

Optical flow based super-resolution: A probabilistic approach

Rik Fransens^{a,*}, Christoph Strecha^a, Luc Van Gool^{a,b}

^a *ESAT-PSI, University of Leuven, Belgium*

^b *Computer Vision Group (BIWI), ETH Zuerich, Switzerland*

Received 25 March 2005; accepted 21 September 2005

Available online 30 January 2007

Communicated by Arthur E. C. Pece

Abstract

This paper deals with the computation of a single super-resolution image from a set of low-resolution images, where the motion fields are not constrained to be parametric. In our approach, the inversion process, in which the super-resolved image is inferred from the input data, is interleaved with the computation of a set of dense optical flow fields. The case of arbitrary motion presents several significant challenges. First of all, the super-resolution setting dictates that the optic flow computations must be very precise. Furthermore, we have to consider the possibility that certain parts of the scene, which are visible in the super-resolved image, are occluded in some of the input images. Such occlusions must be identified and dealt with in the restoration process. We propose a Bayesian approach to tackle these problems. In this framework, the input images are regarded as sub-sampled and noisy versions of the unknown high-quality image. Also, the input data is considered incomplete, in the sense that we do not know which pixels from the evolving super-resolution image are occluded in particular images from the input set. This will be modelled by introducing so-called visibility maps, which are treated as hidden variables. We describe an Expectation-Maximisation (EM) algorithm, which iterates between estimating values for the hidden quantities, and optimising the flow-fields and the super-resolution image. The approach is illustrated with a synthetic and two challenging real-world examples.

© 2006 Elsevier Inc. All rights reserved.

Keywords: Super-resolution; Optical flow; Visibility computation; EM

1. Introduction

The computation of super-resolution (SR) is an important problem that has a wide range of applications, e.g. in the areas of medical imaging, remote sensing, forensic imaging etc.. The goal of SR is to infer a single high quality and high pixel-rate image from a set of low-resolution input images. These input images should represent the same scene, but with a different relative motion between each of the cameras and the object of interest. Due to this relative motion, similar image irradiances will, after spatial integration on the cameras' sensors, result in slightly different sample values. It is exactly these sampling differences that make the computation of SR possible.

Inferring a super-resolved image is essentially an inversion process, in which some image formation model is run backwards. In this process, the apparent motion of the scene (optical flow) must be compensated, i.e. semantically corresponding pixels in the input images should be brought into alignment. Next, the effects of other phenomena, such as spatial integration, optical blur, motion blur, photometric transformations and measurement noise, must be accounted for. Being an inversion process, it is natural to study the problem in a probabilistic setting, where Bayes' rule allows us to evaluate the image formation model backwards. Most existing algorithms therefore are, or can be interpreted as, maximum a posteriori (MAP) estimations [6,7,5,8,10,12]. In Baker and Kanade [3], it is shown that SR-computation is an ill-conditioned problem. For large magnification factors, a huge number of solutions satisfy the reconstruction constraints. Therefore, in most approaches a smoothness

* Corresponding author.

E-mail addresses: Rik.Fransens@esat.kuleuven.be (R. Fransens), Christoph.Strecha@esat.kuleuven.be (C. Strecha), Luc.VanGool@esat.kuleuven.be (L. Van Gool).

prior is imposed on the solution. These authors also introduce context-dependent priors [2,3], leading to the so-called *hallucination* algorithm. An essential part of any SR-algorithm is motion estimation. The motion fields are typically assumed to take a simple parametric form, like an affine transformation [9] or a planar homography [4]. Only few algorithms, e.g. [1], consider the more general case of free-form optical flow.

In this paper, the SR-problem is addressed from a probabilistic point of view, where the motion fields are not constrained to be parametric. This significantly complicates the problem, because optical flow computation is an ill-posed problem, and strong regulatory priors need to be imposed on its solution. Furthermore, we have to consider the possibility that certain parts of the scene, which are visible from the view point of the SR-image, are occluded in some of the input images. Starting from a generative model of image formation, we bottom-up derive a MAP-approach to SR. This leads to an Expectation-Maximisation (EM) algorithm, which iterates between (i) photometric occlusion detection, and (ii) optimisation of the flow-fields and SR-image. The estimated occlusions are also used to define a simple, yet effective, anisotropic optical flow regulariser. The performance of this approach is demonstrated with several examples.

2. Generative model

Suppose we are given a set of $2N + 1$ low-resolution images \mathcal{I}_i , $i \in [-N, \dots, N]$, which associate a 2D-coordinate \mathbf{x} with an image value $\mathcal{I}_i(\mathbf{x})$. If we are dealing with colour images, this value is a 3-vector and for intensity images it is a scalar. The images could be originating from a video stream, in which case i is a (discrete) time instance. Our goal is to estimate a single high-quality image \mathcal{J}_0 , taken from the point of view of \mathcal{I}_0 , but with a higher pixel-resolution. Let l^2 and h^2 be the sizes of the low-resolution images and the super-resolution image, respectively. We define the magnification factor m as h/l , where $m \geq 1$. Typical values for m are 2, 3 and 4.

In this paper, we consider the case in which the apparent motion of the pixel values in the sequence of input images can not be described parametrically. Rather, we assume that a position \mathbf{x} in \mathcal{J}_0 is mapped onto a corresponding position $\mathcal{T}_i(\mathbf{x})$ in \mathcal{I}_i by a free-form transformation \mathcal{T}_i . This transformation can be decomposed into an optic flow field \mathcal{F}_i and a down-size operator \mathcal{D} , giving $\mathcal{T}_i = \mathcal{D} \circ \mathcal{F}_i$. The optic flow field \mathcal{F}_i maps positions $[x, y]$ onto $[x + u_i(x, y), y + v_i(x, y)]$, where $u_i()$ and $v_i()$ are expressed in the coordinate frame of \mathcal{J}_0 . Note that, by choice of reference, \mathcal{F}_0 is the identity transform. The operator \mathcal{D} , on the other hand, compensates for the magnification factor and maps $[x, y]$ onto $[x/m, y/m]$. The resulting transformation becomes:

$$\mathcal{T}_i : \mathcal{R}^2 \rightarrow \mathcal{R}^2 : \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} (x + u_i(x, y))/m \\ (y + v_i(x, y))/m \end{bmatrix}. \quad (1)$$

Sometimes, we will also find it convenient to use the inverse mapping \mathcal{T}_i^{-1} , which relates a position \mathbf{x} in \mathcal{I}_i to a corresponding position $\mathcal{T}_i^{-1}(\mathbf{x})$ in \mathcal{J}_0 . The inverse mapping is given by:

$$\mathcal{T}_i^{-1} : \mathcal{R}^2 \rightarrow \mathcal{R}^2 : \begin{bmatrix} x \\ y \end{bmatrix} \mapsto \begin{bmatrix} mx + u_i^{-1}(mx, my) \\ my + v_i^{-1}(mx, my) \end{bmatrix}. \quad (2)$$

These transformations are graphically depicted in Fig. 2.

In the case of general motion, we have to consider the possibility that certain parts of the scene, which are visible in the high-quality image \mathcal{J}_0 , are occluded in some of the input images. Such occlusions must be identified and properly dealt with in the restoration process. This will be modelled by introducing a set of visibility maps $\mathcal{V}_i(\mathbf{x})$, which signal whether or not a scene point X that projects onto \mathbf{x} in \mathcal{J}_0 , is also visible in image \mathcal{I}_i . The values $\mathcal{V}_i(\mathbf{x})$ are binary random variables which are either 1 or 0, corresponding to visibility or occlusion, respectively. By choice of reference, $\mathcal{V}_0(\mathbf{x}) = 1$. The visibility maps $\mathcal{V}_{i \neq 0}(\mathbf{x})$ are hidden variables, and their values must be inferred from the input images.

2.1. Imaging model

When computing super-resolution, the goal is to integrate the information from all input images into a single super-resolved image. We assume that the observed low-resolution images are generated by the high-resolution image \mathcal{J}_0 as follows. First, the relative motion between \mathcal{I}_i and \mathcal{J}_0 is realised by applying \mathcal{F}_i^{-1} to \mathcal{J}_0 . Next, the resulting image is convolved with a point spread function p . We take p to be an isotropic Gaussian with known width λ_p , i.e. we do not model possible motion blur and only take optical blur into account. Finally, a sampling operator is applied, which spatially integrates a region of the transformed \mathcal{J}_0 into a single low-resolution pixel. This operator is modelled as a square block-filter s of size $m \times m$. We also assume that the low-resolution images are subject to measurement noise ϵ , which is assumed to be normally distributed with zero mean and covariance Σ . More formally, the image formation model can be written as:

$$\begin{aligned} \mathcal{I}_i(\mathbf{x}) &= s * p * \mathcal{J}_0(\mathcal{T}_i^{-1}(\mathbf{x})) + \epsilon \\ \epsilon &\sim \mathcal{N}(\mathbf{0}, \Sigma). \end{aligned} \quad (3)$$

In what follows, we write g for $s * p$, i.e. the effects of s and p are combined in a single filter g . We formulated the model continuously, however, the low-resolution image can be generated by applying Eq. (3) to discrete locations \mathbf{x} in \mathcal{I}_i . The model is graphically depicted in Fig. 1. Note that this model is only valid for pixels $\mathcal{I}_i(\mathbf{x})$ which are visible from the point of view of \mathcal{J}_0 . Estimating the super-resolution image can now be formally stated as finding those values $\mathcal{J}_0(\mathbf{x})$ which make the input set $\mathcal{I}_i(\mathbf{x})$, restricted to the regions visible from \mathcal{J}_0 , most probable under the generative model.

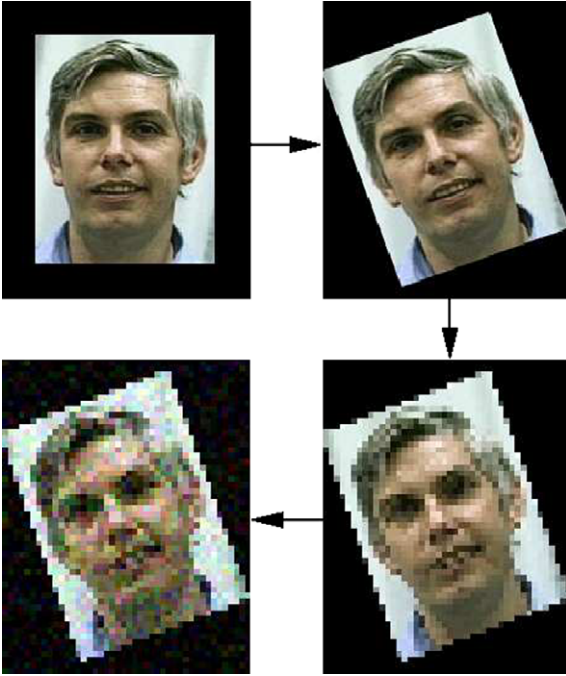


Fig. 1. Generative model. Different stages in the image formation model. The super-resolved image is shown in the left top. The image is geometrically warped according to transformation \mathcal{F}_i^{-1} (top right). The result is low-pass filtered by the smoothed block-filter g and sampled on a discrete lattice (bottom right). Finally, zero mean, normally distributed noise is added (bottom left) (after [4]).

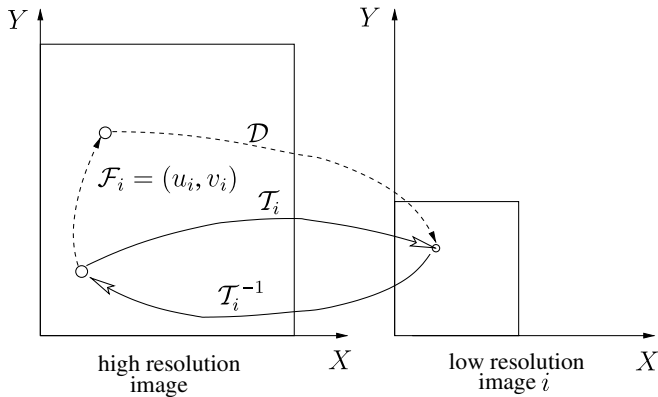


Fig. 2. Transformations. Graphical representation of the different transformations. \mathcal{F}_i is the optic flow vector between corresponding points, and is defined in the coordinate system of the SR-image \mathcal{J}_0 . \mathcal{D} compensates for the magnification. \mathcal{T}_i maps a pixel from \mathcal{J}_0 onto the corresponding location in \mathcal{I}_i , whereas \mathcal{T}_i^{-1} is the inverse operation.

2.2. Data likelihood and priors

The imaging model describes how the set of input images $\mathcal{I} = \{\mathcal{I}_i\}$ is generated, and is parametrised by the unknown quantities $\theta = \mathcal{J}_0, \mathcal{F}_i$ and Σ . Furthermore, we have introduced a set of hidden variables $\mathcal{V} = \{\mathcal{V}_{i \neq 0}\}$, which indicate at which locations the model is valid. In a Bayesian framework, the super-resolution problem consists in finding the value of θ that maximises the posterior

probability $p(\theta|\mathcal{I})$. According to Bayes' rule, this posterior can be written as:

$$\begin{aligned} p(\theta|\mathcal{I}) &\propto p(\mathcal{I}|\theta)p(\theta) \\ &= \sum_{\mathcal{V}} p(\mathcal{I}|\theta, \mathcal{V})p(\theta|\mathcal{V})p(\mathcal{V}), \end{aligned} \quad (4)$$

where we have conditioned the data likelihood and the prior on the hidden variables \mathcal{V} . Note that the sum ranges over all possible configurations of the visibility maps \mathcal{V}_i . Assuming a uniform prior on \mathcal{V} , Eq. (4) can be further simplified to:

$$p(\theta|\mathcal{I}) \propto \sum_{\mathcal{V}} p(\mathcal{I}|\theta, \mathcal{V})p(\theta|\mathcal{V}). \quad (5)$$

Under the assumption that the image noise is i.i.d. for all pixels in all images, the conditional data likelihood $p(\mathcal{I}|\theta, \mathcal{V})$ can be written as the product of all individual pixel probabilities:

$$p(\mathcal{I}|\theta, \mathcal{V}) = \prod_{i=-N}^N \prod_{\mathbf{x}} p(\mathcal{I}_i(\mathcal{T}_i(\mathbf{x}))|\theta, \mathcal{V}_i(\mathbf{x})). \quad (6)$$

Here, \mathbf{x} runs over all h^2 positions of the super-resolution image \mathcal{J}_0 . Given the current estimates for \mathcal{J}_0 and the noise variance Σ , the data likelihood becomes:

$$\begin{aligned} p(\mathcal{I}|\theta, \mathcal{V}) &= \prod_i \prod_{\mathbf{x}} \left[\frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{m}_i^T \Sigma^{-1} \mathbf{m}_i\right) \right]^{\mathcal{V}_i(\mathbf{x})} \left[\frac{1}{C} \right]^{(1-\mathcal{V}_i(\mathbf{x}))}, \end{aligned} \quad (7)$$

where $\mathbf{m}_i = \mathcal{I}_i(\mathcal{T}_i(\mathbf{x})) - g * \mathcal{J}_0(\mathbf{x})$, C is the volume of the colour space, and the variable d in the normalisation constant denotes the dimensionality of \mathbf{m}_i . The individual pixel likelihoods consist of the product of two terms, of which only one is active, depending on the local value of the visibility. If $\mathcal{V}_i(\mathbf{x}) = 1$, the pixel likelihood is measured as the value of the noise distribution, evaluated at the difference between corresponding positions in the low-pass filtered high-resolution image \mathcal{J}_0 and the input image \mathcal{I}_i . If $\mathcal{V}_i(\mathbf{x}) = 0$, i.e. when the location $\mathcal{T}_i(\mathbf{x})$ in \mathcal{I}_i is not visible from \mathcal{J}_0 , the pixel likelihood is measured as the value $1/C$ of the uniform distribution.

The formulation of an appropriate prior is slightly more complicated. We can factorise $p(\theta|\mathcal{V})$ as the product of an image dependent and a flow-field dependent part:

$$p(\theta|\mathcal{V}) \propto p(\mathcal{J}_0) \prod_{i=-N}^N p(\mathcal{F}_i|\mathcal{V}_i). \quad (8)$$

To arrive at this expression, we have silently assumed that (i) the optical flow-fields are a priori independent from each other,¹ and (ii) the super-resolution image is a priori independent from the measurement noise and possible occlu-

¹ Clearly, this assumption is not restrictive enough if the input images originate from a video sequence. In such case, we could impose a temporal smoothness constraint on the flow fields.

sions. The flow-field priors $p(\mathcal{F}_i|\mathcal{V}_i)$ will be modelled as an exponential density distribution of the form $\exp(-R(\mathcal{F}_i, \mathcal{V}_i)/\lambda_f)$. Here, λ_f is a parameter which controls the width of the distribution, and $R(\mathcal{F}_i, \mathcal{V}_i)$ is a ‘regulariser’. From such a regulariser we expect that it reflects our prior belief that the world is essentially simple. For a locally smooth solution \mathcal{F}_i in the neighbourhood of a particular point \mathbf{x} , its value should approach zero, making such a solution very likely. Vice-versa, large fluctuations of the optical flow field should result in large values for the regulariser, making such solutions less likely. Alternatively, at image positions corresponding to scene depth discontinuities and occlusions, large flow discontinuities should not be made a priori unlikely. We use the current visibility estimates to signal such situations, and model the flow priors as follows:

$$p(\mathcal{F}_i|\mathcal{V}_i) \propto \exp\left(-\frac{1}{\lambda_f} \sum_{\mathbf{x}} \mathcal{V}_i(\mathbf{x})(\|\nabla u(\mathbf{x})\|^2 + \|\nabla v(\mathbf{x})\|^2)\right). \quad (9)$$

A prior on \mathcal{J}_0 can be formulated in a similar way. Given the ill-conditioned nature of the SR-problem [3], we constraint the solution to be smooth. Put differently, we want to impose spatial correlation between neighbouring locations in \mathcal{J}_0 . To this end, we define the following distribution for \mathcal{J}_0 :

$$p(\mathcal{J}_0) \propto \exp\left(-\frac{1}{\lambda_j} \sum_{\mathbf{x}} \|\nabla \mathcal{J}_0(\mathbf{x})\|^2\right). \quad (10)$$

In what follows, we will refer to this type of prior as a ‘type I’ prior. We also experimented with a second type of prior, which expresses that the deviation of the super-resolved image from the average of the input images, transformed to the coordinate frame of the super-resolved image, should be spatially smooth. The average of the input images, denoted by \mathcal{J}_0^* , is defined as follows:

$$\mathcal{J}_0^*(\mathbf{x}) = \frac{\sum_i \mathcal{V}_i(\mathbf{x})\mathcal{I}_i(\mathcal{T}(\mathbf{x}))}{\sum_i \mathcal{V}_i(\mathbf{x})}. \quad (11)$$

The prior distribution for \mathcal{J}_0 now becomes:

$$p(\mathcal{J}_0) \propto \exp\left(-\frac{1}{\lambda_j} \sum_{\mathbf{x}} \|\nabla(\mathcal{J}_0 - \mathcal{J}_0^*)(\mathbf{x})\|^2\right). \quad (12)$$

In what follows, we will refer to this type of prior as a ‘type II’ prior. A similar approach was followed by Capel et al. [4], however, they use a median image instead.

3. An EM solution

We now turn back to the optimisation of θ . Instead of maximising the posterior in Eq. (5), we minimise its negative logarithm:

$$\hat{\theta}_{MAP} = \arg \min_{\theta} \left\{ -\log \sum_{\mathcal{V}} p(\mathcal{I}|\theta, \mathcal{V})p(\theta|\mathcal{V}) \right\}. \quad (13)$$

The sum in Eq. (13) ranges over all possible configurations of the visibility maps \mathcal{V}_i . Every visibility map except \mathcal{V}_0 (whose elements are all one) consists of h^2 binary RVs, so the total number of configurations is 2^{2Nh^2} . Even for modest size images this is an incredibly large number, hence direct optimisation of the right-hand side of Eq. (13) is infeasible. The Expectation-Maximisation (EM) algorithm offers a solution to this problem. It produces a sequence of estimates $\{\hat{\theta}^{(t)}, t = 0, 1, 2, \dots\}$ by alternating the following two steps:

E-step. On the $(t+1)^{\text{th}}$ iteration, the conditional expectation of $-\log p(\mathcal{I}|\theta, \mathcal{V})p(\theta|\mathcal{V})$ is computed, where the expectation is w.r.t. the posterior distribution of the hidden variables \mathcal{V} . Because the assumption was made that all visibility maps are a priori equally likely (i.e. no smoothness prior was imposed on these maps), this posterior factorises over all pixel locations:

$$p(\mathcal{V}|\mathcal{I}, \hat{\theta}^{(t)}) = \prod_{i \neq 0} \prod_{\mathbf{x}} f(\mathcal{V}_i(\mathbf{x})|\mathcal{I}_i(\mathcal{T}_i(\mathbf{x})), \hat{\theta}^{(t)}), \quad (14)$$

where f is a Bernoulli distribution over the values $\{0,1\}$:

$$f(\mathcal{V}_i(\mathbf{x})|\mathcal{I}_i(\mathcal{T}_i(\mathbf{x})), \hat{\theta}^{(t)}) = \begin{cases} q_i(\mathbf{x}) & \mathcal{V}_i(\mathbf{x}) = 1 \\ 1 - q_i(\mathbf{x}) & \mathcal{V}_i(\mathbf{x}) = 0 \end{cases}. \quad (15)$$

Here, $q_i(\mathbf{x})$ is a scalar in the range $[0,1]$ which depends on the image value $\mathcal{I}_i(\mathcal{T}_i(\mathbf{x}))$ and the current estimates of the unknowns $\hat{\theta}^{(t)}$. According to Bayes’ rule, the probability of a pixel being visible can be expanded as follows:

$$\begin{aligned} q_i(\mathbf{x}) &= \Pr(\mathcal{V}_i(\mathbf{x}) = 1|\mathcal{I}_i(\mathcal{T}_i(\mathbf{x})), \hat{\theta}^{(t)}) \\ &= \frac{p(\mathcal{I}_i(\mathcal{T}_i(\mathbf{x}))|\mathcal{V}_i(\mathbf{x}) = 1, \hat{\theta}^{(t)})}{p(\mathcal{I}_i(\mathcal{T}_i(\mathbf{x}))|\mathcal{V}_i(\mathbf{x}) = 1, \hat{\theta}^{(t)}) + p(\mathcal{I}_i(\mathcal{T}_i(\mathbf{x}))|\mathcal{V}_i(\mathbf{x}) = 0, \hat{\theta}^{(t)})}, \end{aligned} \quad (16)$$

where we have assumed equal priors on the probability of a pixel being visible or not. The PDF $p(\mathcal{I}_i(\mathbf{x})|\mathcal{V}_i(\mathbf{x}) = 1, \hat{\theta}^{(t)})$ is given by the value of the noise distribution evaluated over the colour-difference between $\mathcal{I}_i(\mathcal{T}_i(\mathbf{x}))$ and $g * \mathcal{J}_0(\mathbf{x})$, where the estimates of \mathcal{J}_0 , Σ and \mathcal{F}_i obtained in the t -th iteration are used. The second PDF is, in principle, given by the uniform distribution over the colour range. However, it is possible to use a more informative background distribution when estimating a pixel’s visibility probability. We provide a *global* estimate for the distribution of occluded pixels by building a histogram of the colour-values in \mathcal{J}_0 which are currently invisible. This is merely the histogram of \mathcal{J}_0 where the contribution of each pixel is weighted by $(1 - q_i(\mathbf{x}))$, and where we use the estimated probabilities from the previous iteration. This is graphically depicted in Fig. 3. Note that, if a particular pixel in \mathcal{J}_0 is marked as not-visible, in the next iterations this will automatically decrease the visibility estimates of all similarly coloured pixels. This makes sense from a

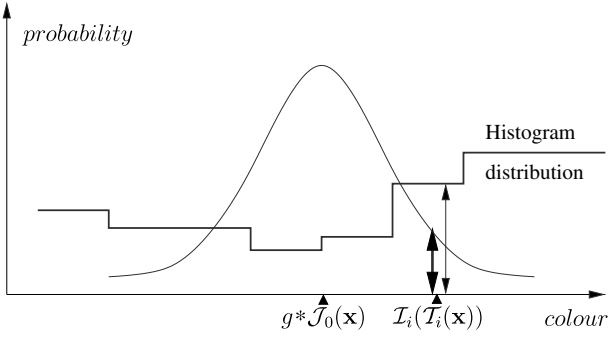


Fig. 3. Visibility computation. The probability of location \mathbf{x} in \mathcal{J}_0 being visible in \mathcal{I}_i , can be computed from the value of the normal distribution, centered around $g * \mathcal{J}_0(\mathbf{x})$, evaluated at $\mathcal{I}_i(\mathcal{T}_i(\mathbf{x}))$ (bold arrow), and the value of the histogram-based estimate, evaluated at the same position (thin arrow).

perceptual point of view, and has a regularising effect on the visibility maps.

If the posterior $p(\mathcal{V}|\mathcal{I}, \hat{\theta}^{(t)})$ factorises like in Eq. (14), the E-step amounts to replacing $\mathcal{V}_i(\mathbf{x})$ by $q_i(\mathbf{x})$. This leads (up to a constant) to the following energy:

$$E(\theta) = \frac{1}{2} \sum_i \sum_{\mathbf{x}} q_i(\mathbf{x}) (\mathbf{m}_i^T \Sigma^{-1} \mathbf{m}_i + \log((2\pi)^d |\Sigma|)) + \frac{1}{\lambda_f} \sum_i \sum_{\mathbf{x}} q_i(\mathbf{x}) (\|\nabla u(\mathbf{x})\|^2 + \|\nabla v(\mathbf{x})\|^2) + \frac{1}{\lambda_j} \sum_{\mathbf{x}} \|\nabla \mathcal{J}_0(\mathbf{x})\|^2 + \sum_i \sum_{\mathbf{x}} (1 - q_i(\mathbf{x})) \frac{1}{C}. \quad (17)$$

M-step. At the M-step, the intent is to compute values for θ that minimise the energy in Eq. (17), given the current estimates of \mathcal{V}_i . This is achieved by setting the parameters θ to the appropriate root of the derivative equation, $\partial E(\theta)/\partial \theta = \mathbf{0}$. We will start by deriving the update equations for the image related quantities Σ and \mathcal{J}_0 , and then proceed with the optimisation of the flow-fields \mathcal{F}_i . The update equation for Σ , which is derived in the Appendix A, is given by:

$$\Sigma \leftarrow \frac{\sum_i \sum_{\mathbf{x}} q_i(\mathbf{x}) \mathbf{m}_i(\mathbf{x}) \mathbf{m}_i(\mathbf{x})^T}{\sum_i \sum_{\mathbf{x}} q_i(\mathbf{x})}. \quad (18)$$

Here, \mathbf{x} runs over all positions in the high-resolution image and $\mathbf{m}_i(\mathbf{x})$ is the difference between corresponding positions in the low-pass filtered high-resolution image \mathcal{J}_0 and the input image \mathcal{I}_i . Note that the contribution of every pixel-difference is weighted by the current visibility estimate of its location.

To derive an update equation for the super-resolution image, we collect the \mathcal{J}_0 -dependent terms from Eq. (17) and represent them in a vector-matrix notation. First of all, the pixels (and colour-values) from the input images \mathcal{I}_i are rearranged in lexicographic order, which gives the dl^2 -dimensional vectors I_i . Similarly, the high-resolution image \mathcal{J}_0 and the visibility maps \mathcal{V}_i are represented by

the dh^2 -dimensional vectors J_0 and V_i . Next, \mathcal{J}_0 is transformed to the coordinate frame of \mathcal{I}_i , by applying the operators F_i (geometric warp according to the flow-field), H (optical blur) and D (spatial integration) to J_0 . This operation transforms J_0 to the dl^2 -dimensional vector $DHF_i J_0 = M_i J_0$. Similarly, the visibility vectors V_i are transformed to $M_i V_i$. Finally, the energy terms related to the image prior can be represented in vector-matrix form as follows. Let us define the $dh^2 \times dh^2$ -dimensional matrix operators D_x , D_y , whose action it is to compute a discrete approximation of $\partial \mathcal{J}_0 / \partial x$ and $\partial \mathcal{J}_0 / \partial y$, respectively. Then, $\sum_{\mathbf{x}} \|\nabla \mathcal{J}_0(\mathbf{x})\|^2$ can be approximated as follows:

$$\begin{aligned} \sum_{\mathbf{x}} \|\nabla \mathcal{J}_0(\mathbf{x})\|^2 &\approx (D_x J_0)^T (D_x J_0) + (D_y J_0)^T (D_y J_0) \\ &= J_0^T (D_x^T D_x + D_y^T D_y) J_0 \\ &= J_0^T Q J_0. \end{aligned} \quad (19)$$

Writing this result in Eq. (10), we see that the distribution for J_0 is given by:

$$p(J_0) \propto \exp\left(-\frac{1}{\lambda_j} J_0^T Q J_0\right), \quad (20)$$

where Q acts as an inverse covariance matrix. In other words, the effect of the super-resolution image prior can be interpreted as defining a zero-mean multivariate Gaussian distribution on the overall (discrete-lattice) image \mathcal{J}_0 , in which neighbouring pixels are correlated with one another. The energy w.r.t. \mathcal{J}_0 now becomes:

$$E(J_0) = J_0^T Q J_0 + \sum_i (M_i J_0 - I_i)^T W_i S^{-1} (M_i J_0 - I_i), \quad (21)$$

where S^{-1} is a block-diagonal matrix with diagonal entries Σ^{-1} , and W_i is a diagonal weighting matrix whose entries are $M_i V_i$. These weights are merely the visibility estimates, but now expressed in the coordinate system of \mathcal{I}_i . Note that the parameter λ_j is absorbed into Q . Differentiation of Eq. (21) w.r.t. J_0 and setting the result to zero, leads to the following linear system:

$$\left(Q + \sum_i M_i^T W_i S^{-1} M_i\right) J_0 = \sum_i M_i^T W_i S^{-1} I_i. \quad (22)$$

If we use the type II prior (Eq. 12), it is straightforward to verify that the linear system becomes:

$$\left(Q + \sum_i M_i^T W_i S^{-1} M_i\right) J_0 = \sum_i M_i^T W_i S^{-1} I_i + Q J_0^*. \quad (23)$$

A closed form expression for the optimal J_0 is readily derived, by computing the inverse of the left-hand side operator and multiplying it with the vector on the right. However, this is not possible in practise because of the large dimensions of the system. Instead, we compute J_0 by preconditioned conjugate gradient descent.

For the update of the flow-fields \mathcal{F}_i , we apply a differential optical flow strategy similar to [11]. In this framework, the functions u_i and v_i (horizontal and vertical motion fields

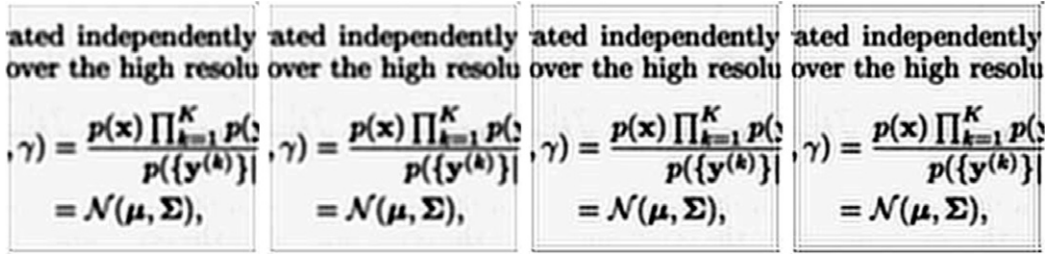


Fig. 4. Synthetic data. Super-resolution reconstructions for increasing values of the SR-smoothness prior parameter λ_j . From left to right, $\lambda_j = 1.0 \times 10^{-4}$, 1.25×10^{-4} , 2.5×10^{-4} , 5.0×10^{-4} , respectively. Note that when λ_j increases, the prior distribution on \mathcal{J}_0 gets wider and consequently the smoothness constraint is relaxed. As a result, the reconstruction gets sharper, but simultaneously high frequency and boundary artefacts start to emerge.

of \mathcal{F}_i) that minimise the energy functional in Eq. (17) are determined. The minimising flow-field satisfies the Euler–Lagrange equations, which leads to the following set of (anisotropic) diffusion equations:

$$\begin{aligned} \frac{\partial u_i}{\partial t} &= q_i \mathbf{m}_i^T \Sigma^{-1} \frac{\partial \mathcal{I}_i}{\partial x} \frac{1}{m} - \frac{1}{\lambda_f} \operatorname{div}(q_i \nabla u_i) \\ \frac{\partial v_i}{\partial t} &= q_i \mathbf{m}_i^T \Sigma^{-1} \frac{\partial \mathcal{I}_i}{\partial y} \frac{1}{m} - \frac{1}{\lambda_f} \operatorname{div}(q_i \nabla v_i). \end{aligned} \quad (24)$$

Here, $\partial \mathcal{I}_i / \partial x$ and $\partial \mathcal{I}_i / \partial y$ are d -vectors which contain the spatial gradients of the colour-bands of \mathcal{I}_i . The diffusion equations are solved by means of implicit discretisation [13].

We end this section with a general overview of the algorithm. At initialisation time, the visibility estimates $q_0(\mathbf{x})$ are set to 1.0, and the estimates $q_{i \neq 0}(\mathbf{x})$ are set to 0.5. The noise covariance Σ is initialised to $\operatorname{diag}(2.0)$, and both the SR-image \mathcal{J}_0 and SR-image prior \mathcal{J}_0^* are initialised with an up-sampled version of \mathcal{I}_0 . Finally, the flow-fields are computed by solving the anisotropic diffusion equations, and the histograms of the currently invisible pixels are initialised with the histogram of \mathcal{J}_0 . Next, the E and M-step are iterated until the maximum relative change of \mathcal{J}_0 is smaller than 0.001. At the M-step, the order of updating the unknowns is of importance, because in the update equations the unknowns are functions of one another. We first solve for the flow fields \mathcal{F}_i , and then we update \mathcal{J}_0^* , Σ and \mathcal{J}_0 in the respective order. Finally, we recompute the histograms which are needed in the next E-step.

4. Experiments

The algorithm was tested on a synthetic and two real world examples. In the experiment on synthetic data, a small fragment of scanned text was taken as input. Next, the image formation model was applied, i.e. the image is geometrically transformed, low-pass filtered and sub-sampled. The flow-field is constant, with a translation of [0.27, 0.17] pixels per time-step. The resulting sequence contains 32 images of size 50×50 . The 16th image was taken as the reference view \mathcal{I}_0 from which the high-resolution image \mathcal{J}_0 was computed. We applied a magnification factor $m = 4$. The purpose of this experiment is to quantify

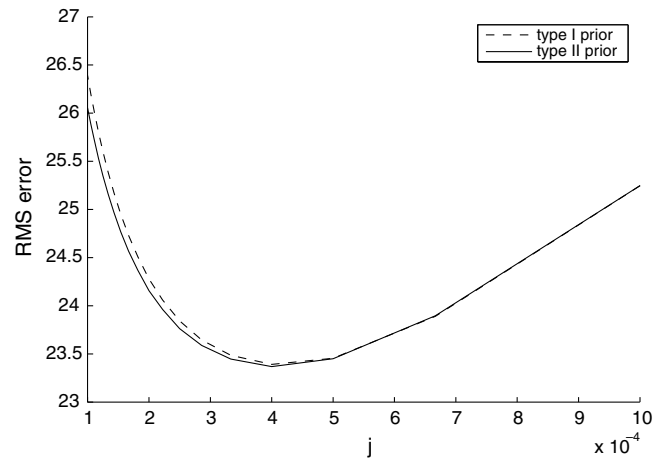


Fig. 5. Synthetic data. The effect of the SR-smoothness prior parameter λ_j on the RMS error, for the type I prior (dashed curve) and the type II prior (solid curve). For both types of prior, the RMS error goes through a minimum at $\lambda_j \approx 4.0 \times 10^{-4}$. The difference between both types of prior, in terms of the RMS error, is not very large, with a slight advantage for the type II prior.

the effect of both types of super-resolution prior on the reconstruction. To this end, we plotted the root mean squared error (RMS error) of the reconstructions obtained with type I prior (Eq. (10)) and type II prior (Eq. (12)) against λ_j , which is varied in the range $[10^{-4} \dots 10^{-3}]$.² In all experiments with synthetic data, the optical flow prior parameter λ_f was set to 0.001. The results are shown in Fig. 5. From this figure, it can be seen that the difference between both types of prior is not very large, with a slight advantage for the type II prior. Furthermore, for both types of prior, the RMS error goes through a minimum at $\lambda_j \approx 4.0 \times 10^{-4}$. Some of the reconstructions, obtained with the type II prior, are shown in Fig. 4. From this figure, it can be observed that when the smoothness constraint is relaxed, the reconstruction gets sharper but simultaneously high frequency and boundary artefacts start to emerge. Finally, the low-resolution reference image and the super-resolution reconstruction with minimal RMS error are

² Grey values are in the range $[0 \dots 255]$ and the video frame rate is 12.5 frames per second. The indicated values of λ_f and λ_j have units frame^{-2} and $\text{grey-level}^2/\text{pixel}^2$.

shown together in Fig. 6. Obviously, the readability of letters and symbols has improved significantly.

We performed two experiments with real-world data. In a first experiment, a scene was recorded with independent camera translation and object movement. From the video sequence, 15 consecutive images were selected, depicting a person in front of a book shelf, from which we segmented 15 windows of size 50×60 . The 8th image was taken as the reference view, from which the high-resolution image was computed. We applied a magnification factor $m = 3$. Some of the input images are shown in the top row of Fig. 7. This is a challenging test case because of the poor lighting conditions. There are strong specular reflections, e.g. on the face and shoulders, which violate the optical flow constant brightness assumption. These deviations from the ideal conditions should be captured by the model’s noise term, and by a reduced ‘visibility’ estimate of the outlier pixels. In this experiment, the type II SR-prior was used, and

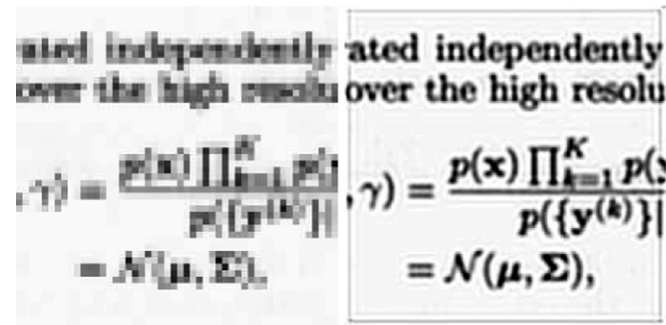


Fig. 6. Synthetic data. Low-resolution input image \mathcal{I}_0 (left) and its super-resolution reconstruction (right) for $\lambda_j = 4.0 \times 10^{-4}$.

the free parameters were set to $\lambda_j = 0.001$ and $\lambda_f = 0.01$. The results of the visibility computations and the flow-field estimates are shown in the middle and bottom row of Fig. 7. Note that the occlusions and depth discontinuities, at the visual hull of the head and the books in the background, are detected well. The low-resolution reference view and a detail view of the head, together with their corresponding super-resolution reconstructions, are shown in Fig. 8. Again, we notice an improved quality, particularly for the facial features of the depicted person.

In a second experiment, the goal is to improve the readability of the licence plate of a car. The camera is static and positioned alongside the road. The cars passing by stay in view for several seconds, so sufficient video frames can be gathered for SR-reconstruction. From the video stream, 24 consecutive images were selected, from which we segmented windows of size 72×72 , covering the rear end of the car. A magnification factor of $m = 4$ was applied. Some of the input images are shown in Fig. 9. Note that, even though the licence plate area is actually planar in 3D, it is not possible to register the corresponding image patches by a parametric transformation (i.e. a planar homography) because of their limited extent. For example, in the reference view, the licence plate covers an area of merely 12×36 pixels. In this experiment, the type II SR-prior was used, and the free parameters were set to $\lambda_j = 0.001$ and $\lambda_f = 0.01$. The SR-reconstruction is shown in Fig. 10, and a detail view is shown in Fig. 11. The improvement of readability is obvious. Especially the last digit (‘8’) and the country id (‘B’) are fully disambiguated, which can be crucial for the success of subsequent analysis by an OCR-algorithm.

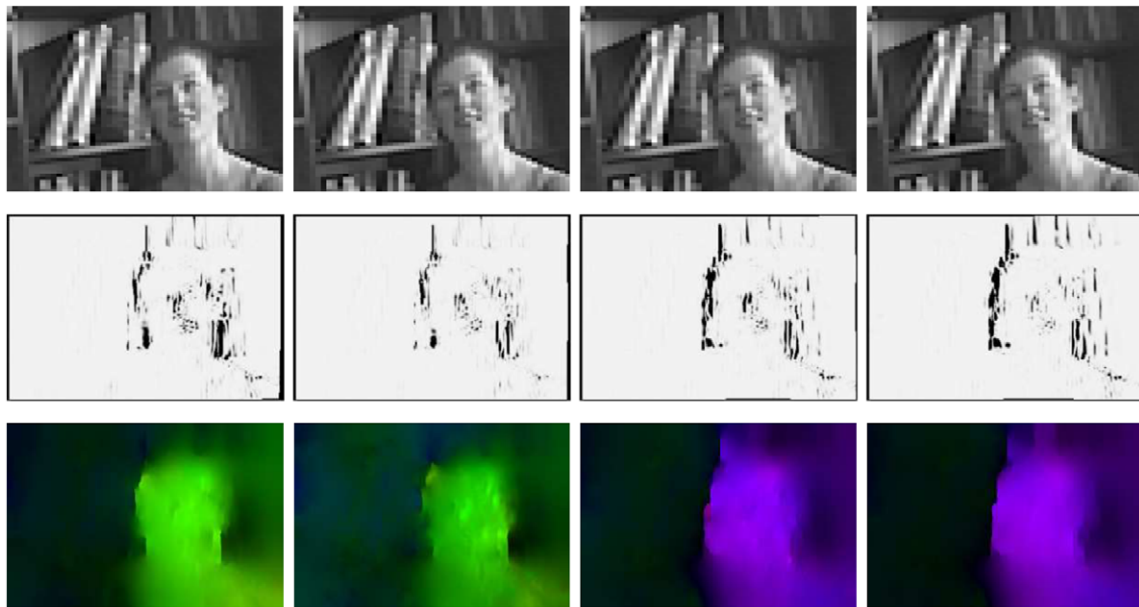


Fig. 7. Real data. Camera translation with independently moving head. (Top row) Some of the input images. (Middle row) Corresponding visibility maps related to the high resolution image shown in Fig. 8. The occlusions and depth discontinuities at the visual hull of the head and the books in the background are detected well. Also within the face, the areas nearby the bridge of the nose and the mouth have a lowered visibility. (Bottom row) Optical flow fields from the high resolution image towards the input images. Flow direction is encoded by colour hue, whereas flow magnitude is encoded by colour saturation. Note that the flow is discontinuous at positions with lowered visibility estimates.



Fig. 8. Real data. Original low-resolution image and a detail view of the head (left) and the corresponding super-resolution reconstructions (right).



Fig. 9. Real data. Static camera with moving cars. Some of the 24 low-resolution input images. The purpose is to enhance the licence plate of the car in the front.



Fig. 10. Real data. Detail view of the rear of the car in the original low-resolution image (left) and the corresponding super-resolution reconstruction (right).



Fig. 11. Real data. Detail view of the licence plate in the original low-resolution image (top) and the corresponding super-resolution reconstruction (bottom).

5. Conclusions

In this paper, a probabilistic approach to optical flow based super-resolution was presented. Starting from a generative model of image formation, a MAP-solution was derived. Several priors were introduced, to alleviate the ill-posedness of the optical flow problem, and the poor conditioning of the super-resolution problem. The flow-fields are regularised by a visibility-based anisotropic operator. The super-resolved image, on the other hand, is constrained to be spatially smooth, by the introduction of an isotropic Gaussian process. This leads to an EM-algorithm, which iterates between (i) photometric occlusion detection, and (ii) estimation of the noise covariance matrix and optimisation of the flow-fields and super-resolution image.

The combination of the likelihood terms and prior terms give rise to a rather involved energy formulation. The estimate of the noise covariance matrix is given in closed form. For the optimisation of this energy w.r.t. the super-resolution image, the relevant terms are isolated and turned into a set of linear equations. In our implementations, we use a sparse conjugate gradient algorithm to solve this system. For the optimisation of the energy w.r.t. the flow-fields, on the other hand, we use the Euler–Lagrange formalism to derive a set of coupled, anisotropic diffusion equations. These equations are iteratively solved by means of implicit discretisation.

The algorithm has two free parameters, λ_f and λ_j , which control the degree of smoothness of the optical flow fields and the super resolved image. Not surprisingly, both parameters result from our prior beliefs we incorporated in the algorithm, and as such they can be considered unavoidable. In the experiments, λ_j was set to 0.001, which gave satisfying results in all experiments. The optimal value of λ_f , however, is problem dependent. If the motion fields are known to be smooth, e.g. when the objects in the scene are known to consist of large planar regions, the optical

flow solution should be constrained to be smooth. If, on the other hand, the motion fields are known to be irregular, the smoothness constraint should be relaxed. In our experiments, λ_f was set to 0.01.

In the presented algorithm, visibilities are computed photometrically with respect to all (sub-)pixel locations in the set of low-resolution input images. These estimates can be interpreted as a measure for how well the input data is explained by the generative model, given the current estimates of the optical flow fields, the noise distribution and the super-resolution image. We can therefore frame our algorithm as a (regularised) iteratively reweighted least squares solution to the SR-constraints. The explicit computation of noise and visibilities has a balancing effect on the optical flow computations. The noise estimate has a global (image location independent) effect. When the noise level is high, the influence of all optical flow matching terms decreases relative with respect to the optical flow regularisation term. As a result, the flow-fields \mathcal{F}_i will be more strongly regularised, i.e. will become more smooth. This is important at initialisation time, when the flows \mathcal{F}_i have not yet converged to their final solution. The visibility estimates, on the other hand, act on a local level. If at a particular image location visibility is high, regularisation becomes locally more important, which tends to smoothen the flow at that position.

We have not yet implemented a temporal smoothness constraint on \mathcal{F}_i . However, if the input images originate from a video stream, smooth temporal flow is a very reasonable prior assumption. This is readily introduced in the presented framework, and will be the object of our future research.

Appendix A

We now derive the update equation Eq. (18) for the noise covariance matrix. It is obtained by setting Σ to the root of the derivative equation $\partial E / \partial \Sigma = \mathbf{0}$. Equivalently, we can identify Σ for which $\partial E / \partial \Sigma^{-1} = \mathbf{0}$, which is more easy to solve. Collecting all Σ -dependent parts from the energy in Eq. (17) and setting the derivative to zero yields the following:

$$\frac{\partial}{\partial \Sigma^{-1}} \sum_i \sum_{\mathbf{x}} q_i \left[\mathbf{m}_i^T \Sigma^{-1} \mathbf{m}_i + \log((2\pi)^d |\Sigma|) \right] = \mathbf{0}. \quad (\text{A.1})$$

The first term in Eq. (A.1) can be written as follows:

$$\mathbf{m}_i^T \Sigma^{-1} \mathbf{m}_i = \text{tr}(\Sigma^{-1} \mathbf{m}_i \mathbf{m}_i^T),$$

where $\text{tr}(\cdot)$ represents the trace operation. The second term can be written as:

$$\log((2\pi)^d |\Sigma|) = \log(2\pi)^d - \log(|\Sigma^{-1}|).$$

Eq. (A.1) now becomes:

$$\sum_i \sum_{\mathbf{x}} q_i \left[\frac{\partial \text{tr}(\Sigma^{-1} \mathbf{m}_i \mathbf{m}_i^T)}{\partial \Sigma^{-1}} - \frac{\partial \log(|\Sigma^{-1}|)}{\partial \Sigma^{-1}} \right] = \mathbf{0}. \quad (\text{A.2})$$

For symmetric matrices \mathbf{A} and \mathbf{B} , the following properties hold:

$$\frac{\partial \text{tr}(\mathbf{AB})}{\partial \mathbf{A}} = 2\mathbf{B} - \text{diag}(\mathbf{B})$$

$$\frac{\partial \log(|\mathbf{A}|)}{\partial \mathbf{A}} = 2\mathbf{A}^{-1} - \text{diag}(\mathbf{A}^{-1}).$$

Therefore, Eq. (A.2) can be written as:

$$\begin{aligned} & \sum_i \sum_{\mathbf{x}} q_i \left[\frac{\partial \text{tr}(\boldsymbol{\Sigma}^{-1} \mathbf{m}_i \mathbf{m}_i^T)}{\partial \boldsymbol{\Sigma}^{-1}} - \frac{\partial \log(|\boldsymbol{\Sigma}^{-1}|)}{\partial \boldsymbol{\Sigma}^{-1}} \right] \\ &= \sum_i \sum_{\mathbf{x}} q_i [2(\mathbf{m}_i \mathbf{m}_i^T - \boldsymbol{\Sigma}) - \text{diag}(\mathbf{m}_i \mathbf{m}_i^T - \boldsymbol{\Sigma})] \\ &= 2\mathbf{M} - \text{diag}(\mathbf{M}) = \mathbf{0} \end{aligned}$$

where we substituted \mathbf{M} for $\sum_i \sum_{\mathbf{x}} q_i (\mathbf{m}_i \mathbf{m}_i^T - \boldsymbol{\Sigma})$. Furthermore, if $2\mathbf{M} - \text{diag}(\mathbf{M}) = \mathbf{0}$, it follows that $\mathbf{M} = \mathbf{0}$ so we can write:

$$\sum_i \sum_{\mathbf{x}} q_i (\mathbf{m}_i \mathbf{m}_i^T - \boldsymbol{\Sigma}) = \mathbf{0}.$$

Solving for $\boldsymbol{\Sigma}$ finally gives:

$$\boldsymbol{\Sigma} = \frac{\sum_i \sum_{\mathbf{x}} q_i \mathbf{m}_i \mathbf{m}_i^T}{\sum_i \sum_{\mathbf{x}} q_i}.$$

References

- [1] S. Baker, T. Kanade, Super-resolution optical flow, Technical Report CMU-RI-TR-99-36, Carnegie Mellon University, 1999.
- [2] S. Baker, T. Kanade, Hallucinating Faces, Technical Report CMU-RI-TR-99-32, Carnegie Mellon University, 1999.
- [3] S. Baker, T. Kanade, Limits on super-resolution and how to break them, IEEE Conference on Computer Vision and Pattern Recognition 2 (2000) 372–379.
- [4] D. Capel, A. Zisserman, Computer Vision Applied to Super Resolution, IEEE Signal Processing Magazine (2003) 75–86.
- [5] P. Cheeseman, B. Kanefsky, R. Kraft, J. Stutz, R. Hanson, Super-resolved surface reconstruction from multiple images, Technical Report FIA-94-12, NASA Ames Research Center, 1994.
- [6] M. Elad, A. Feuer, Super-resolution restoration of an image sequence—adaptive filtering approach, IEEE Transactions on Image Processing 8 (3) (1999) 387–395.
- [7] M. Elad, A. Feuer, Super-resolution restoration of image sequences, IEEE Transactions on Pattern Analysis and Machine Intelligence 21 (9) (1999) 817–834.
- [8] R.C. Hardie, K.J. Barnard, E.E. Armstrong, Joint MAP registration and high-resolution image estimation using a sequence of undersampled images, IEEE Transactions on Image Processing 6 (12) (1997) 1621–1633.
- [9] M. Irani, S. Peleg, Improving resolution by image registration, Graphical Models and Image Processing 53 (1991) 231–239.
- [10] R. Schultz, R. Stevenson, Extraction of high-resolution frames from video sequences, IEEE Transactions on Image Processing 5 (6) (1996) 996–1011.
- [11] C. Strecha, R. Fransens, L. Van Gool, A Probabilistic Approach to Large Displacement Optical Flow and Occlusion Detection, Statistical Methods in Video Processing, ECCV 2004 Workshop SMVP, 2004.
- [12] M.E. Tipping, C.M. Bishop, Bayesian image super-resolution, in: S. Becker, S. Thrun, K. Obermayer (Eds.), Advances in Neural Information Processing Systems, 15, MIT Press, 2003.
- [13] J. Weickert, B.M. ter Haar Romeny, M.A. Viergever, Efficient and reliable schemes for nonlinear diffusion filtering, IEEE Transactions on Image Processing 7 (3) (1998) 398–410.