

A Probabilistic Approach to Large Displacement Optical Flow and Occlusion Detection

Christoph Strecha¹, Rik Fransens¹, and Luc Van Gool¹

K.U.Leuven, Belgium

firstname.surname@esat.kuleuven.ac.be,

WWW home page: <http://www.esat.kuleuven.ac.be/psi/visics>

Abstract. This paper deals with the computation of optical flow and occlusion detection in the case of large displacements. We propose a Bayesian approach to the optical flow problem and solve it by means of differential techniques. The images are regarded as noisy measurements of an underlying 'true' image-function. Additionally, the image data is considered incomplete, in the sense that we do not know which pixels from a particular image are occluded in the other images. We describe an EM-algorithm, which iterates between estimating values for all hidden quantities, and optimizing the current optical flow estimates by differential techniques. The Bayesian way of describing the problem leads to more insight in existing differential approaches, and offers some natural extensions to them. The resulting system involves less parameters and gives an interpretation to the remaining ones. An important new feature is the photometric detection of occluded pixels. We compare the algorithm with existing optical flow methods on ground truth data. The comparison shows that our algorithm generates the most accurate optical flow estimates. We further illustrate the approach with some challenging real-world examples.

1 Introduction

A fundamental problem in the processing of image sequences is the computation of optical flow. Optical flow is caused by the time-varying projection of objects onto a possibly moving image plane. It is therefore the most general transformation, assigning a two dimensional displacement vector to every pixel. There exist other image motion representations that restrict optical flow to parametric models (affine, homography) or to one dimensional disparities as in the case of stereo. Optical flow estimation has been extensively studied. The most common methods are differential, frequency-based, block matching and correlation-based methods. We refer the interested reader to Barron *et al.* [4] for a description and comparison of important optical flow techniques.

In this paper we focus on differential techniques and investigate the explicit detection of occluded pixels. Differential techniques were, in the context of optical flow computation, introduced by Horn and Schunck [7]. They are based on the Optical Flow Constraint (OFC) equation that relates spatial and temporal brightness gradients to the two components of the optical flow field. Additional constraints on this displacement field have to be added to overcome the ill-posedness of the problem. Originally Horn

and Schunck used an isotropic smoothness constraint, which is not able to deal with discontinuities. Other researchers investigated anisotropic versions of smoothness based on image gradients [9,2] and optical flow gradients [5,3]. In these approaches, new parameters controlling the strength of anisotropy are introduced. A taxonomy of different smoothness constraints, of which some are also used in nonlinear diffusion filtering, is presented in [12]. Proesmans *et al.* [10] present a third way to introduce anisotropy. They define a consistency function that depends on the difference of matching image one to image two and vice versa (so-called *forward-backward* matching). In their work, occluded pixels are characterized by a low value of the consistency function. A similar forward-backward strategy for occlusion detection, but with structure tensor regularization, is used by Alvarez *et al.* [1]. However, these authors require the forward-backward optical flow computation *and* additional parameters for solving the occlusion problem.

We formulate the correspondence problem in a probabilistic framework. This results in an EM algorithm, whose maximization step involves diffusion equations similar to the ones previously described. However, the probabilistic description now gives an interpretation to the most important parameter (λ) which controls the balancing between image matching and flow field smoothness. The formulation leads naturally to the detection of occlusions based on the image values themselves, which prevents the computation of the two (forward-backward) optical flow fields. Our algorithm needs no additional parameters for detecting occlusions, and the relative contribution of the spectral components to the matching term is handled automatically.

This paper is organized as follows. We proceed with the explanation of the probabilistic framework and its EM-based solution. Section 3 deals with the actual optical flow computation. We introduce two algorithms that differ by their anisotropic smoothness term and discuss the implications of the probabilistic view to the matching term. Experiments on ground truth and real data are discussed in section 4, where we compare our two algorithms with an implementation of Alvarez *et al.* [2].

2 Probabilistic account for optical flow and occlusions estimation

Suppose we are given two images \mathcal{I}_1 and \mathcal{I}_2 , 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.¹ Our goal is to estimate the two-dimensional displacement or optical flow field $\mathcal{F}(\mathbf{x})$ such that $\mathcal{I}_1(\mathbf{x})$ is brought into correspondence with $\mathcal{I}_2(\mathbf{x} + \mathcal{F}(\mathbf{x}))$.

Faithful to the Bayesian philosophy, we regard the input images as noisy measurements of an unknown image irradiance \mathcal{I}_1^* . This allows us to write:

$$\mathcal{I}_1(\mathbf{x}) = \mathcal{I}_1^*(\mathbf{x}) + \epsilon \quad (1)$$

$$\mathcal{I}_2(\mathbf{x} + \mathcal{F}(\mathbf{x})) = \mathcal{I}_1^*(\mathbf{x}) + \epsilon \quad (2)$$

$$\epsilon \sim \mathcal{N}(\mathbf{0}, \Sigma) \quad (3)$$

¹ In fact one could add other features such as filter responses to the image [5]. We continue the discussion for general n -band images.

where $\mathcal{N}()$ is a normal noise distribution with zero-mean and covariance matrix Σ , which we assume to be equal for all images. Both the irradiance or 'true' image \mathcal{I}_1^* and Σ are unknown, and their estimation becomes part of the optimization procedure.

A major complication for larger displacements is the occlusion problem, which arises from the fact that not all parts of the scene, which are visible in a particular image, are also visible in the other images due to occlusion. When computing image correspondences, such occluded regions should be identified and excluded from the matching procedure. This will be modelled by introducing a visibility map $\mathcal{V}(\mathbf{x})$, which signal whether a scene point \mathbf{X} that projects onto \mathbf{x} in \mathcal{I}_1 is also visible in image \mathcal{I}_2 or not. Every element of $\mathcal{V}(\mathbf{x})$ is a binary random variable which is either 1 or 0, corresponding to visibility or occlusion, respectively. $\mathcal{V}(\mathbf{x})$ is a hidden variable, and its value must be inferred from the input images.

Estimating the optical flow field $\mathcal{F}(\mathbf{x})$ can now be formally stated as finding those flow values which make the image correspondences $\mathcal{I}_1^*(\mathbf{x}) \Leftrightarrow \mathcal{I}_2(\mathbf{x} + \mathcal{F}(\mathbf{x}))$ restricted to $\mathcal{V}(\mathbf{x}) = 1$, most probable.

2.1 MAP estimation

We are now facing the hard problem of estimating the unknown quantities $\theta = \mathcal{F}, \mathcal{I}_1^*$ and Σ given the images \mathcal{I}_1 and \mathcal{I}_2 . Furthermore, we have introduced the unobservable or hidden variables \mathcal{V} , 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 $p(\theta|\mathcal{I}_1, \mathcal{I}_2)$. According to Bayes' rule, this posterior can be written as:

$$p(\theta|\mathcal{I}_1, \mathcal{I}_2) = \frac{\int p(\mathcal{I}_1, \mathcal{I}_2|\theta, \mathcal{V})p(\theta|\mathcal{V})p(\mathcal{V})d\mathcal{V}}{p(\mathcal{I}_1, \mathcal{I}_2)}, \quad (4)$$

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. Hence, we will try to optimize the numerator only. 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:

$$q(\theta|\mathcal{I}_1, \mathcal{I}_2) = p(\mathcal{I}_1, \mathcal{I}_2|\theta, \hat{\mathcal{V}})p(\theta|\hat{\mathcal{V}}) \quad (5)$$

The a-posteriori probability of θ is proportional to the product of two terms: the data-likelihood $p(\mathcal{I}_1, \mathcal{I}_2|\theta, \hat{\mathcal{V}})$ 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 both views, the data likelihood L can be written as the product of all individual pixel probabilities:

$$L = \prod_{\mathbf{x}} p(\mathcal{I}_1(\mathbf{x})|\theta) \prod_{\mathbf{x}} p(\mathcal{I}_2(\mathbf{x} + \mathcal{F}(\mathbf{x}))|\theta), \quad (6)$$

where the second product is restricted to those \mathbf{x} for which $\mathcal{V}(\mathbf{x}) = 1$. Note that, by definition, all pixels in \mathcal{I}_1^* are visible in \mathcal{I}_1 . Given the current estimate of the 'true' image $\mathcal{I}_1^*(\mathbf{x})$ and the noise distribution Σ , we can further specify the likelihood by the normal distribution \mathcal{N} :

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

where $\mathbf{m}_1(\mathbf{x}) = \mathcal{I}_1^*(\mathbf{x}) - \mathcal{I}_1(\mathbf{x})$ and $\mathbf{m}_2(\mathbf{x}) = \mathcal{I}_1^*(\mathbf{x}) - \mathcal{I}_2(\mathbf{x} + \mathcal{F}(\mathbf{x}))$ are the differences of the true image function with the input images \mathcal{I}_1 and \mathcal{I}_2 , the latter one estimated at the current value of \mathcal{F} . 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 marginalize P as the product of a displacement \mathcal{F} dependent and image dependent part:

$$P = p(\mathcal{F} | \mathcal{I}_1^*, \Sigma) p(\mathcal{I}_1^*, \Sigma). \quad (8)$$

Assuming we have no prior preference for the image related parameters, i.e. assuming a uniform prior over \mathcal{I}_1^* and Σ , this can be rewritten as:

$$P = p(\mathcal{F} | \mathcal{I}_1^*, \Sigma) c \quad (9)$$

where c is an appropriate constant.

The displacement prior $p(\mathcal{F} | \mathcal{I}_1^*, \Sigma)$ will be modelled as an exponential density distribution of the form $\exp(-R(\mathcal{I}_1^*, \mathcal{F})/\lambda)$. Here, λ is a parameter which controls the width of the distribution, and $R(\mathcal{I}_1^*, \mathcal{F})$ is a data-driven 'regularizer'. From such a regularizer we expect that it reflects our prior belief that the world is essentially simple, i.e. for a locally smooth solution \mathcal{F} 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. Furthermore, the regularizer should be data-driven: if the image \mathcal{I}_1^* suggests a discontinuity, i.e. by the presence of a high image gradient at a particular point \mathbf{x} , a large discontinuity at \mathbf{x} should not be made a-priori unlikely. Such regularizers are commonly used in the PDE-community [2,9], where they serve as *image driven anisotropic diffusion operators* in optical flow or edge-preserving smoothing computations. The regularizer is given by:

$$R(\mathcal{I}_1^*, \mathcal{F}) = \nabla \mathcal{F}^T T(\nabla \mathcal{I}_1^*) \nabla \mathcal{F}. \quad (10)$$

Here, $T(\nabla \mathcal{I}_1^*)$ is a diffusion tensor defined by:

$$T(\nabla \mathcal{I}_1^*) = \frac{1}{|\nabla \mathcal{I}_1^*|^2 + 2\nu^2} \left(\nabla \mathcal{I}_1^{*\perp} \nabla \mathcal{I}_1^{*\perp T} + \nu^2 \mathbf{1} \right), \quad (11)$$

where $\mathbf{1}$ is the identity matrix, ν is a parameter controlling the degree of anisotropy and $\nabla \mathcal{I}_1^{*\perp}$ is the vector perpendicular to $\nabla \mathcal{I}_1^*$. For color images, the tensor is defined as the sum of the 3 individual color channel tensors. $R(\mathcal{I}_1^*, \mathcal{F})$ is low when $\nabla \mathcal{F}$ is parallel

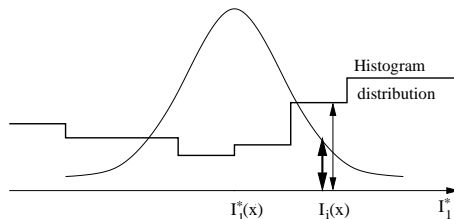


Fig. 1. Visibility estimation: The probability of $\mathcal{I}_i(\mathbf{x})$ being visible, is proportional to its value under the Gauss-curve, and the probability of $\mathcal{I}_i(\mathbf{x})$ being invisible is measured as the value under the histogram-based estimator.

to $\nabla \mathcal{I}_1^*$, which is exactly the desired behavior. Note that, by making the displacement prior dependent on \mathcal{I}_1^* , it implicitly also makes it dependent on the original image data. Strictly speaking, this violates the Bayesian principle that priors should not be estimated from the data. In practice, however, it leads to more sensible solutions than setting them arbitrarily, or using so-called *conjugate* priors, whose main justification comes from computational simplicity [11].

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

$$E[\theta] = \frac{1}{2} \sum_{i=1}^2 \sum_{\mathbf{x}} \mathcal{V}(\mathbf{x}) [\mathbf{m}_i(\mathbf{x})^T \Sigma^{-1} \mathbf{m}_i(\mathbf{x}) + \log((2\pi)^{\frac{d}{2}} |\Sigma|)] + \frac{1}{\lambda} R(\mathcal{I}_1^*(\mathbf{x}), \mathcal{F}(\mathbf{x})) \quad (12)$$

2.2 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 variable $\mathcal{V}(\mathbf{x})$, 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 $E[\mathcal{V}|\mathcal{I}_1^*, \Sigma, \mathcal{F}] \equiv \Pr(\mathcal{V}=1|\mathcal{I}_1^*, \Sigma, \mathcal{F})$. According to Bayes' rule, the latter probability can be expressed as:

$$\Pr(\mathcal{V}=1|\mathcal{I}_1^*, \Sigma, \mathcal{F}) = \frac{p(\mathcal{F}|\mathcal{V}=1, \mathcal{I}_1^*, \Sigma)}{p(\mathcal{F}|\mathcal{V}=1, \mathcal{I}_1^*, \Sigma) + p(\mathcal{F}|\mathcal{V}=0, \mathcal{I}_1^*, \Sigma)}, \quad (13)$$

where we have assumed equal priors on the probability of a pixel being visible or not. Given the current estimate of θ , the PDF $p(\mathcal{F}|\mathcal{V}=1, \mathcal{I}_1^*, \Sigma)$ is given by the value of the noise distribution evaluated over the color-difference between $\mathcal{I}_1^*(\mathbf{x})$ and $\mathcal{I}_2(\mathbf{x} + \mathcal{F}(\mathbf{x}))$:

$$p(\mathcal{F}|\mathcal{V}=1, \mathcal{I}_1^*, \Sigma) = \mathcal{N}(\mathbf{m}_2; \mathbf{0}, \Sigma). \quad (14)$$

1. Initialize $\mathcal{V} = 0.5$ 2. Loop over pyramidal images M-step until convergence Compute \mathcal{I}_1^* and Σ by eq.(16) Compute \mathcal{F} by solving the diffusion equation (see section 3) E-Step Estimate new visibilities \mathcal{V} by eq. (15) 3. Rescale, goto next pyramidal level

Table 1. *Outline of the algorithm*

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}_2 , looks like. We provide a *global* estimate for the PDF of occluded pixels by building a histogram of the color-values in \mathcal{I}_1^* which are currently invisible. This is merely the histogram of \mathcal{I}_1^* where the contribution of each pixel is weighted by $(1 - \mathcal{V}(\mathbf{x}))$. Note that, if a particular pixel in \mathcal{I}_1^* 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. The update equations for $\mathcal{V}(\mathbf{x})$ are now:

$$\mathcal{V} \leftarrow \frac{\mathcal{N}(\mathbf{m}_2; \mathbf{0}, \Sigma)}{\mathcal{N} + \text{HIST}_{\mathcal{I}_1^*, (1-\mathcal{V})}(\mathcal{I}_1^*)}, \quad (15)$$

where \mathcal{N} is evaluated as in (14). This is graphically depicted in fig. (1).

M-step At the M-step, the intent is to compute values for θ that maximizes (12), given the current estimates of \mathcal{V} . This is achieved by setting the parameters θ to the appropriate root of the derivative equation, $\partial E(\theta)/\partial \theta = 0$.

For the image related parameters \mathcal{I}_1^* and Σ , a closed form expressions for the roots can be derived and the update equations are:

$$\begin{aligned} \mathcal{I}_1^*(\mathbf{x}) &\leftarrow \frac{1}{1 + \mathcal{V}(\mathbf{x})} (\mathcal{I}_1(\mathbf{x}) + \mathcal{V}(\mathbf{x})\mathcal{I}_2(\mathbf{x} + \mathcal{F}(\mathbf{x}))) \\ \Sigma &\leftarrow \frac{1}{\sum_x (1 + \mathcal{V}(\mathbf{x}))} \sum_x \mathbf{m}_1(\mathbf{x})\mathbf{m}_1(\mathbf{x})^T + \mathcal{V}(\mathbf{x})\mathbf{m}_2(\mathbf{x})\mathbf{m}_2(\mathbf{x})^T. \end{aligned} \quad (16)$$

In order to arrive at these expressions, we ignored the effects of \mathcal{I}_1^* and Σ on the regularization term. This is admissible because their influence on R is small compared to their influence on the matching term. Σ is only indirectly related to R through the computation of the visibility maps, which have an effect on R via the computation of \mathcal{I}_1^* . The image \mathcal{I}_1^* has an effect on R via its gradient, which is used to define a quadratic norm on the depth gradient (10). Changes of \mathcal{I}_1^* will therefore only have a minor influence on R .

However, for the update of the optical flow field \mathcal{F} we are not so lucky, because \mathcal{F} strongly influences both the matching and the regularization term. To minimize E w.r.t. \mathcal{F} , we solve the diffusion equation that can be derived from eq. 12 using the Euler-Lagrange formalism. The outline of the overall algorithm is given in table 1.

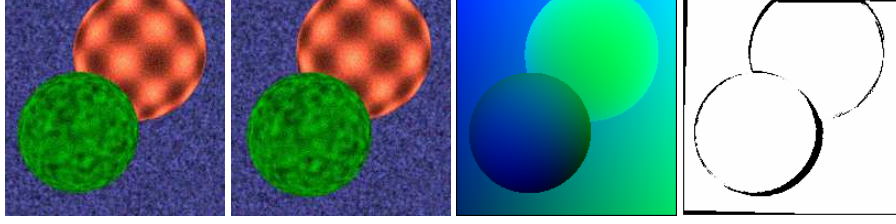


Fig. 2. One of the synthetic scenes used in the experiments. Left two images: input; middle right: ground truth optical flow; right: ground truth occlusions

3 Variational Optical flow estimation

In this section we will derive two update equations for the optical flow field. The first one, with image based anisotropic smoothness, follows directly from the previous considerations. The second one has a visibility based anisotropic smoothness term. We will also discuss the differences between these approaches and other differential optical flow techniques that have the ability to deal with large displacements.

3.1 Image induced anisotropy

Consider the main result of the first section - the energy in eq.(12) - where we only consider the terms depending on $\mathcal{F}(\mathbf{x})$:

$$E[\mathcal{F}(\mathbf{x})] = \sum_{\mathbf{x}} \mathcal{V}(\mathbf{x}) \mathbf{m}_2(\mathbf{x})^T \Sigma^{-1} \mathbf{m}_2(\mathbf{x}) + \frac{1}{\lambda} \nabla \mathcal{F}(\mathbf{x})^T T(\nabla \mathcal{I}_1^*) \nabla \mathcal{F}(\mathbf{x}). \quad (17)$$

The visibilities \mathcal{V} are fixed for a particular instance of the M-step. For a given maximization step the three unknowns are estimated in turn while keeping the others fixed. In that case the minimum of the above energy is given by the Euler-Lagrange equation. To allow for large displacements, we split the value of $\mathcal{F}(\mathbf{x})$ into a current and a residual estimate [10,2], i.e. $\mathcal{F}(\mathbf{x}) = \mathcal{F}_0(\mathbf{x}) + \mathcal{F}_r(\mathbf{x})$. Cutting of terms $\geq O(\mathcal{F}_r^2)$ from the Taylor expansion of $\mathbf{m}_2(\mathbf{x})$, we get:

$$\mathbf{m}_2(\mathbf{x}) = \mathcal{I}_1^* - \mathcal{I}_2(\mathbf{x} + \mathcal{F}_0(\mathbf{x})) - \frac{\partial \mathcal{I}_2(\mathbf{x} + \mathcal{F}_0(\mathbf{x}))}{\partial \mathbf{x}} (\mathcal{F}(\mathbf{x}) - \mathcal{F}_0(\mathbf{x})) \quad (18)$$

The Euler Lagrange equation leads to the following diffusion equation:

$$\frac{\partial \mathcal{F}(\mathbf{x})}{\partial t} = \text{div}(T(\nabla \mathcal{I}_1^*) \nabla \mathcal{F}(\mathbf{x})) - \lambda \mathcal{V}(\mathbf{x}) \left(\frac{\partial \mathcal{I}_2(\mathbf{x} + \mathcal{F}_0(\mathbf{x}))}{\partial \mathbf{x}} \right)^T \Sigma^{-1} \mathbf{m}_2(\mathbf{x}). \quad (19)$$

We now compare this result with two other important optical flow (OF) approaches, described in [2,10]. First of all, the data term in eq.(19) now contains Σ^{-1} , which performs a global, relative weighting of the different spectral components. Apart from this, Σ^{-1} also globally weights the importance of the matching term w.r.t. the smoothness

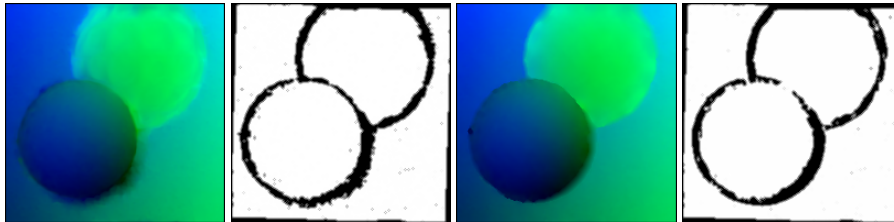


Fig. 3. Resulting optical flow fields and visibilities using the 2nd (left two images) and 3rd (right two images) algorithm, $\log(\lambda) = -7$, histogram size = 8^3 .

term. More image noise decreases the norm of Σ^{-1} . This automatically results in a more smooth solution, which is a desirable mechanism. A further modification of the data term is the local weighting of the image value differences by the visibilities $\mathcal{V}(\mathbf{x})$. When a pixel receives a low visibility score, the smoothness term locally gets more important. This avoids that wrongly matched occluded pixels pull, by the action of the smoothness term, neighboring unoccluded pixels in the wrong direction. Another difference is related to the model image \mathcal{I}_1^* . Instead of comparing $\mathcal{I}_1(\mathbf{x})$ with $\mathcal{I}_2(\mathbf{x} + \mathcal{F}_0(\mathbf{x}))$, which is the usual practice in OF computation, the Bayesian framework tells us to use $\mathcal{I}_1^*(\mathbf{x})$ instead. This results again in a visibility dependent weighting². The algorithm has, similar to Alvarez *et al.*[2], two free parameters. They are ν , which controls the degree of anisotropy in eq. (11), and λ , which controls the width (hence the importance) of regularization prior. The smoothness term in eq. (19) is conceptually similar to the one used by Alvarez *et al.*. However, their regularizer is linear whereas ours is non-linear and \mathcal{I}_1^* dependent.

3.2 Visibility induced anisotropy

In the second algorithm, we employ a smoothness term related to the work by Proesmans *et al.*[10], which outperformed other optical flow algorithms in a recent benchmark [8]. In [10], the basic idea is to compute optical flow from \mathcal{I}_1 to \mathcal{I}_2 and vice-versa, which are called forward and backward diffusion respectively. Ideally corresponding forward and backward flow vectors should sum up to zero. The consistency of the forward-backward match is measured by two other diffusion equations. These equations assume spatial smoothness of the consistency values and constraint them to the range [0..1]. Anisotropy is realized using these local consistency estimates. This approach needs optical flow computation in two directions (forward/backward) as well as the two other consistency diffusion equations with their own parameters. Instead of using these geometric consistencies, we propose to use the visibilities $\mathcal{V}(\mathbf{x})$ as a photometric consistency measure. Only the smoothness term in eq. (19) changes and we get:

$$\frac{\partial \mathcal{F}(\mathbf{x})}{\partial t} = \text{div}(\mathcal{V}(\mathbf{x}) \nabla \mathcal{F}(\mathbf{x})) - \lambda \mathcal{V}(\mathbf{x}) \left(\frac{\partial \mathcal{I}_2(\mathbf{x} + \mathcal{F}_0(\mathbf{x}))}{\partial \mathbf{x}} \right)^T \Sigma^{-1} \mathbf{m}_2(\mathbf{x}). \quad (20)$$

² recall from eq. 16, that $\mathcal{I}_1^*(\mathbf{x}) - \mathcal{I}_2(\mathbf{x} + \mathcal{F}_0(\mathbf{x})) = (\mathcal{I}_1(\mathbf{x}) - \mathcal{I}_2(\mathbf{x} + \mathcal{F}_0(\mathbf{x}))) / (1 + \mathcal{V}(\mathbf{x}))$

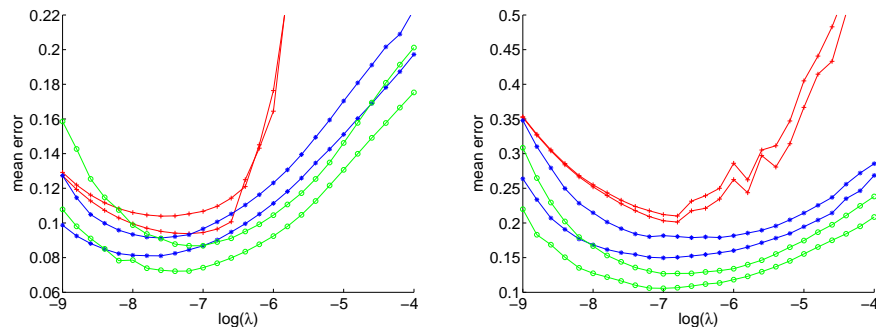


Fig. 4. Mean error of all visible pixels for three algorithms on two synthetic scenes (for 2 different amounts of Gaussian noise) as a function of λ . The different algorithms are shown in 1st (red, +), 2nd (blue, *), 3rd (green, o).

This smoothness term blocks diffusion from places with high visibility estimates to places with lower ones. Typically, at initialisation time visibilities start out isotropically distributed while at the end pixels with low visibilities tend to cluster near discontinuities and occlusions. The algorithm has only one free parameter, being λ .

We will end with a final note on the convergence properties of the algorithms. Dempster *et al.* [6] have shown that, for Maximum-Likelihood (ML) estimation, each iteration of EM is guaranteed to increase the data-likelihood, which drives the parameters θ to a local optimum of L . In this work, we have included a prior on the unknown variables, so for the moment we can not make such strong claims. However, various trials on different data sets have confirmed the robust behavior of the two proposed algorithms.

4 Experiments and discussion

We tested our two optical flow algorithms, together with our implementation of Alvarez *et al.* [2], on two synthetic scenes. To each of these scenes, different amounts of Gaussian noise were added. One of these scenes, together with the ground truth displacements and occlusions, is shown in fig. (2). The three evaluated algorithms are:

- (1) image based anisotropic optical flow, our implementation of [2]
- (2) image based anisotropic optical flow, with probabilistic matching term eq. (19)
- (3) visibility based anisotropic optical flow, with probabilistic matching term eq. (20)

Fig. (3) shows the result of the 2nd and 3rd algorithm for the scene in fig. (2). The optical flow field is displayed color coded. We use the green channel for the horizontal and the blue channel for the vertical component of the flow field. The mean error (mean distance of ground truth with estimated optical flow) of all unoccluded pixels is shown in fig. (4) as a function of the λ parameter. The left hand plot of fig. (4) displays the results of the three algorithms on a scene with moving camera and a rotating and translating ball (not shown due to space limitations). We added Gaussian noise of variance

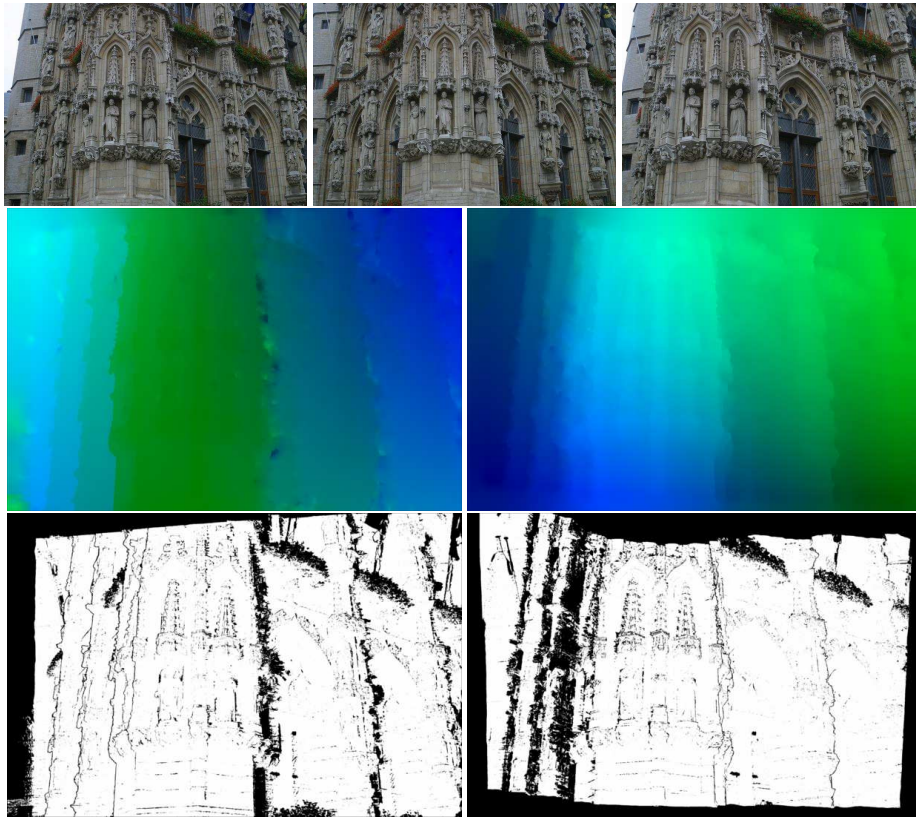


Fig. 5. City hall scene: top: input images 1, 2 and 3; middle: independently computed optical flow from \mathcal{I}_1 to \mathcal{I}_2 (left) and from \mathcal{I}_1 to \mathcal{I}_3 (right), color coding as explained in the text; bottom: visibility maps for these two flow computations.

1 and 3 to the three color channels, so every algorithm appears twice in the plots. The right hand plot in fig. (4) displays the results on the scene shown in fig. (2) under the same experimental conditions. This scene is more complex than the first one.

In all synthetic experiments our probabilistic matching term performs better than the algorithm of Alvarez *et al.*[2]. Concerning the dependency on λ , one can see that Alvarez *et al.*(red,+) shows the expected behavior w.r.t. the amount of noise added. When the amount of noise increases, the optimal λ -value decreases, indicating the need for stronger regularisation. Our algorithms ((blue,*) and (green,o)) show an opposite behavior. This is probably due to the fact that the ML-estimate of Σ in eq. (16) underestimates the true noise level when only two images are used. In conclusion, all methods have a noise-level dependent and scene dependent optimal λ . This dependency can only be resolved with prior knowledge of the complexity of the scene.

When comparing our two algorithms, one can observe better results for the 3rd algorithm, which incorporates the visibility based anisotropic smoothness term (eq. 20).

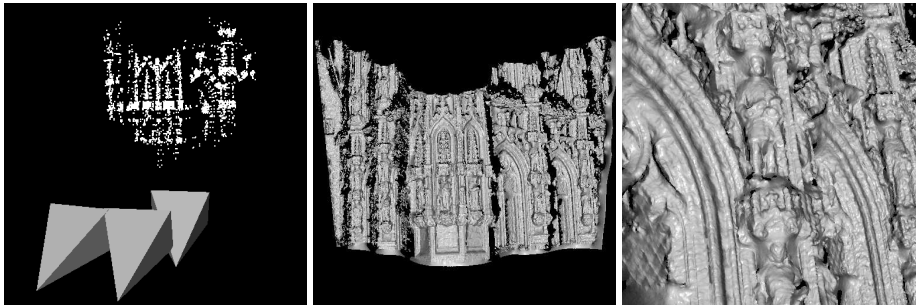


Fig. 6. 3D reconstructions from optical flow estimation: left: result of the self-calibration; middle, right: two rendered un-textured views

The effect is even more outspoken when the complexity of the scene, i.e. the amount of occlusions, increases. This result suggests that occlusions (visibilities) are estimated well in the probabilistic framework. This can also be observed from fig. (3) and fig. (5). By coupling the anisotropy to the visibilities instead of strong image gradients, especially nearby discontinuities, better flow estimates are obtained.

In two real data experiments, we show the results of our best performing 3rd algorithm. The city hall scene is characterized by large displacements and a strong impact of occlusions. Three images of resolution 3072×2048 have been used to compute optical flow and occlusions from image \mathcal{I}_1 to \mathcal{I}_2 and from image \mathcal{I}_1 to \mathcal{I}_3 (top images in fig. (5)) independently. The resulting optical flow fields and the visibility maps are shown in fig. (5). Because we have no ground truth for these image pairs, we decided to evaluate the quality of the results by computing the 3D reconstruction. After processing the two image pairs we self-calibrated the scene using all optical flow correspondences from \mathcal{I}_1 to the other images. Only scene points with high visibility ($\mathcal{V}(\mathbf{x}) \geq 0.5$) to both views and with high spatial gradient were used. This results in the external and internal calibration of the cameras and a sparse set 3-D points. These are shown together with the camera positions on the left image in fig. (6). Using this camera calibration and the optical flow correspondences from \mathcal{I}_1 to \mathcal{I}_2 we computed the 3-D points, now for all pixels with high visibility. The result is shown as an untextured 3-D mesh rendered from two different virtual camera positions (right two images in fig. (6)). We wish to stress that this approach is purely for evaluation purpose. In real applications, one would of course restrict the correspondence search to the epipolar lines. However, even without epipolar constraints, the 2D optical flow matches result in a reliable 3D reconstruction, which is an indication for the quality of the computed flow field.

The visibility maps in fig. (5) show that the occluded pixels and discontinuities are accurately detected. Furthermore, they also show low visibility values for the specularities in the window closest to the viewer and for the flowers at the right top of the images. This is because specularities are classified as outliers given the local estimate of \mathcal{I}_1^* and the global estimate of Σ . As a result, at these positions the importance of the matching term decreases and the flow field estimate is locally strongly regularized.

Also pixel discretisation errors at high frequency detail (e.g. for the flowers) can bring about a similar effect.

5 Conclusions

We have presented two differential optical flow algorithms that simultaneously estimate image displacements and occlusions, as well as the noise distribution and denoised image, as part of an EM algorithm. Starting from relatively straightforward probabilistic considerations, we arrive at an energy formulation with a strong intuitive appeal. The energy often taken as a point of departure in other differential optical flow approaches, turns out to be a special case of this result. More specifically, it can be derived from our formulation by assuming unit strength noise, full visibility and by setting the unknown 'true' irradiance equal to the first image. The estimation of visibilities (occlusions) is naturally incorporated into the algorithm, similar in flavour to outlier detection in iteratively reweighted least squares estimation. This is a rather old concept, however, its use in optical flow computation and occlusion detection is new. Noticably, our best performing algorithm does not introduce additional parameters to realize anisotropic diffusion and to detect occluded pixels.

References

1. L. Alvarez, R. Deriche, T. Papadopoulo, and J. Snchez: Symmetrical dense optical flow estimation with occlusions detection. *ECCV* **1** (2002) 721–735
2. L. Alvarez, J. Weickert, and J. Snchez: Reliable estimation of dense optical flow fields with large displacements. *IJCV* **39** (2000) 41-56
3. G. Aubert, R. Deriche, and P. Kornprobst: Computing optical flow via variational techniques. *SIAM Journal on Applied Mathematics* **60**(1) (1999) 156–182
4. J. Barron, D. Fleet, and S. Beauchemin: Performance of optical flow techniques. *IJCV* **12**(1) (1994) 43–77
5. T. Brox, A. Bruhn, N. Papenberg, and J. Weickert: High accuracy optical flow estimation based on a theory for warping. *ECCV* **4**(2004) 25–36
6. A. P. Dempster, N. M. Laird, and D. B. Rubin: Maximum likelihood from incomplete data via the em algorithm. *J. R. Statist. Soc. B* **39** (1977) 1–38
7. B.K.P. Horn and B.G. Schunck: Determining optical flow. *Artificial Intelligence* **17**(1981)185–204
8. B. McCane, K. Novins, D. Crannitch, and B. Galvin: On benchmarking optical flow. *Computer Vision and Image Understanding* **84**(1) (2001) 126–143
9. H. H. Nagel: Constraints for the estimation of displacement vector fields from image sequences. *Proc. Int. Joint Conf. Artificial Intell.*(1983) 945–951
10. M. Proesmans, L. Van Gool, E. Pauwels, and A. Oosterlinck: Determination of optical flow and its discontinuities using non-linear diffusion. *ECCV* **2** (1994) 295–304
11. N. Vasconceles and A. Lippman: Empirical Bayesian EM-based Motion Segmentation. *CVPR* (1997) 527–532
12. J. Weickert and T. Brox: Diffusion and regularization of vector- and matrix-valued images. *Inverse Problems, Image Analysis, and Medical Imaging. Contemporary Mathematics* **313**(2002)