Illumination-invariant optical flow: Application to endoscopic image mosaicing
Dinh Hoan Trinh, Walter Blondel, Dominique Lamarque, Christian Daul

To cite this version:

HAL Id: hal-01592523
https://hal.archives-ouvertes.fr/hal-01592523
Submitted on 25 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.
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 facilitate 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 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 $P_I(x)$ denote small patches centered at pixel $x$ in image $I$. These patches have a size $(2k+1) \times (2k+1)$ pixels with $k$ a positive integer. Descriptor $D(P_I(x))$ “encodes” the values of the pixels in patch $P_I(x)$ and is given as a vector in a space $\mathbb{R}^m$ ($m > 0$). $D(P_I(x))$ can be used in the descriptor-based OF variational model defined by

$$\min_u |E(u) = E_{reg}(u) + \lambda E_{data}(I_s, I_t, u)|$$

where

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

$$E_{reg} = \sum_{x \in \Omega} \sum_{x' \in N_x} w_{x,x'} \|u_x - u_{x'}\|_1.$$  \hspace{1cm} (4)

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

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

where $\sigma_1$ and $\sigma_2$ are parameters controlling the similarity measure, and $L(x)$ is the color vector in the CIE Lab space.

The minimization of energy $E(u)$ 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:

$$D_{MLDP}(P_I(x)) = \begin{bmatrix} \text{sgn}(M_1 \otimes P_I(x)) \\ \text{sgn}(M_2 \otimes P_I(x)) \\ \vdots \\ \text{sgn}(M_8 \otimes P_I(x)) \end{bmatrix},$$

where $M_1, M_2, \ldots, M_8$ are the Kirsch edge kernels used to compute edge responses in eight directions (see Fig. 1), $P_I(x)$ is a patch of size $3 \times 3$ and centered at pixel $x$, and $\text{sgn}(\nu)$ is a sign function defined by $\text{sgn}(\nu) = 1$ if $\nu > 0$ and $\text{sgn}(\nu) = 0$ 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:

$$P_{I_t}(x + u_x) = a_x P_{I_s}(x) + b_x,$$

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

$$D(a_x P_{I_s}(x) + b_x) = D(P_{I_s}(x))$$

for all pixels $x$ in image $I$, and for all $a_x \in \mathbb{R}^{>0}, b \in \mathbb{R}$.

If one considers the MLDP descriptor in Eq. (6), by representing $P_I(x)$ as a vector of its intensity values $P_I(x) = [I(x_1), I(x_2), \ldots, I(x_9)]^T$ and $M_i (i = 1, 2, \ldots, 8)$ as a vector of the coefficient $M_i(x) = [a_{i,1}, a_{i,2}, \ldots, a_{i,9}]^T$, one gets:

$$M_i \otimes P_I(x) = \sum_{j=1}^{9} a_{i,j} I(x_j).$$

Then,

$$M_i \otimes (a_x P_{I_s}(x) + b_x) = a_x \sum_{j=1}^{9} a_{i,j} I(x_j) + b_x \sum_{j=1}^{9} a_{i,j}.$$  \hspace{1cm} (10)

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

$$M_i \otimes (a_x P_{I_s}(x) + b_x) = a_x M_i \otimes P_I(x), \forall i.$$  \hspace{1cm} (11)

Because $a_x$ is assumed to be a positive number,

$$\text{sgn}(M_i \otimes (a_x P_{I_s}(x) + b_x)) = \text{sgn}(M_i \otimes P_I(x)), \forall i.$$  \hspace{1cm} (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 $\text{sgn}$ 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:

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

\hspace{1cm} (13)

Taking the ratio of the respective sides of equalities in both Eqs. (14) and Eq. (15) leads to
\[ D(a_x P_l(x) + b_x) = D(P_l(x)) \]
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 \( \sigma_1 \) and \( \sigma_2 \) in Eq. (5) were set to 3 and 5, respectively. The \( \lambda \)-parameter in Eq. (2) and the scale factor (denoted by \( P_{y_s} \)) in the coarse-to-fine strategy were experimentally adjusted using the Middlebury training set [13]. For each descriptor, the parameter pair (\( \lambda, P_{y_s} \)) leading to the most accurate OF results was chosen for all the tests in this section. The values of pair (\( \lambda, P_{y_s} \)) 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.

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 in 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 discontinuities 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


