

A Probabilistic Approach to Optical Flow based Super-Resolution

Rik Fransens¹, Christoph Strecha¹, Luc Van Gool^{1,2}

¹ESAT-PSI, University of Leuven, Belgium

²Computer Vision Group (BIWI), ETH Zuerich, Switzerland

{rik.fransens, christoph.strecha}@esat.kuleuven.ac.be, vangool@vision.ee.ethz.ch

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 modeled by introducing so-called visibility maps, which are treated as hidden variables. We describe an EM-algorithm, which iterates between estimating values for the hidden quantities, and optimizing the flow-fields and the super-resolution image. The approach is illustrated with a synthetic and a challenging real-world example.

I. 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 are 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, photo-

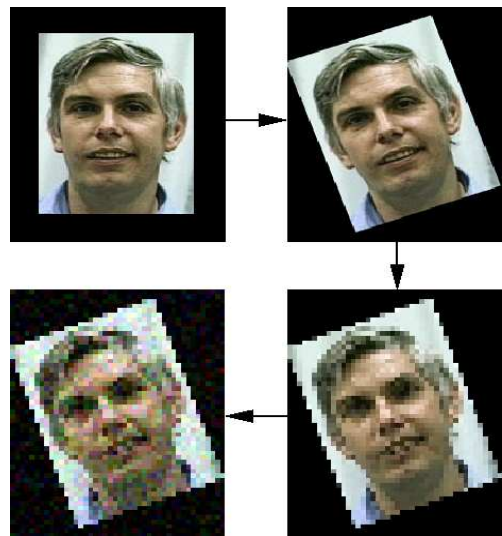


Fig. 1. **Generative model** Different stages in the image formation model. The super-resolved image is shown in the left top. The image is geometrically warped according to transformation \mathcal{F}_i^{-1} (top right). The result is low-pass filtered by the smoothed block-filter g and sampled on a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

metric 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 *et al.* [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 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 Estimation-Maximization (EM) algorithm, which iterates between (i) photometric occlusion detection, and (ii) optimization of the flow-fields and SR-image. The estimated occlusions are also used to define a simple, yet effective, anisotropic optical flow regularizer. The performance of this approach is demonstrated with several examples.

II. PROBLEM STATEMENT

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 color 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 figure 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 modeled 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.

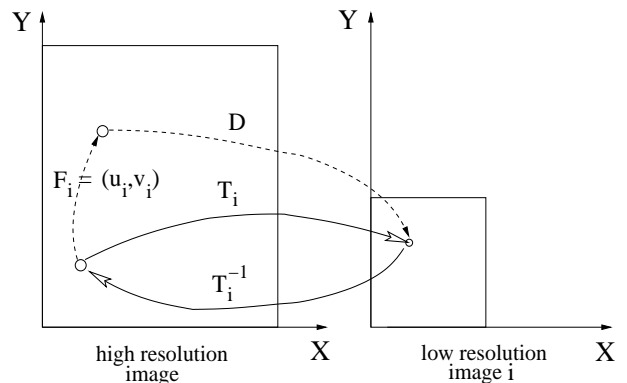


Fig. 2. **Transformations** Graphical representation of the different transformations. \mathcal{F}_i is the optic flow vector between corresponding points, and is defined in the coordinate system of the SR-image \mathcal{J}_0 . \mathcal{D} compensates for the magnification. \mathcal{T}_i maps a pixel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

A. Generative 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 compensated 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 modeled 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.

B. MAP estimation

We are now facing the hard problem of estimating the unknown quantities $\theta = \mathcal{J}_0, \mathcal{F}_i$ and Σ given the input images \mathcal{I}_i . Furthermore, we have introduced a set of hidden variables $\mathcal{V} = \mathcal{V}_{i \neq 0}$, which must also be inferred over the course of the optimization. In a Bayesian framework, the optimal value for θ is the one that maximizes the posterior probability $p(\theta | \mathcal{I}_i)$. According to Bayes' rule, this posterior can be written as:

$$p(\theta | \mathcal{I}_i) = \frac{\int p(\mathcal{I}_i | \theta, \mathcal{V}) p(\theta | \mathcal{V}) p(\mathcal{V}) d\mathcal{V}}{p(\mathcal{I}_i)}, \quad (4)$$

where we have conditioned the data likelihood and the prior on the hidden variables \mathcal{V} . The denominator or 'evidence' is merely the integral of the numerator over all possible values of θ and can be ignored in the maximization problem. In order to find the most probable value for θ , we need to integrate over all possible values of \mathcal{V} which is computationally intractable. Instead, we assume that the probability density function (PDF) of \mathcal{V} is peaked about a single value, i.e. $p(\mathcal{V})$ is a Dirac-function centered at this value. This leads to an Estimation-Maximization (EM) based solution, which iterates between (i) estimating values for \mathcal{V} , given the current estimate of θ , and (ii) maximizing the posterior probability of θ , given the current estimate of \mathcal{V} . A more detailed description of this procedure will be given later. So, given a current estimate $\hat{\mathcal{V}}$ for the hidden variables, we want to optimize:

$$q(\theta | \mathcal{I}_i) = p(\mathcal{I}_i | \theta, \hat{\mathcal{V}}) p(\theta | \hat{\mathcal{V}}) \quad (5)$$

The a-posteriori probability of θ is proportional to the product of two terms: the data-likelihood $p(\mathcal{I}_i | \theta, \hat{\mathcal{V}})$ and a prior $p(\theta | \hat{\mathcal{V}})$, which we call L and P , respectively. We now discuss both terms in turn.

Under the assumption that the image noise is i.i.d. for all pixels in all images, the data likelihood L can be written as the product of all individual pixel probabilities:

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

Here, \mathbf{x} runs over all h^2 positions of the super-resolution image \mathcal{J}_0 , and the product is restricted to those \mathbf{x} for which $\mathcal{V}_i(\mathbf{x}) = 1$. Given the current estimates for \mathcal{J}_0 and the noise distribution Σ , we can further specify L to be:

$$L = \prod_{i=-N}^N \prod_{\mathbf{x}} \frac{1}{(2\pi)^{d/2} |\Sigma|^{1/2}} \exp\left(-\frac{1}{2} \mathbf{m}_i^T \Sigma^{-1} \mathbf{m}_i\right), \quad (7)$$

where $\mathbf{m}_i = \mathcal{I}_i(\mathcal{T}_i(\mathbf{x})) - g * \mathcal{J}_0(\mathbf{x})$. This is the difference between corresponding positions in the low-pass filtered high-resolution image \mathcal{J}_0 and the input image \mathcal{I}_i . The variable d in the normalization constant denotes the dimensionality of \mathbf{m}_i .

The formulation of an appropriate prior is slightly more complicated. We can factorize P as the product of a flow-field dependent and an image dependent part:

$$P = c p(\mathcal{J}_0) \prod_i p(\mathcal{F}_i | \hat{\mathcal{V}}_i), \quad (8)$$

where c is an appropriate constant. To arrive at this expression, we have silently assumed that (i) the optical flow-fields are a-priori independent from each other¹, and (ii) the super resolution image is a-priori independent from the measurement noise and possible occlusions. The flow-field priors $p(\mathcal{F}_i | \hat{\mathcal{V}}_i)$ will be modeled as an exponential density distribution of the form $\exp(-R(\mathcal{F}_i)/\lambda_f)$. Here, λ_f is a parameter which controls the width of the distribution, and $R(\mathcal{F}_i)$ is a 'regularizer'. From such a regularizer we expect that it reflects our prior belief that the world is essentially simple. For a locally smooth solution \mathcal{F}_i in the neighborhood 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 regularizer, 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 prior as follows:

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

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 neighboring locations in \mathcal{J}_0 . To this end, we define the following distribution for \mathcal{J}_0 :

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

¹Clearly, 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.

We can now turn back to the optimization of θ . Instead of maximizing the posterior in (5), we minimize its negative logarithm. This leads (up to a constant) to the following energy:

$$E[\theta] = \frac{1}{2} \sum_i \sum_{\mathbf{x}} \mathcal{V}_i(\mathbf{x}) [\mathbf{m}_i^T \Sigma^{-1} \mathbf{m}_i + \log((2\pi)^{\frac{d}{2}} |\Sigma|)] + \frac{1}{2\lambda_f} \sum_i \sum_{\mathbf{x}} \mathcal{V}_i(\mathbf{x}) [R(\mathcal{F}_i)] + \frac{1}{\lambda_j} \sum_{\mathbf{x}} \|\nabla \mathcal{J}_0(\mathbf{x})\|^2. \quad (11)$$

Interestingly, 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 neighboring pixels are correlated with one another. Suppose we rearranged the pixels in \mathcal{J}_0 in lexicographic order, giving the h^2 -dimensional vector J_0 . Next, we define a $h^2 \times h^2$ -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:

$$\begin{aligned} \sum_{\mathbf{x}} \|\nabla \mathcal{J}_0(\mathbf{x})\|^2 &\approx (D_x J_0)^T (D_x J_0) + (D_y J_0)^T (D_y J_0) \\ &= J_0^T (D_x^T D_x + D_y^T D_y) J_0 \\ &= J_0^T Q J_0. \end{aligned} \quad (12)$$

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

$$p(J_0) = \exp\left(-\frac{1}{\lambda_j} J_0^T Q J_0\right), \quad (13)$$

where Q acts as an inverse covariance matrix. Because this distribution has zero mean, the regularizer tends to remove all high frequency content. Therefore, we also experimented with a second type of prior, where an up-sampled version of \mathcal{I}_0 is used as the mean of the prior distribution. A similar approach was followed by Capel *et al.* [4], however, they use a median image instead.

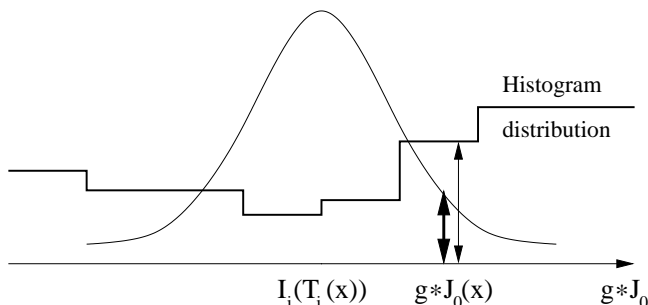


Fig. 3. **Visibility computation** The probability of $g^* \mathcal{J}_0(\mathbf{x})$ being visible in \mathcal{I}_i , can be computed from its value under the normal distribution, centered around $\mathcal{I}_i(\mathcal{T}_i(\mathbf{x}))$ (bold arrow), and its value under the histogram-based estimate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

C. An EM solution

In the previous paragraph, an energy equation, w.r.t. the unknown quantities θ , was derived. This energy corresponds to the negative logarithm of the posterior distribution of θ , given the current estimate of the hidden variable \mathcal{V} . Now we will derive the EM-equations, which iterate between the estimation of \mathcal{V} and the minimization of $E(\theta)$.

E-step On the $(k+1)^{th}$ iteration, the hidden variables $\mathcal{V}_i(\mathbf{x})$ are replaced by their conditional expectation given the data, where we use the current estimates $\theta^{(k)}$ for θ . The expected value for the visibility is given by $E[\mathcal{V}_i | \mathcal{J}_0, \Sigma, \mathcal{F}_i] \equiv \Pr(\mathcal{V}_i = 1 | \mathcal{J}_0, \Sigma, \mathcal{F}_i)$. According to Bayes' rule, the latter probability can be expressed as:

$$\Pr(\mathcal{V}_i = 1 | \mathcal{J}_0, \Sigma, \mathcal{F}_i) = \frac{p(\mathcal{J}_0 | \mathcal{V}_i = 1, \Sigma, \mathcal{F}_i)}{p(\mathcal{J}_0 | \mathcal{V}_i = 1, \Sigma, \mathcal{F}_i) + p(\mathcal{J}_0 | \mathcal{V}_i = 0, \Sigma, \mathcal{F}_i)}, \quad (14)$$

where we have assumed equal priors on the probability of a pixel being visible or not. Given the current estimate of θ , the PDF $p(\mathcal{J}_0 | \mathcal{V}_i = 1, \Sigma, \mathcal{F}_i)$ is given by the value of the noise distribution evaluated over the color-difference between $\mathcal{I}_i(\mathcal{T}_i(\mathbf{x}))$ and $g^* \mathcal{J}_0(\mathbf{x})$. The second PDF is more difficult to estimate, because it is hard to say what the color distribution of a pixel, which has no real counter-part in \mathcal{I}_i , looks like. We provide a *global* estimate for the PDF of occluded pixels by building a histogram of the color-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 $(1 - \mathcal{V}_i(\mathbf{x}))$. This is graphically depicted in figure 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 colored pixels. This makes sense from a perceptual point of view, and has a regularizing effect on the visibility maps.

M-step At the M-step, the intent is to compute values for θ that maximize (11), given the current estimates of \mathcal{V}_i . This is achieved by setting the parameters θ to the appropriate root of the derivative equation, $\partial E(\theta) / \partial \theta = 0$. We will start by deriving the update equations for the image related quantities Σ and \mathcal{J}_0 , and then proceed with the optimization of the flow-fields \mathcal{F}_i . For Σ , the update equation is given by:

$$\Sigma \leftarrow \frac{\sum_i \sum_{\mathbf{x}} \mathcal{V}_i \mathbf{m}_i(\mathbf{x}) \mathbf{m}_i(\mathbf{x})^T}{\sum_i \sum_{\mathbf{x}} \mathcal{V}_i(\mathbf{x})}, \quad (15)$$

where \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

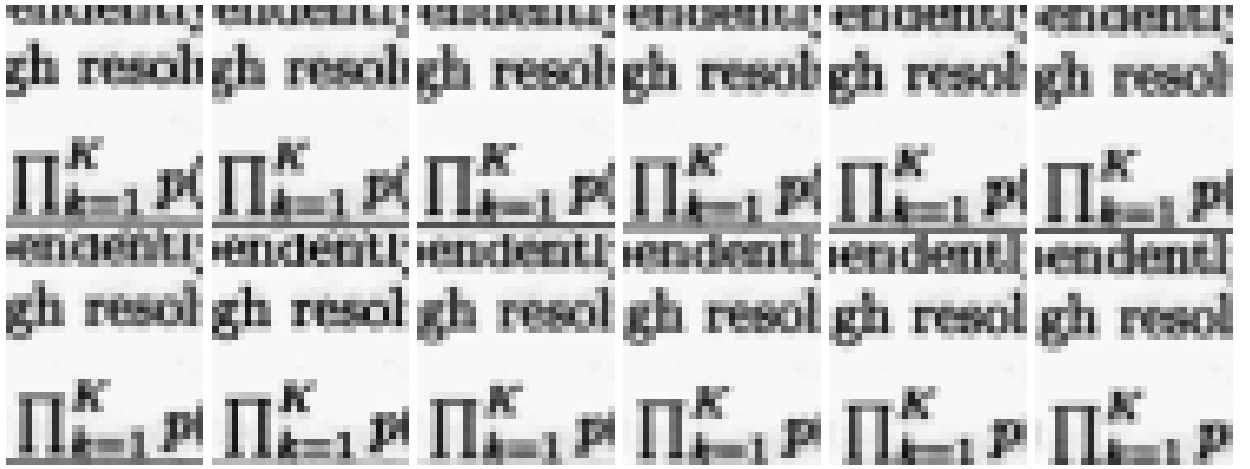


Fig. 4. **Synthetic data** Some of the 16 low-resolution input images. The motion field has a constant velocity of $[0.3, 0.15]$ pixels per time step.

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 (11) and represent them in a vector-matrix notation. First of all, the pixels (and color-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 are written as $J_0^T Q J_0$ (see Eq.12), where λ_j is absorbed into Q . The energy w.r.t. \mathcal{J}_0 now becomes:

$$E[J_0] = J_0^T Q J_0 + \sum_i (M_i J_0 - I_i)^T W_i S^{-1} (M_i J_0 - I_i), \quad (16)$$

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 . Differentiation of Eq.(16) w.r.t. J_0 and setting the result to zero, leads to the following over-determined linear system:

$$(Q + \sum_i M_i W_i S^{-1})^T J_0 = \sum_i (W_i S^{-1})^T I_i. \quad (17)$$

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

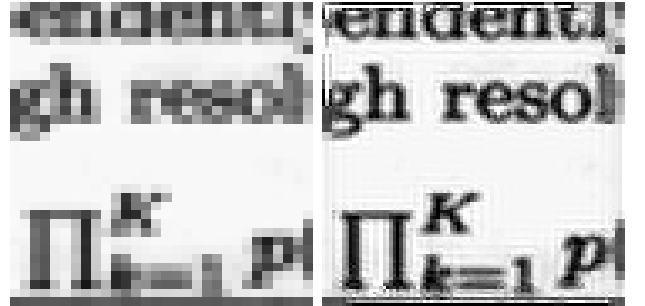


Fig. 5. **Synthetic data** Original low resolution image (left) and the corresponding super-resolution reconstruction (right).

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 of \mathcal{F}_i) that minimize the energy functional in Eq.(11) are determined. The minimizing flow-field satisfies the Euler-Lagrange equations, which leads to the following set of (anisotropic) diffusion equations:

$$\begin{aligned} \frac{\partial u_i}{\partial t} &= \mathcal{V}_i \mathbf{m}_i^T \Sigma^{-1} \frac{\partial \mathcal{I}_i}{\partial x} \frac{1}{m} - \frac{1}{\lambda_f} \text{div}(\mathcal{V}_i \nabla u_i) \\ \frac{\partial v_i}{\partial t} &= \mathcal{V}_i \mathbf{m}_i^T \Sigma^{-1} \frac{\partial \mathcal{I}_i}{\partial y} \frac{1}{m} - \frac{1}{\lambda_f} \text{div}(\mathcal{V}_i \nabla v_i). \end{aligned} \quad (18)$$

Here, $\partial \mathcal{I}_i / \partial x$ and $\partial \mathcal{I}_i / \partial y$ are d -vectors which contain the spatial gradients of the color-bands of \mathcal{I}_i . We omitted the dependencies of \mathbf{x} for notational clarity. The diffusion equations are solved by means of implicit discretization [13].

III. EXPERIMENTS

We tested our algorithm on a synthetic and a real world example. In the experiment on synthetic data, we took a small fragment of scanned text as input. Next,

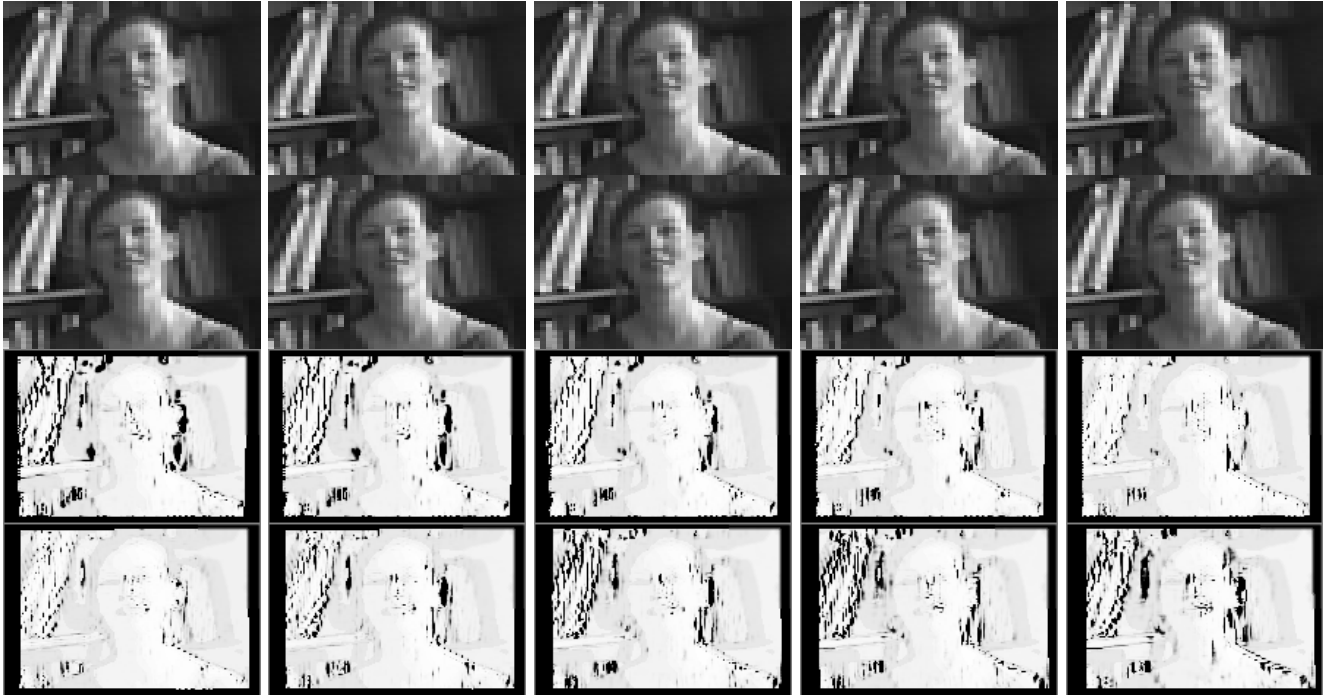


Fig. 6. **Real data** Camera translation with independently moving head. Top: input images, Bottom: corresponding visibility maps related to the reconstruction image shown in Fig.(7). The occlusions and depth discontinuities at the visual hull of the head and the books in the background are detected well. Also within the face, the areas nearby the bridge of the nose and the mouth have a lowered visibility.

we applied the image formation model, i.e. the image was geometrically transformed, low-pass filtered and sub-sampled. We applied a constant motion field of $[0.3, 0.15]$ pixels per time-step. The resulting sequence contains 16 images of size 30×35 (rows \times cols), which are shown in Fig.(4). The 8th image was taken as the reference view \mathcal{I}_0 , from which we compute the high resolution image \mathcal{J}_0 . We applied a magnification factor $m = 3$. For the super-resolution image smoothness constraint we used the data driven prior from Eq.(13), i.e. \mathcal{I}_0 was up-sampled with factor 3 and served as the mean of the spatial Gaussian distribution on \mathcal{J}_0 . The SR-smoothness prior parameter λ_j and the optical flow prior parameter λ_f were set to 1000 and 0.001, respectively. The SR-result is shown in Fig.(5). Obviously, the readability of letters and symbols has improved significantly.

In the experiment with real-world data, we recorded a scene with independent camera translation and object movement. We selected 15 consecutive images from a video sequence, from which we segmented 15 windows of size 60×88 (rows \times cols). The 8th image was taken as the reference view, from which we compute the high resolution image. The input images, except for the reference view, are shown in the two top rows of Fig.(6). This is a challenging test case because of the poor lighting conditions. There are strong specular reflections, e.g. on



Fig. 7. **Real data** Original low resolution image (left) and the corresponding super-resolution reconstruction (right).

the face and shoulders, which violate the optical flow constant brightness-assumptions. 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. For this experiment, we used a similar \mathcal{I}_0 -based prior as described above. The free parameters were set to $\lambda_j = 1000$ and $\lambda_f = 0.01$. The results of the visibility computations are shown in the two bottom rows of Fig.(6). Note that the occlusions and depth discontinuities, at the visual hull of the head and the books in the background, are detected well. The SR-result is shown in Fig.(7). Again, we notice an improved quality, particularly for the facial features and the book shelf in the background.

IV. CONCLUSIONS

In this paper, we presented a probabilistic approach to optical flow based super-resolution. 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 regularized 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 Estimation-Maximization algorithm, which iterates between (i) photometric occlusion detection, and (ii) estimation of the noise covariance matrix and optimization 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 optimization of this energy w.r.t. the super-resolution image, the relevant terms are isolated and turned into an overdetermined set of linear equations. In our implementations, we use a sparse conjugate gradient algorithm to solve this system. For the optimization 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 discretization.

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 our experiments, λ_j was set to 1000, 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 algorithm, the 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 (regularized) iteratively reweighted least squares solution to the SR-constraints.

The explicit computation of noise and visibilities also 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 regularization term. As a result, the flow-fields \mathcal{F}_i will be more strongly regularized, i.e. will become more smooth. This is important at initialization time, when the flows \mathcal{F}_i have not yet converged to a sensible solution. The visibility estimates, on the other hand, act on a local level. If at a particular image location visibility is low, regularization 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.

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," CVPR, Vol.2, pp.372-379, 2000.
- [4] D. Capel, A. Zisserman, "Computer Vision Applied to Super Resolution," IEEE Signal Processing Magazine, pp.75-86, 2003.
- [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, Vol.8(3), pp.387-395, 1999.
- [7] M. Elad, A. Feuer, "Super-resolution restoration of image sequences," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol.21(9), pp.817-834, 1999.
- [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, Vol.6(12), pp.1621-1633, 1997.
- [9] M. Irani, S. Peleg, "Improving resolution by image registration," GMIP, 53, pp.231-239, 1991.
- [10] R. Schultz, R. Stevenson, "Extraction of high-resolution frames from video sequences," IEEE Transactions on Image Processing, Vol.5(6), pp.996-1011, 1996.
- [11] C. Strecha, R. Fransens, L. Van Gool, "A Probabilistic Approach to Large Displacement Optical Flow and Occlusion Detection," to appear in SMVP, 2004.
- [12] M.E. Tipping, C.M. Bishop, "Bayesian image super-resolution," In S. Becker, S. Thrun, and K. Obermayer (Eds.), Advances in Neural Information Processing Systems 15, 2003. MIT Press.
- [13] J. Weickert, B.M. ter Haar Romeny, M.A. Viergever, "Efficient and Reliable Schemes for Nonlinear Diffusion Filtering," IEEE Transactions on Image Processing, Vol.7(3), pp.398-410, 1998.