



# Comparison of the Optical Flow Quality for Video Denoising

Nelson Monzón<sup>1</sup> <sup>a</sup> and Javier Sánchez<sup>2</sup> <sup>b</sup>

<sup>1</sup>*CMLA, École Normale Supérieure, Université Paris-Saclay, France*

<sup>2</sup>*CTIM, Department of Computer Science, University of Las Palmas de Gran Canaria, Spain*

**Keywords:** Optical Flow, Video Denoising, Motion Trajectories, Temporal Filtering.

**Abstract:** Video denoising techniques need to understand the motion present in the scenes. In the literature, many strategies guide their temporal filters according to trajectories controlled by optical flow. However, the quality of these flows is rarely investigated. In fact, there are very few studies that compare the behavior of denoising proposals with different optical flow algorithms. In that direction, we analyze several methods and their performance using a general pipeline that reduces the noise through an average of the pixel's trajectories. This ensures that the denoising strongly depends on the optical flow. We also analyze the behavior of the methods at occlusions and illumination changes. The pipeline incorporates a process to get rid of these effects, so that they do not affect the comparison metrics. We are led to propose a ranking of optical flows methods depending on their efficiency for video denoising, that mainly depends on their complexity.

## 1 INTRODUCTION

Video denoising (Boulanger et al., 2007; Arias and Morel, 2018) is a key problem in image processing. It generally requires a temporal noise filtering guided by motion trajectories. In this regard, optical flow estimation (Horn and Schunck, 1981; Lucas and Kanade, 1981) is quite useful because it provides consistent information of the apparent displacement of the pixels through an image sequence. Therefore, the computed solutions can be used to describe a trajectory that guides the filtering according to the vector fields it provides.

Several works have been published to improve the optical flow computation in noisy images (Spies and Schar, 2001; Schar and Spies, 2005), which is one of the main challenges nowadays. Besides, the current literature about video denoising assumes that the motion estimation between frames is an advantage (Liu and Freeman, 2010; Buades et al., 2016), without exploring its influence in depth.


Nevertheless, in spite of research works like (Larsson and Söderström, 2015), there are not many articles that study the influence of the motion trajectories in video denoising. This is important as pointed out in (Ehret et al., 2018), where the authors observe the limitations of their results due to the optical flow


method. In this sense, this work compares different algorithms and analyze their influence when used for video denoising. The purpose is not to improve the current denoising methods but to observe the behavior of accurate motion trajectories in the final results.

We use a denoising framework that averages the pixel's intensity through the trajectories obtained with the corresponding optical flow method. The reason behind this strategy is to ensure that the accuracy of the denoising depends basically on the vector fields. We use a centered average to reduce the impact of possible changes in the context of a scene. In our experiments, we compare the optical flow methods proposed by Horn and Schunck (Horn and Schunck, 1981), Brox *et al.* (Brox et al., 2004), Zach *et al.* (Zach et al., 2007) and Monzón *et al.* (Monzón et al., 2016). Arguably, an algorithm that uses temporal information could be helpful to obtain consistent trajectories. Thus, we also include in our experiments the temporal method proposed in Sánchez *et al.* (Sánchez et al., 2013) and a temporal extension of the Brox *et al.* method.

## 2 DENOISING FRAMEWORK

Next, we briefly describe the main features of the denoising framework used to compare the influence of the optical flow methods in noise reduction.

<sup>a</sup>  <https://orcid.org/0000-0003-0571-9068>

<sup>b</sup>  <https://orcid.org/0000-0001-8514-4350>

In the first step, we introduce additive white Gaussian noise (AWGN), with standard deviation  $\sigma$ , to the input sequence. Then, we compute the optical flows for the whole noisy sequence in both directions. For each image  $i$ , the neighboring frames  $[i - n, i + n]$  are warped to the central image according to the computed flows, using bicubic interpolation. The value of  $n$  must be small in practice due to changes in pixel intensities, occlusions or brightness changes that worsen the denoising.

We must notice that motion and optical flow estimation are not equivalent goals (Verri and Poggio, 1989; Sellent et al., 2012). The physical 3D motions do not always match with the apparent displacement described by the optical flow in the image plane. This is evident in many situations like occlusions, illumination changes, shadows, etc. In these cases, the true motion is not a good choice for video denoising.

Figure 1 shows an example using the eighth frame of the *Alley1* sequence from the *Sintel* dataset (Butler et al., 2012). The first row shows the central image while the second row depicts the denoised one. The latter has been obtained according to the trajectories described by the ground truths. The red and green rectangles makes zoom in some regions from the results of the first column.

Comparing both images, we observe a “halo effect” in the denoised image. This typically occurs in occluded regions, where it is not possible to find the corresponding pixels in all the frames. The image mask on the the third row represents the pixels where the color difference is higher than a given threshold. Here, it is easy to observe the occluded regions as well as other regions affected by other effects.

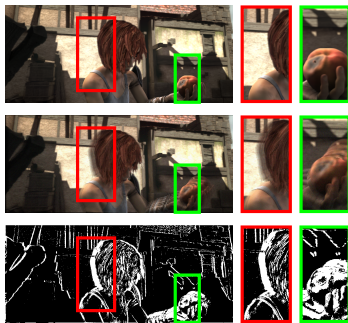


Figure 1: *Alley1* and its denoised image. The first column shows the central frame, the denoised image without discarding pixels and the image mask. The second and third columns show several details of the images at the first column. The “halo effect” appears in the occluded regions, which are detected on the mask.

In our experimental framework, the original sequences are randomly disturbed by adding white Gaussian noise (AWGN) according to a known stan-

dard deviation,  $\sigma$ . We discard the pixels for which the quadratic difference between the original and the warped images are bigger than  $2 \cdot \sigma^2$ . In these cases, the noisy pixels are used for the final result and the image mask is activated in those positions as in the last column of Fig. 1.

Figure 2 shows a diagram of the framework. The denoised image is calculated as an average of the color values in each pixel. Additionally, we estimate an error function as the difference between the denoised values and the pixels in the central image. This allows us to detect pixels that suffer from occlusions or brightness changes, or that are not correctly matched by the optical flow method. Finally, we obtain the root mean square error (RMSE) and peak signal-to-noise ratio (PSNR) between the denoised and the original clean image.

The video denoising framework also calculates the  $RMSE_m$  and  $PSNR_m$  removing the pixels of the mask, and the percentage of pixels (density) used in the average. We also calculate the ratio between the  $RMSE_m$  and the density as  $RMSE_d = \frac{RMSE_m}{density}$ .

### 3 EXPERIMENTS

We compare the performance of several optical flow methods for video denoising and also the results given by the ground truth motions (GT). Our experiments include the algorithms published in (Monzón et al., 2016; Sánchez et al., 2013; Sánchez et al., 2013; Meinhardt-Llopis et al., 2013) of the original methods (Monzón et al., 2016; Brox et al., 2004; Zach et al., 2007; Horn and Schunck, 1981), respectively. We shall refer to them as RDPOF, ROF, TV-L1 and HS, respectively. The method of Sánchez *et al.* (Sánchez et al., 2013) and the temporal extension of ROF are named as TCOF and  $ROF_T$ . The standard datasets only provide the forward optical flows, so we calculate the backward motions using the method proposed in (Sánchez et al., 2015).

First, we compare the results for a given frame using the sequences of *Alley1* (Fig. 3) and *Bandage1* (Fig. 4) from the *Sintel* dataset. Figure 5 shows the results from a video of a person moving his right arm. The number of images of the sequences is 50 and the size is  $1024 \times 436$ , while the *Arm* video contains 255 frames and its size is  $320 \times 240$ . The level of noise introduced in these images is for  $\sigma = 10$ .

In each figure, the first row contains two consecutive noisy frames, the original image and the best denoised result. The following rows show the results using the ground truth, RDPOF, ROF, TV-L1, HS,  $ROF_T$  and TCOF, respectively. The first col-



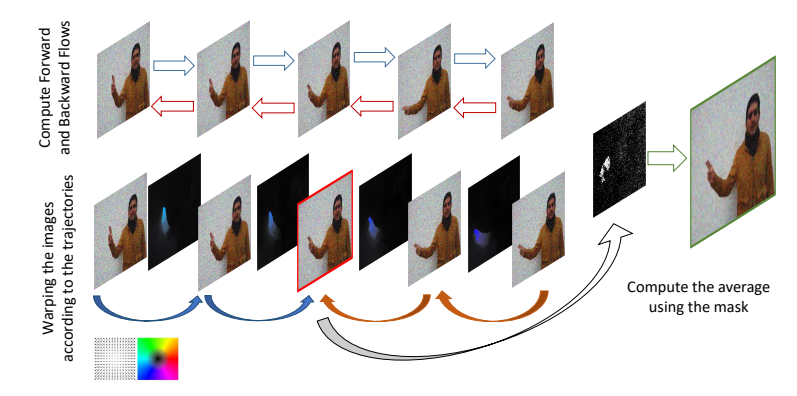


Figure 2: Pipeline for video denoising: First, we add white Gaussian noise (AWGN) to the image sequence; second, the optical flow is calculated in both directions between consecutive frames; then, each image is denoised with the information of several neighbor images; and, finally, we compute the root mean square error (RMSE) with respect to the original image. In this figure, we illustrate an example with 2 frames in each direction.

umn depicts the calculated flow from the central to its successive frame. The second column shows the error image between the original and the denoised images, taking into account the occlusion mask, which is shown in the third column. The fourth one shows the error image without using the mask. Below the images, we show the average angular error (AAE) and the end-point error (EPE), that are standard metrics to measure the optical flow quality (Baker et al., 2007), the parameters used, and the corresponding PSNR, RMSE and *density*, in each case. In these experiments we used 9 frames ( $n = 4$ ).

Analyzing the results, we observe that the  $PSNR_m$  and  $RMSE_m$  using the masks are similar but the AAE and EPE are very different. The RMSE difference between the ground truth and the other optical flow results is around 5%, except for the temporal methods. This means that they are equally good for denoising purposes.

We also notice that, interestingly, the images calculated without the masks achieve good results in spite of the fact that the vector fields are clearly different in several methods. For instance, the flows found for HS in Fig. 4 are evidently worse than those obtained with the RDPOF and ROF methods. However, the differences in the denoising results are not significant. The results for the ground truth are the worst due to the halo effect produced at occlusions.

Table 1 shows the numerical results of the same sequences but for the whole videos. The  $\sigma$  values are 10 and 50. It also includes the average of the runtime required by the optical flow methods. We write in boldface the best RMSE for each sequence.

In general, the temporal methods provide the higher errors and require the biggest execution times, specially TCOF. On the other hand, the results con-

firm that the best optical flows are not necessarily the best for denoising. Furthermore, optical flow fields that are clearly worse compared to the ground truth, are not so different for video denoising.

In our opinion, the most desired method for denoising is TV-L1, although its results are not always the best. The error differences with respect to the other alternatives are not significant. Besides, if we take into account the speed, this method is by far the best choice.

## 4 CONCLUSION

In this work, we analyzed the influence of several optical flow methods for noise reduction. Experiments prove that optical flow is very useful but its accuracy is not as relevant as expected. We observed that the methods yielding the best AAE and EPE are not necessarily associated with the best denoising results. This can be easily observed with the ground truth motion, where the denoised sequences are usually the worst. Here it is important to distinguish between optical flow and motion estimation: The motion of real scenes generates trajectories that do not necessarily preserve the intensity of the pixels along the sequence, while the optical flow finds correspondences based on the similarity of the intensities. For this reason, the results of such different strategies, like Horn and Schunck or TV-L1, are very similar.

At the lights of our results, we may conclude that the best strategy for video denoising is the TV-L1 method. This is not for the accuracy of its flow fields in general, but for its fast running times.

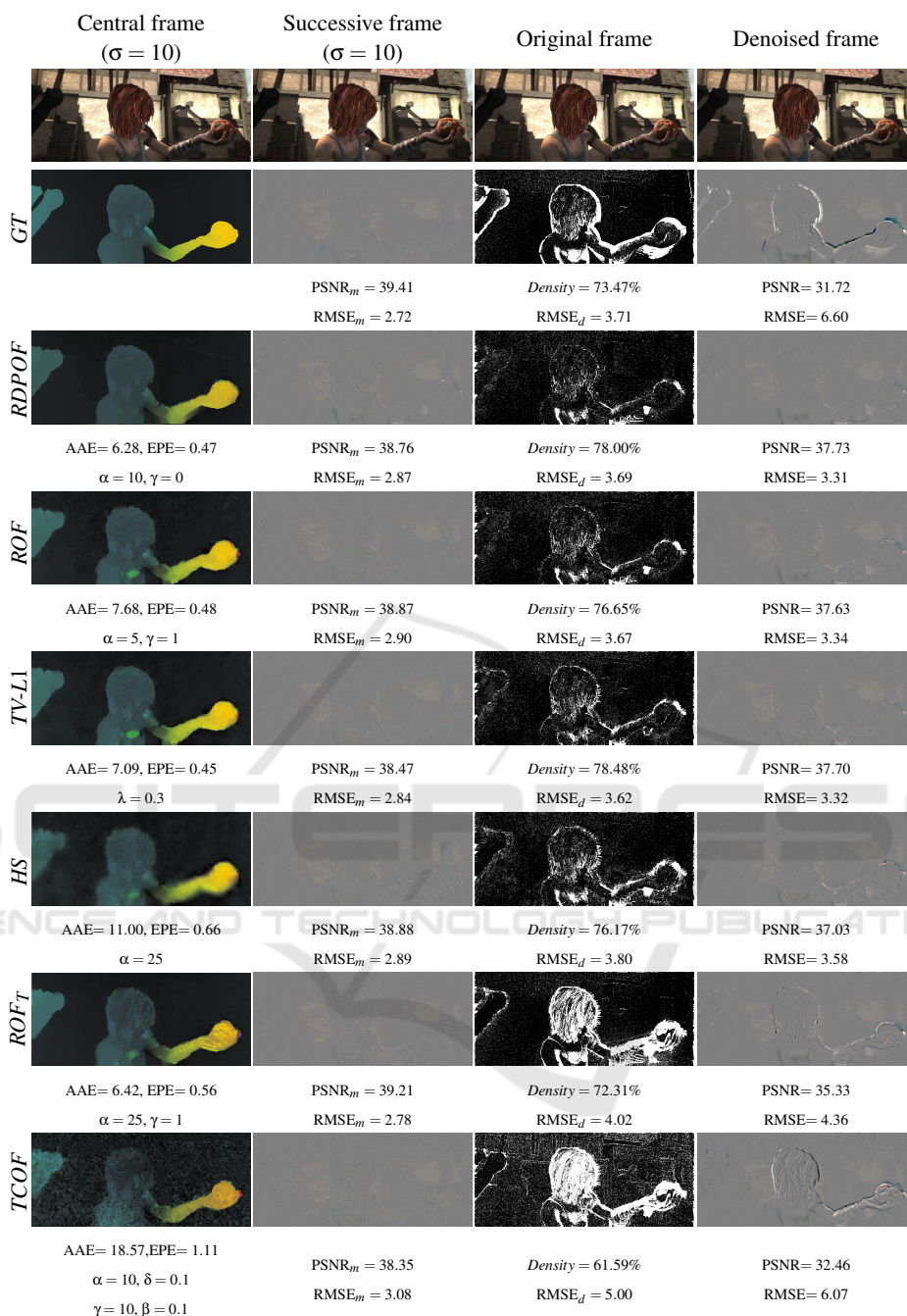


Figure 3: Denoising results for the *Alley1* sequence. In the first row, two images of the *Alley1* sequence with Gaussian noise of  $\sigma = 10$ , the original image, and the denoised result are shown. In the second row, we show the ground truth (GT) motion, the error image between the original frame and the denoised result using the mask (third column), and the error image without the mask. The rest of the rows correspond to the results of the RDPOF, ROF, TV-L1, HS, ROF<sub>T</sub> and TCOF methods, respectively. See the text for more explanations.

## ACKNOWLEDGEMENTS

This work has been partly financed by Office of Naval research grant N00014-17-1-2552, DGA Astrid

project “filmer la Terre” n° ANR-17-ASTR-0013-01, DGA Defals challenge n° ANR-16-DEFA-0004-01. Special thanks to Prof. Jean-Michel Morel for his guidance and help during this research work.

	Central frame ( $\sigma = 10$ )	Successive frame ( $\sigma = 10$ )	Original frame	Denoised frame
<b>GT</b>				
		PSNR <sub>m</sub> = 38.38 RMSE <sub>m</sub> = 3.07	Density = 48.55% RMSE <sub>d</sub> = 6.32	PSNR = 29.50 RMSE = 8.53
<b>RDPOF</b>				
	AAE = 13.52, EPE = 0.53 $\alpha = 5, \gamma = 0$	PSNR <sub>m</sub> = 38.30 RMSE <sub>m</sub> = 3.10	Density = 69.76% RMSE <sub>d</sub> = 4.44	PSNR = 36.74 RMSE = 3.71
<b>ROF</b>				
	AAE = 15.37, EPE = 0.57 $\alpha = 5, \gamma = 1$	PSNR <sub>m</sub> = 38.44 RMSE <sub>m</sub> = 3.04	Density = 70.23% RMSE <sub>d</sub> = 4.33	PSNR = 36.74 RMSE = 3.71
<b>TV-L1</b>				
	AAE = 21.68, EPE = 0.76 $\lambda = 0.3$	PSNR <sub>m</sub> = 38.78 RMSE <sub>m</sub> = 2.95	Density = 73.35% RMSE <sub>d</sub> = 4.03	PSNR = 36.50 RMSE = 3.84
<b>HS</b>				
	AAE = 26.16, EPE = 1.06 $\alpha = 10$	PSNR <sub>m</sub> = 38.45 RMSE <sub>m</sub> = 3.04	Density = 73.31% RMSE <sub>d</sub> = 4.15	PSNR = 36.30 RMSE = 3.90
<b>ROF<sub>T</sub></b>				
	AAE = 12.65, EPE = 0.52 $\alpha = 5, \gamma = 0$	PSNR <sub>m</sub> = 38.85 RMSE <sub>m</sub> = 2.91	Density = 62.25% RMSE <sub>d</sub> = 4.67	PSNR = 34.56 RMSE = 4.76
<b>TCOF</b>				
	AAE = 21.74, EPE = 0.84 $\alpha = 50, \delta = 0.01$ $\gamma = 10, \beta = 0$	PSNR <sub>m</sub> = 38.73 RMSE <sub>m</sub> = 2.94	Density = 61.74% RMSE <sub>d</sub> = 4.77	PSNR = 33.63 RMSE = 5.30

Figure 4: Denoising results for the *Bandage1* sequence. In the first row, two images of the *Bandage1* sequence with Gaussian noise of  $\sigma = 10$ , the original image, and the denoised result are shown. In the second row, we show the ground truth (GT) motion, the error image between the original frame and the denoised result using the mask (third column), and the error image without the mask. The rest of the rows correspond to the results of the RDPOF, ROF, TV-L1, HS, ROF<sub>T</sub> and TCOF methods, respectively. See the text for more explanations.



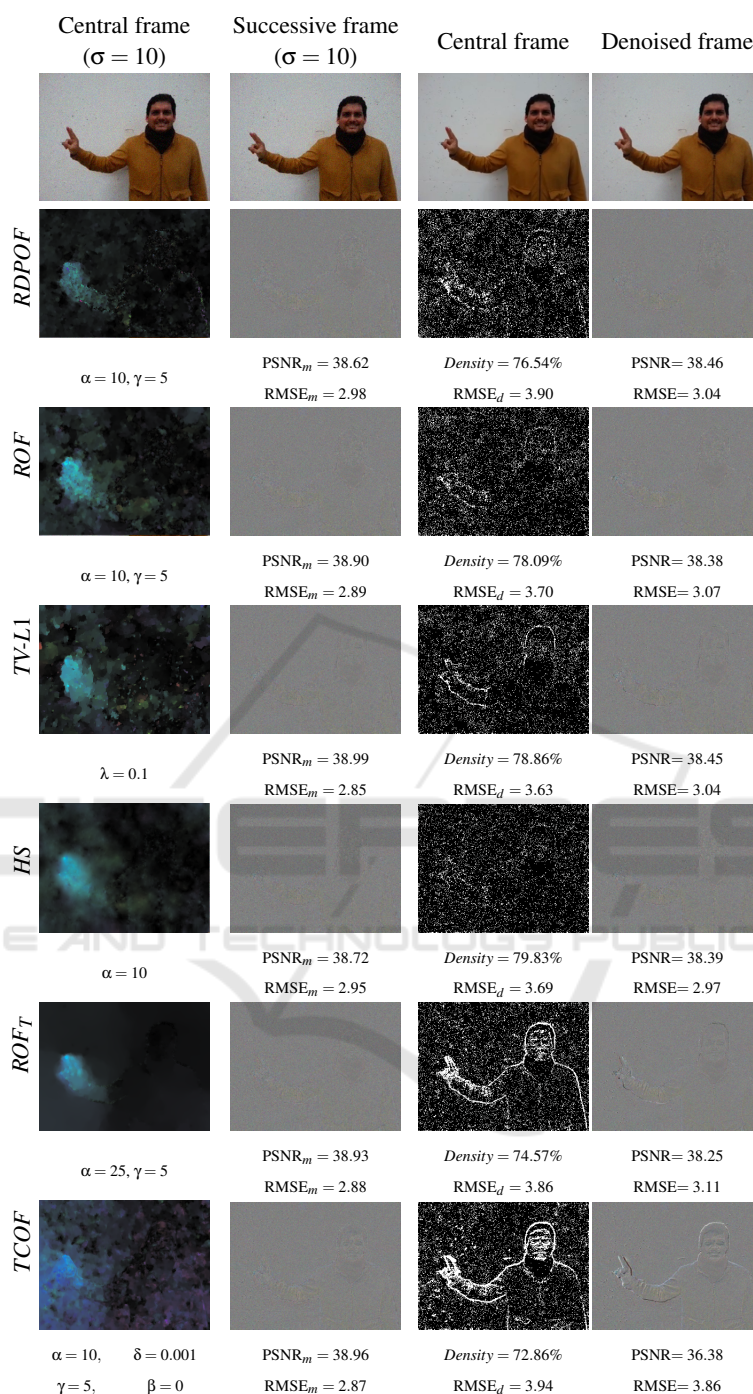


Figure 5: Denoising results for the *Arm* sequence. In the first row, two images of the *Arm* sequence with Gaussian noise of  $\sigma = 10$ , the original image, and the denoised result are shown. In the second row, we show the ground truth (GT) motion, the error image between the original frame and the denoised result using the mask (third column), and the error image without the mask. The rest of rows correspond to the results of the RDPOF, ROF, TV-L1, HS, ROF<sub>T</sub> and TCOF methods, respectively. See the text for more explanations.



Table 1: Best AAE, EPE, average runtime to obtain the optical flows, RMSE, RMSE<sub>m</sub> and densities (%) found for the complete videos of *Alley1*, *Bandage1* and the *Arm* sequence. The noise values are  $\sigma = 10$  and  $\sigma = 50$ .

		<i>Alley1</i> ( $\sigma = 10$ )	<i>Alley1</i> ( $\sigma = 50$ )	<i>Bandage1</i> ( $\sigma = 10$ )	<i>Bandage1</i> ( $\sigma = 50$ )	<i>Arm</i> ( $\sigma = 10$ )	<i>Arm</i> ( $\sigma = 50$ )
GT	RMSE	7.60	16.07	8.53	18.50	-	-
	RMSE <sub>m</sub>	2.89	14.16	3.24	15.21	-	-
	Densities	70.14%	81.69%	41.96%	68.10%	-	-
RDPOF	AAE	7.82 <sup>o</sup>	14.23 <sup>o</sup>	15.80 <sup>o</sup>	29.48 <sup>o</sup>	-	-
	EPE	0.61	1.11	1.23	1.89	-	-
	Time	3.6(s)	3.85(s)	3.57(s)	3.87(s)	0.6(s)	0.8(s)
	RMSE	<b>3.73</b>	15.25	<b>4.15</b>	15.09	3.13	14.12
	RMSE <sub>m</sub>	3.05	14.77	3.26	14.69	3.00	13.85
	Densities	77.37%	83.89%	65.91%	84.08%	76.31%	83.51%
ROF	AAE	7.51 <sup>o</sup>	10.76 <sup>o</sup>	16.01 <sup>o</sup>	26.68 <sup>o</sup>	-	-
	EPE	0.60	0.94	1.20	1.64	-	-
	Time	5.43(s)	6.39(s)	5.67(s)	6.39(s)	1.6(s)	2(s)
	RMSE	3.77	14.80	4.34	14.96	3.11	13.93
	RMSE <sub>m</sub>	3.02	14.30	3.21	14.53	2.91	13.81
	Densities	76.65%	83.75%	64.52%	83.50%	77.15%	84.81%
TV-L1	AAE	7.02 <sup>o</sup>	14.48 <sup>o</sup>	21.69 <sup>o</sup>	28.64 <sup>o</sup>	-	-
	EPE	0.58	0.99	2.11	1.69	-	-
	Time	2(s)	2.6(s)	2(s)	2.2(s)	0.25(s)	0.21(s)
	RMSE	3.87	<b>14.78</b>	4.44	<b>14.94</b>	3.09	<b>13.77</b>
	RMSE <sub>m</sub>	2.96	14.37	3.11	14.48	2.88	13.72
	Densities	75.67%	84.12%	69.27%	83.58%	78.16%	84.91%
HS	AAE	9.91 <sup>o</sup>	21.12 <sup>o</sup>	22.15 <sup>o</sup>	33.53 <sup>o</sup>	-	-
	EPE	0.75	1.31	1.43	2.04	-	-
	Time	4.8(s)	6.5(s)	6.51(s)	4.83(s)	5.35(s)	5.86(s)
	RMSE	4.26	14.98	4.84	15.23	<b>3.07</b>	13.98
	RMSE <sub>m</sub>	3.00	14.53	3.16	14.70	2.96	13.76
	Densities	73.97%	84.10%	68.14%	83.38%	79.83%	84.62%
ROF <sub>T</sub>	AAE	6.81 <sup>o</sup>	12.11 <sup>o</sup>	15.87 <sup>o</sup>	26.41 <sup>o</sup>	-	-
	EPE	0.64	0.94	1.41	1.77	-	-
	Time	6.9(s)	7.47(s)	7.32(s)	7.68(s)	1.23(s)	1.82(s)
	RMSE	5.33	15.1	6.58	15.72	3.39	14.02
	RMSE <sub>m</sub>	2.91	14.43	3.14	14.72	2.91	13.71
	Densities	70.40%	83.31%	52.41%	81.27%	75.26%	83.28%
TCOF	AAE	19.80 <sup>o</sup>	24.33 <sup>o</sup>	27.62 <sup>o</sup>	30.45 <sup>o</sup>	-	-
	EPE	1.26	1.72	2.25	2.83	-	-
	Time	88.52(s)	75.52(s)	79.18(s)	52.12(s)	14.27(s)	10.18(s)
	RMSE	5.75	15.95	8.55	16.60	3.55	14.28
	RMSE <sub>m</sub>	3.18	15.01	3.05	15.24	2.88	13.90
	Densities	65.06%	82.39%	45.75%	79.83%	74.19%	83.87%

## REFERENCES

- Arias, P. and Morel, J.-M. (2018). Video denoising via empirical bayesian estimation of space-time patches. *Journal of Mathematical Imaging and Vision*, 60(1):70–93.
- Baker, S., Scharstein, D., Lewis, J., Roth, S., Black, M., and Szeliski, R. (2007). A database and evaluation methodology for optical flow. In *International Conference on Computer Vision (ICCV 2007)*.
- Boulanger, J., Kervrann, C., and Bouthemy, P. (2007). Space-time adaptation for patch-based image sequence restoration. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(6):1096–1102.
- Brox, T., Bruhn, A., Papenberg, N., and Weickert, J. (2004). High accuracy optical flow estimation based on a the-

- ory for warping. In Pajdla, T. and Matas, J., editors, *European Conference on Computer Vision (ECCV)*, volume 3024 of *LNCS*, pages 25–36, Prague, Czech Republic. Springer.
- Buades, A., Lisani, J., and Miladinović, M. (2016). Patch-based video denoising with optical flow estimation. *IEEE Transactions on Image Processing*, 25(6):2573–2586.
- 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.
- Ehret, T., Morel, J., and Arias, P. (2018). Non-local kalman: A recursive video denoising algorithm. In *2018 25th IEEE International Conference on Image Processing (ICIP)*, pages 3204–3208.
- Horn, B. and Schunck, B. (1981). Determining optical flow. *MIT Artificial Intelligence Laboratory*, 17:185–203.
- Larsson, M. and Söderström, L. (2015). Analysis of optical flow algorithms for denoising. Student Paper.
- Liu, C. and Freeman, W. T. (2010). A high-quality video denoising algorithm based on reliable motion estimation. In Daniilidis, K., Maragos, P., and Paragios, N., editors, *Computer Vision – ECCV 2010*, pages 706–719, Berlin, Heidelberg. Springer Berlin Heidelberg.
- Lucas, B. D. and Kanade, T. (1981). An iterative image registration technique with an application to stereo vision. In *Proceedings of the 7th international joint conference on Artificial intelligence - Volume 2*, pages 674–679, San Francisco, CA, USA. Morgan Kaufmann Publishers Inc.
- Meinhardt-Llopis, E., Sánchez, J., and Kondermann, D. (2013). Horn-Schunck Optical Flow with a Multi-Scale Strategy. *Image Processing On Line*, 3:151–172.
- Monzón, N., Salgado, A., and Sánchez, J. (2016). Regularization strategies for discontinuity-preserving optical flow methods. *IEEE Transactions on Image Processing*, 25(4):1580–1591.
- Monzón, N., Salgado, A., and Sánchez, J. (2016). Robust Discontinuity Preserving Optical Flow Methods. *Image Processing On Line*, 6:165–182.
- Sánchez, J., Meinhardt-Llopis, E., and Facciolo, G. (2013). TV-L1 Optical Flow Estimation. *Image Processing On Line*, 3:137–150.
- Sánchez, J., Monzón, N., and Salgado, A. (2013). Robust optical flow estimation. *Image Processing On Line*, 2013:252–270. <http://dx.doi.org/10.5201/ipol.2013.21>.
- Sánchez, J., Salgado, A., and Monzón, N. (2013). Optical flow estimation with consistent spatio-temporal coherence models. In *International Conference on Computer Vision Theory and Applications (VISAPP)*, pages 366–370. Institute for Systems and Technologies of Information, Control and Communication.
- Sánchez, J., Salgado, A., and Monzón, N. (2015). Computing inverse optical flow. *Pattern Recognition Letters*, 52:32 – 39.
- Scharr, H. and Spies, H. (2005). Accurate optical flow in noisy image sequences using flow adapted anisotropic diffusion. *Sig. Proc.: Image Comm.*, 20:537–553.
- Sellent, A., Kondermann, D., Simon, S. J., Baker, S., Dedeoglu, G., Erdler, O., Parsonage, P., Unger, C., and Niehsen, W. (2012). Optical flow estimation versus motion estimation.
- Spies, H. and Scharr, H. (2001). Accurate optical flow in noisy image sequences. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 1, pages 587–592 vol.1.
- Verri, A. and Poggio, T. A. (1989). Motion field and optical flow: Qualitative properties. *IEEE Trans. Pattern Anal. Mach. Intell.*, 11:490–498.
- Zach, C., Pock, T., and Bischof, H. (2007). A Duality Based Approach for Realtime TV-L1 Optical Flow. In Hamprecht, F. A., Schnörr, C., and Jähne, B., editors, *Pattern Recognition*, volume 4713 of *Lecture Notes in Computer Science*, chapter 22, pages 214–223. Springer Berlin Heidelberg, Berlin, Heidelberg.