



# Variational Temporal Optical Flow for Multi-exposure Video

Onofre Martorell<sup>1</sup><sup>a</sup> and Antoni Buades<sup>1</sup><sup>b</sup>

*Institute of Applied Computing and Community Code (IAC3) and with the Dept. of Mathematics and Computer Science, Universitat de les Illes Balears, Cra. de Valldemossa km. 7.5, E-07122 Palma, Spain*

**Keywords:** Optical Flow, Motion Estimation, Variational Methods, Multi-exposure Videos.

**Abstract:** High Dynamic Range (HDR) reconstruction for multi-exposure video sequences is a very challenging task. One of its main steps is registration of input frames. We propose a novel variational model for optical flow estimation for multi-exposure video sequences. We introduce data terms for consecutive and non consecutive frames, the latter term comparing frames with the same exposure. We also compute forward and backward flow terms for the current frame, naturally introducing temporal regularization. We present a particular formulation with sequences with two exposures, that can be extended to larger number of exposures. We compare the proposed method with state of the art variational models.

## 1 INTRODUCTION


Multi-exposure fusion (MEF) and High Dynamic Range (HDR) imaging are two techniques to combine several images of the same scene taken with different exposure times into a single image. Conventional cameras are not able to capture in one shot the whole large range of natural light present in the scene. Although both families of methods seek the same goal, the main procedure used is different. On one hand, in HDR, the irradiance values of each image need to be obtained. To do that, the camera response function (CRF) has to be estimated, generally using the method proposed by Debevec and Malik (Debevec and Malik, 2008). After the fusion, the result obtained by a HDR method has to be converted into a standard 8 bit image, for visualization in common displays. This process is known as tone-mapping (Reinhard et al., 2005). On the contrary, MEF avoids the CRF estimation and the tone-mapping by directly fusing the images in the standard 8bit domain.


There exists a wide literature on MEF and HDR. The most well-known method in MEF is the one proposed by Mertens et al. (Mertens et al., 2009) and in HDR, a widely used method is the one proposed by Debevec and Malik (Debevec and Malik, 2008). In order to apply both methods, the image sequence needs to be static, that is, the objects in the scene do not move and the camera is fixed. This is not the case

for many sequences, and the application of such methods introduce ghosting effects. To avoid these artefacts, methods might be adapted to non-static image sequences (Martorell et al., 2019), (Ma et al., 2017), (Hu et al., 2013), (Khan et al., 2006).

A natural extension of MEF and HDR imaging algorithms are MEF and HDR video, respectively. There is a lack of specialized hardware to acquire pleasant videos with large range of colors directly or it is too expensive to be used in daily-used devices such as mobile phones because they require complex sensors or optical systems (Tocci et al., 2011)(Zhao et al., 2015). The best way to overcome this problem is by capturing videos where each frame is taken with alternating exposure times (generally 2 or 3) and produce a video with a high range of colors by processing the frames offline using MEF or HDR techniques. Figure 1 shows some examples of multi-exposure video sequences.

The transition from MEF or HDR imaging to the corresponding video method is not straightforward. First, only two or three exposures are available, meaning that for some regions we will only dispose of under or over exposed images. This is not the case for still image combination where five or more exposures might be available. Second, when a single image needs to be synthesized, the image with the best exposure of the initial set can be chosen as reference and the rest be registered according to this reference one. However, in HDR and MEF video, methods need to be applied taking as reference each one of the frames,

<sup>a</sup>  <https://orcid.org/0000-0002-9071-778X>

<sup>b</sup>  <https://orcid.org/0000-0001-9832-3358>

which are generally darker or brighter than a middle-exposed image.

Registration is an important step in video fusion and the overall result depends critically on it. Most HDR video methods try to estimate the optical flow between the central frame and the neighbouring ones, which were taken with a different exposure time. In two exposure video sequences, this means the flow with the previous and posterior frames is computed. Methods additionally compute the flow between the previous and posterior frames which have the same exposure. Then methods select among computed flow locally depending if the regions is or not saturated.

We propose a variational method for optical flow with two alternating exposures. We combine in a single functional the forward and backward flow and take advantage that the previous and posterior frames have the same exposure. All computed flows are defined taking the central image grid as reference, allowing to naturally include temporal regularization.

This paper is organized as follows. In Section 2 we present the existing literature on registration for multi-exposure video sequences and variational optical flow with temporal smoothness constraints. In Section 3 we propose a new optical flow variational model for triplets of images in a video. In Section 4 we compare the obtained results with existing optical flow methods. Finally, in Section 5 we draw some conclusions and future work.

## 2 RELATED WORK

### 2.1 HDR Video

It is not surprising that no methods are proposed for dealing with multi-exposure sequences directly in the input 8 bit domain. MEF methods perform worse than HDR when few exposures are available, which is the case of video. However, there are many works which deal with HDR for video sequences. We will make a review of those methods, focusing on the techniques used to register the input images. All those methods use an estimation of the CRF to transform the input images from an LDR domain to an HDR domain. These radiance frames are then normalized by dividing by its exposition time. In this way, different frames have the same irradiance value at corresponding pixels.

The common step for reducing ghosting artifacts is registering the frames prior to their combination. The classical method for HDR video registration is the use of optical flow, e.g Lukas-Kanade, used by Kang et al. (Kang et al., 2003), block matching, used

by Mangiat et al. (Mangiat and Gibson, 2010) or the method proposed by Liu (Liu et al., 2009), used by Kalantari et al. (Kalantari et al., 2013). The more recent Li et al. (Li et al., 2016) first separate foreground and background before registering the frames.

Musil et al. (Musil et al., 2020) present a fast ghost-free HDR acquisition algorithm for static cameras. For each frame, they compute a certainty map based on the difference of pixel values with its neighboring frames in the radiance domain. This permits the removal of moving objects and the fusion to obtain the final HDR image.

Van Vo and Lee (Van Vo and Lee, 2020) is the only method that takes into account the temporal information: they divide the motion estimation step into two phases: on one hand, they perform optical flow estimation of well-exposed areas in the descriptor domain. On the other hand, they perform a superpixel-based motion estimation on poorly exposed areas by computing the optical flow field from the backward frame to the forward frame. The optical flow from central to forward and backward frames is half of the computed flow.

There exist learning-based methods for HDR video synthesis (Kalantari and Ramamoorthi, 2019) (Chen et al., 2021). In both cases, they compute the optical flow of the input frames by training a CNN coarse-to-fine architecture based on the ones by Ranjan and Black (Ranjan and Black, 2017) and Wang et al. (Wang et al., 2017). Training is performed into an end-to-end way, meaning that only the error of the final HDR sequence is used for defining the loss.

### 2.2 Variational Optical Flow with Temporal Regularization

The first variational formulation of the optical flow problem was presented by Horn and Schunk (Horn and Schunck, 1981). They proposed to find the optical flow field  $\mathbf{w} = (u, v)$  between a pair of images,  $I_0$  and  $I_1$ , by minimizing an energy functional formed by two different terms: the data and the smoothness terms. The first one imposes the brightness constancy constraint, that is, the gray values of corresponding pixels are the same; and the second one penalizes the high variations of the optical flow field. This second term was written as

$$\int_{\Omega} (|\nabla u(\mathbf{x})|^2 + |\nabla v(\mathbf{x})|^2) d\mathbf{x}, \quad (1)$$

with  $\mathbf{x} = (x, y)$  being a two-dimensional point.

Weickert and Schnörr (Weickert and Schnörr, 2001) proposed the first method introducing a temporal derivative in the energy functional. Instead of computing the optical flow between a pair of frames,

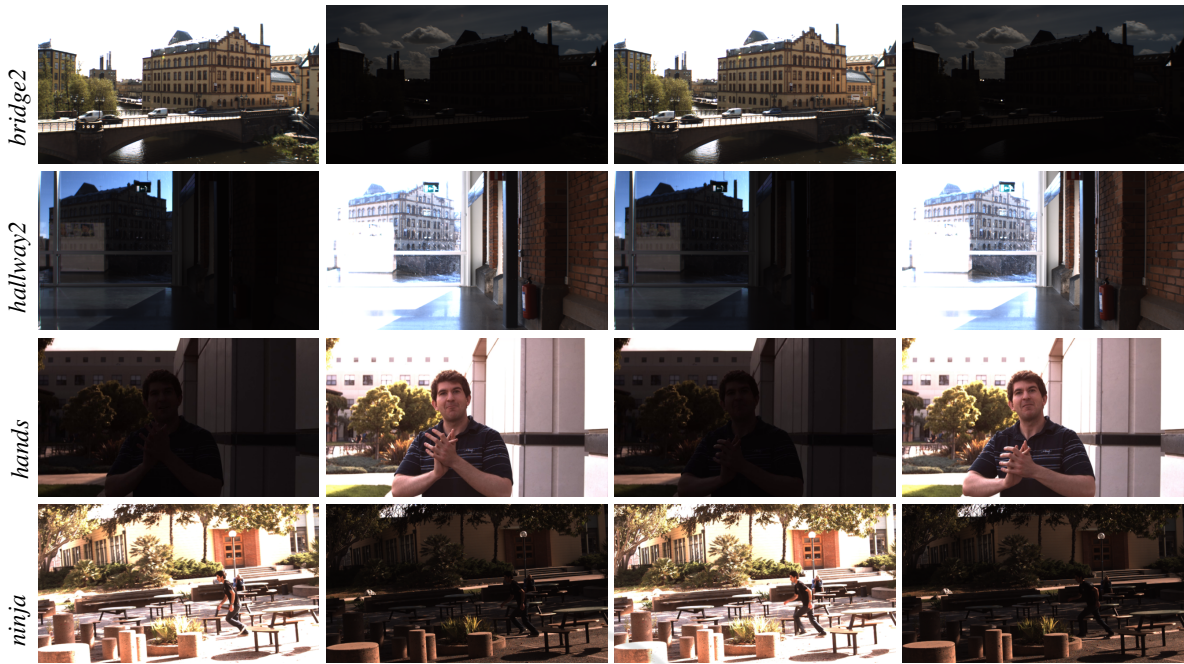


Figure 1: Frames of multi-exposure video sequences with 2 exposure times used for comparison. Sequences obtained from Kalantari et al. (Kalantari et al., 2013) and Li et al. (Li et al., 2016) datasets.

they considered a video sequence  $\{I_t \mid t = 0, \dots, N\}$ . They used the data term proposed in (Horn and Schunck, 1981) but used a 3D gradient operator for the smoothness term:

$$\int_{\Omega \times [0, N]} \Psi(\|\nabla_3 u(\mathbf{x})\| + \|\nabla_3 v(\mathbf{x})\|) d\mathbf{x}, \quad (2)$$

with  $\mathbf{x} = (x, y, t)$  being a three-dimensional point, where the third coordinate corresponds to the time. Brox et al. (Brox et al., 2004) adopted the same regularity term as Weickert and Schnörr but introduced a new data term based on the correspondence of gradients. In both cases, this regularization operator may fail in presence of large and complex motions.

Salgado et al. (Salgado and Sánchez, 2007) propose to separate the temporal regularization constraint from the spatial one in (2). They used the same spatial regularization term as in (1) and proposed the temporal term

$$\int_{\Omega} \sum_{t=0}^{N-1} \Psi(\|\mathbf{w}_t(\mathbf{x}) - \mathbf{w}_{t+1}(\mathbf{x} + \mathbf{w}_t(\mathbf{x}))\|^2) d\mathbf{x} + \int_{\Omega} \sum_{t=1}^N \Psi(\|\mathbf{w}_t(\mathbf{x}) - \mathbf{w}_{t-1}(\mathbf{x} + \mathbf{w}_{t-1}^*(\mathbf{x}))\|^2) d\mathbf{x},$$

where  $\mathbf{w}_{t-1}^*(\mathbf{x})$  is the backward flow. This new temporal term is nonlinear but permits to better deal with large displacements. Moreover, for small displacements, this term can be seen as the discretization of the temporal derivative presented previously. Years

later, Sanchez et al. (Sánchez et al., 2013) improved the method by introducing a second order temporal smoothness constraint at the PDE level:

$$\sum_{t=1}^{N-1} \Psi'(\|\mathbf{w}_{t-1}(\mathbf{x} + \mathbf{w}_{t-1}^*(\mathbf{x})) - \mathbf{w}_{t+1}(\mathbf{x} + \mathbf{w}_t(\mathbf{x}))\|^2) \cdot (\mathbf{w}_{t-1}(\mathbf{x} + \mathbf{w}_{t-1}^*(\mathbf{x})) - 2\mathbf{w}_t + \mathbf{w}_{t+1}(\mathbf{x} + \mathbf{w}_t(\mathbf{x}))). \quad (3)$$

Volz et al. (Volz et al., 2011) proposed a novel parametrization of optical flow fields to encourage temporal regularization along motion trajectories. With this new parametrization, the nonlinearity form Sanchez et al. disappears and the minimization becomes simpler. As Sanchez et al., they also used a second order regularization, in this case added directly in the energy functional:

$$\int_{\Omega} \sum_{t=0}^{N-1} \Psi(\|\mathbf{w}_{t+1}(\mathbf{x}) - \mathbf{w}_t(\mathbf{x})\|^2) d\mathbf{x} + \int_{\Omega} \sum_{t=1}^{N-1} \Psi(\|\mathbf{w}_{t+1}(\mathbf{x}) - 2\mathbf{w}_t(\mathbf{x}) + \mathbf{w}_{t-1}(\mathbf{x})\|^2) d\mathbf{x}. \quad (4)$$

### 3 OPTICAL FLOW WITH TEMPORAL CONSTRAINTS

Let  $\{I_t \mid t = 0, \dots, N\}$  be a sequence of images, where images are taken with 2 different alternating exposure

times. Figure 1 shows some examples of frames of the videos used in the experimentation section. At each frame position we dispose of a single exposure. Hence, we need to compute optical flow between consecutive frames of different exposure.

Let  $I_t$  be the reference frame and  $I_{t+1}$  and  $I_{t-1}$  the previous and posterior frames, respectively. We present the method for this triplet of images, and it is repeated for each  $t \in \{1, \dots, N-1\}$ .

It must be noted that most variational optical flow models rely on the brightness constancy assumption, which states that two corresponding pixels have the same color. However, this assumption is not satisfied with multi-exposure images. To overcome this issue, we apply photometric calibration to the pairs  $(I_t, I_{t+1})$  and  $(I_t, I_{t-1})$  separately by using the method proposed by Martorell et al. (Martorell et al., 2019). As most algorithms, we choose to modify the color of the darker images of the sequence, to look alike the neighbouring brighter ones. Brighter images might have large saturated areas while this is not the case of darker ones.

The used method for photometric calibration requires that the input images are roughly registered. Hence, we first globally register the neighbour frames with respect to the reference one. We use a global affine transformation. This is estimated using SIFT matches (Lowe, 1999) and RANSAC strategy (Fischler and Bolles, 1981). After this initial registration we apply the photometric calibration from Martorell et al. (Martorell et al., 2019).

Once the images are photometrically calibrated and globally registered, we can estimate the optical flow between neighbouring frames using a variational model.

### 3.1 Proposed Variational Model

Let  $I_{t-1}, I_t$  and  $I_{t+1}$  be 3 consecutive frames from a multi-exposure video sequence and let  $\mathbf{w}_F(\mathbf{x})$  be the optical flow field from  $I_t(\mathbf{x})$  to  $I_{t+1}(\mathbf{x})$  and  $\mathbf{w}_B(\mathbf{x})$  the optical flow field from  $I_t(\mathbf{x})$  to  $I_{t-1}(\mathbf{x})$ . We want to jointly estimate the optical flow fields from the central frame to the other two by taking into account the video nature of the input data. Apart from the well-known data and spatial smoothness terms, we want to impose two new aspects in the minimization. On one hand, we want that the motion trajectories are smooth along the video sequence. On the other hand, we want to match pixels from  $I_{t-1}$  to  $I_{t+1}$ .

Most methods presented in Section 2.2 parametrize  $\mathbf{w}_B(\mathbf{x})$  as the optical flow from  $I_{t-1}(\mathbf{x})$  to  $I_t(\mathbf{x})$ , but this forces them to minimize the energy functional twice, in order to warp both neighbour

frames. In our case, we follow the parametrization from Volz et al. (Volz et al., 2011), which parametrizes both optical flow fields with regard to the central frame and allows to warp both neighbour frames by minimizing just one energy functional.

Most methods for HDR video compute separately the registration fields  $\mathbf{w}_F(\mathbf{x})$  and  $\mathbf{w}_B(\mathbf{x})$ . However, these estimations are prone to fail at saturated areas. To solve this, Van Vo and Lee (Van Vo and Lee, 2020) also compute the optical flow from  $I_{t-1}(\mathbf{x})$  to  $I_{t+1}(\mathbf{x})$  and combine it heuristically with  $\mathbf{w}_F(\mathbf{x})$  and  $\mathbf{w}_B(\mathbf{x})$ . In contrast, the proposed method introduces the comparison of non-consecutive frames in the energy functional to address this problem.

With all that, we propose to minimize the following energy functional to estimate the optical flow fields:

$$E(\mathbf{w}_B, \mathbf{w}_F) = \alpha \left( \mathcal{S}(\mathbf{w}_B) + \mathcal{S}(\mathbf{w}_F) \right) + \delta \mathcal{T}(\mathbf{w}_B, \mathbf{w}_F) + \beta \left( \gamma_F \mathcal{D}_F(\mathbf{w}_F) + \gamma_B \mathcal{D}_B(\mathbf{w}_B) + \lambda \mathcal{J}(\mathbf{w}_B, \mathbf{w}_F) \right) \quad (5)$$

where variables  $\beta$ ,  $\gamma_F$ ,  $\gamma_B$ ,  $\lambda$ ,  $\alpha$  and  $\delta$  are the weights that measure the trade-off of each term and

$$\mathcal{D}_F(\mathbf{w}_F) = \int_{\Omega} \Psi \left( (\hat{I}_{t+1,t}(\mathbf{x} + \mathbf{w}_F) - \hat{I}_{t,t+1}(\mathbf{x}))^2 \right) d\mathbf{x} \quad (6)$$

$$\mathcal{D}_B(\mathbf{w}_B) = \int_{\Omega} \Psi \left( (\hat{I}_{t-1,t}(\mathbf{x} - \mathbf{w}_B) - \hat{I}_{t,t-1}(\mathbf{x}))^2 \right) d\mathbf{x} \quad (7)$$

$$\mathcal{J}(\mathbf{w}_B, \mathbf{w}_F) = \int_{\Omega} \Psi \left( (I_{t+1}(\mathbf{x} + \mathbf{w}_F) - I_{t-1}(\mathbf{x} - \mathbf{w}_B))^2 \right) d\mathbf{x} \quad (8)$$

$$\mathcal{S}(\mathbf{w}) = \int_{\Omega} \Psi \left( \|\nabla \mathbf{w}_F\|^2 \right) + \Psi \left( \|\nabla \mathbf{w}_B\|^2 \right) d\mathbf{x} \quad (9)$$

$$\mathcal{T}(\mathbf{w}_B, \mathbf{w}_F) = \int_{\Omega} \Psi \left( \|\mathbf{w}_F - \mathbf{w}_B\|^2 \right) d\mathbf{x} \quad (10)$$

where  $\hat{I}_{kj}$  and  $\hat{I}_{jk}$  are pairs of images  $I_k$  and  $I_j$  photometrically calibrated according to the criteria presented at the beginning of the section. Finally,  $\Psi(s^2) = \sqrt{s^2 + \epsilon^2}$  with  $\epsilon = 0.00001$  is a robust convex function.

The proposed energy functional contains several terms which serve for different purposes, let's analyse in detail all the terms of the energy functional:

- (i)  $\mathcal{D}_F(\mathbf{w}_F)$ : data term to match pixels from  $\hat{I}_{t,t+1}$  to  $\hat{I}_{t+1,t}$  through optical flow field  $\mathbf{w}_F$ .
- (ii)  $\mathcal{D}_B(\mathbf{w}_B)$ : data term to match pixels from  $\hat{I}_{t,t-1}$  to  $\hat{I}_{t-1,t}$  through optical flow field  $\mathbf{w}_B$ . Since  $\mathbf{w}_B$



is computed with regard to the central frame, we need to subtract the optical flow to the coordinates to get the right registration.

- (iii)  $\mathcal{J}(\mathbf{w}_B, \mathbf{w}_F)$ : data term to match pixels from  $I_{t+1}$  and  $I_{t-1}$  through optical flow fields  $\mathbf{w}_t^B$  and  $\mathbf{w}_t^F$ . Since both optical flow fields have the same reference, we can use this term easily, without introducing nonlinearities, which could make the minimization more difficult.
- (iv)  $\mathcal{S}(\mathbf{w})$ : spatial smoothness term for optical flow.
- (v)  $\mathcal{T}(\mathbf{w}_B, \mathbf{w}_F)$ : first order temporal smoothness term to regularize motion trajectories. Since both optical flow fields have the same reference, we just need to subtract them, not as in Sanchez et al. (Sánchez et al., 2013), where they needed to compose optical flow fields to write this term.

### 3.2 Minimization

The minimization of the energy functional (5) at each frame  $I_t$  is done by finding the corresponding Euler-Lagrange equation and solving the obtained linear system of equations with a successive over-relaxation (SOR) scheme. The minimization is embedded in a pyramidal structure to better deal with large displacements: given a number of scales  $N_{scales}$ , and a downsampling factor  $v$ , the input images at scale  $s = 0, \dots, N_{scales} - 1$  are the input images at scale  $s - 1$  convolved with a Gaussian kernel and downsampled by a factor  $v$ . We start estimating the optical flow at scale  $s = N_{scales} - 1$ , then we upsample the output flow and we repeat the minimization at scale  $s - 1$  by using as initialization the upsampled optical flow.

## 4 RESULTS

In this section we compare the proposed fusion algorithm with other variational optical flow methods. We compare our method against Brox et al. (Brox et al., 2004) with and without temporal smoothness constraints. In both cases, we use the implementation available at ipol.im (Sánchez Pérez et al., 2013). We modified the code of the method with temporal constraints to be able to deal with pairs of photometrically calibrated images, as in the proposed model. For simplicity, we name the methods `brox_spatial` and `brox_temporal`, respectively. In both methods we need to compute two flows: in `brox_spatial` we need to compute separately the flow from  $I_t$  to  $I_{t+1}$  and from  $I_t$  to  $I_{t-1}$ . On the other hand, `brox_temporal` model gives always the optical flow from one frame to the next

Table 1: Mean error of 4 frames from each sequence. In bold, the method with lowest error for each sequence.

	Brox et al.	Brox et al. temporal	Ours
<i>hands</i>	<b>10.67</b>	13.68	11.99
<i>ninja</i>	23.30	32.10	<b>20.88</b>
<i>bridge</i>	15.25	18.95	<b>13.93</b>
<i>hallway</i>	<b>10.71</b>	12.59	<b>10.71</b>
<i>parkinglot</i>	24.30	31.67	<b>22.94</b>
Average	14.04	18.17	<b>13.41</b>

one. Hence, we need to apply the model to the sequence  $\{I_{t-1}, I_t, I_{t+1}\}$  and to  $\{I_{t+1}, I_t, I_{t-1}\}$  to get  $\mathbf{w}_F$  and  $\mathbf{w}_B$ .

We evaluate our algorithm by using several multi-exposure video sequences available online: Kalantari et al. (Kalantari et al., 2013) have available online <sup>1</sup> a dataset of 7 sequences, 4 of them with 2 exposures. We use sequences *hands* and *ninja* from this dataset. Moreover, Li et al. (Li et al., 2016) have available online <sup>2</sup> a dataset with 4 video sequences with 2 exposure times. From this dataset, we use sequences *bridge*, *hallway* and *parkinglot*.

Our results were computed using the same parameters for all tests. We set weights  $\beta$ ,  $\alpha$  and  $\delta$  to 1, 13 and 0.5, respectively. The values of  $\gamma_F$ ,  $\gamma_B$  and  $\lambda$  have been set taking into account the kind of input data used: in multi-exposure videos sequences with 2 exposure times, frames  $I_{t-1}$  and  $I_{t+1}$  have the same exposure times, hence, their comparison will be more reliable compared to the ones of consecutive frames, which have different exposure times. Because of that, we have decided to give more weight to the term  $\mathcal{J}(\mathbf{w}_B, \mathbf{w}_F)$  by setting  $\lambda = 1$  and  $\gamma_F = \gamma_B = 0.5$ .

We first evaluate our algorithm quantitatively in 4 images of different sequences. We compute all needed optical flow fields at each frame, we then compute the registration error for each frame and we finally do the mean to get a measure of the error on each sequence. For each frame  $I_t$ , we compute the error as

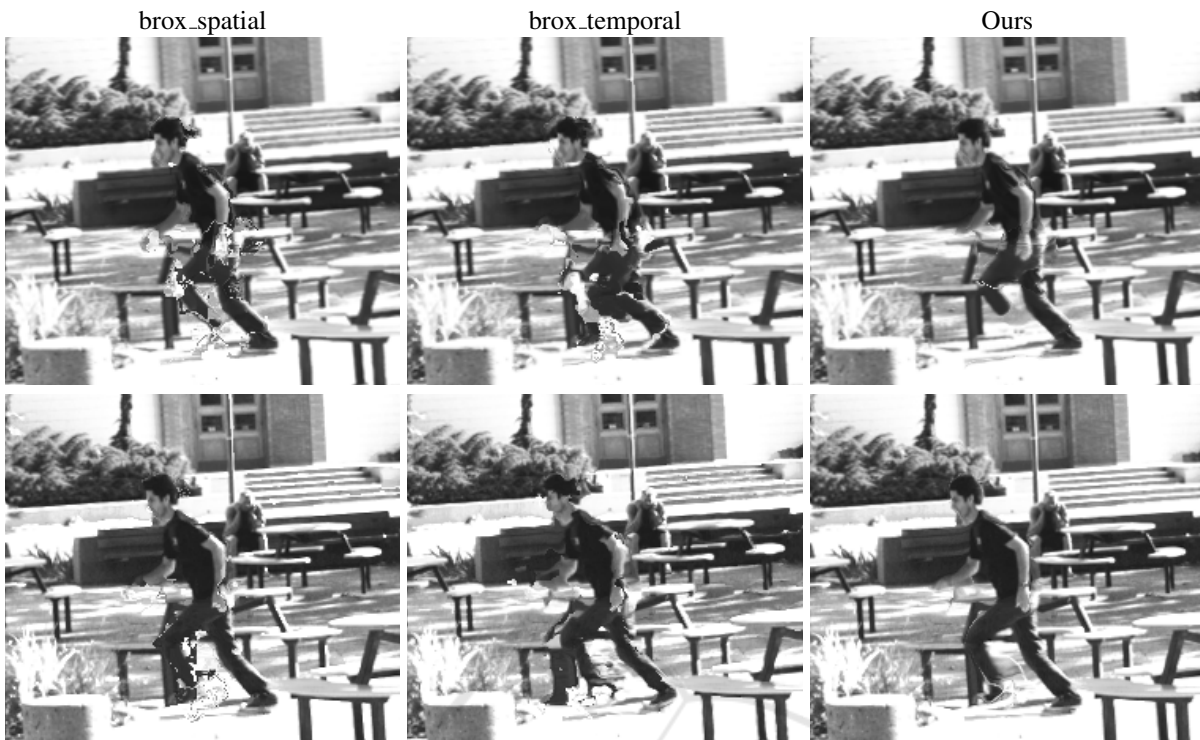
$$\sum_{\mathbf{x} \in \tilde{\Omega}} (\hat{I}_{t+1,t}(\mathbf{x} + \mathbf{w}_F) - \hat{I}_{t,t+1}(\mathbf{x}))^2 + \sum_{\mathbf{x} \in \tilde{\Omega}} (\hat{I}_{t-1,t}(\mathbf{x} - \mathbf{w}_B) - \hat{I}_{t,t-1}(\mathbf{x}))^2, \quad (11)$$

where  $\tilde{\Omega}$  is the discrete domain of the image.

Table 1 shows the results on 5 multi-exposure sequences, where the method with lowest error is highlighted in bold. As we can see, our methods obtains the lowest error in 4 sequences. In *hands* sequence we

<sup>1</sup> <https://web.ece.ucsb.edu/~psen/PaperPages/HDRVideo>

<sup>2</sup> <http://signal.ee.psu.edu/research/MAPHDR.html>

Figure 2: Warped images for two consecutive frames of *ninja* sequence.

get the second best error, and in *hallway* sequence we get the same error as *brox\_spatial*.

We also evaluate our results qualitatively. Figure 2 shows a detail of  $I_{t-1}(\mathbf{x} - \mathbf{w}_B)$  for two consecutive frames of *ninja* sequence. As it can be seen, our method performs better than the compared methods.

We also evaluate our optical flow method by finding pixels with inaccurate displacement: we perform a left-right consistency on the forward flow as

$$LR(\mathbf{x}) = \max \left( \|\mathbf{w}_F^t(\mathbf{x}) + \mathbf{w}_B^{t+1}(\mathbf{x} + \mathbf{w}_F^t)\|, \right. \quad (12)$$

$$\left. \|\mathbf{w}_B^t(\mathbf{x}) + \mathbf{w}_F^{t-1}(\mathbf{x} + \mathbf{w}_B^t)\| \right) \quad (13)$$

$$M_F^{lr}(\mathbf{x}) = \begin{cases} 1 & \text{if } LR(\mathbf{x}) < 2 \\ 0 & \text{otherwise} \end{cases}. \quad (14)$$

and the left-right consistency check on the backward flow  $M_B^{lr}$  is computed analogously. Since we have two masks, we combine them to get a final mask as

$$M^{lr}(\mathbf{x}) = M_F^{lr}(\mathbf{x}) \cdot M_B^{lr}(\mathbf{x}). \quad (15)$$

We apply the HDR algorithm from Khan et al. (Khan et al., 2006) to the triplet of aligned images to get a fusion with a large range of colors.

Figure 3 shows the result of fusing the three registered images using the HDR algorithm from Khan et al. (Khan et al., 2006) as well as the left-right consistency mask superimposed in two consecutive

frames of *ninja* sequence. As it can be seen, our method has a better left-right consistency and alignment. *Brox\_spatial* has more errors at registration: the stairs in the background at the second row are not perfectly aligned and the man has registration errors between the legs. Moreover, *brox\_temporal* has a small misalignment on the right side of the second image, making it look blurry. Finally, our method has less errors due to left-right consistency.

## 5 CONCLUSIONS

We have presented a new variational model for optical flow estimation in multi-exposure sequences. We use temporal constraints to take into account the video nature of input images. We have applied photometric calibration to obtain pairs of images with the same color, in order to satisfy brightness constancy on the optical flow model. Qualitative and quantitative performance show that our method performs better than other optical flow variational models.

In future work, we plan to adapt the proposed method for multi-exposure video sequences with 3 exposure times. In this case, forward and backward frames do not have the same exposure time, hence we need to use more frames in order to add more data

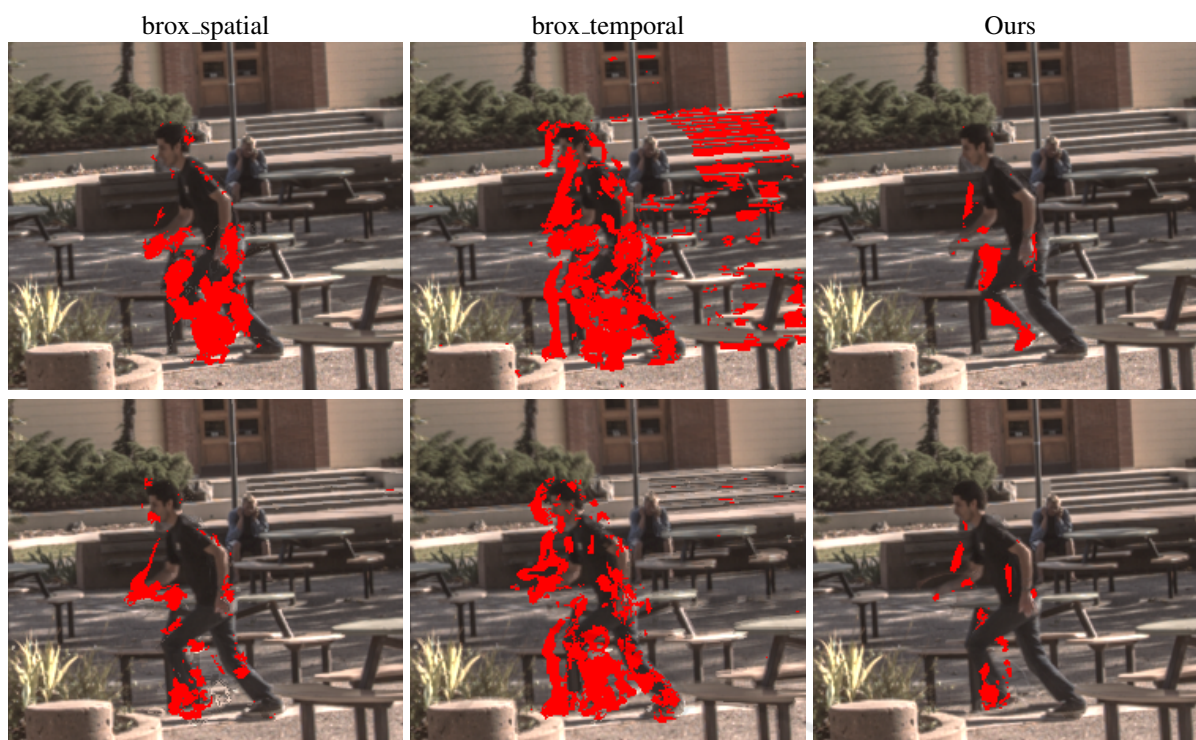


Figure 3: HDR fusion of registered images and left-right consistency mask superimposed in two consecutive frames of *ninja* sequence.

terms like the one that matches pixels from forward and backward frames.

Moreover, as seen in the results shown, our method still has some registration errors, that should be solved so as to get a good looking final fused image. To do that, we plan to apply some processing techniques to remove them.

## ACKNOWLEDGEMENTS

The authors acknowledge the Ministerio de Ciencia, Innovación y Universidades (MCIU), the Agencia Estatal de Investigación (AEI) and the European Regional Development Funds (ERDF) for its support to the project TIN2017-85572-P.

## REFERENCES

- Brox, T., Bruhn, A., Papenbergh, N., and Weickert, J. (2004). High accuracy optical flow estimation based on a theory for warping. In *European conference on computer vision*, pages 25–36. Springer.
- Chen, G., Chen, C., Guo, S., Liang, Z., Wong, K.-Y. K., and Zhang, L. (2021). Hdr video reconstruction: A coarse-to-fine network and a real-world benchmark

dataset. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2502–2511.

- Debevec, P. E. and Malik, J. (2008). Recovering high dynamic range radiance maps from photographs. In *ACM SIGGRAPH 2008 classes*, page 31. ACM.
- Fischler, M. A. and Bolles, R. C. (1981). Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. *Communications of the ACM*, 24(6):381–395.
- Horn, B. K. and Schunck, B. G. (1981). Determining optical flow. *Artificial intelligence*, 17(1-3):185–203.
- Hu, J., Gallo, O., Pulli, K., and Sun, X. (2013). Hdr deghosting: How to deal with saturation? In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1163–1170.
- Kalantari, N. K. and Ramamoorthi, R. (2019). Deep hdr video from sequences with alternating exposures. In *Computer Graphics Forum*, volume 38, pages 193–205. Wiley Online Library.
- Kalantari, N. K., Shechtman, E., Barnes, C., Darabi, S., Goldman, D. B., and Sen, P. (2013). Patch-based high dynamic range video. *ACM Trans. Graph.*, 32(6):202–1.
- Kang, S. B., Uyttendaele, M., Winder, S., and Szeliski, R. (2003). High dynamic range video. In *ACM Transactions on Graphics (TOG)*, volume 22, pages 319–325. ACM.
- Khan, E. A., Akyuz, A. O., and Reinhard, E. (2006). Ghost removal in high dynamic range images. In *2006 In-*



- ternational Conference on Image Processing, pages 2005–2008. IEEE.
- Li, Y., Lee, C., and Monga, V. (2016). A maximum a posteriori estimation framework for robust high dynamic range video synthesis. *IEEE Transactions on Image Processing*, 26(3):1143–1157.
- Liu, C. et al. (2009). *Beyond pixels: exploring new representations and applications for motion analysis*. PhD thesis, Massachusetts Institute of Technology.
- Lowe, D. G. (1999). Object recognition from local scale-invariant features. In *Proceedings of the seventh IEEE international conference on computer vision*, volume 2, pages 1150–1157. Ieee.
- Ma, K., Li, H., Yong, H., Wang, Z., Meng, D., and Zhang, L. (2017). Robust multi-exposure image fusion: A structural patch decomposition approach. *IEEE Trans. Image Processing*, 26(5):2519–2532.
- Mangiat, S. and Gibson, J. (2010). High dynamic range video with ghost removal. In *Applications of Digital Image Processing XXXIII*, volume 7798, page 779812. International Society for Optics and Photonics.
- Martorell, O., Sbert, C., and Buades, A. (2019). Ghosting-free dct based multi-exposure image fusion. *Signal Processing: Image Communication*, 78:409–425.
- Mertens, T., Kautz, J., and Van Reeth, F. (2009). Exposure Fusion: A Simple and Practical Alternative to High Dynamic Range Photography. *Computer Graphics Forum*.
- Musil, M., Nosko, S., and Zemcik, P. (2020). De-ghosted hdr video acquisition for embedded systems. *Journal of Real-Time Image Processing*, pages 1–10.
- Ranjan, A. and Black, M. J. (2017). Optical flow estimation using a spatial pyramid network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4161–4170.
- Reinhard, E., Ward, G., Pattanaik, S., and Debevec, P. (2005). *High Dynamic Range Imaging: Acquisition, Display, and Image-Based Lighting (The Morgan Kaufmann Series in Computer Graphics)*. Morgan Kaufmann Publishers Inc., San Francisco, CA, USA.
- Salgado, A. and Sánchez, J. (2007). Temporal constraints in large optical flow estimation. In *International Conference on Computer Aided Systems Theory*, pages 709–716. Springer.
- Sánchez, J., Salgado de la Nuez, A. J., and Monzón, N. (2013). Optical flow estimation with consistent spatio-temporal coherence models.
- Sánchez Pérez, J., Monzón López, N., and Salgado de la Nuez, A. (2013). Robust Optical Flow Estimation. *Image Processing On Line*, 3:252–270.
- Tocci, M. D., Kiser, C., Tocci, N., and Sen, P. (2011). A versatile hdr video production system. *ACM Transactions on Graphics (TOG)*, 30(4):1–10.
- Van Vo, T. and Lee, C. (2020). High dynamic range video synthesis using superpixel-based illuminance-invariant motion estimation. *IEEE Access*, 8:24576–24587.
- Volz, S., Bruhn, A., Valgaerts, L., and Zimmer, H. (2011). Modeling temporal coherence for optical flow. In *2011 International Conference on Computer Vision*, pages 1116–1123. IEEE.
- Wang, T.-C., Zhu, J.-Y., Kalantari, N. K., Efros, A. A., and Ramamoorthi, R. (2017). Light field video capture using a learning-based hybrid imaging system. *ACM Transactions on Graphics (TOG)*, 36(4):1–13.
- Weickert, J. and Schnörr, C. (2001). Variational optic flow computation with a spatio-temporal smoothness constraint. *Journal of mathematical imaging and vision*, 14(3):245–255.
- Zhao, H., Shi, B., Fernandez-Cull, C., Yeung, S.-K., and Raskar, R. (2015). Unbounded high dynamic range photography using a modulo camera. In *2015 IEEE International Conference on Computational Photography (ICCP)*, pages 1–10. IEEE.