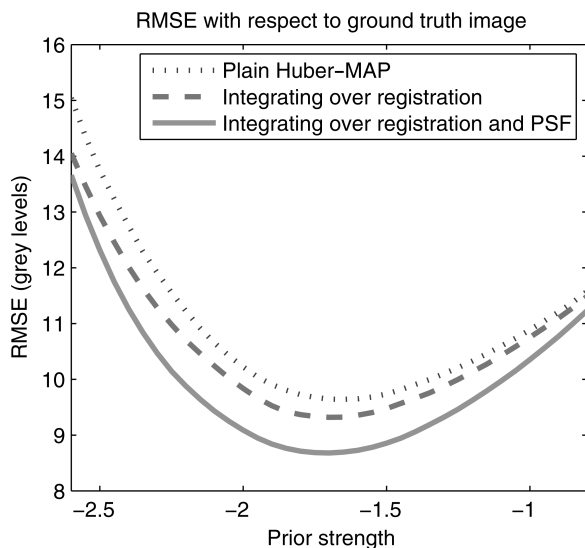


FIGURE 6. Face dataset. Plot showing the variation of RMSE with prior strength for the standard Huber-prior MAP super-resolution method, the approach integrating over θ and λ , and the approach integrating over γ as well as the geometric and photometric registrations. As expected, the result which integrates over the uncertainty in all the estimates produces a lower-error reconstruction than the standard method.

was intentionally a challenging dataset with relatively high noise levels, both face reconstructions exhibit some noise around the cheek region. While the absolute differences between these two images are subtle, improvements in the reconstruction of the cheek to the left and around the eye to the right of the image may be discernible in the registration–marginalizing case, which leads to its lower error than the other methods when compared to ground truth.

4.4. Real data

The final experiment uses real data with a 2D translation motion model and a 2-parameter lighting model exactly as above; the low-resolution images appear in Fig. 8. Homographies were provided with the data, but were not used. Instead, an iterative illumination-based registration was used on the sub-region of the images chosen for super-resolution, and this agreed with the provided homographies to within a few hundredths of a pixel.

Super-resolution images were created for a number of image prior strengths, and equivalent values to those quoted in [7] were selected for the Huber–MAP recovery, following a subjective evaluation of other possible parameter settings. For the registration–marginalizing approach, a similar parameter error distribution as that used in the synthetic experiments was assumed. Finally, Tipping and Bishop’s method was extended to cover the illumination model and used to register and super-resolve the dataset, using the same PSF standard deviation (0.4 low-resolution pixels) as the other methods.

The three sets of results on the real data sequence are shown in Fig. 9. To facilitate a better comparison, a sub-region of each is expanded to make the letter details clearer. The Huber prior tends to make the edges unnaturally sharp, though it is very successful at regularizing the solution elsewhere. Between the Tipping and Bishop image and the registration–marginalizing approach, the text appears more clear in our method, and the regularization in the constant background regions is slightly more successful. Also note



FIGURE 7. Super-resolving the face dataset. (a) Interpolated low-resolution image. (b) Best (minimum MSE) image found using the Huber method. (c) Super-resolution result using our approach of integrating over θ , λ and γ . Again, the difference in results is subtle, but still perceptible. For the images show, the Huber RMSE is 9.64, whereas the registration–marginalizing RMSE is 8.68.

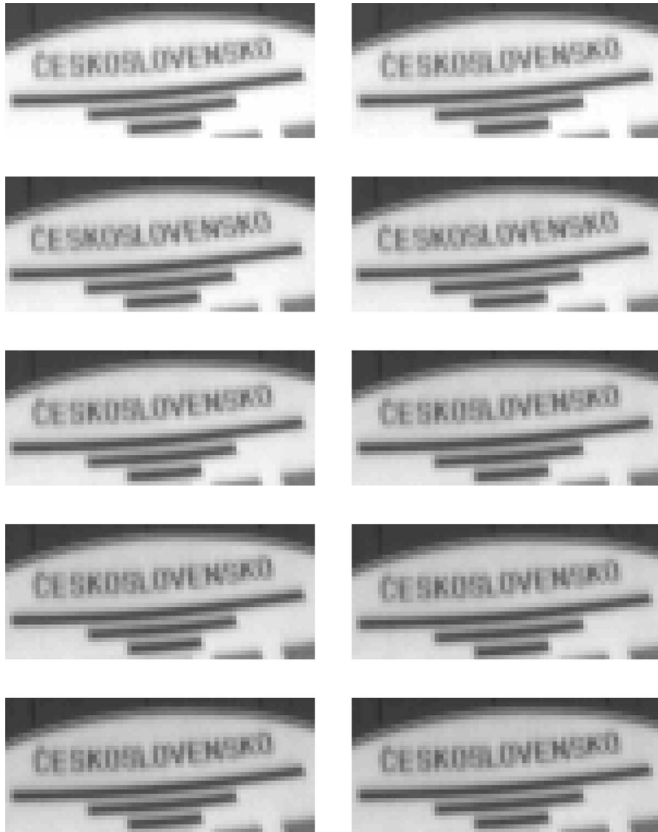


FIGURE 8. The 10 input images in the real dataset. Photometric difference are clearly visible, with the later images being darker in general.



FIGURE 9. Super-resolving the main dataset. (a) The full super-resolution output from our algorithm. (b) Detailed region of the central letters, again with our algorithm. (c) Detailed region of the regular Huber-MAP super-resolution image, using parameter values suggested in [7], which are also found to be subjectively good choices. The edges are slightly artificially crisp, but the large smooth regions are well regularized. (d) Close-up of letter detail for comparison with Tipping and Bishop's method of marginalization. The Gaussian form of their prior leads to a more blurred output, or one that over-fits to the image noise on the input data if the prior's influence is decreased.

that the Gaussian prior on the image-marginalizing method is zero-mean, so in this case having a strong enough prior to suppress the background noise has also biased the output image towards the mid-grey zero value, making the white regions appear darker than they do in the other methods.

5. DISCUSSION

It is possible to interpret the extra terms introduced into the objective function in the derivation of the registration-marginalizing method as an extra regularizer term or image prior. Considering (25), the first two terms are identical to the standard MAP super-resolution problem using a Huber image prior. The two additional terms constitute an additional distribution over \mathbf{x} in the cases where \mathbf{S} is not dominated by \mathbf{V} ; as the distribution over θ and λ tightens to a single point, the terms tend to constant values.

The intuition behind the method's success is that this extra prior resulting from the final two terms of (25) will favour image solutions which are not acutely sensitive to minor adjustments in the image registration. The images of Fig. 10 illustrate the type of solution which would score poorly. To create the figure, one dataset was used to produce two

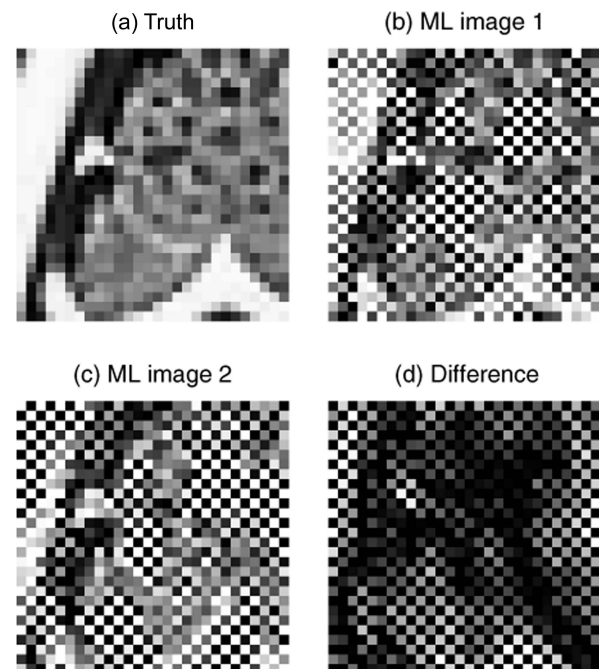


FIGURE 10. An example of the effect of tiny changes in the registration parameters. (a) Ground truth image from which a 16-image low-resolution dataset was generated. (b and c) Two ML super-resolution estimates. In both cases, the same dataset was used, but the registration parameters were perturbed by an *i.i.d.* vector with standard deviation of just 0.04 low-resolution pixels. (d) The difference between the two solutions. In all these images, values outside the valid image intensity range have been rounded to white or black values.

super-resolved images, using two independent sets of registration parameters which were randomly perturbed by an *i.i.d.* Gaussian vector with a standard deviation of only 0.04 low-resolution pixels. The chequer-board pattern typical of ML super-resolution images can be observed, and the difference image on the right shows the drastic contrast between the two image estimates.

The ability of the registration–marginalizing method to suppress the chequer-board pattern can also be observed in the results of the experiment with the butterfly dataset; in Fig. 2, the chequer-board-like noise on the regular Huber-MAP super-resolution images is significantly diminished in the registration–marginalizing method’s outputs.

6. CONCLUSION

In this paper, we have outlined two contrasting Bayesian approaches to image super-resolution, and presented the latest work in integrating over the imaging parameters as the ‘nuisance variables’ in the super-resolution problem.

The registration–marginalizing approach to super-resolution shows several advantages over Tipping and Bishop’s original image-integrating algorithm. These are a formal treatment of registration uncertainty, the use of a much more realistic image prior, and the computational speed and memory efficiency relating to the smaller dimension of the space over which we integrate. The results on real and synthetic images with this method show an advantage over the popular MAP approach, and over the result from Tipping and Bishop’s method, largely owing to our more favourable prior over the super-resolution image.

Note that while the examples in this paper are confined to a translation-only motion model, there is no constraint in the mathematical derivation which prevents it from being applied to more complex parametric motion models such as affine or planar projective homographies.

This paper has demonstrated a quantitative improvement in super-resolution image quality can be achieved by the registration–marginalizing method, compared to the known ground truth image on synthetic datasets. The results on the ‘Czech’ data example show that the method is capable of handling the errors and uncertainty introduced in real practical systems.

ACKNOWLEDGEMENTS

The real dataset used in the results section is due to Tomas Pajdla and Daniel Martinec, CMP, Prague, and is available at <http://www.robots.ox.ac.uk/~vgg/data/>

FUNDING

This work was funded in part by EC Network of Excellence PASCAL.

REFERENCES

- [1] Irani, M. and Peleg, S. (1990) Super resolution from image sequences. *Proc. Int. Conf. Pattern Recognition*, vol. 2, pp. 115–120.
- [2] Salient, Stills. <http://www.salientstills.com/>.
- [3] Hardie, R.C., Barnard, K.J. and Armstrong, E.E. (1997) Joint map registration and high-resolution image estimation using a sequence of undersampled images. *IEEE Trans. Image Process.*, **6**, 1621–1633.
- [4] Robinson, D. and Milanfar, P. (2004) Fundamental performance limits in image registration. *IEEE Trans. Image Process.*, **13**, 1185–1199.
- [5] Tipping, M.E. and Bishop, C.M. (2003) Bayesian image super-resolution. In Becker, S., Thrun, S. and Obermayer, K. (eds) *Advances in Neural Information Processing Systems 15*. MIT Press, Cambridge, MA.
- [6] Baker, S. and Kanade, T. (2002) Limits on superresolution and how to break them. *IEEE Trans. Pattern Anal. Mach. Intell.*, **24**, 1167–1183.
- [7] Capel, D. (2004) *Image Mosaicing and Super-resolution (Distinguished Dissertations)*. Springer, ISBN: 1852337710.
- [8] Elad, M. and Feuer, A. (1999) Super-resolution reconstruction of image sequences. *IEEE Trans. Pattern Anal. Mach. Intell.*, **21**, 817–834.
- [9] Elad, M. and Hel-Or, Y. (2001) A fast super-resolution reconstruction algorithm for pure translational motion and common space invariant blur. *IEEE Trans. Image Process.*, **10**, 1187–93.
- [10] Farsiu, S., Elad, M. and Milanfar, P. (2006) A practical approach to super-resolution. *Proc. SPIE: Visual Communications and Image Processing*, San-Jose.
- [11] Nguyen, N., Milanfar, P. and Golub, G. (2001) Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement. *IEEE Trans. Image Process.*, **10**, 1299–1308.
- [12] Wang, Q., Tang, X. and Shum, H. (2005) Patch based blind image super resolution. *Proc. 10th Int. Conf. Computer Vision*, Beijing, China, pp. 709–716.
- [13] Pickup, L.C., Roberts, S.J. and Zisserman, A. (2006) Optimizing and learning for super-resolution. In Chantler, M.J., Trucco, E. and Fisher, R.B. (eds) *Proc. British Machine Vision Conf.*, vol. 3, pp. 439–448. Edinburgh, UK.
- [14] Schultz, R.R. and Stevenson, R.L. (1994) A Bayesian approach to image expansion for improved definition. *IEEE Trans. Image Process.*, **3**, 233–242.
- [15] Pickup, L.C., Capel, D.P., Roberts, S.J. and Zisserman, A. (2006) Bayesian image super-resolution, continued. In Schölkopf, B., Platt, J. and Hoffman, T. (eds) *Advances in*

- Neural Information Processing Systems 19*, December, pp. 1089–1096. MIT Press, Cambridge, MA.
- [16] Borman, S. (2004) Topics in multiframe superresolution restoration. PhD Thesis, University of Notre Dame Notre Dame, Indiana.
- [17] Fergus, R., Singh, B., Hertzmann, A., Roweis, S.T. and Freeman, W. (2006) Removing camera shake from a single photograph. *ACM Trans. Graphics, SIGGRAPH 2006 Conf. Proc.*, Boston, MA, **25**, PP. 787–794.
- [18] Molina, R., Vega, M., Abad, J. and Katsaggelos, A.K. (2003) Parameter estimation in bayesian high-resolution image reconstruction with multisensors. *IEEE Trans. Image Process.*, **12**, 1655– 1667.
- [19] Pickup, L.C. (2007) *Gradient derivation for parametermarginalizing image super-resolution*. Technical report. University of Oxford.
- [20] Irani, M. and Peleg, S. (1991) Improving resolution by image registration. *Graph. Models Image Process.*, **53**, 231–239.
- [21] Hartley, R.I. and Zisserman, A. (2004) *Multiple View Geometry in Computer Vision* (2nd edn) Cambridge University Press. ISBN: 0521540518.
- [22] Kundur, D. and Hatzinakos, D. (1996) Blind image deconvolution. *IEEE Signal Process. Mag.*, **13**, 43–46.
- [23] Nabney, I. (2002) *Netlab Algorithms for Pattern Recognition*. Springer.