

Research Article

Overcoming Registration Uncertainty in Image Super-Resolution: Maximize or Marginalize?

Lyndsey C. Pickup, David P. Capel, Stephen J. Roberts, and Andrew Zisserman

Information Engineering Building, Department of Engineering Science, Parks Road, Oxford OX1 3PJ, UK

Received 15 September 2006; Accepted 4 May 2007

Recommended by Russell C. Hardie

In multiple-image super-resolution, a high-resolution image is estimated from a number of lower-resolution images. This usually involves computing the parameters of a generative imaging model (such as geometric and photometric registration, and blur) and obtaining a MAP estimate by minimizing a cost function including an appropriate prior. Two alternative approaches are examined. First, both registrations and the super-resolution image are found *simultaneously* using a joint MAP optimization. Second, we perform Bayesian integration over the unknown image registration parameters, deriving a cost function whose only variables of interest are the pixel values of the super-resolution image. We also introduce a scheme to learn the parameters of the image prior as part of the super-resolution algorithm. We show examples on a number of real sequences including multiple stills, digital video, and DVDs of movies.

Copyright © 2007 Lyndsey C. Pickup et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

1. INTRODUCTION

Multiframe image super-resolution refers to the process by which a set of images of the same scene are fused to produce an image or images with a higher spatial resolution, or with more visible detail in the high spatial frequency features [1]. The limits on the resolution of the original imaging device can be improved by exploiting the relative subpixel motion between the scene and the imaging plane. Applications are common, with everything from holiday snaps and DVD frames to satellite terrain imagery providing collections of low-resolution images to be enhanced, for instance to produce a more aesthetic image for media publication [2, 3], object or surface reconstruction [4], or for higher-level vision tasks such as object recognition or localization [5]. Figure 1 shows examples from a still camera and a DVD movie.

In previous work, a few methods have assumed no scene motion, and use other cues such as lighting or varying zoom [6]. However, the vast majority of current super-resolution methods do assume motion, and either preregister the inputs using standard registration techniques, or assume that a perfect registration is given a priori [1, 7], before carrying out the super-resolution estimate. However, the steps taken in super-resolution are seldom truly independent, and this is too often ignored in current super-resolution techniques

[1, 7–12]. In this work we will develop two algorithms which consider the problem in a more unified way.

The first approach is to estimate a super-resolution image at the same time as finding the low-resolution image registrations. This *simultaneous* approach offers visible benefits on results obtained from real data sequences. The registration model is fully projective, and we also incorporate a photometric model to handle brightness changes often present in images captured in a temporal sequence. This makes the model far more general than most super-resolution approaches. In contrast to *fixed-registration* methods—that is, those like [7, 13], which first estimate and freeze the registration parameter values before calculating the super-resolution image—we make use of the high-resolution image estimate common to all the low-resolution images to improve the registration estimate.

An alternative approach, and the second one we explore, is to *marginalize* over the unknown registration parameters. This leads to a super-resolution algorithm which takes into account the residual uncertainty in any image registration estimate [14], taking the Bayesian approach of integrating these unknown parameters out of the problem. We demonstrate results on synthetic and real image data which shows improved super-resolution results compared to the standard fixed registration approach.

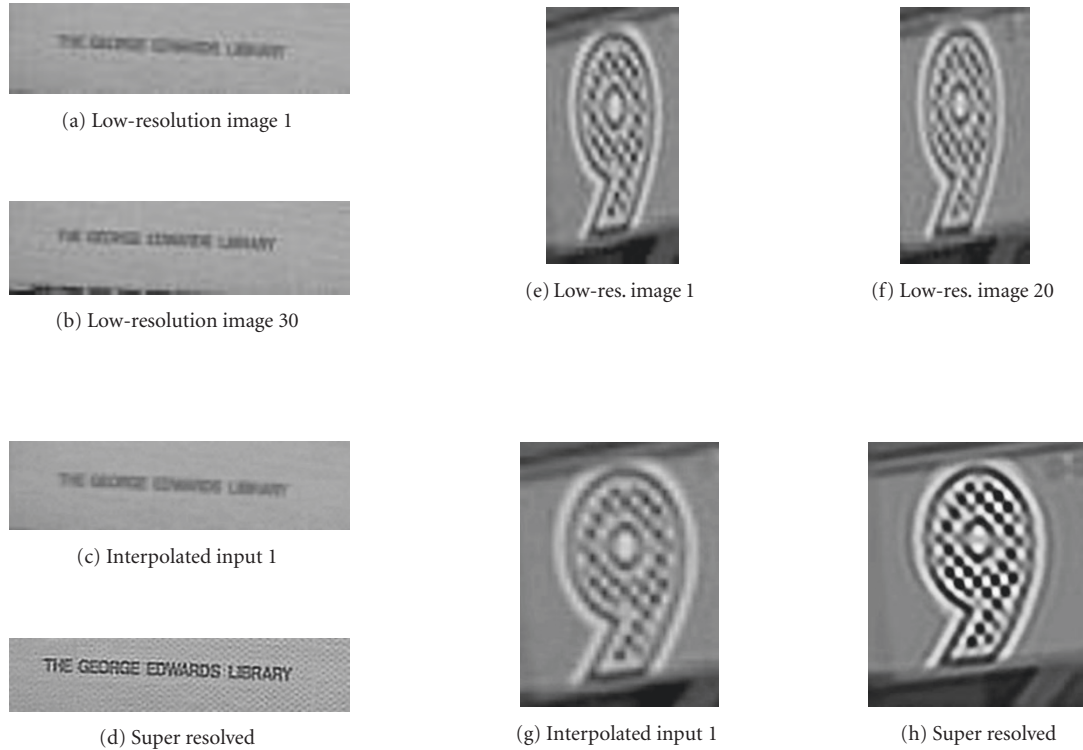


FIGURE 1: Examples of simultaneous MAP super-resolution. (a), (b) Two close-ups from a 30-frame digital camera sequence; (c) first image interpolated into high-resolution frame; (d) simultaneous super-resolution output; (e), (f) two close-ups from a set of 29 DVD movie frames; (g) first image interpolated into high-resolution frame (at corrected aspect ratio); (h) simultaneous super-resolution output.

The third component of this work introduces a scheme by which the parameters of an image prior can be learnt in the super-resolution framework even when there is possible mis-registration in the input images. Poorly chosen prior values will lead to ill-conditioned systems or to overly-smooth super-resolution estimates. Since the best values for any particular problem depend heavily on the statistics of the image being super resolved and the characteristics of the input dataset, having an online method to tune these parameters to each problem is important.

The super-resolution model and notation are introduced in Section 2, followed by the standard *maximum a posteriori* (MAP) solution, and an overview of the ways in which it is extended in this paper. The simultaneous registration and super-resolution approach is developed in Section 3, and this is followed by the learning of the prior parameters, which is incorporated into the algorithm to give a complete simultaneous approach. Section 4 develops the marginalization approach by considering how to integrate over the registration parameters.

Results on several challenging real datasets are used to illustrate the efficacy of the joint MAP technique in Section 5, as well as an illustration using synthetic data. Results using the marginalization super-resolution algorithm are shown for a subset of these datasets in Section 6. A discussion of both approaches and concluding remarks are given in Section 7.

1.1. Background

The work of Hardie et al. [5] has previously examined the joint MAP image registration and super-resolution approach, but with a much more limited model. The high-resolution estimate is used to update the image registrations, but the motion model is limited to shifts on a quantized grid (a $1/4$ -pixel spacing is used in their implementation), so registration is a search across grid locations, which would quickly become infeasible with more degrees of freedom.

Tipping and Bishop [15] marginalize out the high-resolution image to learn a Euclidean registration directly, but with such a high computational cost that their inputs are restricted to 9×9 pixels. We suggest it is more desirable to integrate over the registration parameters rather than the super-resolution image, because it is the registration that constitutes the “nuisance parameters,” and the super-resolution image that we wish to estimate.

With reference to learning the image prior, the generalized cross validation (GCV) work of Nguyen et al. [12] learns a regularization coefficient based on the data. All three of the above approaches [5, 12, 15] rely on Gaussian image priors, whereas a considerable body of super-resolution research has demonstrated that there are many families of priors more suitable for image super-resolution [13, 16–20]. In the following work, we use a more realistic image prior, not a Gaussian.

Preliminary versions of the algorithms presented here appear in [21, 22].

2. THE ANATOMY OF MULTIFRAME SUPER-RESOLUTION

A high-resolution scene \mathbf{x} , with N pixels, is assumed to have generated a set of K low-resolution images $\mathbf{y}^{(k)}$, each with M pixels. For each image, the warping, blurring, and subsampling of the scene is modelled by an $M \times N$ sparse matrix $\mathbf{W}^{(k)}$ [15, 18], and a global affine photometric correction results from addition and multiplication across all pixels by scalars $\lambda_\alpha^{(k)}$ and $\lambda_\beta^{(k)}$, respectively [18]. Thus the generative model is

$$\mathbf{y}^{(k)} = \lambda_\alpha^{(k)} \mathbf{W}^{(k)} \mathbf{x} + \lambda_\beta^{(k)} \mathbf{1} + \boldsymbol{\epsilon}^{(k)}, \quad (1)$$

where $\boldsymbol{\epsilon}^{(k)}$ represents noise on the low-resolution image, and consists of i.i.d. samples from a zero-mean Gaussian with precision β (equivalent to std $\sigma_N = \beta^{-1/2}$), and images \mathbf{x} and $\mathbf{y}^{(k)}$ are represented as vectors. The transform that maps between the frame of \mathbf{x} and that of $\mathbf{y}^{(k)}$ is assumed to be parameterized by some vector $\boldsymbol{\theta}^{(k)}$ (e.g., rotations, or an eight-parameter projective transform), so $\mathbf{W}^{(k)}$ is a function of $\boldsymbol{\theta}^{(k)}$ and of the image point-spread function (PSF), which accounts for blur introduced by the camera optics and physical imaging process. Given $\{\mathbf{y}^{(k)}\}$, the goal is to recover \mathbf{x} , without any explicit knowledge of $\{\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}, \sigma_N\}$.

For an individual low-resolution image $\mathbf{y}^{(k)}$, given registrations and \mathbf{x} , the probability of having observed that image is

$$\begin{aligned} p(\mathbf{y}^{(k)} \mid \mathbf{x}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}) \\ = \left(\frac{\beta}{2\pi}\right)^{M/2} \exp\left\{-\frac{\beta}{2} \left\| \mathbf{y}^{(k)} - \lambda_\alpha^{(k)} \mathbf{W}(\boldsymbol{\theta}^{(k)}) \mathbf{x} - \lambda_\beta^{(k)} \mathbf{1} \right\|_2^2\right\}, \end{aligned} \quad (2)$$

which comes from (1), and from the assumption of Gaussian noise. Other noise model choices lead to slightly different expressions, like the L_1 norm model of [19].

The vector \mathbf{x} yielding the maximal value of $p(\mathbf{y}^{(k)} \mid \mathbf{x}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)})$ would be the *maximum likelihood* (ML) solution to the problem. However, the super-resolution problem is almost always poorly conditioned, so a prior over \mathbf{x} is usually required to avoid solutions which are subjectively very implausible to the human viewer.

We choose a prior based on the Huber function, which here will be applied to directional image gradients of the super-resolution image. The Huber function takes a parameter α , and for each directional image gradient z , it is defined:

$$\rho(z, \alpha) = \begin{cases} z^2 & \text{if } |z| < \alpha, \\ 2\alpha|z| - \alpha^2 & \text{otherwise.} \end{cases} \quad (3)$$

The set of directional image gradients in the horizontal, vertical, and two diagonal directions at all pixel locations in \mathbf{x} is denoted by $\mathcal{G}(\mathbf{x})$, and the prior probability of a high-resolution image \mathbf{x} is then

$$p(\mathbf{x}) = \frac{1}{\mathcal{Z}_x} \exp\left\{-\frac{\nu}{2} \sum_{z \in \mathcal{G}(\mathbf{x})} \rho(z, \alpha)\right\}, \quad (4)$$

where ν is the prior strength parameter and \mathcal{Z}_x is a normalization constant. The penalty for an individual directional gradient estimate z is quadratic for small values of z , which encourages smoothness, but the penalty is linear (i.e., less than quadratic) if z is large, which penalizes edges less severely than a Gaussian.

In the next two sections, we will overview and contrast the simultaneous *maximum a posteriori* and marginalization approaches to the super-resolution problem. These two approaches will then be developed in Sections 3 and 4, respectively.

2.1. Simultaneous maximum a posteriori super-resolution

The *maximum a posteriori* (MAP) solution is found using Bayes' rule,

$$p(\mathbf{x} \mid \{\mathbf{y}^{(k)}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\}) = \frac{p(\mathbf{x}) \prod_{k=1}^K p(\mathbf{y}^{(k)} \mid \mathbf{x}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)})}{p(\{\mathbf{y}^{(k)}\} \mid \{\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\})}, \quad (5)$$

and by taking logs and neglecting terms which are not functions of \mathbf{x} or the registration parameters, this leads to the objective function

$$\mathcal{F} = \underbrace{\beta \sum_{k=1}^K \left\| \mathbf{y}^{(k)} - \lambda_\alpha^{(k)} \mathbf{W}^{(k)} \mathbf{x} - \lambda_\beta^{(k)} \mathbf{1} \right\|_2^2}_{\text{generative model}} + \underbrace{\nu \sum_{z \in \mathcal{G}(\mathbf{x})} \rho(z, \alpha)}_{\text{prior}}. \quad (6)$$

In fixed-registration MAP super-resolution, \mathbf{W} and $\boldsymbol{\lambda}$ values are first estimated and frozen, typically using a feature-based registration scheme (see, e.g., [7, 23]), then the intensities of the registered images are corrected for photometric differences. The resulting problem is convex in \mathbf{x} , and a gradient descent algorithm, such as *scaled conjugate gradients* (SCG) [24], will easily find the optimum at

$$\frac{\partial \mathcal{F}}{\partial \mathbf{x}} = 0. \quad (7)$$

In the simultaneous MAP approach here, we optimize \mathcal{F} explicitly with respect to \mathbf{x} , the set of geometric registration parameters $\boldsymbol{\theta}$ (which parameterize \mathbf{W}), and the photometric parameters $\boldsymbol{\lambda}$ (composed of the λ_α and λ_β values), at the same time, that is, we determine the point at which

$$\frac{\partial \mathcal{F}}{\partial \mathbf{x}} = \frac{\partial \mathcal{F}}{\partial \boldsymbol{\theta}} = \frac{\partial \mathcal{F}}{\partial \boldsymbol{\lambda}} = 0. \quad (8)$$

The problem in (7) is convex, because \mathcal{F} is a quadratic function of \mathbf{x} . Unfortunately, the optimization in (8) is not necessarily convex with respect to $\boldsymbol{\theta}$. To see this, consider a scene composed of a regularly tiled square texture: any two $\boldsymbol{\theta}$ values mapping two identical tiles onto each other will be equally valid. However, we will show that a combination of good initial conditions and weak priors over the variables of interest allows us to arrive at an accurate solution.

2.2. Marginalization super-resolution

In the approach above, which we term the *joint MAP approach*, we estimate \mathbf{x} by maximizing over $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$. Now in a second approach, the *marginalization approach*, we estimate $p(\mathbf{x} | \{\mathbf{y}^{(k)}\})$ by marginalizing over $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$ instead. In the marginalization approach, a MAP estimate of \mathbf{x} can then be obtained by maximizing $p(\mathbf{x} | \{\mathbf{y}^{(k)}\})$ directly with respect to \mathbf{x} .

Using the identity

$$p(\mathbf{x} | \mathbf{d}) = \int p(\mathbf{x} | \mathbf{d}, \mathbf{t}) p(\mathbf{t}) d\mathbf{t}, \quad (9)$$

the integral over the unknown geometric and photometric parameters, $\{\boldsymbol{\theta}, \boldsymbol{\lambda}\}$, can be written as

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

$$= \int \frac{p(\mathbf{x}) \prod_{k=1}^K p(\mathbf{y}^{(k)} | \mathbf{x}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)})}{p(\{\mathbf{y}^{(k)}\} | \{\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\})} \times p(\{\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\}) d\{\boldsymbol{\theta}, \boldsymbol{\lambda}\} \quad (11)$$

$$= \frac{p(\mathbf{x})}{p(\{\mathbf{y}^{(k)}\})} \int \prod_{k=1}^K p(\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}) \times p(\mathbf{y}^{(k)} | \mathbf{x}, \boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}) d\{\boldsymbol{\theta}, \boldsymbol{\lambda}\}, \quad (12)$$

where expression (11) comes from substituting (5) into (10), and expression (12) uses the assumption that the images are generated independently from the model [15] to take the denominator out of the integral. Details of how this integral is evaluated are deferred to Section 4, but notice that the left-hand side depends only on \mathbf{x} , not the registration parameters $\boldsymbol{\theta}$ and $\boldsymbol{\lambda}$, and that on the right-hand side, the prior $p(\mathbf{x})$ is outside the integral.

3. SIMULTANEOUS SUPER-RESOLUTION WITH MOTION AND PRIOR ESTIMATION

In this section, we fill out the details of the joint MAP image registration and super-resolution approach, and couple it to a scheme for learning the parameters of the image prior, to form our complete simultaneous MAP super-resolution algorithm.

The first key point is that in addition to optimizing the objective function (6) with respect to the super-resolution image estimate \mathbf{x} , we also optimize it with respect to the geometric and photometric registration parameter set $\{\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}\}$. This strategy closely resembles the well-studied problem of bundle adjustment [25], in that the camera parameters and image features are found simultaneously. Because most high-resolution pixels are observed in most frames, the super-resolution problem is closest to the ‘‘strongly convergent camera geometry’’ setup, and conjugate gradient methods are expected to converge rapidly [25].

This optimization of the MAP objective function is interleaved with a scheme to update the values of α and ν which

parameterize the edge-preserving image prior. This overall super-resolution algorithm is assumed to have converged at a point where all parameters change by less than a preset threshold in successive iterations. An overview of the joint MAP algorithm is given in Figure 1, and details of the learning of the prior are given in Section 3.3.

Section 3.1 offers a few comments on model suitability and potential pitfalls. A sensible way of initializing the various parts of the super-resolution problem helps it converge rapidly to good solutions, so initialization details are given in Section 3.2. Finally, Section 3.3 gives details of the iterations used to tune the values of the prior parameters.

3.1. Discussion of the joint MAP model

Errors in either geometric or photometric registration in the low-resolution dataset have consequences for the estimation of other super-resolution components. The uncertainty in localization can give the appearance of a larger point-spread function kernel, because the effects of a scene point on the low-resolution image set is more dispersed. Uncertainty in photometric registration increases the variance of intensity values at each spatial location, giving the appearance of more low-resolution image noise, because low-resolution image values will tend to lie further from the values of the back-projected estimate. Increased noise in turn is an indicator that a change in the prior weighting is required, thus lighting parameters can have a knock-on effect on the image edge appearances.

By far the most difficult component of most super-resolution systems to determine is the point-spread function (PSF), which is of crucial importance, because it describes how each pixel in \mathbf{x} influences pixels in the observed images. Resulting from optical blur in the camera, artifacts in the sensor medium (film or a CCD array), and potentially also through motion during the image exposure, the PSF is almost invariably modelled either as an isotropic Gaussian or a uniform disk in super-resolution, though some authors suggest other functions derived from assumptions on the camera optics and sensor array [9, 16, 26]. The exact shape of the kernel depends on the entire process from photon to pixel.

Identifying and reversing the blur process is the domain of *blind image deconvolution*. Approaches based on generalized cross-validation [27] or maximum likelihood [28] are less sensitive to noise than other available techniques [29], and both have direct analogs in current super-resolution work [12, 15]. Because of the parametric nature of both sets of algorithms, neither is truly capable of recovering an arbitrary point-spread function. With this in mind, we choose a few sensible forms of PSF and concentrate on super-resolution which handles mismatches between the true and assumed PSF as gracefully as possible.

3.2. Initialization and implementation details

There are convenient initializations for the geometric and photometric registrations and for the high-resolution image \mathbf{x} , which by itself even gives a quick and reasonable super-resolution estimate. Input images are assumed to be

- (1) Initialize PSF, image registrations, super-resolution image and prior parameters according to Section 3.2.
- (2)
 - (a) (Re)-sample the set of validation pixels (see Section 3.3).
 - (b) Update α and ν (prior parameters) using cross-validation-style gradient descent (see Section 3.3). This includes a few steps of a suboptimization of \mathcal{F} with respect to \mathbf{x} .
 - (c) Optimize \mathcal{F} (6) jointly with respect to \mathbf{x} (super-resolution image), λ (photometric transform), and $\boldsymbol{\theta}$ (geometric transform). For SCG, the gradient expressions are given in (15) and (17).
- (3) If the maximum absolute change in α , ν , or any element of \mathbf{x} , λ , or $\boldsymbol{\theta}$ is above preset convergence thresholds, return to (2).

ALGORITHM 1: Basic structure of the multiframe super-resolution algorithm with simultaneous image registration and learning of prior parameter values.

preregistered by a standard algorithm such as RANSAC [23] so that points at the image centres correspond to within a small number of low-resolution pixels.

The image registration problem itself is not convex, and repeating textures can cause naive intensity-based registration algorithms to fall into a local minimum, though when initialized sensibly, very accurate results are obtained. The pathological case where the footprints of the low-resolution images fail to overlap in the high-resolution frame can be avoided by adding an extra prior term to \mathcal{F} to penalize large deviations in the registration parameters from the initial registration estimate.

The initial registration estimate (both geometric and photometric) is refined by optimizing the MAP objective function \mathcal{F} with respect to the registration parameters, but using a cheap over-smooth approximation to \mathbf{x} , known as the *average image*, \mathbf{a} [18]. Since \mathbf{a} is a function of the registration parameters, it is recalculated at each step. Details of the average image are given in Section 3.2.1, and the derivatives expressions for the simultaneous optimization method are given in (see Section 3.2.2).

Once $\{\boldsymbol{\theta}^{(k)}, \lambda^{(k)}\}$ have been estimated, the value of \mathbf{a} can be used as an initial estimate for \mathbf{x} , and then the scaled conjugate gradients algorithm is applied to the ML cost function (the first term of \mathcal{F}), but terminated after around $K/4$ steps, before the instabilities dominate because there is no prior. This gives a sharper result than initializing with \mathbf{a} as in [18]. When only a few images are available, a more stable ML solution can be found by using a constrained optimization to bound the pixel values so they must lie in the permitted image intensity range.

In our system, the elements of \mathbf{x} are scaled to lie in the range $[-1/2, 1/2]$, and the geometric registration is decompose into a “fixed” component, which is the initial mapping from $\mathbf{y}^{(k)}$ to \mathbf{x} , and a projective correction term, which is itself decomposed into constituent shifts, rotations, axis scalings, and projective parameters, which are the $\boldsymbol{\theta}$ parameters, then concatenated with λ to give one parameter vector. This is then “whitened” to be zero mean and have a std of 0.35 units, which is approximately the standard deviation of \mathbf{x} . The prior over registration values suggested above is achieved simply by penalizing large values in this registration vector.

Boundary conditions are treated as in [15], making the super-resolution image big enough so that the PSF kernel associated with any low-resolution pixel under any expected

registration is adequately supported. Gradients with respect to \mathbf{x} and λ can be found analytically, and those with respect to $\boldsymbol{\theta}$ are found numerically.

Finally, the prior parameters are initialized to around $\alpha = 0.01$ and $\nu = 0.1$. We work with $\log \alpha$ and $\log \nu$, since any real value for these log quantities gives a positive value for ν and α , which we require for the prior. For the PSF, a Gaussian with $\text{std} \approx 0.45$ low-resolution pixels is reasonable for in-focus images, and a disk of radius upwards of 0.8 is suitable for slightly defocused scenes.

3.2.1. The average image

The average image \mathbf{a} is a stable though excessively smooth approximation to \mathbf{x} [18]. 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

$$\mathbf{a} = \mathbf{S}^{-1} \mathbf{W}^T \boldsymbol{\Lambda}_\alpha^{-1} (\mathbf{y} - \boldsymbol{\Lambda}_\beta), \quad (13)$$

where \mathbf{W} , \mathbf{y} , $\boldsymbol{\Lambda}_\alpha$, and $\boldsymbol{\Lambda}_\beta$ are the stacks of the K groups of $\mathbf{W}^{(k)}$, $\mathbf{y}^{(k)}$, $\lambda_\alpha^{(k)} \mathbf{I}$, and $\lambda_\beta^{(k)} \mathbf{I}$, respectively, and \mathbf{S} 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. Using \mathbf{a} in place of \mathbf{x} , we optimize the first term of \mathcal{F} with respect to $\boldsymbol{\theta}$ and λ only. This provides a good estimate for the registration parameters, without requiring \mathbf{x} or the prior parameters.

3.2.2. Gradient expressions for the simultaneous method

Defining the model fit error for the k th image as $\mathbf{e}^{(k)}$, so that

$$\mathbf{e}^{(k)} = \mathbf{y}^{(k)} - \lambda_\alpha^{(k)} \mathbf{W}^{(k)} \mathbf{x} - \lambda_\beta^{(k)} \mathbf{1}, \quad (14)$$

then the gradient of the objective function \mathcal{F} (6) with respect to the super-resolution estimate \mathbf{x} can be computed as

$$\frac{\partial \mathcal{F}}{\partial \mathbf{x}} = -2\beta \sum_{k=1}^K \lambda_\alpha^{(k)} \mathbf{W}^{(k)T} \mathbf{e}^{(k)} - 2\nu \mathbf{D}^T \rho'(\mathbf{D}\mathbf{x}, \alpha), \quad (15)$$

where $\mathbf{D}\mathbf{x}$ is a vector comprising all the elements of $\mathcal{G}(\mathbf{x})$, and \mathbf{D} itself is a large sparse matrix. For each directional gradient

element z , the corresponding gradient element of the prior term is given by

$$\rho'(z, \alpha) = \begin{cases} 2z, & \text{if } |z| \leq \alpha, \\ 2\alpha \operatorname{sign}(z), & \text{otherwise.} \end{cases} \quad (16)$$

The gradients of the objective function with respect to the registration parameters are given by

$$\begin{aligned} \frac{\partial \mathcal{F}}{\partial \theta_i^{(k)}} &= -2\beta \sum_{\text{elements}} \left[\lambda_\alpha^{(k)} \mathbf{e}^{(k)} \mathbf{x}^T \odot \frac{\partial \mathbf{W}^{(k)}}{\partial \theta_i^{(k)}} \right], \\ \frac{\partial \mathcal{F}}{\partial \lambda_\alpha^{(k)}} &= -2\beta \mathbf{x}^T \mathbf{W}^{(k)} \mathbf{e}^{(k)}, \\ \frac{\partial \mathcal{F}}{\partial \lambda_\beta^{(k)}} &= -2\beta \sum_i^M e_i^{(k)}, \end{aligned} \quad (17)$$

where \odot is the Hadamard (element-wise) matrix product.

The \mathbf{W} matrix represents the composition of spatial blur, decimation, and resampling of the high-resolution image in the frame of the low-resolution image, so even for a relatively simple motion model (such as an affine homography with 6 degrees of freedom per image in the geometric registration parameters), it is quicker to calculate the partial derivative with respect to the parameters, $\partial \mathbf{W}^{(k)} / \partial \theta_i^{(k)}$, using a central difference approximation than to evaluate explicit derivatives using the chain rule.

3.3. Learning the prior parameters with possible registration error

It is necessary to determine ν and α of the Huber function of (4) while still in the process of converging on the estimates of \mathbf{x} , $\boldsymbol{\theta}$, and $\boldsymbol{\lambda}$. This is done by removing some *individual low-resolution pixels* from the problem, solving for \mathbf{x} using the remaining pixels, then projecting this back into the original image frames to determine its quality by the withheld validation pixels using a robust L_1 norm. The selected α and ν should minimize this cross-validation error.

This defines a subtly different cross-validation approach to those used previously for image super-resolution, because validation pixels are selected at random from the collection of $K \times M$ *individual linear equations* comprising the overall problem, rather than from the K *images*. This distinction is important when uncertainty in the registrations is assumed, since validation *images* can be misregistered in their entirety. Assuming independence of the registration error on each frame given \mathbf{x} , the pixel-wise validation approach has a clear advantage.

In determining a search direction in (ν, α) -space, \mathcal{F} can be optimized with respect to \mathbf{x} , starting with the current \mathbf{x} estimate, for *just a few steps* to determine whether the parameter combination improves the estimate. This intermediate optimization does not need to run to convergence in order to provide a gradient direction worthy of exploration. This is much faster than the usual approach of running a complete optimization for a number of parameter combinations, especially useful if the initial estimate is poor. An arbitrary

5% of pixels are used for validation, ignoring regions within a few pixels of edges, to avoid boundary complications, and because inputs are centred on the region of interest.

4. THE MARGINALIZATION APPROACH

We now turn our attention to handling residual registration uncertainty by considering distributions over possible registrations, then integrating these out of the problem. A set of equations depending only upon the super-resolution estimate \mathbf{x} , the input images $\{\mathbf{y}^{(k)}\}$, and a starting estimate of the registration parameter distributions are used to refine the super-resolution estimate without having to maintain a registration estimate.

When the registration is known approximately, for instance by preregistering inputs (as described in Section 3.2), the uncertainty can be modeled as a Gaussian perturbation about the mean estimate $[\bar{\boldsymbol{\theta}}^{(k)T}, \bar{\lambda}_\alpha^{(k)}, \bar{\lambda}_\beta^{(k)}]^T$ for each image's parameter set,

$$\begin{bmatrix} \boldsymbol{\theta}^{(k)} \\ \lambda_\alpha^{(k)} \\ \lambda_\beta^{(k)} \end{bmatrix} = \begin{bmatrix} \bar{\boldsymbol{\theta}}^{(k)} \\ \bar{\lambda}_\alpha^{(k)} \\ \bar{\lambda}_\beta^{(k)} \end{bmatrix} + \boldsymbol{\delta}^{(k)}, \quad (18)$$

$$\boldsymbol{\delta}^{(k)} \sim \mathcal{N}(\mathbf{0}, \mathbf{C}), \quad (19)$$

$$p(\boldsymbol{\theta}^{(k)}, \boldsymbol{\lambda}^{(k)}) = \left(\frac{|\mathbf{C}^{-1}|}{(2\pi)^n} \right)^{1/2} \exp \left\{ -\frac{1}{2} \boldsymbol{\delta}^{(k)T} \mathbf{C}^{-1} \boldsymbol{\delta}^{(k)} \right\}. \quad (20)$$

In order to obtain an expression for $p(\mathbf{x} | \{\mathbf{y}^{(k)}\})$ from (2), (4), and (20), the parameter variations $\boldsymbol{\delta}^{(k)}$ must be integrated out of the problem, and details of this are given in the following subsection. The diagonal matrix \mathbf{C} is constructed to reflect the confidence in each parameter estimate. This might mean a standard deviation of a tenth of a low-resolution pixel on image translation parameters, or a few grey levels' shift on the illumination model, for instance.

4.1. Marginalizing over registration parameters

We now give details of how the integral is evaluated. With reference to (12), substituting in (2), (4), and (20), the integral performed is

$$\begin{aligned} p(\mathbf{x} | \{\mathbf{y}^{(k)}\}) &= \frac{1}{p(\{\mathbf{y}^{(k)}\})} \left(\frac{\beta}{2\pi} \right)^{KM/2} \left(\frac{b}{2\pi} \right)^{Kn/2} \frac{1}{Z_x} \\ &\times \exp \left\{ -\frac{\nu}{2} \sum_{z \in \mathcal{G}(\mathbf{x})} \rho(z, \alpha) \right\} \\ &\times \int \exp \left\{ -\sum_{k=1}^K \left[\frac{\beta}{2} r^{(k)} + \frac{1}{2} \boldsymbol{\delta}^{(k)T} \mathbf{C}^{(k)-1} \boldsymbol{\delta}^{(k)} \right] \right\} \mathbf{d}\boldsymbol{\delta}, \end{aligned} \quad (21)$$

where

$$\begin{aligned} r^{(k)} &= \|\mathbf{e}^{(k)}\|_2^2, \\ \boldsymbol{\delta}^T &= [\boldsymbol{\delta}^{(1)T}, \boldsymbol{\delta}^{(2)T}, \dots, \boldsymbol{\delta}^{(K)T}], \end{aligned} \quad (22)$$

and all the λ and $\boldsymbol{\theta}$ parameters are functions of $\boldsymbol{\delta}$ as in (18).

Expanding the data error term in the exponent for each low-resolution image as a second-order Taylor series about the estimated geometric registration parameter yields

$$r^{(k)}(\boldsymbol{\delta}) \approx F^{(k)} + \mathbf{G}^{(k)T} \boldsymbol{\delta} + \frac{1}{2} \boldsymbol{\delta}^{(k)T} \mathbf{H}^{(k)} \boldsymbol{\delta}^{(k)}. \quad (23)$$

Values for F , \mathbf{G} , and \mathbf{H} in our implementation are found numerically (for geometric registrations) or analytically (for the photometric parameters) from \mathbf{x} and $\{\mathbf{y}^{(k)}, \boldsymbol{\theta}^{(k)}, \lambda_\alpha^{(k)}, \lambda_\beta^{(k)}\}$. Thus the whole exponent of (21), f , becomes

$$\begin{aligned} f &= \sum_{k=1}^K \left(-\frac{\beta}{2} F^{(k)} - \frac{\beta}{2} \mathbf{G}^{(k)T} \boldsymbol{\delta}^{(k)} \right. \\ &\quad \left. - \frac{1}{2} \boldsymbol{\delta}^{(k)T} \left[\frac{\beta}{2} \mathbf{H}^{(k)} + \mathbf{C}^{-1} \right] \boldsymbol{\delta}^{(k)} \right) \\ &= -\frac{\beta}{2} F - \frac{\beta}{2} \mathbf{G}^T \boldsymbol{\delta} - \frac{1}{2} \boldsymbol{\delta}^T \left[\frac{\beta}{2} \mathbf{H} + \mathbf{V}^{-1} \right] \boldsymbol{\delta}, \end{aligned} \quad (24)$$

where the omission of image superscripts indicates stacked matrices, and \mathbf{H} is therefore a block-diagonal $nK \times nK$ sparse matrix, and \mathbf{V} consists of the repeated diagonal of \mathbf{C} .

Finally, letting $\mathbf{S} = (\beta/2)\mathbf{H} + \mathbf{V}^{-1}$,

$$\int \exp\{f\} \mathbf{d}\boldsymbol{\delta} = \exp\left\{-\frac{\beta}{2} F\right\} \int \exp\left\{-\frac{\beta}{2} \mathbf{G}^T \boldsymbol{\delta} - \frac{1}{2} \boldsymbol{\delta}^T \mathbf{S} \boldsymbol{\delta}\right\} \mathbf{d}\boldsymbol{\delta} \quad (25)$$

$$= \exp\left\{-\frac{\beta}{2} F\right\} (2\pi)^{nK/2} |\mathbf{S}|^{-1/2} \exp\left\{\frac{\beta^2}{8} \mathbf{G}^T \mathbf{S}^{-1} \mathbf{G}\right\}. \quad (26)$$

The objective function, \mathcal{L} to be minimized with respect to \mathbf{x} , is obtained by taking the negative log of (21), using the result from (26), and neglecting the constant terms:

$$\mathcal{L} = \frac{\nu}{2} \rho(\mathbf{D}\mathbf{x}, \alpha) + \frac{\beta}{2} F + \frac{1}{2} \log |\mathbf{S}| - \frac{\beta^2}{8} \mathbf{G}^T \mathbf{S}^{-1} \mathbf{G}. \quad (27)$$

This can be optimized using SCG [24], noting that the gradient can be expressed:

$$\begin{aligned} \frac{\mathbf{d}\mathcal{L}}{\mathbf{d}\mathbf{x}} &= \frac{\nu}{2} \mathbf{D}^T \frac{\mathbf{d}}{\mathbf{d}\mathbf{x}} \rho(\mathbf{D}\mathbf{x}) + \frac{\beta}{2} \frac{\mathbf{d}F}{\mathbf{d}\mathbf{x}} - \frac{\beta^2}{4} \mathbf{G}^T \mathbf{S}^{-1} \frac{\mathbf{d}\mathbf{G}}{\mathbf{d}\mathbf{x}} \\ &\quad + \left[\frac{\beta}{4} \text{vec}(\mathbf{S}^{-1})^T + \frac{\beta^3}{16} (\mathbf{G}^T \mathbf{S}^{-1} \otimes \mathbf{G}^T \mathbf{S}^{-1}) \right] \frac{\mathbf{d} \text{vec} \mathbf{H}}{\mathbf{d}\mathbf{x}}, \end{aligned} \quad (28)$$

where \otimes is the Kronecker product and vec is the operation that vectorizes a matrix. 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.

4.2. Discussion of the marginalization approach

It is possible to interpret the extra terms introduced into the objective function in the derivation of the marginalization method as an extra regularizer term or image prior. Considering (27), 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 $\boldsymbol{\theta}$ and λ tightens to a single point, the terms tend to constant values.

The intuition behind the method's success (see Section 6) is that this prior will favor image solutions which are not acutely sensitive to minor adjustments in the image registration. The images of Figure 2 illustrate the type of solution which would score poorly. To create the figure, one dataset was used to produce two 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 checkerboard 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.

4.3. Implementation details for parameter marginalization

The terms of the Taylor expansion are found using a mixture of analytic and numerical gradients. Notice that the value F is simply the reprojection error of the current estimate of \mathbf{x} at the mean registration parameter values, and that gradients of this expression with respect to the λ parameters, and with respect to \mathbf{x} can both be found analytically. To find the gradient with respect to a geometric registration parameter $\theta_i^{(k)}$, 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 using around 10 iterations of SCG to find the maximum likelihood solution evaluated at these mean parameters. Additionally, pixel values are scaled to lie between $-1/2$ and $1/2$, and the ML solution is bounded to lie within these values in order to curb the severe overfitting usually observed in ML super-resolution results.

5. EXPERIMENTAL RESULTS FOR SIMULTANEOUS MAP APPROACH

The performance of simultaneous registration, super-resolution, and prior updating is evaluated using real data from a variety of sources. Using the scaled conjugate gradients (SCG) implementation from Netlab [24], rapid convergence is observed up to a point, beyond which a slow steady decrease in \mathcal{F} gives no subjective improvement in the solution, but this can be avoided by specifying sensible convergence criteria.

The joint MAP results are contrasted with a fixed-registration approach, where registrations between the inputs are found then fixed before the super-resolution process.

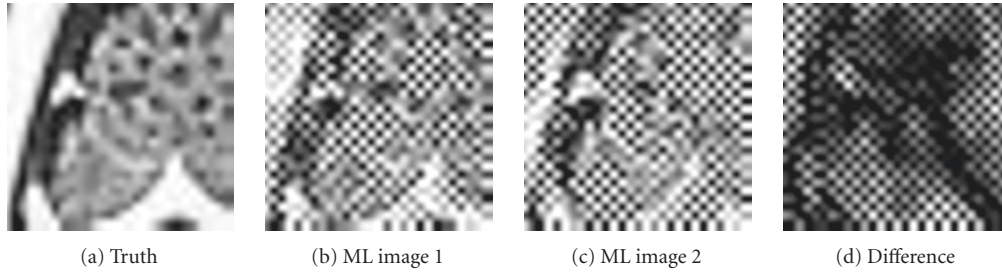


FIGURE 2: An example of the effect of tiny changes in the registration parameters. (a) Ground truth image from which a 16-image low-resolution dataset was generated. (b), (c) Two ML super-resolution estimates. In both cases, the same dataset was used, but the registration parameters were perturbed by 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.

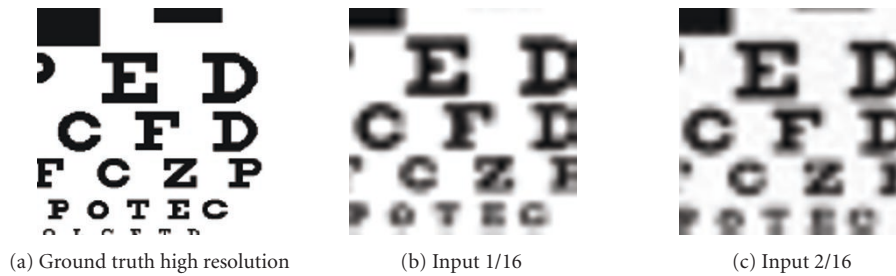


FIGURE 3: Synthetic data: (a) ground truth image. (b), (c) Two example low-resolution images of 30×30 pixels, with clearly different geometric and photometric registrations.

This fixed registration is found using the method described in Section 3.2, and then (6) is optimized with respect *only* to \mathbf{x} to obtain a high-resolution estimate.

Synthetic dataset

Experiments are first performed on synthetic data, generated using the generative model (1) applied to a ground truth image at a zoom factor of 4, with each pixel being corrupted by additive Gaussian to give a SNR of 30 dB. Values for a shift-only geometric registration, $\boldsymbol{\theta}$, and a 2D photometric registration $\boldsymbol{\lambda}$ are sampled independently from uniform distributions. The ground truth image and two of the low-resolution images generated by the forward model are shown in Figure 3. The mean intensity is clearly different, and the vertical shift is easily observed by comparing the top and bottom edge pixels of each low-resolution image.

An initial registration was then carried out using an iterative intensity-based scheme which optimized both geometric and photometric parameters. This initial “fixed” registration differs from the ground truth by an average of 0.0142 pixels, and 1.00 grey levels for the photometric shift. Allowing the joint MAP super-resolution algorithm to update this registration while super resolving the image resulted in registration errors of just 0.0024 pixels and 0.28 grey levels given the optimal prior settings (see below and Figure 4).

We now sweep through values of the prior strength parameter ν , keeping the Huber parameter α set to 0.04. The noise precision parameter β is chosen so that the noise is

assumed to have a standard deviation of 5 grey levels. For each value of ν , both the fixed-registration and the joint MAP methods are applied to the data, and the root mean square error (RMSE) compared to the ground truth image is calculated.

The RMSE compared to the ground truth image for both the fixed registration and the joint MAP approach are plotted, in Figure 4, along with a curve representing the performance if the ground truth registration is known. The prior strength represented on the horizontal axis is $\log_{10}(\nu/\beta)$. Examples of the improvement in geometric and photometric registration parameters are also shown.

Note that we have not learned the prior values in this synthetic-data experiment, in order to plot how the value of ν affects the output. We now evaluate the performance of the whole simultaneous super-resolution algorithm, including the learning of the ν and α values, on a selection of real sequences.

Surrey library sequence

The camera motion is a slow pan through a small angle, and the sign on a wall is illegible given any one of the inputs alone. A small interest area of size 25×95 pixels is highlighted in the first of the 30 frames. Gaussian PSFs with std = 0.375, 0.45, 0.525 are selected, and used in both algorithms. There are 77003 elements in \mathbf{y} , and \mathbf{x} has 45936 elements with a zoom factor of 4. \mathbf{W} has around 3.5×10^9 elements, of which around 0.26% are nonzero with the smallest of these

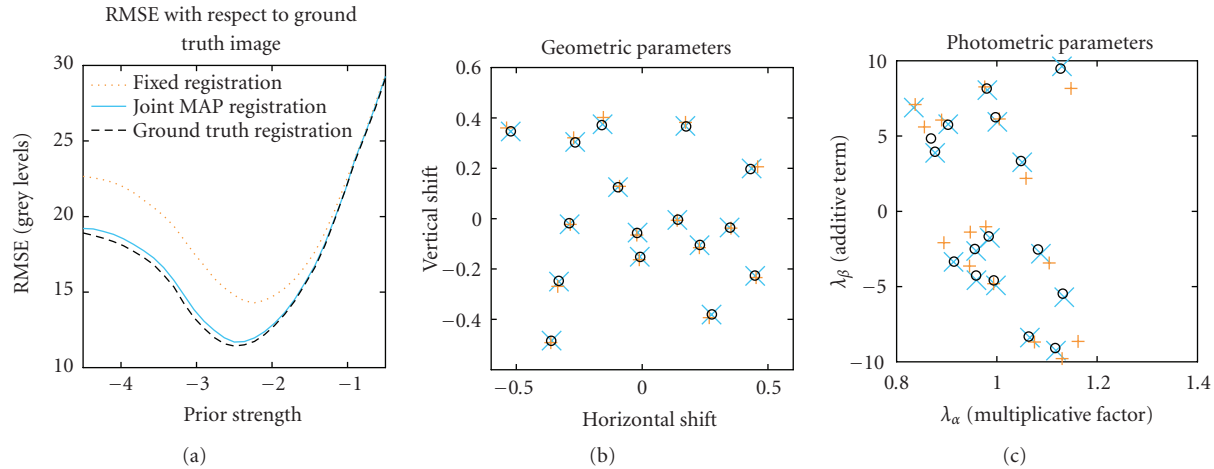


FIGURE 4: Synthetic data plots. (a) RMSE compared to ground truth, plotted for the fixed and joint MAP algorithms, and for the Huber super-resolution image found using the ground truth registration. (b), (c) plots showing the registration values for the initial (orange “+”), joint MAP (blue “x”) and ground truth (black “o”) registrations. In most cases, the joint MAP registration value is considerably closer to the true value than the initial “fixed” value is.

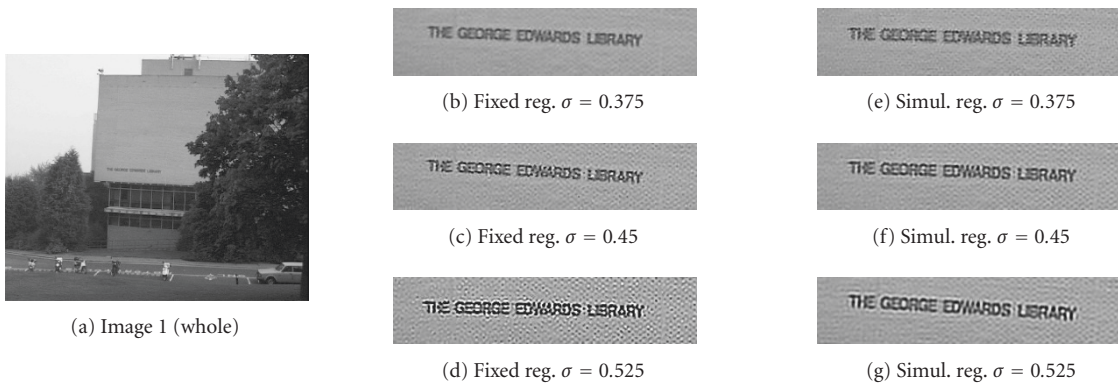


FIGURE 5: Surrey library sequence. (a) One of the 30 original images. (b), (c), (d) Super-resolution found using fixed registrations. (e), (f), (g) Super-resolution images using simultaneous MAP algorithm. Detailed regions of two of the low-resolution images can be seen in Figures 1(a), 1(b).

PSF kernels, and 0.49% with the largest. Most instances of the simultaneous algorithm converge in 2 to 5 iterations. Results are shown in Figure 5, showing that while both algorithms perform well with the middle PSF size, the simultaneous-registration algorithm handles deviations from this optimum more gracefully.

“Československo” sequence

The ten images in this sequence were captured on a rig which constrained the motion to be pure translation, though photometric differences are very apparent in the input images. Gaussian PSFs with $\text{std} = 0.325, 0.40, 0.475$ are used in both super-resolution algorithms. The results are shown in Figure 6, and the lines and text are much more clearly defined in the super-resolution version.

Eye-test card sequence

The second real-data experiment uses just 10 images of an eye-test card, captured using a webcam. The card is tilted and rotated slightly, and image brightness varies as the lighting and camera angles change. Gaussian PSFs with $\text{std} = 0.30, 0.375, 0.45$ are used in both super-resolution algorithms. The results are shown in the left portion of Figure 7. Note that the last row is illegible in the low-resolution images, but can be read in the super-resolution images.

Camera “9” sequence

The model is adapted to handle DVD input, where the aspect ratio of the input images is 1.25 : 1, but they represent 1.85 : 1 video. The correction in the horizontal scaling is incorporated into the “fixed” part of the homography representation, and the PSF is assumed to be radially symmetric. This avoids

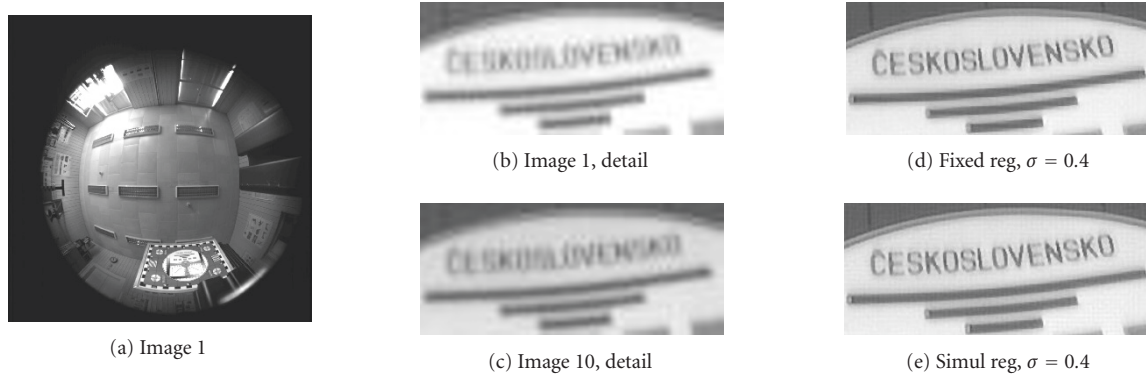


FIGURE 6: “Československo” sequence. (a) The first image in the sequence. (b), (c) details of the region of interest in the first and last low-resolution images. (d) Super-resolution found using fixed registrations. (e) Super-resolution images using simultaneous MAP algorithm.

an undesirable interpolation of the inputs prior to super resolving, which would lose high-frequency information, and also avoids working with squashed super-resolution images throughout the process, which would violate the assumption of an isotropic prior over \mathbf{x} . In short, we do not scale any of the images, but instead work with inputs and outputs at different aspect ratios.

The Camera “9” sequence consists of 29 I-frames¹ from the movie *Groundhog Day*. An on-screen hand-held TV camera moves independently of the real camera, and the logo on the side is chosen as the interest region. Disk-shaped PSFs with radii of 1.0, 1.4, and 1.8 pixels are used. In both the eye-test card and Camera “9” sequences, the simultaneously optimized super-resolution images again appear subjectively better to the human viewer, and are more consistent across different PSFs.

Lola Rennt sequences

Finally, results obtained from difficult DVD input sequences that were taken from the movie *Lola Rennt* are shown in Figure 8. In the “cars” sequence, there are just 9 I-frames showing a pair of cars, and the areas of interest are the car number plates. The “badge” sequence shows the badge of a bank security officer. Seven I-frames are available, but are all dark, making the noise level proportionally very high. Significant improvements at a zoom factor of 4 (in each direction) can be seen.

6. EXPERIMENTAL RESULTS FOR THE MARGINALIZATION APPROACH

The performance of the marginalization approach was evaluated in a similar way to the simultaneous joint MAP method of Section 5. The objective function (27) was optimized directly with respect to the super-resolution image pixels, first

working on synthetic datasets with known ground truth, and then on real-data sequences. Results are compared with the fixed-registration Huber-MAP method, and with the simultaneous joint MAP method.

Synthetic experiments

The first experiment takes a sixteen-image synthetic dataset created from the eyechart image of Figure 3(a). The dataset is generated using the same procedure as already described, except that the subpixel 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 [12], but with additional lighting variation.

The images giving lowest RMS error from each set are displayed in Figure 9. The lowest RMSE for the marginalizing approach is 11.73 grey levels, and the corresponding RMSE for the registration-fixing approach is 14.01. Using the L_1 norm (mean absolute pixel difference), the error is 3.81 grey levels for the fixed-registration approach, and 3.29 for the marginalizing approach proposed here. The standard deviation of the prior over θ is set to 0.004, which is found empirically to give good results. Visually, the differences between the images are subtle, though the bottom row of letters is better defined in the marginalization approach.

The RMSE for three approaches (fixed registration, joint MAP, and marginalizing) is plotted in Figure 10, and again the horizontal axis represents $\log_{10}(\nu/\beta)$. The dotted orange curve reflects the error from the fixed-registration approach using the registration estimated from the low-resolution inputs. Both the joint MAP (blue curve) and marginalization (green curve) approaches obtain lower errors, closer to those obtained if the ground truth registration is known (dashed black curve).

Note that while the lowest error values are achieved using the joint MAP approach, the results using the marginalization approach are obtained using only the initial (incorrect) registration values. The marginalization approach also stays consistently good over a wider range of possible prior values, making it more robust than either of the other methods to

¹ I-frames are encoded as complete images, rather than requiring nearby frames in order to render them.

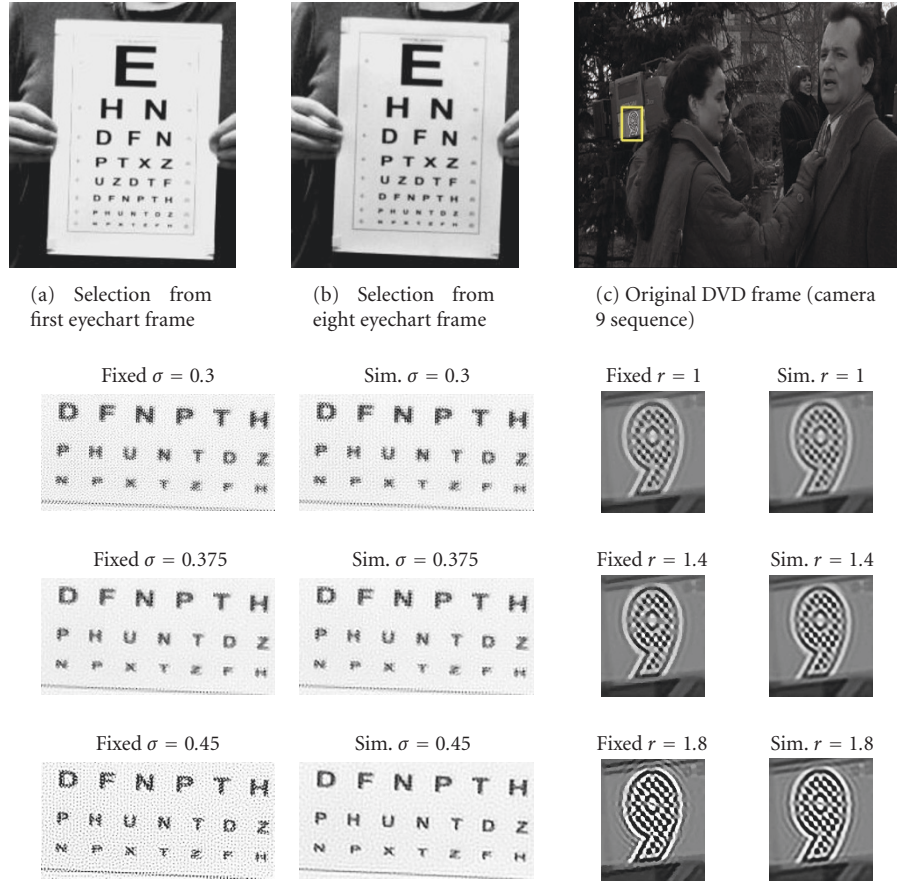


FIGURE 7: Eyechart sequence and camera “9” sequence. (a), (b) Sections of two input images from the 10-frame eye-test card sequence. Notice the card appears much brighter in the left image. (c) Raw DVD frame for camera “9” sequence (see Figure 1 for close-up of interest region). Lower section, first and third columns: results obtained by fixing registration prior to super resolution. Lower section, second and fourth columns: results obtained using the simultaneous approach to optimize the super-resolution image, registration parameters, and prior parameters.

poor estimates of the prior distribution, or of the precision β of the noise on the input dataset.

Real data

We again use the “Československo,” which is seen in Figure 6. Image registration was carried out in the same manner as before, and the geometric parameters agree with the provided homographies to within a few hundreds of a pixel. Super-resolution images were created for a number of ν values, and subjectively, the equivalent values to those quoted in [18] were selected for the Huber recovery. As with the synthetic data, a slightly larger ν value was chosen for the registration-marginalizing output, and a similar covariance of the registration parameters was assumed. We also compare against Tipping and Bishop’s method [15], which we have extended to cover the illumination model and used to register and super resolve the dataset, using the same PSF value (0.4 low-resolution pixels) as the other methods.

The three sets of results on the real-data sequence are shown in the middle and bottom rows of Figure 11. To facil-

itate a better comparison, a subregion of each is expanded to make the letter details clearer. The Huber prior used alone in the fixed-registration method tends to make the edges unnaturally sharp, though it is very successful at regularizing the solution elsewhere. The text in the marginalizing approach’s image appears clearer than the text in the image found using Tipping and Bishop’s method, and the regularization in the constant background regions is slightly more successful.

7. CONCLUSIONS

This work has examined two methods of considering the image registration and other parts of the super-resolution problem at the same time as the high-resolution image estimate, and illustrated these with examples where both methods give improvements in the quality of the resulting super-resolved image.

Firstly, we showed that optimizing the MAP image solution with respect to the low-resolution image registration parameters as well as the high-resolution pixel values yields a better solution than the two-phase register-then-super-



FIGURE 8: Results from the simultaneous super-resolution, image registration, and prior parameter updating scheme applied to the movie *Lola Rennt* on DVD. (a): two raw DVD frames. (b): five low-res frames from each sequence (black car's number plate, white car's number plate, security guard's ID badge). (c): the same image regions super resolved using the simultaneous method. In the case of the security guard's ID badge, intensities have been scaled for ease of viewing. Please refer to the text for notes on the aspect ratios involved when working with DVD frames.

resolve approach conventionally used. In addition to this, our MAP algorithm included a cross-validation step to select a prior distribution appropriate for the scene statistics, which could be incorporated without great additional expense into the iterative recovery algorithm.

Secondly, we developed an alternative approach using Bayesian marginalization within the super-resolution model, with several advantages over Tipping and Bishop's original 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 fixed-registration approach, and over the result from Tipping and Bishop's method, largely

owing to our more favourable prior over the super-resolution image.

In future work, a combination of these methods may prove most accurate, with the registration improvement of the simultaneous method providing a stronger starting point for the marginalization approach, as well as a very accurate initial super-resolution image estimate. It seems plausible that image accuracy could be improved considerably by swapping the Huber prior used here with another more specialized or domain-specific image prior [20, 30].

ACKNOWLEDGMENTS

The "library" data sequence used in Figures 1 and 5 is due to Barbara Levienaise-Obadia, University of Surrey,

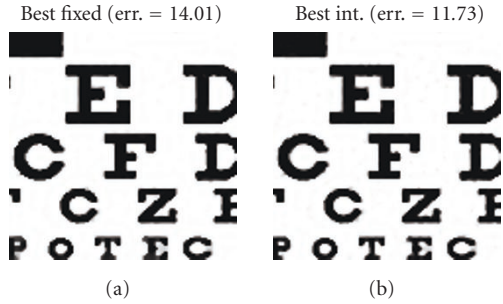


FIGURE 9: Synthetic dataset results. (a) The best (minimum MSE) image from the fixed-registration algorithm, having super resolved the dataset multiple times with different prior strength settings. (b) The best result using our approach of integrating over θ and λ . As well as having a lower RMSE, note the improvement in black-white edge detail on some of the letters on the bottom line.

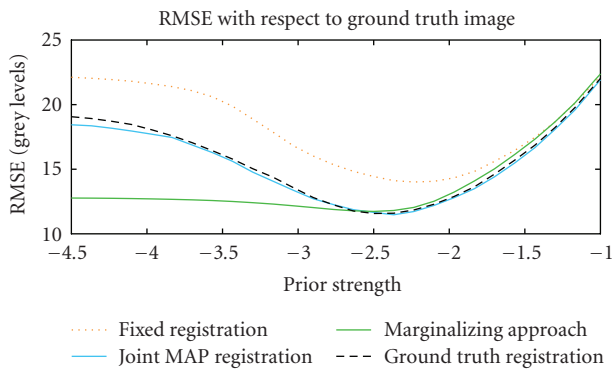


FIGURE 10: Plot showing the variation of RMSE with prior strength for the fixed Huber MAP method and our approach integrating over θ and λ , applied to the synthetic dataset of Figure 9. As well as reaching a lower minimum, the integrating approach appears to be more consistent across variations in prior strength.

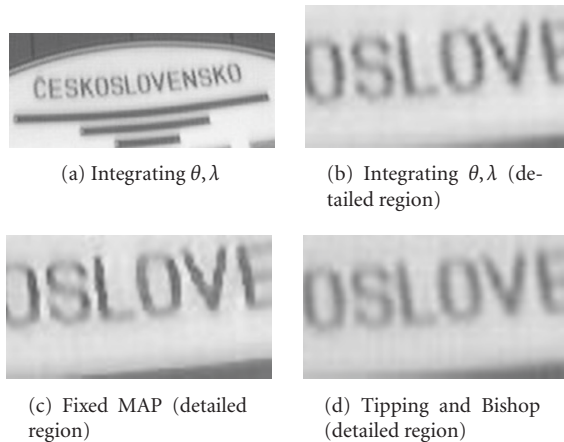


FIGURE 11: (a) The full super-resolution output from our algorithm. (b) Detailed region of the central letters, again with our algorithm. (c) Detailed region of the regular Huber MAP super-resolution image, using parameter values suggested in [18], which are also found to be subjectively good choices. The edges are slightly artificial 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 present on the input data.

and the constrained-motion dataset of Figure 11 is due to Tomas Pajdla and Daniel Martinec, CMP, Prague. Both datasets are available for download from <http://www.robots.ox.ac.uk/~vgg/data>. This work was funded in part by EC Network of Excellence PASCAL and by the EPSRC. David Capel is with 2D3, <http://www.2d3.com>.

REFERENCES

- [1] M. Irani and S. Peleg, "Super resolution from image sequences," in *Proceedings of the 10th International Conference on Pattern Recognition (ICPR '90)*, vol. 2, pp. 115–120, Atlantic City, NJ, USA, June 1990.
- [2] A. J. Patti, M. I. Sezan, and A. M. Tekalp, "Robust methods for high-quality stills from interlaced video in the presence of dominant motion," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 7, no. 2, pp. 328–342, 1997.
- [3] Salient Stills, <http://www.salientstills.com/>.
- [4] P. Cheeseman, B. Kanefsky, R. Kraft, J. Stutz, and R. Hanson, "Super-resolved surface reconstruction from multiple images," in *Maximum Entropy and Bayesian Methods*, G. R. Heidbreder, Ed., pp. 293–308, Kluwer Academic Publishers, Dordrecht, The Netherlands, 1996.
- [5] R. C. Hardie, K. J. Barnard, and E. E. Armstrong, "Joint MAP registration and high-resolution image estimation using a sequence of undersampled images," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1621–1633, 1997.
- [6] M. V. Joshi, S. Chaudhuri, and R. Panuganti, "A learning-based method for image super-resolution from zoomed observations," *IEEE Transactions on Systems, Man, and Cybernetics, Part B*, vol. 35, no. 3, pp. 527–537, 2005.
- [7] D. P. Capel and A. Zisserman, "Automated mosaicing with super-resolution zoom," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '98)*, pp. 885–891, Santa Barbara, Calif, USA, June 1998.
- [8] Y. Altunbasak, A. J. Patti, and R. M. Mersereau, "Super-resolution still and video reconstruction from MPEG-coded video," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 12, no. 4, pp. 217–226, 2002.
- [9] S. Baker and T. Kanade, "Limits on super-resolution and how to break them," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 9, pp. 1167–1183, 2002.
- [10] B. Bascle, A. Blake, and A. Zisserman, "Motion deblurring and super-resolution from an image sequence," in *Proceedings of the 4th European Conference on Computer Vision (ECCV '96)*, vol. 2, pp. 573–582, Springer, Cambridge, UK, April 1996.
- [11] M. Elad and A. Feuer, "Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images," *IEEE Transactions on Image Processing*, vol. 6, no. 12, pp. 1646–1658, 1997.
- [12] N. Nguyen, P. Milanfar, and G. Golub, "Efficient generalized cross-validation with applications to parametric image restoration and resolution enhancement," *IEEE Transactions on Image Processing*, vol. 10, no. 9, pp. 1299–1308, 2001.
- [13] R. R. Schultz and R. L. Stevenson, "A Bayesian approach to image expansion for improved definition," *IEEE Transactions on Image Processing*, vol. 3, no. 3, pp. 233–242, 1994.
- [14] D. Robinson and P. Milanfar, "Fundamental performance limits in image registration," *IEEE Transactions on Image Processing*, vol. 13, no. 9, pp. 1185–1199, 2004.

- [15] M. E. Tipping and C. M. Bishop, "Bayesian image super-resolution," in *Proceedings of Advances in Neural Information Processing Systems 15 (NIPS '02)*, pp. 1279–1286, Vancouver, British Columbia, Canada, December 2002.
- [16] S. Borman, *Topics in multiframe superresolution restoration*, Ph.D. thesis, University of Notre Dame, Notre Dame, Ind, USA, May 2004.
- [17] S. Borman and R. L. Stevenson, "Simultaneous multi-frame MAP super-resolution video enhancement using spatiotemporal priors," in *Proceedings of International Conference on Image Processing (ICIP '99)*, vol. 3, pp. 469–473, Kobe, Japan, October 1999.
- [18] D. P. Capel, *Image Mosaicing and Super-Resolution*, Distinguished Dissertations, Springer, New York, NY, USA, 2004.
- [19] S. Farsiu, M. Elad, and P. Milanfar, "A practical approach to super-resolution," in *Visual Communications and Image Processing*, vol. 6077 of *Proceedings of SPIE*, San Jose, Calif, USA, January 2006.
- [20] L. C. Pickup, S. J. Roberts, and A. Zisserman, "A sampled texture prior for image super-resolution," in *Proceedings of Advances in Neural Information Processing Systems 16 (NIPS '03)*, pp. 1587–1594, Vancouver, British Columbia, Canada, December 2004.
- [21] L. C. Pickup, D. P. Capel, S. J. Roberts, and A. Zisserman, "Bayesian image super-resolution, continued," in *Advances in Neural Information Processing Systems 19*, pp. 1089–1096, Cambridge, Mass, USA, December 2006.
- [22] L. C. Pickup, S. J. Roberts, and A. Zisserman, "Optimizing and learning for super-resolution," in *Proceedings of the 17th British Machine Vision Conference (BMVC '06)*, Edinburgh, UK, September 2006.
- [23] R. I. Hartley and A. Zisserman, *Multiple View Geometry in Computer Vision*, Cambridge University Press, Cambridge, UK, 2nd edition, 2004.
- [24] I. Nabney, *NETLAB: Algorithms for Pattern Recognition*, Springer, New York, NY, USA, 2002.
- [25] B. Triggs, P. F. McLauchlan, R. I. Hartley, and A. W. Fitzgibbon, "Bundle adjustment—a modern synthesis," in *Proceedings of International Workshop on Vision Algorithms on Vision Algorithms: Theory and Practice*, B. Triggs, A. Zisserman, and R. Szeliski, Eds., vol. 1883 of *Lecture Notes in Computer Science*, pp. 298–372, Springer, Corfu, Greece, September 1999.
- [26] R. C. Hardie, K. J. Barnard, J. G. Bogner, E. E. Armstrong, and E. A. Watson, "High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system," *Optical Engineering*, vol. 37, no. 1, pp. 247–260, 1998.
- [27] S. J. Reeves and R. M. Mersereau, "Blur identification by the method of generalized cross-validation," *IEEE Transactions on Image Processing*, vol. 1, no. 3, pp. 301–311, 1992.
- [28] R. L. Lagendijk, A. M. Tekalp, and J. Biemond, "Maximum likelihood image and blur identification: a unifying approach," *Optical Engineering*, vol. 29, no. 5, pp. 422–435, 1990.
- [29] D. Kundur and D. Hatzinakos, "Blind image deconvolution," *IEEE Signal Processing Magazine*, vol. 13, no. 3, pp. 43–64, 1996.
- [30] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

Lyndsey C. Pickup is a Researcher in the Machine Learning Group and Visual Geometry Groups at the University of Oxford. She graduated from Keble College, University of Oxford, with first class honours in engineering and computing science in 2002. Her interests lie in the application of Bayesian methods to computer vision, and specifically in handling noise and uncertainty in super-resolution without oversimplifying the image model.



David P. Capel received the M.Eng. degree in engineering and computing science from Oxford University in 1996. He completed his Ph.D. degree on image mosaicing and super-resolution as part of the Visual Geometry Group, also at Oxford University, in 2001. Since then, he has worked as a Vision Scientist at 2d3 Ltd., contributing to the development of the Emmy award-winning camera tracking software, "boujou." In his current role as Lead Scientist for 2d3's Advanced Imagery Group, he focuses on computer vision applications for aerial imagery. His research interests are in real-time computer vision and video enhancement, sensor fusion for long-range camera tracking, and automatic scene reconstruction.



Stephen J. Roberts' main area of research lies in machine learning approaches to data analysis. He has particular interests in the development of machine learning theory for problems in time series analysis and decision theory. His current research applies Bayesian statistics, graphical models, and information theory to diverse problem domains including mathematical biology, finance, and sensor fusion. He runs the Pattern Analysis and Machine Learning Research Group at the University of Oxford and is a fellow of Somerville College, Oxford.



Andrew Zisserman is the RAE/Microsoft Professor of computer vision at the Department of Engineering Science, University of Oxford, where he heads the Visual Geometry Group. He graduated from the University of Cambridge with a degree in theoretical physics, and for the last 20 years has carried out research in computer vision. He has coauthored and coedited several books on this area. The most recent, *Multiple View Geometry in Computer Vision* (written with Richard Hartley), has now been published as a second edition in paperback and also translated into Chinese. He has been a Program Chair and a General Chair for the IEEE International Conference on Computer Vision. He was elected a fellow of the Royal Society in 2007.