

# Optical Flow Estimation with Uncertainties through Dynamic MRFs

Ben Glocker<sup>1,2\*</sup>, Nikos Paragios<sup>1</sup>, Nikos Komodakis<sup>1,3</sup>,  
Georgios Tziritas<sup>3</sup>, and Nassir Navab<sup>2</sup>

<sup>1</sup> GALEN Group, Mathématiques Appliquées aux Systèmes (MAS), Ecole Centrale Paris, France  
paragios@ecp.fr

<sup>2</sup> Computer Aided Medical Procedures (CAMP), Technische Universität München, Germany  
{glocker, navab}@in.tum.de

<sup>3</sup> Computer Science Department, University of Crete, Greece  
{komod, tziritas}@csd.uoc.gr

## Abstract

*In this paper, we propose a novel dynamic discrete framework to address image morphing with application to optical flow estimation. We reformulate the problem using a number of discrete displacements, and therefore the estimation of the morphing parameters becomes a tractable matching criteria independent combinatorial problem which is solved through the FastPD algorithm. In order to overcome the main limitation of discrete approaches (low dimensionality of the label space is unable to capture the continuous nature of the expected solution), we introduce a dynamic behavior in the model where the plausible discrete deformations (displacements) are varying in space (across the domain) and time (different states of the process - successive morphing states) according to the local uncertainty of the obtained solution.*

## 1. Introduction

Optical flow estimation is among the most challenging problems in computer vision. The task consists of recovering a 2D displacement vector establishing correspondences between the consecutive projections of a 3D patch in the image. The central assumption in such a context is the visual preservation constraint [12], that assumes pixel corresponding to the same 3D projection will have the same intensities. In the most general case, the motion field can either be global or local. Global transformations consist of a parametric function of variables that is applied to the entire domain, and has often as parameters the spatial position of the pixel. Rigid, similarity, affine and quadratic are some examples. These methods are a compromise between low com-

\*Partially supported by Siemens Medical Solutions, Germany.



**Figure 1:** The 10th frame of the Army sequence (upper row) and Mequon sequence (lower row) and our estimated (color-encoded) optical flow fields.

plexity and performance. Dense optical flow methods aim to recover one-to-one correspondences spatially varying between pixels. Variational methods [4, 18], MRF-based approaches [3, 11] as well as statistical techniques [8] were often considered to determine the optical flow estimates.

Dense optical flow estimation is by definition an ill-posed problem since the number of constraints is inferior to the number of variables. Consequently, in order to address the ill-posedness of the problem additional regularization constraints are to be introduced. Despite such additional constraints, the estimation of motion vectors is not trivial either because (i) the optimization process cannot guarantee the optimal solution, or (ii) ambiguities exist in the flow where different displacements locally might correspond to the same error (like in the absence of texture). This is the case for methods based on gradient descent which are sensitive to the initial conditions. On the other hand, in the recent years we have observed the renaissance of discrete

MRFs [9] thanks to the introduction of efficient optimization techniques [27, 3, 24, 6, 15, 17]. In [10] for example a registration approach based on discrete optimization showing promising performance is proposed. However, the use of such models to recover solutions of continuous problems suffer from the discretization bottleneck. In other words, a compromise is to be found between the number of labels (quantization of the search space), the computational complexity as well as the ability to capture a good minimum. Furthermore, towards tractable computational solutions in discrete optimization the form of the objective function is constrained in terms of the regularization term [16] (only sub-modular interaction terms can be considered).

In this paper, we propose a discrete to continuous approach for optical flow estimation that (i) can capture a better minimum, (ii) has incrementally refined precision that is defined locally and varies according to the image structure, (iii) can encode complex interactions between graph nodes, and (v) can complete the task in reasonable computational time. The central idea of our approach is to explore/determine uncertainties in the context of multi-label optimization. Our approach is based on linear programming and explores the duality theorem. Furthermore, it explores the notion of min/max marginals towards recovering successive uncertainty maps which are directly related with the covariance matrix of the obtained solution. Then, it uses these maps within a dynamic model where the set of plausible deformations is varying in space and state of the optimization while being able to self-adjust the precision of the solution according to the observed uncertainties.

There is quite limited work on uncertainty-based morphing focusing mostly on surface matching. In [19] various optimization criteria have been proposed to recover correspondences between geometric features with an explicit estimation of the uncertainty for the resulting transformation. The use of uncertainties to determine the search range within the iterated closest point algorithm has been introduced in [23] while a more advanced shape representation with explicit estimation of uncertainties for thin plate splines was proposed in [20]. Last but not least in [25] uncertainties were considered in a implicit registration framework with surfaces being described as distance maps. These methods are applicable for a specific range of transformations as well as for a given similarity metric. In the image domain, the work presented in [22] assumes displacements to be Gaussian to determine their covariance matrix using the image noise model. In the discrete case, one can only point out recent work on graph-cuts and min/max marginals [14] for the case of segmentation where a binary labeling problem is to be addressed.

The remainder of this paper is organized as follows. In Section 2, we present the generic optical flow estimation framework while in Section 3 we introduce the notion of

variable-precision labeling and the estimation of uncertainties. The nature of the displacement field and the experimental validation are part of Section 4 while discussion concludes the paper.

## 2. Optical Flow Estimation with Uncertainties

The basic idea for optical flow estimation is that image points are translated in a sequence from one frame to the next while assuming brightness constancy for all points,

$$I(\mathbf{x}, f) = I(\mathbf{x} + \mathcal{D}(\mathbf{x}), f + 1), \quad (1)$$

where  $I(\mathbf{x}, f)$  is the image intensity of the pixel  $\mathbf{x}$  in frame  $f$ , and  $\mathcal{D}$  is the 2D displacement vector field.

A common way to formulate the problem of optical flow estimation for two given frames is through the definition of a distance between these images that is to be minimized in the entire domain  $\Omega$ , or

$$E_{\text{data}}(\mathcal{D}) = \int_{\Omega} \underbrace{|I(\mathbf{x}, f) - I(\mathbf{x} + \mathcal{D}(\mathbf{x}), f + 1)|}_{\rho(\mathbf{x})} d\mathbf{x} \quad (2)$$

Recovering the optimal potential of this cost function is not straightforward. Two variables are to be determined while only one constraint is available per pixel. The most basic approach to address this limitation is through the use of a regularization function on space of unknown variables [26], or

$$E_{\text{smooth}}(\mathcal{D}) = \int_{\Omega} \phi(\nabla \mathcal{D}(\mathbf{x})) d\mathbf{x} \quad (3)$$

with  $\phi$  being a convex function imposing smoothness on the displacement field for neighboring pixels. Such a term will make the estimation of the displacement field feasible assuming that an appropriate function  $\rho$  is known.

Gradient descent is the most common approach to perform the optimization, a method that has some strengths and known limitations. One can claim that this approach is convenient and often it is straightforward to implement. On the other hand, since the objective function is non-convex one cannot guarantee that the obtained solution will be the optimal one. Last, but not least gradient numerical manipulation is not straightforward when projecting from the continuous space to the discrete one.

The above observations lead to a natural conclusion that one should seek (i) dimensionality reduction on the degrees of freedom of the motion model, (ii) more efficient optimization techniques both in terms of ability to approach the optimal solution with reasonable computational cost, and (iii) techniques that do not require continuous gradient manipulation in discrete spaces.

### 2.1. Dimensionality Reduction

Our motion model follows the idea of [10]. The authors use this model for image registration mainly in the domain

of medical imaging. It basically consists of a set of control points  $G$  uniformly distributed over the image domain  $\Omega$ . Here, usually the number of control points is much lower than the number of pixels. While moving the control points and thus manipulating the model configuration the underlying image structures get transformed. The displacement of an image pixel  $\mathbf{x}$  can be expressed using a linear or non-linear combination of the control point displacements  $\mathcal{C}$ , or

$$\mathcal{D}(\mathbf{x}) = \sum_{\mathbf{p} \in G} \eta(|\mathbf{x} - \mathbf{p}|) \mathcal{C}(\mathbf{p}) \quad (4)$$

with  $\eta$  being the influence function. In this work we consider cubic B-Splines [21] as basis functions for the motion model. We redefine the matching criteria on the motion model in the following form

$$E_{\text{data}}(\mathcal{D}) = \frac{1}{|G|} \sum_{\mathbf{p} \in G} \int_{\Omega} \hat{\eta}(|\mathbf{x} - \mathbf{p}|) \rho(\mathbf{x}) d\mathbf{x} \quad (5)$$

where  $\hat{\eta}$  is the ‘‘inverse’’ influence function determining the contribution of the image pixel  $\mathbf{x}$  to the propagated data term on the level of control points. For pixel-wise measures (e.g. the sum of absolute differences) this is defined as

$$\hat{\eta}(|\mathbf{x} - \mathbf{p}|) = \frac{\eta(|\mathbf{x} - \mathbf{p}|)}{\int_{\Omega} \eta(|\mathbf{y} - \mathbf{p}|) d\mathbf{y}} \quad (6)$$

which results in equal data terms for Equations (2) and (5). For statistical measures (e.g. correlation coefficient) we define this function differently as

$$\hat{\eta}(|\mathbf{x} - \mathbf{p}|) = \begin{cases} 1, & \text{if } \eta(|\mathbf{x} - \mathbf{p}|) > 0 \\ 0 & \text{if } \eta(|\mathbf{x} - \mathbf{p}|) = 0 \end{cases} \quad (7)$$

which results in the computation of local statistics within each control point influence area.

Such a matching process is purely image based and therefore cannot guarantee smoothness on the displacement field. The use of partial derivatives of the displacements is often considered to determine a measure of smoothness, or

$$E_{\text{smooth}}(\mathcal{D}) = \frac{1}{|G|} \sum_{\mathbf{p} \in G} \int_{\Omega} \hat{\eta}(|\mathbf{x} - \mathbf{p}|) (|\partial_x \mathcal{D}(\mathbf{x})| + |\partial_y \mathcal{D}(\mathbf{x})|) d\mathbf{x} \quad (8)$$

a term motivated by the diffusion energy functional, here used to control the spatial variations of the displacements. Obviously, the smoothness term could also be directly computed on the control point displacements  $\mathcal{C}$ . In Section 3.1, where we will describe our incremental approach it hopefully becomes clear why we choose to propagate the smoothness computation from the image level back to the control point level.

The complete cost function for the optical flow estimation problem defined on our motion model then becomes

$$E_{\text{total}} = E_{\text{data}} + \lambda E_{\text{smooth}} \quad (9)$$

where  $\lambda$  controls the balance between the data and the smoothness term.

## 2.2. Discrete Formulation

We consider a discrete set of labels  $L = \{l^1, \dots, l^m\}$  corresponding to a quantized version of the displacement space  $\Theta = \{\mathbf{d}^1, \dots, \mathbf{d}^m\}$ . A label assignment  $l_{\mathbf{p}}$  to a control point  $\mathbf{p}$  is associated with displacing the point by the corresponding vector  $\mathbf{d}^{l_{\mathbf{p}}}$ . The dense displacement field associated with a certain discrete labeling  $l$  becomes

$$\mathcal{D}(\mathbf{x}) = \sum_{\mathbf{p} \in G} \eta(|\mathbf{x} - \mathbf{p}|) \mathbf{d}^{l_{\mathbf{p}}} \quad (10)$$

This allows us to reformulate the optical flow estimation as a discrete multi-labeling problem modeled in first-order Markov Random Field (MRF), where the goal is to assign individual labels to the MRF nodes. The first-order MRF energy formulation is defined as

$$E_{\text{MRF}}(l) = \sum_{\mathbf{p} \in G} V_{\mathbf{p}}(l_{\mathbf{p}}) + \sum_{\mathbf{p} \in G} \sum_{\mathbf{q} \in \mathcal{N}(\mathbf{p})} V_{\mathbf{p}\mathbf{q}}(l_{\mathbf{p}}, l_{\mathbf{q}}) \quad (11)$$

where  $V_{\mathbf{p}}$  are the unary potentials representing the data term,  $V_{\mathbf{p}\mathbf{q}}$  are the pairwise potentials representing the smoothness term, and  $\mathcal{N}$  represents the neighborhood system of the nodes. We define the unary potentials according to our data term as

$$V_{\mathbf{p}}(l_{\mathbf{p}}) \approx \int_{\Omega} \hat{\eta}(|\mathbf{x} - \mathbf{p}|) \rho(\mathbf{x}) d\mathbf{x}. \quad (12)$$

In general the unary potentials are assumed to be independent which in most cases according to the influence function  $\eta$  is not true. Naturally, neighboring control points influence in an overlapping area of the dense displacement field in a linear or non-linear manner. However, using  $\hat{\eta}$  while propagating the costs to the control point level, reduces the approximation error.

In addition, we define the pairwise potentials as

$$V_{\mathbf{p}\mathbf{q}}(l_{\mathbf{p}}, l_{\mathbf{q}}) = \lambda |(\mathcal{R}(\mathbf{p}) + \mathbf{d}^{l_{\mathbf{p}}}) - (\mathcal{R}(\mathbf{q}) + \mathbf{d}^{l_{\mathbf{q}}})| \quad (13)$$

where  $\mathcal{R}$  back-projects the current displacement field onto the level of control points, or

$$\mathcal{R}(\mathbf{p}) = \int_{\Omega} \hat{\eta}(|\mathbf{x} - \mathbf{p}|) \mathcal{D}(\mathbf{x}) d\mathbf{x}. \quad (14)$$

Such a smoothness term allows for regularization of the displacement field across the domain of the images as well as the time of the optimization process. We should note, that this is different from the smoothness term proposed in [10] which is only able to regularize the single optimization steps independently.

### 2.3. FastPD & Discrete Optimization

For optimizing the above discrete Markov Random Field, we will make use of a recently proposed method, called Fast-PD [17]. This is an optimization technique, which builds upon principles drawn from the duality theory of linear programming in order to derive efficient algorithms that can provide approximately optimal solutions for a wide class of NP-hard MRFs.

Instead of working directly with the discrete MRF optimization problem above, Fast-PD first reformulates that problem as an integer linear programming problem (the primal problem) and also takes the dual of the corresponding LP relaxation. Given these 2 problems, i.e. the primal and the dual, Fast-PD then generates a sequence of integral feasible primal solutions, as well as a sequence of dual feasible solutions. These two sequences of solutions make local improvements to each other until the primal-dual gap (i.e. the gap between the objective function of the primal and the objective function of the dual) becomes small enough. Once this happens, the last generated primal solution is guaranteed to be an approximately optimal solution, i.e. within a certain distance from the optimum (in fact, this distance has been shown to be smaller than the achieved primal-dual gap). This is exactly what the next theorem, also known as the *primal-dual principle*, states.

**Primal-Dual Principle 1** *Consider the following pair of primal and dual linear programs:*

$$\begin{array}{ll} \text{PRIMAL: } \min \mathbf{c}^T \mathbf{x} & \text{DUAL: } \max \mathbf{b}^T \mathbf{y} \\ \text{s.t. } \mathbf{A} \mathbf{x} = \mathbf{b}, \mathbf{x} \geq \mathbf{0} & \text{s.t. } \mathbf{A}^T \mathbf{y} \leq \mathbf{c} \end{array}$$

and let  $\mathbf{x}, \mathbf{y}$  be integral-primal and dual feasible solutions, having a primal-dual gap less than  $f$ , i.e.:

$$\mathbf{c}^T \mathbf{x} \leq f \cdot \mathbf{b}^T \mathbf{y}.$$

Then  $\mathbf{x}$  is guaranteed to be an  $f$ -approximation to the optimal integral solution  $\mathbf{x}^*$ , i.e.,  $\mathbf{c}^T \mathbf{x}^* \leq \mathbf{c}^T \mathbf{x} \leq f \cdot \mathbf{c}^T \mathbf{x}^*$

Note that when this principle is applied to MRF optimization, primal costs corresponds to MRF energies, while dual costs correspond to lower bounds on the optimum MRF energy. Hence, each iteration of the primal-dual schema produces a pair of upper and lower bounds to the optimum MRF energy. It was shown in [17] that each such iteration can be reduced to solving a max-flow problem for a capacitated graph constructed based on the current primal-dual pair  $(\mathbf{x}, \mathbf{y})$ . The resulting flows can be computed very efficiently and tell us how to update both the primal and dual variables in order to get better upper and lower bounds (on the MRF optimum) for the next iteration.

### 3. Morphing and Uncertainties

One of the main limitations of discrete labeling when seeking estimates of continuous variables is the precision, in particular when the range of displacements to be captured is important as in optical flow. One can address this demand through the use of more complete label sets, which however will make the problem far more complex and almost intractable from a computational perspective.

One can first decrease the importance of such a constraint through an incremental approach. To this end, we can consider a fixed set of labels and corresponding displacements, and iteratively determine the additive component of the deformation which will decrease maximally the cost function. Let  $l^0$  being the labeling of the grid that corresponds to no-displacement  $\mathcal{D}^0$ , then for a given set of labels, one can determine the solution to the labeling through the optimum of

$$l^1 = \arg \min_l \sum_{\mathbf{p} \in G} V_{\mathbf{p}}(l_{\mathbf{p}}) + \sum_{\mathbf{p} \in G} \sum_{\mathbf{q} \in \mathcal{N}(\mathbf{p})} V_{\mathbf{p}\mathbf{q}}(l_{\mathbf{p}}, l_{\mathbf{q}}) \quad (15)$$

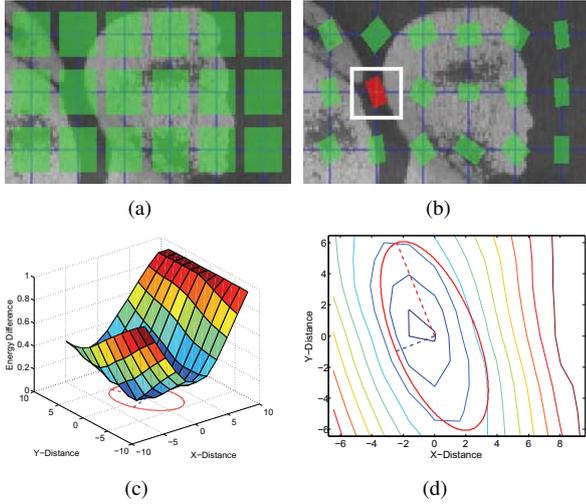
which will produce a 2D displacement field  $\mathcal{D}^1$ . Such a displacement field can be used to morph the frame  $F^0 = I(\mathbf{x}, f)$  to a new frame  $F^1 = I(\mathbf{x} + \mathcal{D}^1(\mathbf{x}), f^1)$ . Incremental morphing consists of recovering an update component  $\mathcal{D}^2$ , which will bring closer  $F^1$  to the frame  $F^* = I(\mathbf{x} + \mathcal{D}(\mathbf{x}), f + 1)$  of the sequence  $I$ . In the most general case, let

- $l^t$  being the labeling that corresponds to the incremental component of the displacements at time  $t$ , and let  $\mathcal{D}_{F^t} = \sum_{\tau=1}^t \mathcal{D}^\tau$  being the accumulative displacements,
- $F^t$  being the  $t$ -morphing of the original frame  $F^0$  using  $\mathcal{D}_{F^t}$ .

Here, it should become clear why we previously defined the smoothness term in Equation (8) on the pixel level, since the control point displacements are only incrementally determined and updated on  $\mathcal{D}_{F^t}$ . Now, one can consider improving the current solution, through the lowest potential of

$$l^{t+1} = \arg \min_l E_{\text{MRF}}(l | F^t, F^*) \quad (16)$$

Such an incremental approach will reduce complexity while preserving the ability to capture important displacements. However, it still suffers from being isotropic since each control point refers to the same label set and the additive displacements remain constant over time. One can either decrease the distance between the set of deformations as a function over time (often done based on heuristics) or determine the uncertainties of the current solution and sample the displacement capture range according to the anisotropy and the scale of the covariance matrix corresponding to them.



**Figure 2:** The displacement capture range (shown in green) at each control point before (a) and after (b) the first cycle using the dynamic local adjustment through uncertainty estimation. The min-marginal maps and estimated covariances of a label set for an exemplary control point (marked in red) are shown in (c) and (d).

### 3.1. Uncertainties & Dynamic Morphing

Let us assume  $\Sigma_d(\mathbf{p})$  being the covariance matrix of a Gaussian distribution explaining the local uncertainty of the flow estimation at a given control point. This can be considered as a good indicator on the direction which potential additive displacements might improve the current estimate as well as the sampling that is to be used along the two axes. Let  $\phi(\mathbf{p}), \phi^\dagger(\mathbf{p})$  being the orientations,  $s_1(\mathbf{p}), s_2(\mathbf{p})$  the relative scales of two axes of  $\Sigma_d(\mathbf{p})$ , and let us consider a label set that consists  $(2n+1) \times (2n+1)$  labels. Then, in order to define an optimal set of labels, one should consider  $(2n+1)$  labels in  $\phi$  corresponding to a uniform sampling along this axis with distance  $d_1(\mathbf{p}) = \frac{s_1(\mathbf{p})}{2n+1}$  (similar definition is considered for the second axis:  $d_2(\mathbf{p}) = \frac{s_2(\mathbf{p})}{2n+1}$ ). Then, the set of non-uniform labels/displacements that best explores the image content is defined through all possible combinations of the two subsets. Such a process will produce a set of labels that is varying across control points and optimally explores the current uncertainties of the solution.

Localizing the global minimum of an objective function  $E$  is equivalent to finding the major mode of a random variable with density  $\exp(-E/\beta)$ . The coefficient  $\beta$  corresponds to the allowable variation in the energy value around the minimum.

Let us consider the optimal solution obtained for a particular control point  $\mathcal{C}(\mathbf{p}) = (\mathcal{C}_x(\mathbf{p}), \mathcal{C}_y(\mathbf{p}))$  at a given time  $t$ . In order to determine the variation of the cost function, one can consider a series of perturbations. Let  $\Delta\mathcal{C} = \{-k\delta\epsilon, \dots, -\delta\epsilon, 0, \delta\epsilon, \dots, k\delta\epsilon\}$  being an one-dimensional set with  $\delta\epsilon \rightarrow 0$  and an incremental displacement  $\lambda_1 \in \Delta\mathcal{C}, \lambda_2 \in \Delta\mathcal{C}$

$$\Delta\mathcal{C}(\mathbf{p}) = \mathcal{C}(\mathbf{p}) + \lambda_1 \mathbf{e}_1 + \lambda_2 \mathbf{e}_2 \quad (17)$$

with  $\mathbf{e}_1 = (d_1(\mathbf{p}), 0)$  and  $\mathbf{e}_2 = (0, d_2(\mathbf{p}))$ . Then, it is obvious that  $\Delta\mathcal{C}(\mathbf{p})$  is part of the label set, that is it exists a label  $l_k$  such that  $\mathbf{d}^{l_k} = \Delta\mathcal{C}(\mathbf{p})$ . Then, a good indicator on the variation/form of the cost function at any given time  $t$  compares the optimal cost (obtained displacement) with the one produced from a slight perturbation. In the context of our approach the optimal labeling is  $l^t$  produces an optimal morphing between  $F^{t-1}$  and  $F^t$  and the obtained solution corresponds to the exact global minimum of:

$$\sum_{\mathbf{p} \in G} V_{\mathbf{p}}(l_{\mathbf{p}}^t | F^{t-1}, F^t) + \sum_{\mathbf{p} \in G} \sum_{\mathbf{q} \in \mathcal{N}(\mathbf{p})} V_{\mathbf{p}\mathbf{q}}(l_{\mathbf{p}}^t) \quad (18)$$

In order to determine the local uncertainty [14], one can consider a min-marginal function that measures information about the minimum values of the energy  $E_{\text{MRMF}}(l^t | F^{t-1}, F^t)$  under different constraints. In our case, such constraints for each control point  $\mathbf{p}$  adopt the optimal labeling and introduce local variations of the  $(\lambda_1, \lambda_2)$  subset.

$$U(\mathcal{C}(\mathbf{p}), \lambda_1, \lambda_2) = \frac{\exp(-\psi_{\mathbf{p}; \Delta\mathcal{C}(\mathbf{p}); F^{t-1}, F^t})}{\sum_{l_{\mathbf{p}} \in L} \exp(-\psi_{\mathbf{p}; \mathbf{d}^{l_{\mathbf{p}}}; F^{t-1}, F^t})} \quad (19)$$

with  $\psi_{\mathbf{p}; \mathbf{d}; F^{t-1}, F^t}$  being the min-marginal energy under the constraint that  $\mathbf{p}$  is forced to be displaced by  $\mathbf{d}$ , or

$$\psi_{\mathbf{p}; \mathbf{d}; F^{t-1}, F^t} = \min_{l, \mathcal{C}(\mathbf{p}) = \mathbf{d}} E_{\text{MRMF}}(l | F^{t-1}, F^t) \quad (20)$$

The exact computation of these costs is feasible without additional cost if  $\lambda_1, \lambda_2$  are constrained to produce displacements that do belong to the existing set.

The variation of these measures form a 2D density and the estimation of the covariance matrix could provide a qualitative interpretation of the results. Intuitively, if we plot the spatial min-marginal energies (see Figure 2) for a grid node we can estimate the uncertainty by measuring the flatness of the plotted curve. As flatter the curve becomes in a particular direction as more uncertain is the control point position. In other words, it would not make much difference in the total energy if this control point is displaced by a non-optimal label. Contrary, if the energy difference is large in the area around a control point we are quite certain about its position. In order to determine these uncertainties we should sample the entire space towards recovering these uncertainties which is rather unrealistic from computational perspective. Still, we can approximate this covariance matrix using the min-marginal map for the label sets (see Figure 2).

### 3.2. FastPD & Min/Max Marginals

In order to determine the uncertainties, inspired from recent publications [13, 14] we use the FastPD to compute the min-marginal energies for every control point and every

displacement in the set of labels. To this end, we make use of the Fast-PD’s ability to accelerate inference on dynamic MRFs (where this term refers to MRFs slowly varying over time). In fact, the computation of min-marginals for a discrete MRF is equivalent to optimizing such a dynamic MRF. To see this, let  $V_{\mathbf{p}}(\cdot)$ ,  $V_{\mathbf{pq}}(\cdot, \cdot)$  be the unary and pairwise potentials of a discrete MRF. To compute the min-marginal for label  $l_{\mathbf{p}} = \alpha$  at node  $\mathbf{p}$ , we simply need to replace the unary potential  $V_{\mathbf{p}}(\cdot)$  at node  $\mathbf{p}$  with the following potential function  $V_{\mathbf{p}}^{\text{new}}(\cdot)$ :

$$V_{\mathbf{p}}^{\text{new}}(l_{\mathbf{p}}) = \begin{cases} V_{\mathbf{p}}(\alpha), & \text{if } l_{\mathbf{p}} = \alpha \\ \infty, & \text{if } l_{\mathbf{p}} \neq \alpha \end{cases}$$

No other changes need to be made to the potential functions of the original MRF. It is then easy to see that the desired min-marginal can be obtained by optimizing the resulting MRF. Hence, to compute all min-marginals, one has to optimize one such MRF for every node  $\mathbf{p}$  and label  $l_{\mathbf{p}}$ . This can be efficiently done using the Fast-PD algorithm, since all these MRFs are almost identical to each other and thus form a very slowly varying dynamic MRF. Although the produced min-marginals will not be exact (since optimizing multi-label MRFs is, in general, NP-hard), nevertheless they will be close to the true marginals, since Fast-PD can guarantee to generate approximately optimal solutions (this is in contrast to using Belief Propagation, in which case no optimality guarantees exist for the computed marginals). Moreover, an upper-bound per marginal can be computed by the Fast-PD algorithm, telling how far the estimated marginal is from the true one. After estimating the min marginals  $\mu_p(\alpha) = \min_{l_{\mathbf{p}}=\alpha} E(l)$ , one can then compute an uncertainty measure  $U$  for any MRF labeling as follows:

$$U_p(\alpha) = \frac{\exp^{-\mu_p(\alpha)}}{\sum_l \exp^{-\mu_p(l)}}$$

## 4. Experimental Validation

So far, our framework has been quite general in terms of the used matching criteria, the motion model, and the parameter configuration. Therefore, we will first present details on our implementation and describe the settings for our experimental validation before presenting the actual results.

### 4.1. Free Form Deformations

We consider Free Form Deformations (FFD) based on cubic B-splines as an interpolation model for the influence function  $\eta$ . FFD are successfully applied in non-rigid image registration (e.g. in [21]). Deformation of an object is achieved by manipulating an underlying mesh of uniformly spaced control points with spacing  $\delta$ . Two-dimensional FFD based on cubic B-Splines are defined on the image do-

main  $\Omega$  as

$$\mathcal{D}(\mathbf{x}) = \sum_{m=0}^3 \sum_{n=0}^3 B_m(u)B_n(v)\mathcal{C}(\mathbf{p}_{i+m,j+n}) \quad (21)$$

where  $i$  and  $j$  denote the index of the control point cell containing  $\mathbf{x} = (x, y)^T$ , and  $u$  and  $v$  are the relative positions of  $x$  and  $y$  inside that cell, e.g.  $i = \lfloor x/\delta \rfloor - 1$  and  $u = x/\delta - (i + 1)$ .  $B_l$  represents the  $l$ th basis function of the B-Spline. A single control point has only local influence on the resulting pixel displacements. A pyramidal approach allows for representing large and small displacements simultaneously.

### 4.2. Matching Criteria

The matching criteria we use is an adapted version of the correlation coefficient (CC) combined with the gradient inner product (GIP), or

$$\rho_{\text{CC+GIP}} = (1 - \gamma) \left( 1 - \left| \frac{\sum_{\Omega} (a - \bar{a})(b - \bar{b})}{\sqrt{\sum_{\Omega} (a - \bar{a})^2 \sum_{\Omega} (b - \bar{b})^2}} \right| \right) + \gamma \sum_{\Omega} \left| \frac{\nabla a}{|\nabla a|} \cdot \frac{\nabla b}{|\nabla b|} \right| \quad (22)$$

where  $a = I(\mathbf{x}, f)$  and  $b = I(\mathbf{x} + \mathcal{D}(\mathbf{x}), f + 1)$ . Such a term will provide photometric as well as a geometric preservation constraints where the parameter  $\gamma$  controls the balance between them. As earlier explained other functions can be used as well. We should not that no particular attention was paid to discontinuities or occlusions.

### 4.3. Optical Flow Estimation

We test our method on image sequences of a recently published database for optical flow [1]. The images are divided into three categories. The first category consists of challenging real data where the ground truth flow fields is achieved from the hidden fluorescent texture. The second category, contains realistic synthetic data generated using computer graphics. This part includes the famous Yosemite sequence. The last category is a set of modified stereo sequences. For detailed information about the data, we refer the reader to [1]. The database consists of two sets, one that is available with ground truth for calibration of the method and a second one where the ground truth is not available to the user. The overall performance is measured only on the second set of sequences. To determine the performance, two common error measures are considered: the angular error (AE) [7], which provides a relative measure of performance, and the endpoint error (EP) computing the residual length of the difference vectors between the estimated flow

Average Angular Error	Army			Mequon			Schefflera			Wooden			Grove			Urban			Yosemite			Teddy		
	all	disc	untext																					
<b>Dynamic MRF</b>	<b>4.95</b>	<b>15.0</b>	<b>4.20</b>	<b>3.71</b>	<b>15.3</b>	<b>2.58</b>	<b>6.77</b>	<b>17.8</b>	<b>3.26</b>	<b>4.50</b>	<b>23.7</b>	<b>2.62</b>	3.95	4.63	4.65	7.67	19.1	7.20	3.63	5.29	4.62	8.32	17.8	8.16
LP-Registration	7.36	16.8	6.30	3.94	13.8	3.00	7.33	17.8	4.43	5.54	24.5	3.57	4.24	4.56	5.63	10.5	21.4	9.57	4.54	5.48	3.95	8.15	17.9	7.82
Black & Anandan 2	7.83	18.7	6.41	9.70	21.9	8.60	13.7	23.7	18.1	10.9	30.0	9.44	4.43	5.23	4.94	7.95	18.2	6.51	2.61	4.44	2.15	8.58	14.3	8.54
2D-CLG	10.1	22.6	7.59	9.84	16.9	11.1	16.9	28.2	18.8	14.1	31.1	13.1	<b>3.66</b>	<b>4.25</b>	<b>4.41</b>	<b>6.69</b>	22.2	6.96	<b>1.76</b>	<b>3.14</b>	<b>1.46</b>	<b>6.29</b>	<b>12.9</b>	<b>5.81</b>
Horn & Schunck	8.01	19.9	8.38	9.13	23.2	7.71	14.2	25.9	14.6	12.4	30.6	11.3	4.44	5.27	4.59	8.25	25.8	8.77	4.01	5.41	1.95	9.16	17.5	8.86
Black & Anandan	8.93	18.5	9.99	12.9	22.4	13.3	15.8	25.9	18.3	13.2	31.8	12.0	5.69	6.35	7.77	9.37	18.8	9.02	3.10	4.88	3.96	13.4	18.3	15.1
Pyramid LK	13.9	20.9	21.4	24.1	23.1	30.2	20.9	29.5	21.9	22.2	34.6	25.0	18.7	22.9	19.9	21.9	26.2	23.5	6.41	7.02	10.8	25.6	31.5	34.5
MediaPlayerTM	18.3	30.8	15.0	17.7	29.2	17.4	19.9	32.7	21.6	26.3	45.9	25.9	7.23	6.95	10.2	19.4	32.9	19.3	12.7	18.7	17.2	17.4	22.9	20.7
Average Endpoint Error	all	disc	untext																					
<b>Dynamic MRF</b>	<b>0.13</b>	<b>0.41</b>	<b>0.11</b>	<b>0.26</b>	<b>1.00</b>	<b>0.18</b>	<b>0.52</b>	<b>1.22</b>	<b>0.29</b>	<b>0.24</b>	<b>1.29</b>	<b>0.15</b>	<b>1.08</b>	<b>1.46</b>	<b>1.25</b>	1.76	2.38	<b>0.92</b>	0.20	0.19	0.40	1.42	2.83	<b>1.58</b>
LP-Registration	0.20	0.46	0.17	0.29	0.93	0.21	0.55	1.22	0.35	0.29	1.32	0.20	1.12	1.48	1.37	2.20	2.52	1.18	0.20	0.19	0.32	<b>1.33</b>	2.79	<b>1.46</b>
2D-CLG	0.28	0.62	0.21	0.67	1.21	0.70	1.12	1.80	0.99	1.07	2.06	1.12	1.19	<b>1.44</b>	1.82	1.55	<b>2.31</b>	0.94	<b>0.10</b>	<b>0.11</b>	0.16	1.38	<b>2.26</b>	1.83
Black & Anandan 2	0.21	0.52	0.17	0.65	1.52	0.58	0.93	1.54	1.03	0.76	1.97	0.73	1.35	1.72	1.67	2.03	2.40	1.04	0.15	0.17	0.26	1.68	2.64	2.06
Horn & Schunck	0.22	0.55	0.22	0.61	1.53	0.52	1.01	1.73	0.80	0.78	2.02	0.77	1.22	1.50	1.73	<b>1.41</b>	2.82	0.98	0.16	0.18	<b>0.15</b>	1.51	2.50	1.88
Black & Anandan	0.24	0.52	0.27	0.95	1.59	1.00	1.14	1.68	1.11	0.94	2.09	0.92	1.53	1.88	2.03	2.21	2.47	1.33	0.20	0.18	0.41	2.45	3.43	3.09
MediaPlayerTM	0.48	0.83	0.39	1.30	2.12	1.26	1.33	2.00	1.25	1.66	3.12	1.65	1.73	1.93	2.45	2.84	3.38	2.02	0.59	0.58	1.08	2.57	3.28	3.20
Pyramid LK	0.39	0.61	0.61	1.67	1.78	2.00	1.50	1.97	1.38	1.57	2.39	1.78	2.91	3.65	3.01	3.29	2.99	2.42	0.30	0.24	0.73	3.80	5.08	4.88

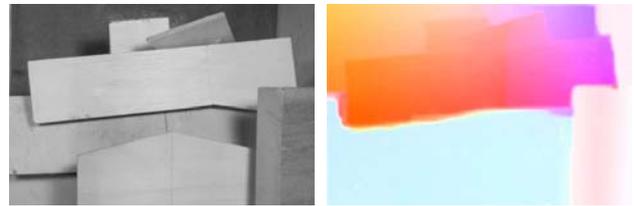
**Table 1:** Results for the angular error (AE) and the endpoint error (EP) for the whole frame (all), the discontinuity parts (disc), and the untextured regions (untext).

and the ground truth. Thanks to [1], we can also report a much more detailed evaluation including errors in discontinuity parts and untextured regions of the image as well as robustness statistics.

The following parameters were used for all experiments, determined towards optimizing the results for the sequences with ground truth available:

- We use a three-level pyramidal approach where the image resolution as well as the FFD resolution is successively increased from one level to another. We start with a 16 pixels control point spacing while reducing it to 8 and finally 4 pixels for the highest resolution. For the Army sequence this results for instance in FFD resolutions of  $38 \times 25$ ,  $74 \times 50$ , and  $147 \times 98$ .
- On each level we perform 5 cycles of the FastPD algorithm. The initial maximum displacement for each level is set to the half of the control point spacing and we perform 5 sampling steps from zero to this maximum in each direction. This results in a total number of  $(5 \cdot 2 + 1)^2 = 121$  labels (including the zero-displacement) for each of the 15 cycles.
- The weighting to balance between the matching criteria and our smoothness penalty is experimentally set to 0.3 for all examples while the weighting within the matching criteria to balance between the photometric (CC) and the geometric term (GIP) is set to 0.45.

We compare our method with the ones being used up to now of this specific database which includes the most closely related approach proposed in [10], called LP-Registration. Our novel approach of integrating the uncertainties where the scaling as well as the orientation of the labels sets is locally estimated after each cycle (see Figure 2) is here denoted as *Dynamic MRF*. Because of the limited space, we refer the reader to the website of the database



**Figure 3:** The 10th frame and our (color-encoded) result for the Wooden sequence.

(<http://vision.middlebury.edu/flow/eval>) to see the input images of all sequences used in our paper as well as the references to the other methods.

#### 4.4. Evaluation

We perform the evaluation of our algorithm on eight image sequences, again divided into the three categories. We use the Army sequence, the Mequon sequence, the Schefflera sequence, and the Wooden sequence from the hidden fluorescent texture category. Additionally, we estimate the optical flow on the Grove sequence, the Urban sequence, and the Yosemite sequence from the realistic synthetic category and on a modified stereo dataset called the Teddy sequence. In [1], the complexity and difficulty of the hidden fluorescent texture sequences is mentioned. Still, we achieve very good results for these datasets compared to other state of the art algorithms (including Bruhn *et al.* [5] and Black and Anandan [2]). The performance of the other algorithms is assessed by the maintainers of the database. For detailed information, again we refer to [1]. A part of our evaluation results are summarized in Table 1. Our method *Dynamic MRF* is in average ranked first on all statistical evaluations, including the robustness statistics RX [1]. The detailed evaluation can be found on <http://vision.middlebury.edu/flow/eval>. However, on the Yosemite sequence our results are not competitive to recently proposed methods [4, 18] which achieve an average AE of less than 1.0. Still, we should note that we did not

perform an explicit optimization on the parameters for specific data to minimize the AE as it is for instance mentioned in [18]. In stereo we could achieve only average results. Here, the estimation of the label set orientations might be a drawback. Slight orientation changes on the label sets are unwanted in stereo settings. The largest errors in all data sets occur in the discontinuity parts which seems to be natural when using FFD. Our color-encoded estimated flow fields for three of the eight datasets are shown in Figure 1 and 3. The estimation of the optical flow for an image pair takes about 45 seconds (Intel Mobile 2.16 GHz).

## 5. Discussion

In this paper we have proposed a novel approach to optical flow estimation. Our approach is based on a MRF, uses principles from linear programming to determine the optimal solution and dynamically determines the precision of solution using local uncertainties. It is flexible in the choice of the matching criteria since no differentiation of the cost function is needed, and provides quantitative and qualitative results. Experimental validation and comparisons with the state of the art methods demonstrate the potentials of our approach. Future work consists of addressing novel matching criteria. Non-uniform configurations of the control points together with different models of the influence function will be investigated. The hope is to create a motion model which is able to provide smooth results in untextured regions while at the same time it preserves discontinuities. Furthermore, we would like to investigate more efficient means to address the sub-optimality of the obtained solution. The use of priority-based belief propagation networks is an very challenging and promising research direction.

## References

- [1] S. Baker, D. Scharstein, J. Lewis, S. Roth, M. J. Black, and R. Szeliski. A database and evaluation methodology for optical flow. In *ICCV*, 2007.
- [2] M. J. Black and P. Anandan. The robust estimation of multiple motions: parametric and piecewise-smooth flow fields. *CVIU*, 63(1):75–104, 1996.
- [3] Y. Boykov, O. Veksler, and R. Zabih. Fast approximate energy minimization via graph cuts. *PAMI*, 23(11):1222–1239, Nov. 2001.
- [4] T. Brox, A. Bruhn, N. Papenberger, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *ECCV*, 2004.
- [5] A. Bruhn, J. Weickert, and C. Schnörr. Lucas/kanade meets horn/schunck: Combining local and global optic flow methods. *IJCV*, 61(3):211–231, 2005.
- [6] P. F. Felzenszwalb and D. P. Huttenlocher. Efficient belief propagation for early vision. *IJCV*, 70(1):41–54, 2006.
- [7] D. Fleet and A. Jepson. Computation of component image velocity from local phase information. *IJCV*, 5:77–104, 1990.
- [8] D. Fleet and Y. Weiss. *Handbook of Mathematical Models in Computer Vision*, chapter Optical Flow Estimation, pages 239–256. Springer, 2006.
- [9] S. Geman and D. Geman. Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. In *PAMI*, volume 6, pages 721–741, 1984.
- [10] B. Glocker, N. Komodakis, N. Paragios, G. Tziritas, and N. Navab. Inter and intra-modal deformable registration: Continuous deformations meet efficient optimal linear programming. In *IPMI*, 2007.
- [11] F. Heitz and P. Bouthemy. Multimodal estimation of discontinuous optical flow using markov random fields. In *PAMI*, 1993.
- [12] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 17:185–204, 1981.
- [13] P. Kohli and P. Torr. Efficiently solving dynamic markov random fields using graph cuts. In *ICCV*, 2005.
- [14] P. Kohli and P. H. Torr. Measuring uncertainty in graph cut solutions: Efficiently computing min-marginal energies using dynamic graph cuts. In *ECCV*, 2006.
- [15] V. Kolmogorov. Convergent tree-reweighted message passing for energy minimization. *PAMI*, 28(10):1568–1583, October 2006.
- [16] V. Kolmogorov and R. Zabih. What energy functions can be minimized via graph cuts? *PAMI*, 26(2):147–159, Feb 2004.
- [17] N. Komodakis, G. Tziritas, and N. Paragios. Fast, approximately optimal solutions for single and dynamic mrf. In *CVPR*, 2007.
- [18] T. Nir, A. M. Bruckstein, and R. Kimmel. Over-parameterized variational optical flow. *IJCV*, 2007.
- [19] X. Pennec and J.-P. Thirion. A framework for uncertainty and validation of 3-d registration methods based on points and frames. In *IJCV*, 1997.
- [20] K. Rohr, M. Fornefett, and H. S. Stiehl. Approximating thin-plate splines for elastic registration: Integration of landmark errors and orientation attributes. In *IPMI*, 1999.
- [21] D. Rueckert, L. Sonoda, C. Hayes, D. Hill, M. Leach, and D. Hawkes. Nonrigid registration using free-form deformations: application to breast mr images. *TMI*, 18(8):712–721, 1999.
- [22] E. Simoncelli. Bayesian multi-scale differential optical flow. In *Handbook of Computer Vision and Applications*, 1999.
- [23] C. Stewart, C.-L. Tsai, and B. Roysam. The dual bootstrap icp algorithm with application to retinal image registration. In *TMI*, 2003.
- [24] M. Tappen and W. Freeman. Comparison of graph cuts with belief propagation for stereo, using identical mrf parameters. In *ICCV*, 2003.
- [25] M. Taron, N. Paragios, and M.-P. Jolly. Modelling shapes with uncertainties: Higher order polynomials, variable bandwidth kernels and non parametric density estimation. In *ICCV*, 2005.
- [26] A. Tikhonov. Ill-posed problems in natural sciences, 1992. Coronet.
- [27] O. Veksler. *Efficient graph-based energy minimization methods in computer vision*. PhD thesis, Cornell University, 1999.