Illumination-invariant Optical Flow: Application to Endoscopic Image Mosaicing

Dinh Hoan TRINH[1], Walter BLONDEL[1], Dominique LAMARQUE[2], Christian DAUL[1]

[1]Université de Lorraine and CNRS, CRAN, UMR 7039
2 avenue de la Forêt de Haye, 54518 Vandœuvre-lès-Nancy cedex, France

[2]Hôpital Ambroise Paré
9 Avenue Charles de Gaulle, 92100 Boulogne-Billancourt, France

dinh-hoan.trinh@univ-lorraine.fr, walter.blondel@univ-lorraine.fr
dominique.lamarque@aphp.fr, christian.daul@univ-lorraine.fr

Résumé – Cette contribution traite du calcul du flot optique dans des scènes faiblement texturées et affectées par des changements d'illuminations importants. Un nouveau descripteur invariant à l'illumination est proposé dans le terme d'attache aux données de la fonctionnelle à minimiser. Un modèle local de changement d'illumination a été utilisé pour prouver théoriquement que le descripteur est invariant à des changements complexes d'éclairage. Des tests avec des images simulées et des données endoscopiques réelles (cliniques) permettent d'évaluer la précision du flot optique et de mettre en lumière la robustesse induite par le descripteur.

Abstract – This contribution deals with the optical flow computation in weakly textured scenes affected by strong illumination changes. The data-term is based on a new illumination invariant patch-based descriptor which also preserves the texture information. An illumination change model was used to theoretically prove that the descriptor is invariant to complex illumination changes. Tests with complicated simulated images and challenging clinical endoscopic data allowed us to assess the accuracy of the OF and to highlight the robustness induced by the descriptor.

1 Introduction

Optical flow (OF) plays a key role in numerous computer vision applications. Computing precise OF fields under complicated illumination conditions is at stake and frequently required in the medical field. Although OF determination has been studied for decades, accurate OF estimation is still a challenge when mosaics have to be constructed from endoscopic video-sequences (e.g, in gastroscopy [1] or cystoscopy [2]). Although this contribution focuses on endoscopic images, the proposed OF algorithm can deal with many other complex scenes.

In gastroscopy, an endoscope is used to scan the inner surface of the stomach. By extending the field of view, mosaicing of gastroscopic images should facilitates the diagnosis and follow-up of stomach inflammatory diseases. The flow field gives the correspondence of homologous pixels between image pairs of the video-sequence and and its computation is a crucial step for the data registraton and mosaicing [3].

As illustrated in Figs. 4(a)-4(e), gastroscopic images show low or no texture and changing illumination conditions between the acquisitions. Moreover, images can be affected by specular reflections or blurred. For such scenes, the OF methods based on matching sparse keypoints [4] may not achieve the necessary registration performances in terms of robustness and accuracy. Dense OF approaches often lead to higher accuracy than sparse OF methods, especially when images are weakly textured. Thus, dense OF estimation based on a variational model is more appropriate to gastroscopic images.

Variational OF models can be formulated as follows. Given a source image I_s and a target image I_t, the dense flow field $\mathbf{u} = (u_x, u_y)$ between I_s and I_t is computed by minimizing

$$E(\mathbf{u}) = E_{reg}(\mathbf{u}) + \lambda E_{data}(I_s, I_t, \mathbf{u}), \quad (1)$$

where E_{reg} is a regularization term that assumes smoothness of solution $\mathbf{u}$, E_{data} is a data-term that measures the similarity of the pixels in I_s and I_t, and $\lambda > 0$ is a parameter controlling the relative importance of these two terms. In Eq. (1), E_{data} plays the key role in handling illumination changes. Numerous improved OF methods have been proposed [1,5–10] since the first variational methods of Horn and Schunck [11] and Nagel [12]. Recently, several methods [1, 8, 10] using local descriptors in the data-term have proven their robustness in estimating OF under varying illumination conditions.

This contribution proposes a variational OF method based on a novel illumination-invariant descriptor. The illumination invariance of the proposed descriptor is theoretically proven based on a linear illumination change model. Experimental results on the Middlebury data-base [13] and on real gastroscopic images show that the proposed method achieves high accuracy OF estimation for scenes with few (or no) textures and strong illumination changes.

2 Proposed OF method

2.1 Descriptor-based variational OF model

Let $P_I(\mathbf{x})$ denote small patches centered at pixel $\mathbf{x}$ in image I. These patches have a size $(2k+1) \times (2k+1)$ pixels with k a positive integer. Descriptor $\mathbf{D}(P_I(\mathbf{x}))$ "encodes" the values of the pixels in patch $P_I(\mathbf{x})$ and is given as a vector in a space $\mathbb{R}^m$ $(m > 0)$. $\mathbf{D}(P_I(\mathbf{x}))$ can be used in the descriptor-based OF variational model defined by

$$\min_{\mathbf{u}} [E(\mathbf{u}) = E_{reg}(\mathbf{u}) + \lambda E_{data}(I_s, I_t, \mathbf{u})] \tag{2}$$

where

$$E_{data} = \sum_{\mathbf{x}\in\Omega} \|\mathbf{D}(P_{I_s}(\mathbf{x})) - \mathbf{D}(P_{I_t}(\mathbf{x}+\mathbf{u_x}))\|^2, \tag{3}$$

$$E_{reg} = \sum_{\mathbf{x}\in\Omega} \sum_{\mathbf{x}'\in\mathcal{N}_{\mathbf{x}}} w_{\mathbf{x}}^{\mathbf{x}'} \|\mathbf{u_x} - \mathbf{u_{x'}}\|_1. \tag{4}$$

In the data-term defined by Eq. (3), Ω stands for the image domain and $\mathbf{u_x}$ denotes the motion vector at pixel $\mathbf{x}$ between I_s and I_t. The regularization term in Eq. (4) is a non-local total variation of flow field $\mathbf{u}$. E_{reg} is computed with the set of pixels defined by neighborhoods $\mathcal{N}_{\mathbf{x}}$ centered on $\mathbf{x}$ and using weights $w_{\mathbf{x}}^{\mathbf{x}'}$ depending on the similarity of pixels $\mathbf{x}$ and $\mathbf{x}'$. Similarly to [8], the weights $w_{\mathbf{x}}^{\mathbf{x}'}$ are defined as follows :

$$w_{\mathbf{x}}^{\mathbf{x}'} = \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma_1^2} - \frac{\|L(\mathbf{x}) - L(\mathbf{x}')\|^2}{2\sigma_2^2}\right), \tag{5}$$

where σ_1 and σ_2 are parameters controlling the similarity measure, and $L(\mathbf{x})$ is the color vector in the CIE Lab space.

The minimization of energy $E(\mathbf{u})$ defined by Eqs. (2)-(4) is performed with the projected-proximal-point algorithm [14]. Interested readers can refer to [8] for more details about the general OF scheme used in this contribution.

2.2 New illumination-invariant descriptor

The proposed descriptor is inspired by a binary descriptor (MLDP, [10]) which is a modified (M) version of the local directional pattern (LDP) presented in [15]. MLDP is defined by :

$$\mathbf{D}_{MLDP}(P_I(\mathbf{x})) = \begin{bmatrix} \mathrm{sgn}(M_1 \otimes P_I(\mathbf{x})) \\ \mathrm{sgn}(M_2 \otimes P_I(\mathbf{x})) \\ \vdots \\ \mathrm{sgn}(M_8 \otimes P_I(\mathbf{x})) \end{bmatrix}, \tag{6}$$

where $M_1, M_2, \ldots, M_8$ are the Kirsch edge kernels used to compute edge responses in eight directions (see Fig. 1), $P_I(\mathbf{x})$ is a patch of size 3×3 and centered at pixel $\mathbf{x}$, and $\mathrm{sgn}(\upsilon)$ is a sign function defined by $\mathrm{sgn}(\upsilon) = 1$ if $\upsilon > 0$ and $\mathrm{sgn}(\upsilon) = 0$ otherwise. Operator $\otimes$ in Eq. (6) gives the sum of the element-wise products of two matrices. The performance of descriptor MLDP was experimentally validated by the high ranking of the OF method in [10] on the Middlebury [1] and the KITTI [2] benchmarks. Although results in [10] showed that MLDP may handle illumination, it is however sensitive to strong illumination changes [1]. In the aim of improving illumination invariance of the latter, let us first assume that in small patches the illumination change between (I_s, I_t) pairs can be modelled by the following linear model :

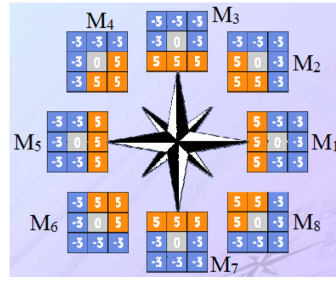


FIGURE 1 – Kirsch kernels for the MLDP descriptor.

$$P_{I_t}(\mathbf{x}+\mathbf{u_x}) = a_{\mathbf{x}} P_{I_s}(\mathbf{x}) + b_{\mathbf{x}}, \tag{7}$$

where $P_{I_s}(\mathbf{x})$ and $P_{I_t}(\mathbf{x}+\mathbf{u_x})$ are two corresponding patches in I_s and I_t, and $a_{\mathbf{x}} \in \mathbb{R}_{>0}, b \in \mathbb{R}$ ($a_{\mathbf{x}}$ is positive since the intensity values of pixels are non-negative). Therefore, a descriptor $\mathbf{D}$ is invariant to illumination changes when

$$\mathbf{D}(a_{\mathbf{x}} P_I(\mathbf{x}) + b_{\mathbf{x}}) = \mathbf{D}(P_I(\mathbf{x})) \tag{8}$$

for all pixels $\mathbf{x}$ in image I, and for all $a_{\mathbf{x}} \in \mathbb{R}_{>0}, b \in \mathbb{R}$.

If one considers the MLDP descriptor in Eq. (6), by representing $P_I(\mathbf{x})$ as a vector of its intensity values $P_I(\mathbf{x}) = [I(\mathbf{x}_1), I(\mathbf{x}_2), \ldots, I(\mathbf{x}_9)]^T$ and M_i $(i = 1, 2, \ldots, 8)$ as a vector of the coefficient $M_i(\mathbf{x}) = [\alpha_{i,1}, \alpha_{i,2}, \ldots, \alpha_{i,9}]^T$, one gets :

$$M_i \otimes P_I(\mathbf{x}) = \sum_{j=1}^{9} \alpha_{i,j} I(\mathbf{x}_j). \tag{9}$$

Then,

$$M_i \otimes (a_{\mathbf{x}} P_{I_s}(\mathbf{x}) + b_{\mathbf{x}}) = a_{\mathbf{x}} \sum_{j=1}^{9} \alpha_{i,j} I(\mathbf{x}_j) + b_{\mathbf{x}} \sum_{j=1}^{9} \alpha_{i,j}. \tag{10}$$

Note that for all the Kirsch kernels M_i, the sum $\alpha_{i,1} + \alpha_{i,2} + \ldots \alpha_{i,9} = 0$, and thus

$$M_i \otimes (a_{\mathbf{x}} P_I(\mathbf{x}) + b_{\mathbf{x}}) = a_{\mathbf{x}} M_i \otimes P_I(\mathbf{x}), \forall i. \tag{11}$$

Because $a_{\mathbf{x}}$ is assumed to be a positive number,

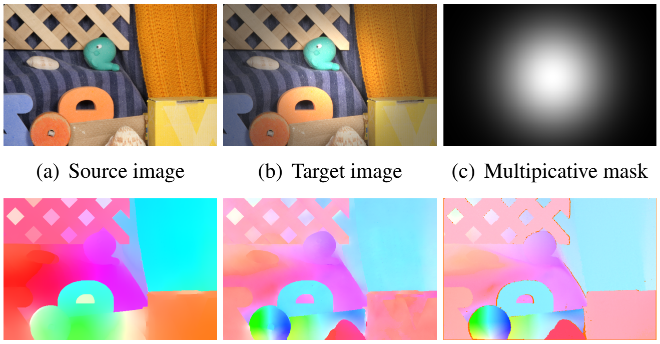
$$\mathrm{sgn}\left(M_i \otimes (a_{\mathbf{x}} P_{I_s}(\mathbf{x}) + b_{\mathbf{x}})\right) = \mathrm{sgn}(M_i \otimes P_I(\mathbf{x})), \forall i. \tag{12}$$

Consequently, MLDP satisfies the illumination change model in Eq. (8) which represents complicated variations between images when the patches are small. This justifies the robustness towards illumination changes of the MLDP descriptor. However, the threshold of the sgn function in Eq. (6) results in the loss of valuable texture or structure information in the patches.

A novel descriptor can be defined as follows to overcome this drawback of the MLDP descriptor :

$$\mathbf{D}(P_I(\mathbf{x})) = \frac{\mathcal{V}(P_I(\mathbf{x}))}{\|\mathcal{V}(P_I(\mathbf{x}))\|}, \tag{13}$$

1. http ://vision.middlebury.edu/flow/eval/
2. http ://www.cvlibs.net/datasets/kitti/index.php



(a) Source image (b) Target image (c) Multipicative mask

(d) MLDP (e) Proposed descriptor (f) Ground truth OF

FIGURE 2 – OF results. In (d)-(f) the classical motion color code is used to visualize the flow fields.

with $\mathcal{V}_{\mathbf{x}} = [M_1 \otimes P_I(\mathbf{x}), M_2 \otimes P_I(\mathbf{x}), \ldots, M_8 \otimes P_I(\mathbf{x})]^T \in \mathbb{R}^8$. From (11), one can deduce

$$\mathcal{V}\left(a_{\mathbf{x}} P_I(\mathbf{x}) + b_{\mathbf{x}}\right) = a_{\mathbf{x}} \mathcal{V}(P_I(\mathbf{x})), \tag{14}$$

$$\|\mathcal{V}(a_{\mathbf{x}} P_I(\mathbf{x}) + b_{\mathbf{x}})\| = a_{\mathbf{x}} \|\mathcal{V}(P_I(\mathbf{x}))\|. \tag{15}$$

Taking the ratio of the respective sides of equalities in both Eqs. (14) and Eq. (15) leads to $\mathbf{D}(a_{\mathbf{x}} P_I(\mathbf{x}) + b_{\mathbf{x}}) = \mathbf{D}(P_I(\mathbf{x}))$. Consequently, the new descriptor in Eq. (13) is also illumination invariant.

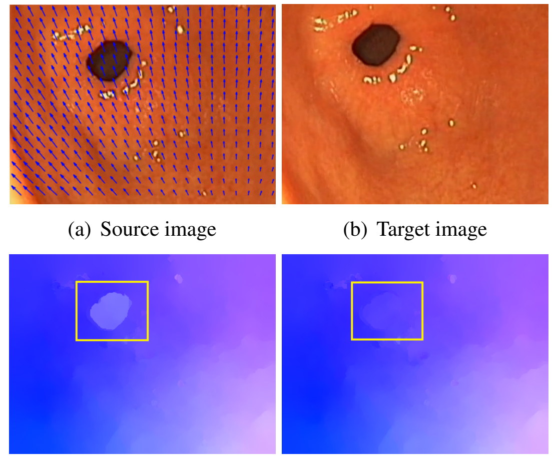
The advantage of this method is two-fold : (i) the use of the Kirsch kernels without thresholding functions preserves local spatial textures due to pixel values changes in the patches, and (ii) the illumination variation model between small patches (see Eq. (7)) facilitates the design of descriptors which are robust against complex illumination changes (see [16]).

3 Results and discussion

This section compares the performances of the proposed descriptor with those of MLDP, both descriptors being used in the same variational OF scheme for a fair result assessment.

The well-known coarse-to-fine warping strategy is used in the OF scheme of Section 2.1 to cope with large displacements. Parameters σ_1 and σ_2 in Eq. (5) were set to 3 and 5, respectively. The λ-parameter in Eq. (2) and the scale factor (denoted by Py_s) in the coarse-to-fine strategy were experimentally adjusted using the Middlebury training set [13]. For each descriptor, the parameter pair (λ, Py_s) leading to the most accurate OF results was chosen for all the tests in this section. The values of pair (λ, Py_s) were set to (10, 0.5) and (40, 0.5) for the MLDP descriptor and the proposed descriptor, respectively.

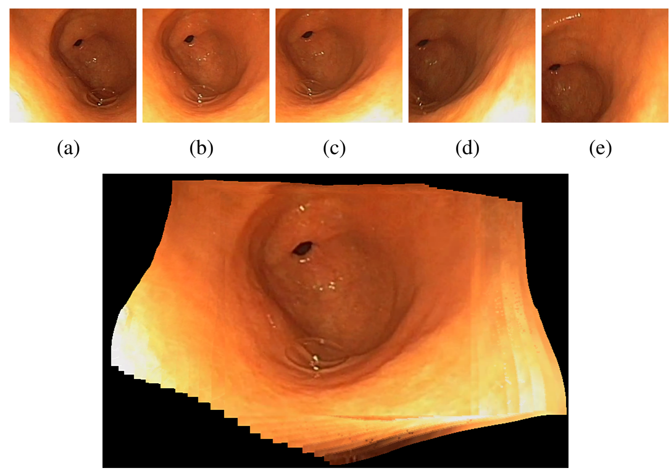
It is worth noticing that all parameters were adjusted with images with weak illumination changes, whereas the tests described in the next sub-sections were performed for strong illumination changes in simulated data with known ground truth (Section 3.1) and real clinical data without ground truth (Section 3.2). This way to proceed was chosen to highlight the robustness of the OF algorithm based on the new descriptor.



(a) Source image (b) Target image

(c) MLDP (d) Proposed descriptor

FIGURE 3 – OF results on a pair of gastroscopic images.



(a) (b) (c) (d) (e)

(f) Mosaiced image

FIGURE 4 – Mosaic built with 21 images of a gastroscopic sequence. Five images of the sequence are given in (a)-(e).

3.1 Tests using simulated illumination changes

For these tests with known ground truth, strong illumination changes were simulated by using the RubberWhale image pair of the Middlebury training set (see Figs. 2(a)-2(b)). Target image I_t (see Fig. 2(b)) was obtained by modifying the color intensities of one of the RubberWhale images with the Gaussian multiplicative mask given in Fig. 2(c) and an additive factor of 20 to simulate the vignetting effect of endoscopes. Source I_s is the other RubberWhale image with unchanged colors.

As visually perceptible, the OF image obtained with the proposed descriptor (see Fig. 2(e)) is significantly closer to the ground truth shown in Fig. 2(f) than the OF of the MLDP descriptor given in Fig. 2(d)). Moreover, the error criteria AEPE (Average End-Point Error) and AAE (Average Angular Error) of MLDP are 5.12 pixels and 19.48 degrees respectively, while the respective values for our proposed descriptor are significantly lower with 0.09 pixels and 2.92 degrees.

3.2 Tests on gastroscopic images

This section gives OF results on gastroscopic images of the pyloric antrum region (stomach) to demonstrate the robustness of the proposed descriptors on complex real scenes.

Fig. 3 shows the vector flow field related to the homologous pixels between the source image in Fig. 3(a) and the target image in Fig. 3(b), the point displacements between the images being caused by both inner stomach and endoscopic camera movements. Although there is a difference in magnitude and direction between the motion vectors of different pixels, the motion fields are smooth for such scenes. Moreover, the motion vectors inside the black "hole" (duodenum) and the flow vectors at the circular border of this "hole" (pyloric sphincter limits) have to be equal since the movement is constant is this region. Visually, the OF result of the proposed descriptor in Fig. 3(d) is smoother than that of descriptor MLDP in Fig. 3(c). As shown in the rectangles of Fig. 3, the flow field of the proposed descriptor is closer to an OF without discontinuities (it is nearly constant onto and around the duodenum black disc) compared to MLDP one.

Fig. 4 shows the mosaicing result computed with a sequence of 21 images. The registration uses the flow fields estimated by the proposed method. The precise mosaic of Fig. 4(f) once again confirms the potential and robustness of the proposed method. The illumination discontinues were intentionally not corrected (as in [17]) to show the image superimposition.

4 Conclusion

This paper presents a method for accurate OF estimation. A new descriptor is introduced which can effectively capture the image spatial structure and is highly invariant to illumination changes. The illumination-invariance of the proposed descriptor was theoretically proven. Experimental results have demonstrated the suitability of the proposed method for registering gastroscopic images and its potential interest for other similar applications.

Acknowledgements

This work was funded by the Agence Nationale de la Recherche (EMMIE project, ANR-15-CE17-0015).

Références

[1] S. Ali, C. Daul, E. Galbrun, and W. Blondel, "Illumination invariant optical flow using neighborhood descriptors," *Computer Vision and Image Understanding*, vol. 145, pp. 95–110, 2016.

[2] Y. Hernandez-Mier, W. Blondel, C.Daul, D. Wolf, and F. Guillemin, "Fast construction of panoramic images for cystoscopic exploration," *Computerized Medical Imaging and Graphics*, vol. 34, no. 7, pp. 579–592, 2010.

[3] T. Weibel, C. Daul, D. Wolf, R. Rösch, and F. Guillemin, "Graph based construction of textured large field of view mosaics for bladder cancer diagnosis," *Pattern Recognition*, vol. 45, no. 12, pp. 4138–4150, 2012.

[4] J. Revaud, P. Weinzaepfel, Z. Harchaoui, and C. Schmid, "Deepmatching : Hierarchical deformable dense matching," *Int. Journal of Computer Vision*, vol. 120, no. 3, pp. 300–323, 2016.

[5] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *ECCV, Czech Republic*, 2004, pp. 25–36.

[6] A. Wedel, T. Brox, C. Zach, H. Bischof, and D. Cremers, "An improved algorithm for tv-l1 optical flow," in *Dagstuhl Motion Workshop*, 2008.

[7] P. Krähenbühl and V. Koltun, "Efficient nonlocal regularization for optical flow," in *ECCV*, 2012, pp. 356–369.

[8] M. Drulea and S. Nedevschi, "Motion estimation using the correlation transform," *IEEE Trans. Image Processing*, vol. 22, no. 8, pp. 3260–3270, 2013.

[9] R. Ranftl, K. Bredies, and T. Pock, "Non-local total generalized variation for optical flow estimation," in *ECCV, Switzerland*, 2014, pp. 439–454.

[10] M. A. Mohamed, H. A. Rashwan, B. Mertsching, M. A. García, and D. Puig, "Illumination-robust optical flow using a local directional pattern," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 24, no. 9, pp. 1499–1508, 2014.

[11] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, no. 1-3, pp. 185–203, 1981.

[12] Hans-Hellmut Nagel, "Constraints for the estimation of displacement vector fields from image sequences," in *The 8th Int. Joint Conf. on Artificial Intelligence. Karlsruhe, FRG*, 1983, pp. 945–951.

[13] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *Int. Journal of Computer Vision*, vol. 92, no. 1, pp. 1–31, 2011.

[14] A. Chambolle and T. Pock, "A first-order primal-dual algorithm for convex problems with applications to imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, no. 1, pp. 120–145, 2011.

[15] M. H. Kabir, T. Jabid, and O. Chae, "A local directional pattern variance (LDPv) based face descriptor for human facial expression recognition," in *IEEE Int. Conf. on Adv. Video and Signal Based Surveil.*, 2010, pp. 526–532.

[16] D. H. Trinh, W. Blondel, and C. Daul, "A general form of illumination-invariant descriptors in variational optical flow estimation," in *IEEE Int. Conf. on Image Processing (ICIP)*, Beijing, China, Sept. 2017.

[17] T. Weibel, C. Daul, D. Wolf, and R. Rösch, "Contrast-enhancing seam detection and blending using graph cuts," in *ICPR*, 2012, pp. 2732–2735.