Jointly Optical Flow and Occlusion Estimation for Images with Large Displacements

Vanel Lazcano[1], Luis Garrido[2] and Coloma Ballester[3]
[1]*NMFE, Universidad Mayor, Avda. Manuel Montt 318, Santiago, Chile*
[2]*DMI, Universitat de Barcelona, Gran Via 585, Barcelona, Spain*
[3]*DTIC, Universitat Pompeu Fabra, Roc Boronat 138, Barcelona, Spain*

Keywords: Optical Flow, Exhaustive Search, Large Displacements, Illumination Changes.

Abstract: This paper deals with motion estimation of objects in a video sequence. This problem is known as optical flow estimation. Traditional models to estimate it fail in presence of occlusions and non-uniform illumination. To tackle these problems we propose a variational model to jointly estimate optical flow and occlusions. The proposed model is able to deal with the usual drawback of variational methods in dealing with large displacements of objects in the scene which are larger than the object itself. The addition of a term that balances gradient and intensities increases the robustness to illumination changes of the proposed model. The inclusion of a supplementary matching obtained by exhaustive search in specific locations helps to follow large displacements.

1 INTRODUCTION

The apparent motion of pixels in a sequence of images is usually called the optical flow. Optical flow computation is one of the most challenging problems in computer vision, especially in real scenarios where occlusions and illumination changes occur. Optical flow has many applications, including autonomous flight of vehicles, insertion of objects on video, video compression and many others. In order to estimate this flow field an energy model is stated, which computes the estimation error of the optical flown. Most of the optical flow methods are grounded on the optical flow constraint. This constraint is based on the *brightness constancy assumption* which states that the brightness or intensity of pixels in the image remains constant from frame to frame along the movement of objects. The optical flow constraint is only suitable when the motion field is small enough or images are very smooth.

Solving the intensity constraint is an ill-posed problem which is usually solved by adding a regularity prior. Then the regularity prior or regularization term added to the energy model allows defining the structure of the motion field and ensures that the optical flow computation is well posed.

In (Horn and Schunck, 1981) was proposed to add to the energy model a quadratic regularization term. Actually, the work of (Horn and Schunck, 1981) was the first one which introduced variational methods to compute dense optical flow. However, the Horn-Schunck model does not cope well with motion discontinuities, is highly sensible to noise in the images. To tackle those drawbacks other regularization terms have been proposed, (Nagel and Ekelman, 1986; Black and Ananda, 1996; Brox et al., 2004; Zach et al., 2007; Werlberger et al., 2009; Sun et al., 2010; Werlberger et al., 2010; Krähenbühl and Koltun., 2012; Xu et al., 2012; Chen et al., 2013; Sánchez et al., 2014; Zimmer et al., 2011; Strekalovskiy et al., 2014; Palomares et al., 2015; Ranftl et al., 2014; Sun et al., 2014). In order to cope with large displacements, optimization typically proceeds in a coarse-to-fine manner (also called a multi-scale strategy).

Optical flow estimation using models based on classical variational models fails if the sequence presents: i) occluded pixels, ii) displacements larger than the size of the objects and iii) changes of illumination. Occlusions produce lack of correspondence between some points in the image sequence. Occluded pixels include pixels of an image frame which are covered by the movement of objects in the following frame. For those occluded pixels, there is no a reliable optical flow. In particular, the brightness constancy assumption is flawed in realistic scenarios, where occlusions occur due to the relative motion between objects in

Lazcano, V., Garrido, L. and Ballester, C.
Jointly Optical Flow and Occlusion Estimation for Images with Large Displacements.
DOI: 10.5220/0006716305880595
In *Proceedings of the 13th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2018) - Volume 5: VISAPP*, pages 588-595
ISBN: 978-989-758-290-5
Copyright © 2018 by SCITEPRESS – Science and Technology Publications, Lda. All rights reserved

the scene or the camera movement, as well as illumination changes. Indeed, shadows or light reflections that appear and move in the image sequence can also make the brightness constancy assumption to fail.

These facts motivate us to consider an alternative to the classical brightness constancy constraint, also consider occlusion estimation and a new term to cope with large displacements. In this paper we extend the model in (Ballester et al., 2012) to a model which is robust to illumination changes and is able to handle large displacements.

2 RELATED WORKS

In (Zach et al., 2007) the authors present an approach to estimate the optical flow that preserves discontinuities and it is robust to noise. In order to compute the optical flow $\mathbf{u} = (u_1, u_2) : \Omega \to \mathbb{R}^2$ between I_0 and I_1, the authors propose to minimize the energy

$$E(u) = \int_\Omega (\lambda |I_0(x) - I_1(x+u)| + |\nabla u_1| + |\nabla u_2|)\, dx, \tag{1}$$

with a relative weight given by the parameter $\lambda > 0$. This variational model is usually called the *TV-L1 formulation*.

Occlusion is a challenging problem in the estimation of the optical flow. Some methods implicitly deal with occlusion by using robust norms terms in the data term while others do an explicit occlusion handling. A first step towards taking into account occlusions was done by jointly estimating forward and backwards optical flow in (Alvarez et al., 2007). Authors argue that at non-occluded pixels forward and backward flows are symmetric. This idea was taken by the authors in (Ince and Konrad., 2008) and they proposed to extrapolate optical flow in occluded areas. In (Xu et al., 2012) a method to estimate occlusions is used. They consider the fact that multiple points mapped by the optical flow to the same point in the following frame (collision) are likely to be occluded.

On the other hand, robustness against illumination changes would be desirable. The gradient of the image is robust to additive illumination changes in images (Brox et al., 2004), and therefore the gradient constancy assumption:

$$\nabla I_0(x) - \nabla I_1(x+u(x)) = 0,$$

may well be included as a new data term in a variational energy in order to compute the optical flow u (Xu et al., 2012).

While traditional methodology works well in cases where small structures move more or less the same way as larger scale structures, the approach fails with large displacement. In recent years this topic has been tackled in interesting approaches. In (Brox et al., 2009), a method for large displacements is proposed that performs region-based descriptor matching. This method estimates correctly large displacement but it can match outliers. (Steinbruecker and Pock, 2009) also proposes a method in order to tackle large displacement. The methodology performs well in real images with large displacements but it presents a lack of subpixel accuracy.

To tackle large displacement (Xu et al., 2012) incorporates matching of SIFT features computed between images of the sequence. The fusion between matching of SIFT features and optical flow estimation is performed using graph cuts.

Recently new models have been proposed in order to handle large displacements in (Weinzaepfel et al., 2013), (Timofte and Van Gool, 2015) , (Kennedy and Taylor, 2015), (Fortun et al., 2016) and (Palomares et al., 2017). These models consider sparse or dense matching using Deep matching algorithm (Weinzaepfel et al., 2013) or motion candidates (Fortun et al., 2016). The principal idea is to give "hints" to the variational optical flow approach by using these sparse matching (Weinzaepfel et al., 2013). In (Kennedy and Taylor, 2015) and (Fortun et al., 2016) the occlusion layer is also estimated.

3 PROPOSED MODEL

We propose a variational model for joint optical flow and occlusion estimation, which considers color image sequences and is able to handle illumination changes as well as large displacements. The ingredients are detailed in the following sections.

3.1 Occlusion Estimation

Inspired by (1) and (Ballester et al., 2012) we present a joint optical flow and occlusions estimation model. The divergence of the motion field can be used to distinguish between different types of motion areas: the divergence of a flow field is negative for occluded areas, positive for dis-occluded, and near zero for the matched areas.

Our model considers three consecutive color frames $I_{-1}, I_0, I_1 : \Omega \to \mathbb{R}^3$ as (Ballester et al., 2012), which we assume to have values in the RBG color space, hence each frame I_i has three color components I_i^1, I_i^2, I_i^3, associated to the red, green and blue channels, respectively. In order to compute the optical flow between I_0, I_1 let $\chi : \Omega \to [0,1]$ be the function modeling

the occlusion mask, so that $\chi = 1$ identifies the occluded pixels, i.e. pixels that are visible in I_0 but not in I_1. Our model is based on the assumptions: (i) pixels that are not visible in frame I_1 are visible in the previous frame of I_0 (let I_{-1} be that frame), (ii) the occluded region given by $\chi = 1$ should be correlated with the region where $\mathrm{div}(u)$ is negative, and (iii) motion of the occluded background area is not fast as the one of the occluding foreground (Ballester et al., 2012). Thus, we propose to compute the optical flow and the occlusion mask by minimizing the energy:

$$E^c(u,\chi) = E^c_d(u,\chi) + E_r(u,\chi) + \frac{\eta}{2}\int_\Omega \chi|u|^2\,dx + \beta\int_\Omega \chi\,\mathrm{div}(u)\,dx,$$

where $E^c_d(u.\chi)$ and $E_r(u,\chi)$ are given by

$$E^c_d(u,\chi) = \lambda\sum_{k=1}^{3}\int_\Omega ((1-\chi)|I^k_0(x) - I^k_1(x+u(x))| + \lambda\sum_{k=1}^{3}\int_\Omega \chi|I^k_0(x) - I^k_{-1}(x-u(x))|)dx.$$

$$E_r(u,\chi) = \int_\Omega g(x)(|\nabla u_1| + |\nabla u_2| + |\nabla\chi|)dx, \quad (2)$$

with $\eta \geq 0$, $\beta > 0$ and $g(x) = \frac{1}{1+\gamma|\nabla I_0(x)|}$, $x \in \Omega$, $\gamma > 0$. Notice that, if $\chi(x) = 0$, then we compare $I^k_0(x)$ and $I^k_1(x+u(x))$. If $\chi(x) = 1$, we compare $I^k_0(x)$ and $I^k_{-1}(x-u(x))$.

3.2 Robustness to Color Changes

The color constancy assumption is frequently violated due to illumination changes, shadows or reflections. A combination of the color constancy assumption and the gradient constancy assumption in the data term seems to be a valuable approach to alleviate this problem (Xu et al., 2012). We extend our color model to consider a combination of intensities and gradients by introducing an adaptive weight map $\alpha : \Omega \to [0,1]$ that allows to balance in an adaptive way the contribution of color and gradient constraints at each point in the image domain (Xu et al., 2012). We propose the following model:

$$E^c_\alpha(u,\chi) = E^c_{d,\alpha}(u,\chi) + E_r(u,\chi) + \frac{\eta}{2}\int_\Omega \chi|u|^2dx + \beta\int_\Omega \chi div(u)dx,$$

where $E^c_{d,\alpha}(u,\chi)$ can be written as:

$$E^c_{d,\alpha}(u,\chi) = \int_\Omega \alpha(x)D_{I,\chi}(u,\chi,x) + \int_\Omega (1-\alpha(x))D_{\nabla I,\chi}(u,\chi,x)dx, \quad (3)$$

and $D_{I,\chi}(u,\chi,x)$ and $D_{\nabla I,\chi}(u,\chi,x)$ are point-wise data costs based on the comparison of color and gradient of the image, respectively. Roughly speaking, $D_{I,\chi}$ contains the comparison $\|I^k_0(x) - I^k_1(x+u)\|$ and $D_{\nabla I,\chi}$ the comparison $\tau\|\nabla I_0(x) - \nabla I_1(x+u)\|$, with $\tau > 0$. Then, the weight map $\alpha(x)$ is defined in (Xu et al., 2012) as

$$\alpha(x) = \frac{1}{1 + e^{\tilde{\beta}(D_{I,\chi}(u,x) - D_{\nabla I,\chi}(u,x))}}, \quad (4)$$

where $\tilde{\beta}$ is a positive constant. Let us comment about the behavior of (4). If the term $D_{I,\chi}(u,x) \gg D_{\nabla I,\chi}(u,x)$, the difference $D_{I,\chi}(u,x) - D_{\nabla I,\chi}(u,x)$ will be positive and the exponential value $e^{\tilde{\beta}(D_{I,\chi}(u,x) - D_{\nabla I,\chi}(u,x))}$ will be large. Then, $\alpha(x)$ will be a small value, say near 0, and the data term will have more confidence on the gradient constancy assumption. On the other hand, if $D_{\nabla I,\chi}(u,x) \gg D_{I,\chi}(u,x)$, the difference $D_{I,\chi}(u,x) - D_{\nabla I,\chi}(u,x)$ will be negative and the exponential value $e^{\tilde{\beta}(D_{I,\chi}(u,x) - D_{\nabla I,\chi}(u,x))}$ will be very small. In other words, the data term will be more confident on the color constancy assumption.

3.3 Large Displacements

To handle large displacements we add to our model a term $\mu\int_\Omega \chi_p c(x)\,|u - u_e|$, where u_e is an optical flow obtained by exhaustive search, χ_p is a characteristic function indicating location where supplementary matching could improve the motion estimation, $c(x)$ is a confidence on the exhaustive matching at x and $\mu > 0$.

Summarising, the proposed model to handle large displacement is:

$$E^c_{\alpha l}(u,\chi) = E^c_{d,\alpha}(u,\chi) + E_r(u,\chi) + \frac{\eta}{2}\int_\Omega \chi|u|^2dx + \beta\int_\Omega \chi div(u)dx + \mu\int_\Omega \chi_p c(x)\,|u - u_e|\,dx, \quad (5)$$

In our implementation and for efficiency reasons, we consider an upper bound for the expected maximum displacement v_{max}.

3.3.1 Confidence Function $c(x)$

We directly integrate exhaustive point correspondences into the variational model and the proposed confidence measure, used to determine the weight given to matching computed by exhaustive search, is

$$c(x) = \left(\frac{d_2 - d_1}{d_1}\right)^2 \left(\frac{E^c_{d\alpha}(u,x)}{E_{exha}(u_e,x)}\right)^2$$

where d_1, d_2 are the distances to the first and second best candidate respectively of the exhaustive search, $E^c_{d\alpha}(u,x)$ is the error defined in (3) and $E_{exha}(u_e,x)$ is the error of the exhaustive search. This measure was used in (Stoll et al., 2012) to validate the correctness of a given optical flow field at each point.

3.3.2 Construction of χ_p

In order to determine specific locations where supplementary matching could improve the motion estimation, we evaluate the data term, at each $x \in \Omega$ with the computed flow u and the occlusion map χ. The idea is that if the value $E^c_{d,\alpha}(u,\chi)$ is large, then the estimation might be improved. Additionally, we consider the smaller eigenvalue $\lambda(x)$ of the structure tensor associated to the image I_0. With these ingredients, the set Ω_{χ_p} where supplementary matching could improve motion estimation is defined as:

$$\Omega_{\chi_p} = \{x \in \Omega \mid E^c_{d\alpha}(u,\chi)(x) > \theta_E \wedge \lambda(x) > \theta_\lambda\}$$

where θ_E and θ_λ are given constants which we will determine empirically and fix for the experiments we performed. That is, if $E^c_{d,\alpha} > \theta_E$, then we assume that the error is large enough to be improved using a supplementary match. The set of points that belong to Ω_{χ_p} define a binary mask, which we denote by $\chi_p : \Omega \to [0,1]$.

3.4 Solving the Model

In order to minimize (5), we relax it and introduce five auxiliary variables v_1, v_2, v_3, v_4, v_5 representing the flow and used to decouple the nonlinear terms, where v_1, v_2, v_3 correspond to the red, green and blue channels, respectively and v_4, v_5 correspond to ∂_x and ∂_y respectively. We penalize the difference between the optical flow u and each of the auxiliary variables v_1, v_2, v_3, v_4, v_5. Thus, to compute the occlusions and the optical flow between I_0, I_1, we propose to minimize the following energy:

$$\widetilde{E}^c_{\alpha,l}(u,\chi,\tilde{v}) = E^c_d(\tilde{v},\chi) + E_r(u,\chi) + \frac{\eta}{2}\int_\Omega \chi\,|\tilde{v}|^2\,dx + \beta\int_\Omega \chi\, div(u)\,dx + \frac{1}{2\theta}\int_\Omega |\tilde{u}-\tilde{v}|^2\,dx, \tag{6}$$

where $|\tilde{v}|^2$ stands for $\sum_{k=1}^{5} |v_k|^2$ and

$$E^c_d(\tilde{v},\chi) = \lambda\int_\Omega (1-\chi)\sum_{k=1}^{3}\left|\rho^k_1(v_k)\right|dx + \lambda\int_\Omega \chi\sum_{k=1}^{3}\left|\rho^k_{-1}(v_k)\right|dx,$$

and ρ^k_i is the linearized version of $I^k_0(x) - I^k_i(x+\varepsilon_i v_k)$ around an approximation u_0 of u, with $i = -1, 1$ and $\varepsilon_{-1} = -1$ and $\varepsilon_1 = 1$, and $k = 1,2,3$ (corresponding to each color channel). The linearization procedure is applied to each $\rho^k_i(x)$.

We minimize $\widetilde{E}^c_{\alpha l}$ in (6) by alternating among the minimization with respect to each variable while keeping the remaining fixed as (Ballester et al., 2012),(Zach et al., 2007). In particular, the minimization of $\widetilde{E}^c_{\alpha l}$ with respect to u, v_k and χ is described in the following propositions.

Proposition 1. *The minimum of $\widetilde{E}^c_{\alpha l}$ with respect to $\mathbf{u} = (u_1, u_2)$ is given by*

$$u_i = \frac{\frac{1}{5}\sum_{k=1}^{5} v^i_k + \theta\,\mathrm{div}(g\xi_i) + \theta\beta\frac{\partial\chi}{\partial x_i} + \mu\theta u^i_e \chi_p c}{1 + \mu\theta\chi_p c}, \tag{7}$$

with i=1,2 and $\mathbf{u_e} = (u^1_e, u^2_e)$. ξ_1 and ξ_2 are computed using the following iterative scheme

$$\xi^{t+1}_i = \frac{\xi^t_i + \frac{\tau_u}{\theta} g\nabla(\frac{1}{5}\sum_{k=1}^{5} v^i_k + \theta\,\mathrm{div}(g\xi^t_i) + \theta\beta\frac{\partial\chi}{\partial x_i})}{1 + \frac{\tau_u}{\theta}\left|g\nabla(\frac{1}{5}\sum_{k=1}^{5} v^i_k + \theta\,\mathrm{div}(g\xi^t_i) + \theta\beta\frac{\partial\chi}{\partial x_i})\right|}, \tag{8}$$

where $\xi^0_i = 0$ and $\tau_u \le 1/8$.

Proposition 2. *Assume that $\chi : \Omega \to \{0,1\}$. The minimum of $\tilde{E}^c_{\alpha,l}$ with respect to $\mathbf{v_k} = (v^1_k, v^2_k)$ is*

$$\mathbf{v_k} = \begin{cases} \eta_i\mathbf{u} - \mu_i\varepsilon_i\alpha(x)\nabla I^k_i(\mathbf{x}^*) & \text{if } \Lambda^k_i(\mathbf{u}) > \mu_i\alpha(x)m^k_i \\ \eta_i\mathbf{u} + \mu_i\varepsilon_i\alpha(x)\nabla I^k_i(\mathbf{x}^*) & \text{if } \Lambda^k_i(\mathbf{u}) < -\mu_i\alpha(x)m^k_i \\ \mathbf{u} - \varepsilon_i\rho^k_i(\mathbf{u})\frac{\nabla I^k_i(\mathbf{x}^*)}{|\nabla I^k_i(\mathbf{x}^*)|^2} & \text{if } |\Lambda^k_i(\mathbf{u})| \le \mu_i\alpha(x)m^k_i, \end{cases}$$

when $i = 1$ and $\varepsilon_1 = 1$, $\eta_1 = 1$, $\mu_1 = \lambda\theta$, $\Lambda^k_1(\mathbf{u}) = \rho^k_1(\mathbf{u})$ when $\chi = 0$, and $i = -1$, $\varepsilon_{-1} = -1$, $\eta_{-1} = \frac{1}{1+\eta\theta}$, $\mu_{-1} = \frac{\lambda\theta}{1+\eta\theta}$, $\Lambda_{-1}(\mathbf{u}) = \rho^k_{-1}(\mathbf{u}) + \frac{\eta\theta}{1+\eta\theta}\mathbf{u}\cdot\nabla I^k_{-1}(\mathbf{x}+\varepsilon_i\mathbf{u}_0)$ when $\chi = 1$. Additionally we create $\mathbf{x}^* = \mathbf{x} + \varepsilon_i\mathbf{u}_0$. The term m^k_i was defined as $m^k_i := |\nabla I^k_i(\mathbf{x}^*)|^2$. Arguments $\mathbf{x}$ in $\mathbf{u}, \mathbf{u}_0$ are omitted.

Once all v_k are computed, we define $F = \lambda A + \frac{1}{5}B$, where

$$A = \left(-\alpha(x)\sum_{k=1}^{3}\left|\rho^k_1\right| - (1-\alpha(x))\sum_{k=4}^{5}|\rho_1(v_k)|\right),$$

$$B = \left(\sum_{k=1}^{5}(v_k - u)^2\right),$$

and $G = \lambda C + \frac{\eta}{5}D$, where,

$$C = \left(-\alpha(x)\sum_{k=1}^{3}\left|\rho^k_{-1}\right| - (1-\alpha(x))\sum_{k=4}^{5}|\rho_{-1}(v_k)|\right),$$

$$D = \left(\sum_{k=1}^{5}(v_k)^2\right).$$

Proposition 3. *Let* $0 < \tau_\psi \tau_\chi < 1/8$. *Given* $\mathbf{u}, \mathbf{v}$, *the minimum* $\tilde{\chi}$ *of* $\tilde{E}^c_\alpha$ *with respect to* χ *can be obtained by the following primal-dual algorithm*

$$\psi^{n+1} = P_B(\psi^n + \tau_\psi\, g\, \nabla\chi^n)$$
$$\chi^{n+1} = P_{[0,1]}\left(\chi^n + \tau_\chi\left(\mathrm{div}(g\psi^{n+1}) - \beta\,\mathrm{div}\mathbf{u} - F - G\right)\right),$$

where $P_B(\psi)$ *denotes the projection of* ψ *on the unit ball of* $\mathbb{R}^2$ *and* $P_{[0,1]}(r) = \max(\min(r,1),0)$, $r \in \mathbb{R}$.

3.5 Algorithm

This section is devoted to present the numerical algorithm for the minimization of (6), including pseudocodes describing its main steps. In particular, Algorithm 1 summarizes our illumination changes and large displacement robust optical flow model presented in section 3.4. The value of $\alpha(x)$, for all $x \in \Omega$ are updated after each propagation of the optical flow to the finer scale, before starting the estimation of the flow field at that scale.

The data attachment $\int c(x)\chi_p(u-u_e)^2$ depends on the confidence value $c(x)$, the mask χ_p and exhaustive matchings u_e. The confidence value is an estimation of the reliability of the exhaustive matchings.

4 DATABASE AND EXPERIMENTS

We evaluate our model in two publicly databases: Middlebury (Scharstein and Szeliski, 2002) and MPI Sintel (Butler et al., 2012). In Figure 1, we show images of the Middlebury dataset. These sequences contain displacements larger than the size of the object and also contain shadows and reflections. Figure 1 shows three consecutive frames of the sequence Beanbags(BB) and DogDance(DD). BB sequence presents balls that move while producing shadows on the T-shirt. In DD sequence the girl moves to the right and the dog moves to the left.

MPI database (Butler et al., 2012) presents long synthetic sequences containing large displacements, blur or reflections, fog and shadows. Moreover, there are two versions of the MPI database: clean and final. The final version is claimed to be more challenging and we take it for our evaluation. Figure 2 displays some examples of the MPI database. There are images with large displacements. In the cave_4 sequence a girl fight with a dragon moving her lance inside a cave in (a), (b), (c). In (d), (e) and (f) the girl moves downward a fruit on her hand.

Input : Three consecutive color frames I_{-1}, I_0, I_1 and u_e
Output: Flow field $\mathbf{u}$ and occlusion layer χ for I_0, and $\alpha(x)$

Compute down-scaled images I^s_{-1}, I^s_0, I^s_1 for $s = 1, \ldots, N_{\text{scales}}$;
Initialize $\mathbf{u}^{N_{\text{scales}}} = \mathbf{v}_k^{N_{\text{scales}}} = 0$, and $\chi^{N_{\text{scales}}} = 0$, $\alpha^{N_{\text{scales}}}(x) = 1.0$, $\gamma = 0$;
for $s \leftarrow N_{\text{scales}}$ **to** 1 **do**
  Compute $\alpha^s(x)$ using (4);
  for $w \leftarrow 1$ **to** N_{warps} **do**
    Compute $I^s_i(\mathbf{x} + \varepsilon_i \mathbf{u}_0(\mathbf{x}))$, $\nabla I^s_i(\mathbf{x} + \varepsilon_i \mathbf{u}_0(\mathbf{x}))$, and ρ_i, $i = -1, 1$;
    $n \leftarrow 0$;
    while $n <$ outer_iterations **do**
      Compute $\mathbf{v_k}^s$ using Proposition 2;
      for $l \leftarrow 1$ **to** inner_iterations_u **do**
        Solve for $\xi_i^{l+1,s}$, $i \in \{1,2\}$, using the fixed point iteration (Proposition 1);
      end
      Compute $\mathbf{u}^s$ using Proposition 1 considering data attachment $\mu \int c(x)\chi_p(u-u_e)$;
      for $m \leftarrow 1$ **to** inner_iterations_χ **do**
        Solve for χ^{m+1} using the primal-dual algorithm (Proposition 3);
      end
    end
  end
  Compute $E^c_{d\alpha}(x)$, $\lambda(x)$;
  Compute $\chi_p(E^c_{d\alpha}(x), \lambda(x), \theta_\lambda, \theta_E)$ implies Ω_{χ_p};
  If $s > 1$ **then** scale-up $\mathbf{u}^s, \mathbf{v}^s, \chi^s$ to $\mathbf{u}^{s-1}, \mathbf{v}^{s-1}, \chi^{s-1}$;
end
$u = u^1$ and $\chi = T_\mu(\chi^1)$

Algorithm 1: Algorithm for illumination changes and large displacement robust optical flow.

5 RESULTS

For all experiments parameters are fixed to: $\theta = 0.40$, $\lambda = 0.60$, $\alpha = 0.0$, $\beta = 1.0$, $\theta_{\lambda 1} = 0.98$ and $\theta_E = 0.98$. The μ decreased its value in each iteration with initial value $\mu_o = 300$ and $\mu_n = (0.6)^n \mu_0$ in the following iterations. For real images we use blocks of 7×7 pixels and for synthetic images we use blocks of 31×31 pixels.

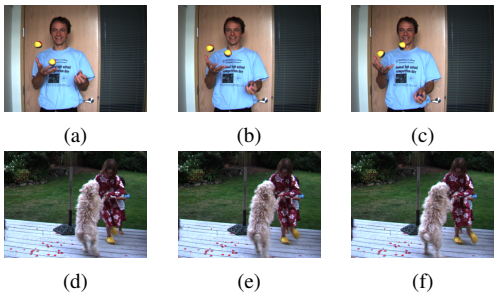

Figure 1: Middlebury BB video containing large displacements, illumination changes, shadows that moves in scene. (a) frame9, (b) frame10, (c) frame11 of the BB sequence. (d) frame09, (e) frame10 and (f) frame11 of the DD sequence.

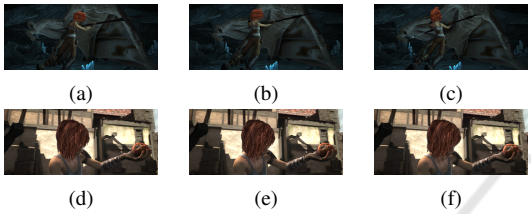

Figure 2: Images of the MPI database. (a) frame13, (b) frame14 and (c) frame15 of cave_4 sequence. (d) frame19, (e) frame20 and (f) frame21 of Alley_4 sequence.

Figure 3 presents the obtained results: (a) color coded estimated optical flow for BB, (b) occluded regions (let us observe how they are correctly estimated, in particular on the face of the man), (c) χ_p for BB (d) the compensated image. (e) color coding scheme. (f) encoded optical flow for DD, (g) the estimated occlusion (notice that the occlusion appears in the right side of the girl and in the left side of the dog), (h) χ_p for DD.

We have divided MPI database in three subsets: large, medium and small displacements. The quantitative obtained results are shown in Table 1. For large displacements we set the parameter $v_{max} = 150$, for medium displacements we set $v_{max} = 40$ and for small displacement we set $v_{max} = 1$. For large displacement we set $\theta_\lambda = 0.50$ and $\theta_e = 0.50$, for medium and small displacement we set $\theta_\lambda = 0.98$ and $\theta_e = 0.98$. The Average End Point Error for the whole database is presented in Table 1.

Table 1: End Point Error obtained by our model in subset: large displacement, medium displacement and small displacement of MPI.

Large		Medium		Small	
EPE	18.82	EPE	1.41	EPE	0.80
Total	Average EPE		7.17		

Let us observe from Table 1 that although the obtained average is $EPE = 18.82$ in Large Displacement videos, the average EPE in all sequence drops to 7.17. If we only consider frames that contains displacements less than 150 pixels the error drops to 8.82 in Table 2. We also show in Table 2 results obtained by DeepFlow in these subsets (Weinzaepfel et al., 2013) and MDPOF (Xu et al., 2012).

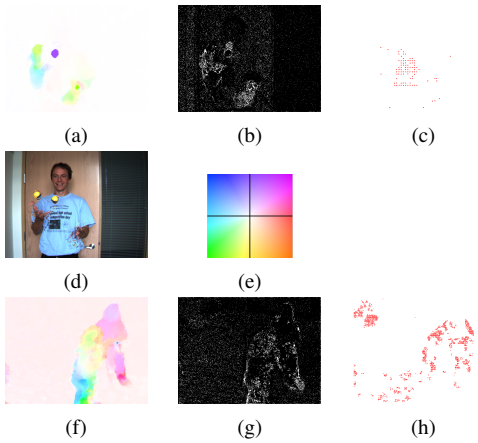

Figure 3: Results obtained in BB and DD sequence. (a) color coded flow field obtained by our model. (b) estimated occlusion mask. (c) χ_p. (d) Compensated image using the occlusion mask, Compensated $= (1-\chi)I_1(x+u(x)) + \chi I_1(x-u(x))$. (e) color code for flow field. (f) color code for DogDance sequence. (g) estimated occlusion mask. (h) χ_p

Table 2: End Point Error obtained by our model in subset of MPI considering displacement < 150 pixels.

Large Displacement	Our model	DeepFlow	MDP-Flow2
Average EPE	8.82	10.61	9.12

In Figure 5 and Figure 4 we show qualitative results obtained for MPI data base.

In Figure 4 we have computed the optical flow between frame 27 and frame 28 of the sequence ambush_7, by considering three frames: frame 26 which is considered to be I_{-1} in our energy model, frame 27 which is I_0 and frame 27 is I_1. Results are shown in Figure 4. Original frames 26, 27 and 28 correspond to subfigures (a), (c) and (e), respectively. This sequence presents small displacement but there is a

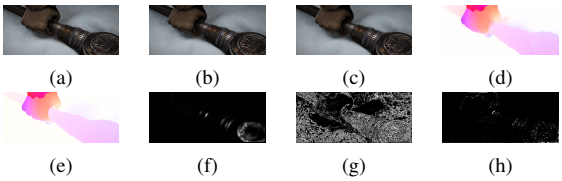

Figure 4: Results obtained by our method in ambush_7 video sequence. (a), (b) and (c): frame 26, 27 and 28, respectively. (b) Optical flow ground truth. (f) Estimated optical flow. (f) $\lambda(x)$. (g) $\alpha(x)$. (h) $\chi(x)$.

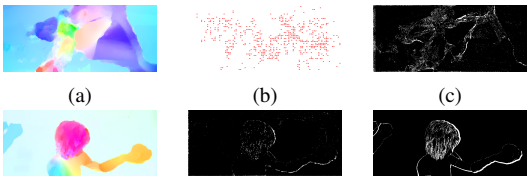

Figure 5: Images of the MPI video database. (a) color coded optical flow , (b) χ_p and (c) estimated occlusions cave4 sequence. (d) color coded optical flow, (e) estimated occlusion and (f) ground truth occlusion.

shadow that moves. The lance in the image presents small texture. This texture in the sequence presents small variations (like noise) frame to frame. In (d) we show the estimated optical flow. We observe that the optical flow was robustly estimated on the snow (where the shadow moves). In (f) we show the minimum eigenvalue of the structure tensor of the frame 27. We observe that the structure lay on the small details of the lance. In (g) we show the adaptive balance term $\alpha(x)$. This show that on the snow region the color constancy constrains does not holds and gradients should be used ($\alpha(x) = 0$). On the other hand where $\alpha(x) = 1.0$ intensity should be used. In (h) we have estimated occlusions on the texture of the lace due to this small variation frame to frame.

In Figure 5 our estimated optical flow is displayed in (a). In (b) we show χ_p indicating the positions where the exhaustive search is incorporated. (c) presents the estimated occlusion layer. (d) color coded optical flow for Alley1 sequence. (e) Estimated occlusion layer. (f) ground truth occlusion layer. Comparing (e) and (f) we see that they are very similar.

Figure 6 shows a comparison on MPI Sintel database. These results are available in the Sintel website (Butler et al., 2012). Our proposal is denoted as **OF_OCC_LD**. Notice that for s0-10 our model is ranked 20 (in brakets) of 110 reported method in the MPI site. For s10-40 our method is ranked 41. Finally for EPEall our method is ranked 97 outperforming TV-L1 which is ranked 102.

5.1 Critical Discussion

MPI test set includes small, medium and large displacements (approx. 400 pixels). For small and medium displacements, our method is ranked 20 and 41, respectively. For large displacements, the position drops to 97 which may well be due to the fact that, for efficiency reasons, in our experiments the large displacement threshold v_{max} (which should be at least 400) was set to 150.

	EPE all	s0-10	s10-40
DeepDiscreteFlow [8]	5.728	0.959	3.072
OF_OCC_LD [20]	10.024	1.064	4.017
SparseFlow [21]	7.851	1.071	3.771
MDP-Flow2 [71]	8.445	1.420	5.449
FALDOI [75]	7.337	1.487	4.355
TV-L1 [79]	10.462	1.551	6.012

Figure 6: Comparative results obtained by our method in MPI test set. EPEall is Endpoint error over the complete frames, s0-10 error over regions with displacements lower than 10 pixels, s0-40 error over regions with displacements between 10 and 40 pixels.

6 CONCLUSIONS

We proposed a variational model to jointly estimate the optical flow and the occlusion layer incorporating the occlusion information in its energy based on the divergence of the flow. The optical flow on visible pixels is forward estimated and while it is backwards estimated on occlude pixels, from three consecutive frames. The proposed robust model handles illuminations changes using a balance term between gradients and intensities improving the performance of the optical flow estimation in scenarios with illumination changes. Thanks to the use of supplementary matches the model is able to capture large displacements, even of small objects. As future work we plan to accelerate the exhaustive matching computation in order to avoid the v_{max} restriction and handle arbitrary largest displacements. It also includes a parallel computation and GPU implementation.

REFERENCES

Alvarez, L., Deriche, R., Papadopoulo, T., and Sanchez, J. (2007). Symmetrical dense optical flow estimation with occlusions detection. *International Journal of Computer Vision*, 75(3):371–385.

Ballester, C., Garrido, L., Lazcano, V., and Caselles, V. (2012). A tv-l1 optical flow method with occlusion detection. In *DAGM/OAGM, LNCS 7476*. Springer Verlag.

Black, M. J. and Ananda, P. (1996). The robust estimation of multiple motions: Parametric and piecewise-smooth flow fields. *Computer vision and image understanding*, 63(1):75–104.

Brox, T., Bregler, C., and Malik, J. (2009). Large displacement optical flow. In *Proceedings of the IEEE Computer Vision and Pattern Recognition*, pages 500–513.

Brox, T., Bruhn, A., Papenberg, N., and Weickert, J. (2004). High accuracy optical flow estimation based on a the-

ory for warping. In *European Conference on Computer Vision (ECCV)*, volume 3024, pages 25–36. Lecture Notes in Computer Science.

Butler, D. J., Wulff, J., Stanley, G. B., and Black, M. J. (2012). A naturalistic open source movie for optical flow evaluation. In A. Fitzgibbon et al. (Eds.), editor, *European Conf. on Computer Vision (ECCV)*, Part IV, LNCS 7577, pages 611–625. Springer-Verlag.

Chen, Z., Jin, H., Lin, Z., Cohen, S., and Wu, Y. (2013). Large displacement optical flow from nearest neighbor fields. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 2443–2450.

Fortun, D., Bouthemy, P., and Kervrann, C. (2016). Aggregation of local parametric candidates with exemplar-based occlusion hangling for optical flow. *Computer Vision and Image Undestanding*, pages 81–94.

Horn, B. K. and Schunck, B. H. (1981). Determining optical flow. *Artificial Intelligence*, 17:185–203.

Ince, S. and Konrad., J. (2008). Occlusion-aware optical flow estimation. *IEEE Transactions on Image Processing*, 17(8):1443–1451.

Kennedy, R. and Taylor, C. J. (2015). Optical flow with geometric occlusion estimation and fusion of multiple frames. In *EMMCVPR 2015, Hong Kong, China, January 13-16, 2015. Proceedings*, pages 364–377. Springer International Publishing.

Krähenbühl, P. and Koltun., V. (2012). Efficient nonlocal regularization for optical flow. In *European Conference on Computer Vision (ECCV)*, pages 356–369. Springer.

Nagel, H.-H. and Ekelman, W. (1986). An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences. *Pattern Analysis and Machine Intelligence*, 6(5):565–593.

Palomares, R. P., Haro, G., and Ballester, C. (2015). A rotation-invariant regularization term for optical flow related problems. *Lectures Notes in Computer Science*, 9007:304–319.

Palomares, R. P., Meinhardt-Llopis, E., Ballester, C., and Haro, G. (2017). Faldoi: A new minimization strategy for large displacement variational optical flow. *Journal of Mathematical Imaging and Vision*, 58(1):27–46.

Ranftl, R., Bredies, K., and Pock, T. (2014). Non-local total generalized variation for optical flow estimation. In *Computer Vision–ECCV*, pages 439–454. Springer.

Sánchez, J., Salgado, A., and Monzón, N. (2014). Preserving accurate motion contours with reliable parameter selection. In *IEEE International Conference on Image Processing (ICIP)*, pages 209–213.

Scharstein, S. and Szeliski, R. (2002). A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. *International journal of computer vision*, 47 :7–42.

Steinbruecker, F. and Pock, T. (2009). Large displacement optical flow computation without warping. In *International Conference on Computer Vision*, pages 1609–1614.

Stoll, M., Volz, S., and Bruhn., A. (2012). Adaptive integration of features matches into variational optical flow methods. In *Proc. of the Asian Conference in Computer Vision (ACCV)*, pages 1–14.

Strekalovskiy, E., Chambolle, A., and Cremers, D. (2014). Convex relaxation of vectorial problems with coupled regularization. *SIAM J. Imaging Sciences*, 7(1):294–336.

Sun, D., Roth, S., and Black, M. J. (2010). Secrets of optical flow estimation and their principles. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2432–2439.

Sun, D., Roth, S., and Black, M. J. (2014). A quantitative analysis of current practices in optical flow estimation and the principles behind them. *International Journal of Computer Vision*, 2 (106):115–137.

Timofte, R. and Van Gool, L. (2015). Sparse flow: Sparse matching for small to large displacement optical flow. *IEEE WACV*, 00:1100–1106.

Weinzaepfel, P., Revaud, J., Harchaoui, Z., and Schmid, C. (2013). Deepflow: Large displacement optical flow with deep matching. In *IEEE International Conference on Computer Vision, Sydney, Australia*, pages 1385–1392.

Werlberger, M., Pock, T., and Bischof, H. (2010). Motion estimation with non-local total variation regularization. In *IEEE Conference on Computer Vision and Pattern Recognition(CVPR)*, pages 2464–2471.

Werlberger, M., Trobin, W., Pock, T., Wedel, A., Cremers, D., and Bischof, H. (2009). Anisotropic huber-l1 optical flow. In *Proceedings of the BMVC*.

Xu, L., Jia, J., and Matsushita, Y. (2012). Motion detail preserving optical flow estimation. In *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, volume 34, pages 1744–1757.

Zach, C., Pock, T., and Bischof, H. (2007). A duality based approach for realtime tv-l1 optical flow. In *Proceedings of the 29th DAGM Conference on Pattern Recognition*, pages 214–223. Berlin, Heidelberg, Springer-Verlag.

Zimmer, H., Bruhn, A., and J., W. (2011). Optic flow in harmony. *International Journal of Computer Vision*, 93(3):368–388.