On illumination-invariant variational optical flow for weakly textured scenes

Dinh Hoan Trinh, Christian Daul

Version of Record: https://www.sciencedirect.com/science/article/pii/S107731421830256X
Manuscript_9b2a5165d5487d24131e6758ad94cf30

Computer Vision and Image Understanding

journal homepage: www.elsevier.com

On Illumination-Invariant Variational Optical Flow for Weakly Textured Scenes

Dinh-Hoan Trinh, Christian Daul**

Université de Lorraine and CNRS, CRAN UMR 7039, 2 avenue de la Forêt de Haye, 54518 Vandœuvre-Lès-Nancy Cedex, France.

ABSTRACT

This paper deals with variational optical flow approaches for motion estimation under varying illumination conditions in weakly textured scenes. It proposes a systematic and complete study on descriptor-based data-terms that lead to a robust variational optical flow model. Unlike the literature which most often only experimentally shows that a descriptor is illumination invariant, this contribution gives a theoretical proof of this invariance. First, a local illumination change model is proposed and used to mathematically check whether a descriptor is invariant or not with respect to illumination variations between images. Then, this contribution proposes two general mathematical formulations which can be used to design a wide variety of new illumination-invariant descriptors. To illustrate the interest of the proposed approach, two novel illumination-invariant descriptors are constructed using the proposed general formulations. Moreover, the performance of the descriptors was evaluated on numerous datasets with known ground truth optical flow, while the robustness of the variational optical flow approach was highlighted using complex medical image sequences without ground truth. These experimental results have shown that data-terms based on the proposed descriptors led to accurate and constant optical flow under varying illumination conditions.

1. Introduction

1.1. Context of the work

Optical flow (OF) is an essential part of various computer vision applications. Despite the numerous and effective methods published during the last decades, OF determination remains challenging due to the data variability arising from both the acquisition conditions and the scenes themselves.

Recently, convolutional neural network (CNN)-based methods such as FlowNet [1], FlowNet2 [2], PWC-Net [3], LiteFlowNet [4] have demonstrated impressive OF results, both in terms of accuracy and from the computation time point of view. However, such learning-based approaches often require the availability of databases with ground truth OF for the CNN training. Moreover, in order to get highly accurate results, the content of the images in the testing phase should be similar to that of the images of the training dataset. In fact, there exist applications in which CNN-based OF methods are quite difficult to be used due to the lack of training datasets. For instance, in medical imaging, it is often difficult to construct training datasets with known ground truth OF.

Alternatively, the patch-matching methods (e.g. DeepFlow [5], FlowFields [6], EpicFlow [7], CPM-Flow [8]) establish homologous point correspondences by matching image patches. Their effectiveness has been shown on well-known benchmarks, namely Middlebury [9], MPI Sintel [10] and KITTI [11, 12]. However, most of images in these benchmarks are with rather numerous and/or well contrasted image primitives or textures. For a particular (but widespread) class of difficult images characterized by a lack of textures and strong illumination changes (such scenes are classical in medical endoscopy for instance), patch-matching methods are often less efficient than

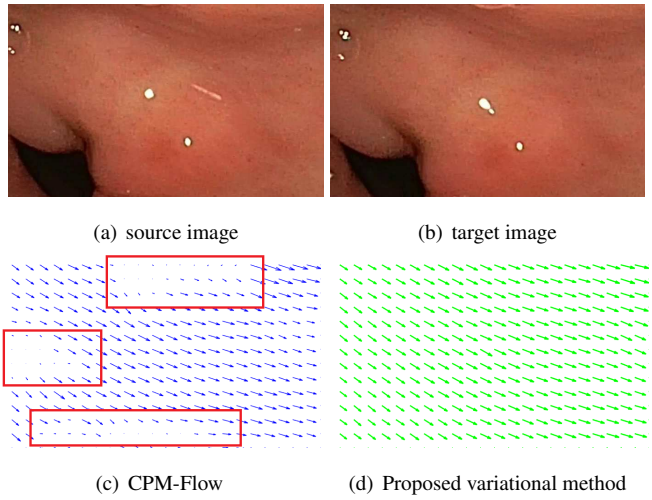
**Corresponding author: Tel.: +33 3 72 74 39 47
e-mail: christian.daul@univ-lorraine.fr (Christian Daul)

variational methods [13, 14, 15, 16].

As an example, Fig. 1 illustrates the OF result obtained by the CPM-Flow patch matching method [8] and the proposed variational method for a pair of gastroscopic images with very weak textures. Figs. 1(a)-1(b) are two consecutive video-frames of the inner stomach wall obtained with the endoscope axis which is relatively perpendicular to the epithelium surface and with a camera translation between the two acquisitions (i.e. with a camera movement without in- and out-plane rotations). In such a situation the flow field vectors should all be almost parallel and their magnitude should only slightly change according to the distance between the camera and the scene points. It is clearly visible in Fig. 1(c) that the vector field obtained with the CPM-Flow method [8] is not uniform. Indeed, the three red image rectangles highlight zones in which there is a large disparity in the vector magnitudes and orientations reveal rather large errors in the OF determination. On the contrary, as shown in Fig. 1(d), with the variational method proposed in this paper, the vector field is quite uniform, both in orientation and magnitude[1].

As illustrated by the previous example, variational methods remain valuable approaches for difficult scenes with few textures and with changing illumination conditions. Moreover, the variational methods can also be used as a refinement step in other approaches such as the FlowNet [1]. Contrary to recent publications which highlighted the interest of CNN-based or patch matching methods, this contribution focuses on the contributing role of variational methods for weakly textured scenes under strong illumination changes. The aim of this study is not to compare the performance of different OF approaches (learning based methods, patch matching approaches or variational methods) since they have all optimal performances for different scene types, and their applicability depends on the scene type (e.g., in medical endoscopy it is not possible to obtain ground truth flow fields for the learning step of CNN-based methods). The goal of this paper is rather to deeply investigate patch-based illumination invariant descriptors leading to robust data-terms in a variational OF model. From the theoretical point of view, the major contribution of this paper lies in the general mathematical formulation of illumination-invariant descriptors that facilitate the design of robust data-terms.

[1]The difference in uniformity of the OF fields of the IIOF-NLDP method and the CPM-Flow method is also shown in the 120 images of the gastroscopic video-sequence (pyloric antrum region of the stomach) provided as a supplementary material of this paper.



(a) source image (b) target image

(c) CPM-Flow (d) Proposed variational method

Fig. 1. OF results of the CPM-Flow method [8] and the proposed method on a pair of gastroscopic images.

1.2. Previous works

The importance of variational OF estimation has been demonstrated since the pioneering work of Horn and Schunck [17]. Since this publication, numerous improvements were presented in the field of variational OF. Interested readers may refer to [9, 18, 19] for a comprehensive overview of OF methods and the general principles behind them.

Variational models for OF estimation can be generally 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} stands for the data-term that measures the similarity of the pixels in I_s and I_t while parameter $\lambda > 0$ controls the relative importance of the data and regularization terms.

The data-terms $E_{data}(I_s, I_t, u)$ proposed in the literature correspond to an energy with one or several components $E_i(I_s, I_t, u)$:

$$E_{data}(I_s, I_t, \mathbf{u}) = \sum_{i=1}^{K} \lambda_i E_i(I_s, I_t, \mathbf{u}). \quad (2)$$

The scalars $\lambda_i > 0$ act as weights controlling the relative importance of the K different data-term parts. Each component $E_i(I_s, I_t, \mathbf{u})$ in (2) is based on a certain constancy assumption.

The classical variational OF methods [17, 20, 21, 22] define the data-term using the well-known brightness constancy assumption (BCA):

$$I_t(\mathbf{x} + \mathbf{u}_\mathbf{x}) = I_s(\mathbf{x}). \quad (3)$$

More specifically,

$$E_{data} = E_{BCA} = \sum_{\mathbf{x}\in\Omega} \Psi(I_s(\mathbf{x}) - I_t(\mathbf{x} + \mathbf{u}_\mathbf{x})), \quad (4)$$

where $\Psi(\cdot)$ is a penalty function and Ω stands for the image domain. It is clear that the BCA is most often not an appropriate assumption since, in many real scenes, the illumination of homologous pixels changes between two images.

Brox *et al.* [23] have shown that the robustness against illumination changes can be improved by combining the BCA with the gradient constancy assumption (GCA, energy E_{GCA}). The two-component data-term becomes:

$$E_{data} = \lambda_1 E_{BCA} + \lambda_2 E_{GCA}, \quad (5)$$

with

$$E_{GCA} = \sum_{\mathbf{x}\in\Omega} \Psi(\|\nabla I_t(\mathbf{x} + \mathbf{u}_\mathbf{x}) - \nabla I_s(\mathbf{x})\|). \quad (6)$$

This popular BCA-GCA data-term was implemented in several optical flow methods [24, 25]. However, Xu *et al.* [13] showed that a simultaneous use of BCA and GCA in all pixels is not optimal in terms of robustness. In fact, an efficient simultaneous use of BCA and GCA is difficult to reach because determining the weights (which give the appropriate relative importance of the BCA and GCA terms) is not easy and is strongly scene dependent. Moreover, the simultaneous use of both constancy assumptions leads also to a computational complexity in solving the optimization problem. Consequently, Xu *et al.* [13] proposed to use a binary mapping method that locally selects (according the pixel values in space Ω) E_{data} based either on BCA or on GCA. Although GCA is less sensitive than BCA to illumination variations, it is only invariant under additive illumination changes. As pointed out in [26] and [27], GCA is not able to compensate for the multiplicative illumination changes. In particular, in scenarios with large illumination changes, GCA is usually not fulfilled since multiplicative illumination changes are significant in such scene conditions.

Other constancy assumptions based on high order derivatives such as the Laplacian and the Hessian have been investigated in [28]. The data-term proposed in [28] is obtained by a linear combination (as in (2)) of BCA, GCA, the Laplacian constancy assumption, and the Hessian constancy assumption. However, the use of higher order derivatives significantly increases both the computational complexity and the sensitivity with respect to noise (data-term based on GCA is less sensitive to noise than that in [28] since it contains lower order derivatives). More generally, a major drawback of a linear combination of constancy assumptions lies in the difficulty to find the weights λ_i in (2) ensuring a robust OF determination for different scenes.

Wedel *et al.* [29] proposed an alternative to higher-order constancy assumptions that preprocesses the images using a structure-texture decomposition. The input image for OF estimation is a linear combination of the structure and texture components, with an emphasis on the texture components. Although this method was successfully used in [18], its computation complexity is high and the weight adjustment (balance between texture and structure) is not trivial for different scenes.

In other approaches [30, 31, 32, 33, 34, 35], the data-term was computed with photometric invariants to improve the robustness of the OF against illumination changes in color image sequences. The overall idea behind photometric invariants is detailed in Mileva *et al.* [32]. These authors proposed a data-term combining several photometric invariants obtained from normalized RGB channels, log-derivatives, from the HSI color space, or from the $r\phi\theta$ color space. However, this model is only applicable to color images. Zimmer *et al.* [34] proposed to use constraint normalization, and a HSV color representation with

higher order constancy assumptions.

Kumar *et al.* [27] assume that the image intensity I is proportional to the product of illumination L with reflection R, i.e. $I \propto L \times R$. The authors proposed to decouple the intensities of images I_s and I_t in a illumination component L and reflection component R. Then, the OF field between I_s and I_t is estimated from the two images $\widetilde{I_s} = \beta \log L_s + \log R_s$ and $\widetilde{I_t} = \beta \log L_t + \log R_t$, with $\beta \in (0, 1)$ since reflectance is invariant to illumination changes. It was demonstrated that the decoupled method combined with the BCA and GCA assumptions achieves improved accuracy under high illumination variation. However, the performance of this method strongly depends on the accurate estimation of the illumination component L. No theoretical basis that proves that this method is invariant to illumination changes was provided.

Haussecker and Fleet [36] introduced a framework that tackle the illumination issue by using explicit models of the underlying physical process that causes illumination changes. In this framework, the OF field and the parameters of the illumination model have to be simultaneously estimated. Unlike [36], Negahdaripour [37] proposed the generalized dynamic image model (GDIM) that describes the illumination changes as a combination of a multiplicative and an additive component regardless of the underlying physical events. Accordingly, the author replaced the classical BCA in (3) by a more general constancy constraint:

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

where the brightness of corresponding pixels in two consecutive images is related via the motion vector $\mathbf{u}_\mathbf{x}$, as well as by the radiometric parameters $a_\mathbf{x}$ and $b_\mathbf{x}$. It is clear that the GDIM model is suited to many scenes due its general nature. Kim *et al.* [38] used this model to construct a robust variational method for OF computation. However, the optimization of the OF with this method is quite complex for a limited effectiveness.

Recently, descriptor-based variational OF methods became very popular because of their effectiveness in tackling the problem of illumination changes. The underlying idea of these methods is to construct a robust data-term by defining illumination-invariant descriptors at each pixel. Some popular descriptor-based approaches are the census transform-based methods [39, 40, 41, 42], the rank transform-based method [43], the correlation-based methods [44, 45, 14, 46], the MLDP method of Mohamed *et al.* [15], and the NND method of Ali *et al.* [16].

1.3. Motivation of the paper

As demonstrated in [16], the descriptor-based methods are able to preserve accuracy of the OF under changing illumination conditions. However, up to now, most of the contributions describing descriptor-based methods only introduce a formulation of the descriptors and use experiments to highlight their performances, without giving a mathematical justification of their invariance according to illumination change models. Moreover, although many of the descriptors are based on similar principles, there is no unified theoretical basis for defining these descriptors.

In preliminary works we proposed a new illumination-invariant descriptor [47] and a general form [48] from which descriptors can be derived. However, no general and detailed mathematical discussion was given about the illumination invariance of descriptors and their design. Moreover, no thorough assessment and comparison of existing and new descriptors were made on reference benchmarks (KITTI, MPI-Sintel, etc.). This paper proposes further developments in the definition of illumination-invariant data terms in variational OF model. The major contributions of this paper are as follows:

1. We use a local illumination change model to describe the illumination changes between homologous images regions. Complex scene illumination changes can be considered when the corresponding neighborhoods in images I_s and I_t are small enough.

2. We propose a unified theoretical basis for defining illumination-invariant descriptors in the data-term of variational approaches. This theoretical basis is used to gives answers to following questions : What explains the robustness against illumination changes of some well-known

models as the census transform, the correlation transform or the complete rank transform? What are their underlying mathematical concepts?

3. We introduce two generalized formulations facilitating the design of illumination-invariant descriptors (a completely novel formulation and a more detailed description of the formulation in [48]).

4. Based on the generalized formulations, we propose two illumination-invariant descriptors (a novel descriptor and an improved mathematical justification of the descriptor in [48]).

5. We also present a comparison and evaluation of illumination-invariant descriptors using the same variational model.

1.4. Paper organization

The remainder of this paper is organized as follows. Section 2 presents the global and general framework for the descriptor-based variational OF methods. Section 3 proposes an illumination change model, as well as a criterion to determine whether a descriptor is invariant or not to illumination changes. This section also recalls some existing descriptors which are robust to varying illumination, and mathematically proves their illumination invariance. Two generalized forms of illumination-invariant descriptors are introduced in Section 4. The general forms given in Section 4 are then used to propose two novel descriptors in Section 5. In Section 6, the performance of the new descriptors is compared to that of reference descriptors of the literature. Section 7 shows the interest of the proposed OF approach in the frame of complex scene mosaicing. The last section gives a general conclusion for this contribution.

2. Descriptor-based Variational Optical Flow Model

In the general energy given in (1), descriptor-based data-terms can be formulated as follows:

$$E_{data} = \sum_{\mathbf{x}\in\Omega} \Psi(\mathbf{D}(P_{I_s}(\mathbf{x})) - \mathbf{D}(P_{I_t}(\mathbf{x}+\mathbf{u}_\mathbf{x}))), \tag{8}$$

where I_s and $I_t : \Omega \to \mathbb{R}$ are source and target images respectively, $\Omega = \{\mathbf{x} = (x, y) \mid 1 \le x \le N, 1 \le y \le M\} \subseteq \mathbb{N}^2$, $\Psi(\cdot)$ is a penalty function and $\mathbf{D}$ is a descriptor that locally characterizes image region similarities in images I_s and I_t. $\mathbf{D}(P_I(\mathbf{x}))$ denotes the feature descriptor of the pixels in the neighborhood of pixel $\mathbf{x}$ in image I, this descriptor being computed with the pixel values of patch $P_I(\mathbf{x})$ centered on $\mathbf{x}$.

While the data-term in (8) will take various forms, the regularization term in (1) is kept constant in this contribution. We use the non-local total variation as in [45, 49, 50, 14, 25] to define the regularization term:

$$E_{reg}(\mathbf{u}) = \sum_{\mathbf{x}\in\Omega} \sum_{\mathbf{x}'\in\mathcal{N}_\mathbf{x}} w_\mathbf{x}^{\mathbf{x}'} \|\mathbf{u}_\mathbf{x} - \mathbf{u}_{\mathbf{x}'}\|_1, \tag{9}$$

where $\mathcal{N}_\mathbf{x}$ is the set of neighbor pixels centered on pixel $\mathbf{x}$ and the weights $w_\mathbf{x}^{\mathbf{x}'}$ depend on the similarity between pixels $\mathbf{x}$ and $\mathbf{x}'$. Similarly to [14], the weights $w_\mathbf{x}^{\mathbf{x}'}$ are defined as follows:

$$w_\mathbf{x}^{\mathbf{x}'} = \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|^2}{2\sigma_1^2} - \frac{\|L(\mathbf{x})-L(\mathbf{x}')\|^2}{2\sigma_2^2}\right), \tag{10}$$

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

According to [14], by defining a weight matrix W in space $\mathbb{R}^{|\Omega|\times|\mathcal{N}|}$,

$$W = \begin{bmatrix} w_{(1,1)}^{\mathbf{x}'\in\mathcal{N}_{(1,1)}} & \cdots & w_{(1,1)}^{\mathbf{x}'\in\mathcal{N}_{(1,1)}} \\ \vdots & \ddots & \vdots \\ w_{(N,M)}^{\mathbf{x}'\in\mathcal{N}_{(N,M)}} & \cdots & w_{(N,M)}^{\mathbf{x}'\in\mathcal{N}_{(N,M)}} \end{bmatrix}, \tag{11}$$

and a linear operator $K : \mathbb{R}^{2\times|\Omega|} \to \mathbb{R}^{2\times|\Omega|\times|\mathcal{N}|}$

$$K\mathbf{u} = \begin{bmatrix} \mathbf{u}_{\mathbf{x}'\in\mathcal{N}_{(1,1)}} - \mathbf{u}_{(1,1)} & \cdots & \mathbf{u}_{\mathbf{x}'\in\mathcal{N}_{(1,1)}} - \mathbf{u}_{(1,1)} \\ \vdots & \ddots & \vdots \\ \mathbf{u}_{\mathbf{x}'\in\mathcal{N}_{(N,M)}} - \mathbf{u}_{(N,M)} & \cdots & \mathbf{u}_{\mathbf{x}'\in\mathcal{N}_{(N,M)}} - \mathbf{u}_{(N,M)} \end{bmatrix},$$

where $|\mathcal{N}|$ is the size of the neighborhoods $\mathcal{N}_\mathbf{x}$, the regularization term (9) can be rewritten as

$$E_{reg} = F(K\mathbf{u}), \tag{12}$$

with $F : \mathbb{R}^{2\times|\Omega|\times|\mathcal{N}|} \to \mathbb{R}$ is a function defined by

$$F(\mathbf{z}) = \|W \cdot \mathbf{z}\|_1. \tag{13}$$

Therefore, flow field $\mathbf{u}$ is computed by solving following optimization problem:

$$\min_\mathbf{u} F(K\mathbf{u}) + \lambda E_{data}(I_s, I_t, \mathbf{u}). \tag{14}$$

This optimization can be effectively implemented using the projected-proximal-point algorithm [51]. The details of the algorithm are beyond the scope of this work. Interested readers can refer to [51, 14] for more details on the mathematical concepts.

The following section details the underlying principles of the descriptor-based data-terms that are robust to illumination changes.

3. Descriptor-based Illumination-Invariant Data-Terms

The descriptor-based data-term in (8) is robust against illumination changes when local descriptor $\mathbf{D}$ is illumination-invariant, i.e., if pixel $\mathbf{x}$ in I_s and pixel $(\mathbf{x}+\mathbf{u}_\mathbf{x})$ in I_t are two corresponding pixels then $\mathbf{D}(P_{I_s}(\mathbf{x})) = \mathbf{D}(P_{I_t}(\mathbf{x}+\mathbf{u}_\mathbf{x}))$.

Local descriptors $\mathbf{D}$ are computed with small patches $P_I(\mathbf{x}_0)$ centered at pixel $\mathbf{x}_0$ in image I and with a size of $(2k+1)\times(2k+1)$ pixels, where k is a positive integer. $P_I(\mathbf{x}_0)$ can be represented by a vector taking the pixel intensity values as components:

$$P_I(\mathbf{x}_0) = [I(\mathbf{x}_0), I(\mathbf{x}_1), \dots, I(\mathbf{x}_n)]^T \in \mathbb{R}^{n+1}, \tag{15}$$

where $n = (2k+1)^2 - 1$ is the number of neighbor pixels of $\mathbf{x}_0$ in the patch. The general formulation of descriptors $\mathbf{D}$ for any pixel $\mathbf{x}_0$ in image I can be written as:

$$\mathbf{D}(P_I(\mathbf{x}_0)) = \begin{bmatrix} f_0\left(P_I(\mathbf{x}_0)\right) \\ f_1\left(P_I(\mathbf{x}_0)\right) \\ \vdots \\ f_m\left(P_I(\mathbf{x}_0)\right) \end{bmatrix} \in \mathbb{R}^m, \tag{16}$$

where $f_i : \mathbb{R}^{n+1} \to \mathbb{R}, i = 0, \dots, m$ are real functions.

The main issue arising when designing an illumination invariant data-term lies in the definition of the functions $f_i(\cdot)$ leading to an accurate OF field.

3.1. Illumination change model and illumination-invariance criterion for descriptors

In numerous scenes (e.g. in outdoor or medical scenes) the illumination changes between consecutive images of a video-sequence are complex and cannot be represented by a global model. For instance, gradient constancy over complete images (or large image parts) is most often not a realistic illumination change assumption since the illumination variation is barely constant for all homologous pixels of two images. Illumination changes between images should rather be locally modeled to be as general as possible in terms of scene types.

Negahdaripour [37] simulated the illumination changes by an affine transformation (see (7)). This pixel-wise model is able to capture complex variations. However, as shown in [38], integrating this model into a variational calculus algorithm leads to a complex optimization problem and high computation times since, besides a flow vector, parameters $a_\mathbf{x}$ and $b_\mathbf{x}$ have to be estimated for each pixel.

An interesting fact is that two consecutive video frames have usually the property of local stationarity such that all pixels in a small image region may share the same parameters in the illumination variation model. Therefore, the assumption is made that the illumination changes of all corresponding pixels in small homologous neighborhoods in two consecutive images can be accurately represented by a model with constant parameters.

Thus, instead of using the linear model in (7) at pixel-level, a patch-based model is used to describe locally the illumination changes between homologous neighborhoods:

$$P_{I_t}(\mathbf{x}+\mathbf{u}_\mathbf{x}) = a_\mathbf{x} P_{I_s}(\mathbf{x}) + b_\mathbf{x}, \tag{17}$$

where $a_\mathbf{x} \in \mathbb{R}_{>0}, b \in \mathbb{R}$, $P_{I_s}(\mathbf{x})$ and $P_{I_t}(\mathbf{x}+\mathbf{u}_\mathbf{x})$ are two corresponding patches in I_s and I_t, and $\mathbf{u}_\mathbf{x}$ is the displacement vector at pixel $\mathbf{x}$ of image I_s. Constant values of parameters $a_\mathbf{x}$ and $b_\mathbf{x}$ in small image regions allows for modelling complex illuminations changes between images.

Unlike in the GDIM model in (7), parameter $a_\mathbf{x}$ in the proposed model is assumed to be greater than 0, since the intensity values of pixels are non-negative. Rather than computing the values of parameters $a_\mathbf{x}$ and $b_\mathbf{x}$, the model in (17) is only used to design illumination independent descriptors.

The criterion allowing to confirm or to contradict the illumination invariance property of descriptor $\mathbf{D}$ is first given by exploiting (17).

Definition 1. *Descriptor* $\mathbf{D}$ *is illumination-invariant if it is in-*

variant with respect to the illumination variation model given in (17). *In other words,* $\mathbf{D}$ *has to satisfy following constraint:*

$$\mathbf{D}(P_I(\mathbf{x}_0)) = \mathbf{D}(a_{\mathbf{x}_0} P_I(\mathbf{x}_0) + b_{\mathbf{x}_0}), \tag{18}$$

for all $a_{\mathbf{x}_0} \in \mathbb{R}_{>0}$, $b_{\mathbf{x}_0} \in \mathbb{R}$.

Equation (18) gives the necessary condition for descriptors to be illumination-invariant.

The literature presents descriptor-based variational OF methods that have only been experimentally proven to be robust to illumination changes [39, 40, 41, 42, 43, 44, 45, 14, 15, 16]. The data-terms of these methods use patch-based descriptors under the form given in (8). In sub-sections 3.2 and 3.3, we rewrite some well-known descriptors and explain how it can be mathematically checked whether these descriptors are invariant or not with respect to the illumination change model proposed in (17).

3.2. Existing Local descriptors

3.2.1. Census transform

The Census transform [39] has been used as an effective solution for computing OF under outdoor lighting conditions [40, 41, 42]. Writing the Census transform under the form given in (16) leads to:

$$\mathbf{D}_{Census}(P_I(\mathbf{x}_0)) = \begin{bmatrix} \text{sgn}(I(\mathbf{x}_0) - I(\mathbf{x}_1)) \\ \text{sgn}(I(\mathbf{x}_0) - I(\mathbf{x}_2)) \\ \vdots \\ \text{sgn}(I(\mathbf{x}_0) - I(\mathbf{x}_n)) \end{bmatrix} \tag{19}$$

with

$$\text{sgn}(\upsilon) = \begin{cases} 1, & \upsilon > 0 \\ 0, & \text{otherwise.} \end{cases} \tag{20}$$

3.2.2. Complete rank transform (CRT)

The CRT descriptor [43] is computed as follows:

$$\mathbf{D}_{CRT}(P_I(\mathbf{x}_0)) = \begin{bmatrix} \sum_{j=0}^{n} \text{sgn}(I(\mathbf{x}_0) - I(\mathbf{x}_j)) \\ \sum_{j=0}^{n} \text{sgn}(I(\mathbf{x}_1) - I(\mathbf{x}_j)) \\ \vdots \\ \sum_{j=0}^{n} \text{sgn}(I(\mathbf{x}_n) - I(\mathbf{x}_j)) \end{bmatrix}. \tag{21}$$

The i-th component of $\mathbf{D}_{CRT}$ gives the number of pixels in $P_I(\mathbf{x}_0)$ having a smaller intensity than pixel $\mathbf{x}_i$.

$$\mathrm{M}_4 = \begin{bmatrix} 5 & 5 & -3 \\ 5 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \quad \mathrm{M}_3 = \begin{bmatrix} 5 & 5 & 5 \\ -3 & 0 & -3 \\ -3 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} -3 & 5 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & -3 \end{bmatrix} = \mathrm{M}_2$$

$$\mathrm{M}_5 = \begin{bmatrix} 5 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & -3 & -3 \end{bmatrix} \quad \begin{bmatrix} I(\mathbf{x}_4) & I(\mathbf{x}_3) & I(\mathbf{x}_2) \\ I(\mathbf{x}_5) & I(\mathbf{x}_0) & I(\mathbf{x}_1) \\ I(\mathbf{x}_6) & I(\mathbf{x}_7) & I(\mathbf{x}_8) \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & 5 \\ -3 & 0 & 5 \\ -3 & -3 & 5 \end{bmatrix} = \mathrm{M}_1$$

$$\mathrm{M}_6 = \begin{bmatrix} -3 & -3 & -3 \\ 5 & 0 & -3 \\ 5 & 5 & -3 \end{bmatrix} \quad \mathrm{M}_7 = \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & -3 \\ 5 & 5 & 5 \end{bmatrix} \quad \begin{bmatrix} -3 & -3 & -3 \\ -3 & 0 & 5 \\ -3 & 5 & 5 \end{bmatrix} = \mathrm{M}_8$$

Fig. 2. Kirsch edge kernels for the LDP and MLDP descriptor computation. Kernel M_i, $i \in \{1, 2, \ldots, 8\}$ is used to determine the gradient component along the i-th direction defined by the line passing through points $\mathbf{x}_0$ and $\mathbf{x}_i$.

3.2.3. Local Directional Pattern (LDP)

The LPD descriptor [52] corresponding to pixel $\mathbf{x}_0$ is a binary code computed for patch $P_I(\mathbf{x}_0)$ of size 3×3 pixels and based on the eight Kirsch edge kernels M_i, $i \in \{1, 2, \ldots, 8\}$ given in Fig. 2. The LDP descriptor is mathematically defined as:

$$\mathbf{D}_{LDP}(P_I(\mathbf{x}_0)) = \begin{bmatrix} \text{sgn}(|M_1 \otimes P_I(\mathbf{x}_0)| - m_k) \\ \text{sgn}(|M_2 \otimes P_I(\mathbf{x}_0)| - m_k) \\ \vdots \\ \text{sgn}(|M_8 \otimes P_I(\mathbf{x}_0)| - m_k) \end{bmatrix} \tag{22}$$

where operator $\otimes$ gives the sum of the element-wise product of two matrices, and m_k is the k-th largest element in array $\{|M_1 \otimes P_I(\mathbf{x}_0)|, |M_2 \otimes P_I(\mathbf{x}_0)|, \ldots, |M_8 \otimes P_I(\mathbf{x}_0)|\}$. As shown in Fig. 2, the edge response in direction i is given by $M_i \otimes P_I(\mathbf{x}_0)$ and $i = 1$ corresponds to the East direction. The edge magnitudes along the i-th direction (defined by pixel pair $\mathbf{x}_0$ and $\mathbf{x}_i$) is used to determine the binary code for pixel $\mathbf{x}_0$.

3.2.4. Modified Local Directional Pattern (MLDP)

Mohamed *et al.* [15] used the descriptor of Section 3.2.3 to obtain the modified local directional pattern (MLDP). Similarly to the LDP descriptor, the MLDP vector corresponds to an 8-bit information generated from the signs of the eight Kirsch edge responses. This descriptor is defined by:

$$\mathbf{D}_{MLDP}(P_I(\mathbf{x}_0)) = \begin{bmatrix} \text{sgn}(M_1 \otimes P_I(\mathbf{x}_0)) \\ \text{sgn}(M_2 \otimes P_I(\mathbf{x}_0)) \\ \vdots \\ \text{sgn}(M_8 \otimes P_I(\mathbf{x}_0)) \end{bmatrix}, \tag{23}$$

where $\text{sgn}(\cdot)$ is a binary function as defined in (20).

3.2.5. Correlation transform

For the correlation transform [14], the descriptor is defined with the mean μ_P and variance σ_P^2 of the $n+1$ pixels intensities of the patch:

$$\mathbf{D}_{Corr}(P_I(\mathbf{x}_0)) = \begin{bmatrix} \frac{I(\mathbf{x}_0)-\mu_P}{\sigma_P} \\ \frac{I(\mathbf{x}_1)-\mu_P}{\sigma_P} \\ \vdots \\ \frac{I(\mathbf{x}_n)-\mu_P}{\sigma_P} \end{bmatrix}, \tag{24}$$

with $\mu_P = \frac{\sum_{i=0}^{n} I(\mathbf{x}_i)}{n+1}$, and $\sigma_P^2 = \frac{\sum_{i=0}^{n} (I(\mathbf{x}_i)-\mu_P)^2}{n+1}$.

3.2.6. Normalized Neighborhood Descriptor (NND)

In order to compute their descriptor, Ali *et al.* [16] used a patch $P_I(\mathbf{x}_0)$ with a minimal size of 5×5 pixels (i.e., $k \geq 2$). At central pixel $\mathbf{x}_0$, a sub-patch $\mathcal{P}_0 \subset P_I(\mathbf{x}_0)$ of size $(2k'+1) \times (2k'+1)$ is defined ($k' > 0$). Here, $\mathbf{x}_0$ is also the central pixel of sub-patch $\mathcal{P}_0$, and the size of $P_I(\mathbf{x}_0)$ and $\mathcal{P}_0$ is constrained by $k = 2k'$. Likewise, there exists a set of $m = (2k'+1)^2 - 1$ sub-patches $\{\mathcal{P}_j \,|\, \mathcal{P}_j \subset P_I(\mathbf{x}_0)\}_{j=1}^{m}$ where $\mathcal{P}_j$ has the same size as $\mathcal{P}_0$. The pixels within sub-patch $\mathcal{P}_j$ are denoted by $\mathbf{x}_j^1, \mathbf{x}_j^2, \ldots, \mathbf{x}_j^{m+1}$. Using theses patches, the NND descriptor is defined as:

$$\mathbf{D}_{NND}(P_I(\mathbf{x}_0)) = \begin{bmatrix} \exp \frac{-\|\mathcal{P}_1-\mathcal{P}_0\|^2}{h_{\mathbf{x}_0}^2} \\ \exp \frac{-\|\mathcal{P}_2-\mathcal{P}_0\|^2}{h_{\mathbf{x}_0}^2} \\ \vdots \\ \exp \frac{-\|\mathcal{P}_m-\mathcal{P}_0\|^2}{h_{\mathbf{x}_0}^2} \end{bmatrix}, \tag{25}$$

where

$$\|\mathcal{P}_j - \mathcal{P}_0\|^2 = \sum_{i=1}^{m+1} (I(\mathbf{x}_j^i) - I(\mathbf{x}_0^i))^2, \tag{26}$$

$$\text{and} \quad h_{\mathbf{x}_0}^2 = \sum_{k=0}^{3} \|\mathcal{P}_{2k+1} - \mathcal{P}_0\|^2/4. \tag{27}$$

In (27), $\mathcal{P}_1, \mathcal{P}_3, \mathcal{P}_5$, and $\mathcal{P}_7$ are 4 sub-patches centered at 4 neighbor pixels of $\mathbf{x}_0$, namely $\mathbf{x}_1, \mathbf{x}_3, \mathbf{x}_5$ and $\mathbf{x}_7$, respectively (refer to Fig. 2 for the relative pixel positions).

3.3. Descriptor Illumination-Invariance Assessment

Until now, the effectiveness in estimating OF under changing illumination conditions has only been shown experimentally for the descriptors listed in Section 3.2. To the best of our knowledge, none of the corresponding contributions have discussed the mathematical concepts explaining why such descriptors can cope with illumination changes. This sub-section mathematically justifies the appropriateness of these descriptors.

Theorem 1. *Descriptors* $\mathbf{D}_{Census}$, $\mathbf{D}_{CRT}$, $\mathbf{D}_{LDP}$, $\mathbf{D}_{MLDP}$, $\mathbf{D}_{Corr}$, *and* $\mathbf{D}_{NND}$ *are illumination-invariant.*

Proof. Denote $\mathbf{D}^i$ as the i-th component of descriptor $\mathbf{D}$ and $\widetilde{P}_I(\mathbf{x}_0) = a_{\mathbf{x}_0} P_I(\mathbf{x}_0) + b_{\mathbf{x}_0}$. One needs to prove that for all $a_{\mathbf{x}_0} \in \mathbb{R}_{>0}$ and $b_{\mathbf{x}_0} \in \mathbb{R}$,

$$\mathbf{D}^i(P_I(\mathbf{x}_0)) = \mathbf{D}^i(\widetilde{P}_I(\mathbf{x}_0)). \tag{28}$$

- For $\mathbf{D}_{Census}$, the i-th component can be written as:

$$\mathbf{D}^i_{Census}(P_I(\mathbf{x}_0)) = \text{sgn}(I(\mathbf{x}_0) - I(\mathbf{x}_i)). \tag{29}$$

Due to $a_{\mathbf{x}_0} > 0$,

$$\begin{aligned} \mathbf{D}^i_{Census}(\widetilde{P}_I(\mathbf{x}_0)) &= \text{sgn}(a_{\mathbf{x}_0}(I(\mathbf{x}_0) - I(\mathbf{x}_i))) \\ &= \mathbf{D}^i_{Census}(P_I(\mathbf{x}_0)). \end{aligned} \tag{30}$$

- The i-th $\mathbf{D}_{CRT}$ component is:

$$\mathbf{D}^i_{CRT}(P_I(\mathbf{x}_0)) = \sum_{j=0}^{n} \text{sgn}(I(\mathbf{x}_i) - I(\mathbf{x}_j)). \tag{31}$$

Because $a_{\mathbf{x}_0} > 0$,

$$\begin{aligned} &\text{sgn}[(a_{\mathbf{x}_0} I(\mathbf{x}_i) + b_{\mathbf{x}_0}) - (a_{\mathbf{x}_0} I(\mathbf{x}_j) + b_{\mathbf{x}_0})] = \\ &= \text{sgn}[a_{\mathbf{x}_0}(I(\mathbf{x}_i) - I(\mathbf{x}_j))] = \text{sgn}(I(\mathbf{x}_i) - I(\mathbf{x}_j)). \end{aligned} \tag{32}$$

Therefore, $\mathbf{D}^i_{CRT}(\widetilde{P}_I(\mathbf{x}_0)) = \mathbf{D}^i_{CRT}(P_I(\mathbf{x}_0))$.
- Similarly, each component of $\mathbf{D}_{LDP}$, can be written as:

$$\mathbf{D}^i_{LDP}(P_I(\mathbf{x}_0)) = \text{sgn}(|M_i \otimes P_I(\mathbf{x}_0)| - m_k^s) \tag{33}$$

where m_k^s is the k-th largest element in array

$$\mathcal{A}_s = \{|M_1 \otimes P_I(\mathbf{x}_0)|, \ldots, |M_8 \otimes P_I(\mathbf{x}_0)|\}. \tag{34}$$

Thus,

$$\mathbf{D}^i_{LDP}(\widetilde{P}_I(\mathbf{x}_0)) = \text{sgn}\left(|M_i \otimes \widetilde{P}_I(\mathbf{x}_0)| - m_k^t\right), \tag{35}$$

where m_k^t is the k-th largest element in array

$$\mathcal{A}_t = \{|M_1 \otimes \widetilde{P}_I(\mathbf{x}_0)|, \ldots, |M_8 \otimes \widetilde{P}_I(\mathbf{x}_0)|\}. \tag{36}$$

As it can be seen in Fig. 2, the sum of the elements in the Kirsch matrix M_i equals zero for all $i = 1, \ldots, 8$. Thus, it is easy to conclude that

$$M_i \otimes [a_{\mathbf{x}_0} P_I(\mathbf{x}_0) + b_{\mathbf{x}_0}] = a_{\mathbf{x}_0} (M_i \otimes P_I(\mathbf{x}_0)). \tag{37}$$

Therefore, $\forall i, j \in \{1, \ldots, 8\}, \forall a_{\mathbf{x}_0} > 0$,

$$\begin{aligned} &\left|M_i \otimes \widetilde{P}_I(\mathbf{x}_0)\right| \geq \left|M_j \otimes \widetilde{P}_I(\mathbf{x}_0)\right| \\ \Leftrightarrow\ &|M_i \otimes P_I(\mathbf{x}_0)| \geq \left|M_j \otimes P_I(\mathbf{x}_0)\right|. \end{aligned} \tag{38}$$

It means that the element order in array $\mathcal{A}_t$ is the same as the element order in the corresponding array $\mathcal{A}_s$. More precisely, if $m_k^s = |M_j \otimes P_I(\mathbf{x}_0)|$ is the k-th largest element in array $\mathcal{A}_s$, then its corresponding $m_k^t = |M_j \otimes \widetilde{P}_I(\mathbf{x}_0)|$ is also the k-th largest element in array $\mathcal{A}_t$. Therefore,

$$\mathbf{D}^i_{LDP}(\widetilde{P}_I(\mathbf{x}_0)) = \mathbf{D}^i_{LDP}(P_I(\mathbf{x}_0)). \tag{39}$$

- For $\mathbf{D}_{MLDP}$, the i-th component can be written as:

$$\mathbf{D}^i_{MLDP}(P_I(\mathbf{x}_0)) = \text{sgn}(M_i \otimes P_I(\mathbf{x}_0)). \tag{40}$$

From (37) it follows:

$$\text{sgn}(M_i \otimes [a_{\mathbf{x}_0} P_I(\mathbf{x}_0) + b_{\mathbf{x}_0}]) = \text{sgn}(a_{\mathbf{x}_0}(M_i \otimes P_I(\mathbf{x}_0))).$$

Therefore, with $a_{\mathbf{x}_0} > 0$

$$\mathbf{D}^i_{MLDP}(\widetilde{P}_I(\mathbf{x}_0)) = \mathbf{D}^i_{MLDP}(P_I(\mathbf{x}_0)). \tag{41}$$

- Component i of descriptor $\mathbf{D}_{Corr}$ becomes:

$$\mathbf{D}^i_{Corr}(P_I(\mathbf{x}_0)) = (I(\mathbf{x}_i) - \mu_P)/\sigma_P. \tag{42}$$

One has

$$\mathbf{D}^i_{Corr}(\widetilde{P}_I(\mathbf{x}_0)) = \left(a_{\mathbf{x}_0} I(\mathbf{x}_i) + b_{\mathbf{x}_0} - \mu_{\widetilde{P}}\right)/\sigma_{\widetilde{P}}, \tag{43}$$

$$\begin{aligned} a_{\mathbf{x}_0} I(\mathbf{x}_i) + b_{\mathbf{x}_0} - \mu_{\widetilde{P}} &= a_{\mathbf{x}_0} I(\mathbf{x}_i) + b_{\mathbf{x}_0} - a_{\mathbf{x}_0}\mu_P - b_{\mathbf{x}_0} \\ &= a_{\mathbf{x}_0}(I(\mathbf{x}_i) - \mu_P), \end{aligned} \tag{44}$$

and

$$\begin{aligned} \sigma_{\widetilde{P}} &= \sqrt{\frac{1}{n+1}\sum_{i=0}^{n}(a_{\mathbf{x}_0} I(\mathbf{x}_i) + b_{\mathbf{x}_0} - \mu_{\widetilde{P}})^2} \\ &= \sqrt{\frac{1}{n+1}\sum_{i=0}^{n} a^2_{\mathbf{x}_0}(I(\mathbf{x}_i) - \mu_P)^2} = a_{\mathbf{x}_0}\sigma_P. \end{aligned} \tag{45}$$

By combining (43), (44), and (45) we deduce

$$\mathbf{D}^i_{Corr}(\widetilde{P}_I(\mathbf{x}_0)) = \mathbf{D}^i_{Corr}(P_I(\mathbf{x}_0)). \tag{46}$$

- Finally, the i-th component of descriptor $\mathbf{D}_{NND}$ is:

$$\mathbf{D}^i_{NND}(P_I(\mathbf{x}_0)) = \exp\left(-\|\mathcal{P}_i - \mathcal{P}_0\|^2/h^2_{\mathbf{x}_0}\right). \tag{47}$$

Let us consider the NND descriptor on $\widetilde{P}_I(\mathbf{x}_0)$. From (26) and (27) one gets

$$\mathbf{D}^i_{NND}(\widetilde{P}_I(\mathbf{x}_0)) = \exp\left(-\|\widetilde{\mathcal{P}}_i - \widetilde{\mathcal{P}}_0\|^2/\tilde{h}^2_{\mathbf{x}_0}\right), \tag{48}$$

where

$$\begin{aligned} \|\widetilde{\mathcal{P}}_i - \widetilde{\mathcal{P}}_0\|^2 &= \sum_{i=1}^{m+1}(a_{\mathbf{x}_0} I(\mathbf{x}^i_j) + b_{\mathbf{x}_0} - a_{\mathbf{x}_0} I(\mathbf{x}^i_0) - b_{\mathbf{x}_0})^2 \\ &= \sum_{i=1}^{m+1} a^2_{\mathbf{x}_0}(I(\mathbf{x}^i_j) - I(\mathbf{x}^i_0))^2 \\ &= a^2_{\mathbf{x}_0}\|\mathcal{P}_i - \mathcal{P}_0\|^2, \end{aligned} \tag{49}$$

and

$$\begin{aligned} \tilde{h}^2_{\mathbf{x}_0} &= \frac{\sum_{k=0}^{3}\|\widetilde{\mathcal{P}}_{2k+1} - \widetilde{\mathcal{P}}_0\|^2}{4} \\ &= a^2_{\mathbf{x}_0}\frac{\sum_{k=0}^{3}\|\mathcal{P}_{2k+1} - \mathcal{P}_0\|^2}{4} = a^2_{\mathbf{x}_0} h^2_{\mathbf{x}_0}. \end{aligned} \tag{50}$$

By combining (48), (49), and (50), one obtains

$$\begin{aligned} \mathbf{D}^i_{NND}(\widetilde{P}_I(\mathbf{x}_0)) &= \exp\frac{-a^2_{\mathbf{x}_0}\|\mathcal{P}_i - \mathcal{P}_0\|^2}{a^2_{\mathbf{x}_0} h^2_{\mathbf{x}_0}} \\ &= \mathbf{D}^i_{NND}(P_I(\mathbf{x}_0)). \end{aligned} \tag{51}$$

In this section it was theoretically proven why descriptors $\mathbf{D}_{Census}$, $\mathbf{D}_{CRT}$, $\mathbf{D}_{LDP}$, $\mathbf{D}_{MLDP}$, $\mathbf{D}_{Corr}$ and $\mathbf{D}_{NND}$ are effective for OF estimation under changing illumination conditions. However, experimental results in [14, 15, 16] demonstrated that there is a significant difference in the accuracy of OF fields estimated by these descriptors. Thus, the design of illumination invariant descriptors leading to accurate flow fields and involving a limited computational complexity remains an open challenge. The next section provides two generalized types of illumination-invariant descriptors that will facilitate the design of new descriptors that are robust to illumination changes.

4. Generalized Formulations for Illumination-Invariant Descriptors

Consider two vectors $\mathbf{v}_s = [v^s_0, v^s_1, \ldots, v^s_n]^T$ and $\mathbf{v}_t = [v^t_0, v^t_1, \ldots, v^t_n]^T$ in $\mathbb{R}^{n+1}$ such that

$$v^t_i = a_{\mathbf{x}} v^s_i + b_{\mathbf{x}}, \forall i = 0, 1, \ldots, n \tag{52}$$

with $a_{\mathbf{x}} \in \mathbb{R}_{>0}$ and $b_{\mathbf{x}} \in \mathbb{R}$. These vectors correspond to the descriptors of patches P_{I_s} and P_{I_t} of the source (I_s) and target (I_t) images, respectively. The relationship between the two vectors given in (52) is based on the illumination change model given in (17).

To propose general forms of illumination-invariant descriptors, let us start with the following lemma.

Lemma 1. *Suppose $\mathbf{v}_s$ and $\mathbf{v}_t$ are two vectors in $\mathbb{R}^{n+1}$ satisfying* (52)*. If $\{\alpha_0, \alpha_1, \ldots, \alpha_n\}$ is a sequence of real numbers such that*

$$\begin{cases} \alpha_0^2 + \alpha_1^2 + \cdots + \alpha_n^2 \neq 0 \\ \alpha_0 + \alpha_1 + \cdots + \alpha_n = 0, \end{cases} \tag{53}$$

then

$$\sum_{i=0}^{n} \alpha_i v_i^t = a_{\mathbf{x}} \sum_{i=0}^{n} \alpha_i v_i^s, \text{ and} \tag{54}$$

$$sgn \sum_{i=0}^{n} \alpha_i v_i^t = sgn \sum_{i=0}^{n} \alpha_i v_i^s. \tag{55}$$

Proof. We have,

$$\begin{aligned} \sum_{i=0}^{n} \alpha_i v_i^t &= \sum_{i=0}^{n} \alpha_i \underbrace{(a_{\mathbf{x}} v_i^s + b_{\mathbf{x}})}_{v_i^t = a_{\mathbf{x}} v_i^s + b_{\mathbf{x}}} \\ &= a_{\mathbf{x}} \sum_{i=0}^{n} \alpha_i v_i^s + b_{\mathbf{x}} \underbrace{\sum_{i=0}^{n} \alpha_i}_{=0}. \end{aligned} \tag{56}$$

From $\sum_{i=0}^{n} \alpha_i = 0$ and $a_{\mathbf{x}} > 0$, one can deduce that

$$\begin{cases} \sum_{i=0}^{n} \alpha_i v_i^t = a_{\mathbf{x}} \sum_{i=0}^{n} \alpha_i v_i^s \\ \text{sgn}\left(\sum_{i=0}^{n} \alpha_i v_i^t\right) = \text{sgn}\left(\sum_{i=0}^{n} \alpha_i v_i^s\right). \end{cases} \tag{57}$$

Using Lemma 1, we can propose two general forms of illumination-invariant descriptors as follows.

4.1. Descriptor Form based on Sign-Invariance

Using Lemma 1, a generalized form of illumination-invariant descriptors can be defined as:

$$\mathbf{D}(P_I(\mathbf{x}_0)) = \begin{bmatrix} \text{sgn}\left(\sum_{j=0}^{n} \alpha_{0,j} I(\mathbf{x}_j)\right) \\ \text{sgn}\left(\sum_{j=0}^{n} \alpha_{1,j} I(\mathbf{x}_j)\right) \\ \vdots \\ \text{sgn}\left(\sum_{j=0}^{n} \alpha_{m,j} I(\mathbf{x}_j)\right) \end{bmatrix}, \tag{58}$$

where $\alpha_i = \{\alpha_{i,0}, \alpha_{i,1}, \ldots, \alpha_{i,n}\}$, $i = 0, 1, \ldots, m$ are the sequences of real numbers satisfying (53).

With (58), it is possible to generate new descriptors by choosing appropriate values for α_i. For instance, when $\alpha_i = \{\alpha_{i,0}, \alpha_{i,1}, \ldots, \alpha_{i,n}\}$ with $\alpha_{i,0} = 1$, $\alpha_{i,i} = -1$, $\alpha_{i,j} = 0$ with $j \notin \{0, i\}$, the functions f_i in (16) are equal to $\text{sgn}\,(I(\mathbf{x}_0) - I(\mathbf{x}_i))$, and one gets (19) of the $\mathbf{D}_{Census}$ descriptor. Likewise, with similar calculations, it is possible to show that descriptors $\mathbf{D}_{CRT}$, $\mathbf{D}_{LDP}$ and $\mathbf{D}_{MLDP}$ respectively defined in (21), (22) and (23), can be derived from the general equation (58).

4.2. Descriptor Form based on Illumination Model Parameter Elimination

This section presents a generalized form of illumination-invariant descriptors based on the elimination of the illumination model parameters $a_{\mathbf{x}_0}$ and $b_{\mathbf{x}_0}$. The aim is to construct descriptor $\mathbf{D}$ such that it verifies (18):

$$\mathbf{D}^i(P_I(\mathbf{x}_0)) = \mathbf{D}^i(a_{\mathbf{x}_0} P_I(\mathbf{x}_0) + b_{\mathbf{x}_0}), \forall i. \tag{59}$$

To this end, the i-th component in $\mathbf{D}$ is defined by:

$$\mathbf{D}^i(P_I(\mathbf{x}_0)) = \Psi\left(\frac{g_{1,i}(P_I(\mathbf{x}_0))}{g_{2,i}(P_I(\mathbf{x}_0))}\right) \tag{60}$$

where $\Psi(\cdot)$ is a non-constant function and $g_{1,i}, g_{2,i} : \mathbb{R}^{n+1} \to \mathbb{R}$ are functions such that

$$g_{1,i}(a_{\mathbf{x}_0} P_I(\mathbf{x}_0) + b_{\mathbf{x}_0}) = h(a_{\mathbf{x}_0}) g_{1,i}(P_I(\mathbf{x}_0)) \tag{61}$$

$$g_{2,i}(a_{\mathbf{x}_0} P_I(\mathbf{x}_0) + b_{\mathbf{x}_0}) = h(a_{\mathbf{x}_0}) g_{2,i}(P_I(\mathbf{x}_0)), \tag{62}$$

with function $h(.)$ taking $a_{\mathbf{x}_0}$ as unique variable. Thus, one has for each component i of a descriptor:

$$\mathbf{D}^i(a_{\mathbf{x}_0} P_I(\mathbf{x}_0) + b_{\mathbf{x}_0}) = \Psi\left(\frac{g_{1,i}(P_I(\mathbf{x}_0))}{g_{2,i}(P_I(\mathbf{x}_0))}\right) = \mathbf{D}^i(P_I(\mathbf{x}_0)),$$

and, consequently, $\mathbf{D}(a_{\mathbf{x}_0} P_I(\mathbf{x}_0) + b_{\mathbf{x}_0}) = \mathbf{D}(P_I(\mathbf{x}_0))$.

As an example, the following function can be used both as $g_{1,i}$ and $g_{2,i}$ in (61) and (62) :

$$g(P_I(\mathbf{x}_0)) = \gamma\left(\sum_{i=1}^{L}\Big(\sum_{j=0}^{n} \alpha_{i,j} I(\mathbf{x}_j)\Big)^{\tau}\right)^{\eta}, \tag{63}$$

where the sequences $\{\alpha_{i,j}\}_{j=0}^{n}$, $i = 0, 1, \ldots, L$ satisfy condition (53), and γ, τ and η are fixed positive numbers. In this case, $g(a_{\mathbf{x}_0} P_I(\mathbf{x}_0) + b_{\mathbf{x}_0}) = (a_{\mathbf{x}_0})^{\tau\eta} g(P_I(\mathbf{x}_0))$.

Descriptors $\mathbf{D}_{Corr}$ and $\mathbf{D}_{NND}$ are two specific cases of the generalized form (60). For example, for descriptor $\mathbf{D}_{Corr}$, one can see that

$$\begin{aligned} \mathbf{D}^i_{Corr}(P_I(\mathbf{x}_0)) &= \frac{I(\mathbf{x}_i) - \mu_P}{\sigma_P} \\ &= \Psi\left(\frac{g_{1,i}(P_I(\mathbf{x}_0))}{g_{2,i}(P_I(\mathbf{x}_0))}\right) \end{aligned} \tag{64}$$

in which functions Ψ, $g_{1,i}$ and $g_{2,i}$ are given by:

$$\begin{cases} \Psi(x) = x \\ g_{1,i}(P_I(\mathbf{x}_0)) = I(\mathbf{x}_i) - \mu_P \\ g_{2,i}(P_I(\mathbf{x}_0)) = \sigma_P = \frac{1}{\sqrt{n+1}}\left(\sum_{i=0}^{n}(I(\mathbf{x}_i) - \mu_P)^2\right)^{\frac{1}{2}}. \end{cases} \tag{65}$$

We have, $I(\mathbf{x}_i) - \mu_P = \sum_{j=0}^{n} \alpha_{i,j} I(\mathbf{x}_i)$ with the sequence $\{\alpha_{i,j}\}_{j=0}^{n}$ is given by

$$\alpha_{i,j} = \begin{cases} \frac{-1}{n+1}, & \text{if } j \neq i \\ \frac{n}{n+1}, & \text{otherwise.} \end{cases} \tag{66}$$

$$\mathrm{M}_4=\begin{bmatrix}2&1&0\\1&0&-1\\0&-1&-2\end{bmatrix}\quad \mathrm{M}_3=\begin{bmatrix}1&2&1\\0&0&0\\-1&-2&-1\end{bmatrix}\quad \begin{bmatrix}0&1&2\\-1&0&1\\-2&-1&0\end{bmatrix}=\mathrm{M}_2$$

$$\mathrm{M}_5=\begin{bmatrix}1&0&-1\\2&0&-2\\1&0&-1\end{bmatrix}\quad \begin{bmatrix}I(\mathbf{x}_4)&I(\mathbf{x}_3)&I(\mathbf{x}_2)\\I(\mathbf{x}_5)&I(\mathbf{x}_0)&I(\mathbf{x}_1)\\I(\mathbf{x}_6)&I(\mathbf{x}_7)&I(\mathbf{x}_8)\end{bmatrix}\quad \begin{bmatrix}-1&0&1\\-2&0&2\\-1&0&1\end{bmatrix}=\mathrm{M}_1$$

$$\mathrm{M}_6=\begin{bmatrix}0&-1&-2\\1&0&-1\\2&1&0\end{bmatrix}\quad \mathrm{M}_7=\begin{bmatrix}-1&-2&-1\\0&0&0\\1&2&1\end{bmatrix}\quad \begin{bmatrix}-2&-1&0\\-1&0&1\\0&1&2\end{bmatrix}=\mathrm{M}_8$$

Fig. 3. Robinson kernels used to define the first descriptor. The patch corresponds to a 3×3 neighborhood around pixel $\mathbf{x}_0$.

It is remarkable that sequence $\{\alpha_{i,j}\}_{j=0}^{n}$ satisfies condition (53) for all $i = 0, 1, \ldots, n$. Therefore, functions $g_{1,i}$ and $g_{2,i}$ in (64) can be rewritten as:

$$g_{1,i}(P_I(\mathbf{x}_0)) = \sum_{j=0}^{n}\alpha_{i,j}I(\mathbf{x}_i) \tag{67}$$

$$g_{2,i}(P_I(\mathbf{x}_0)) = \frac{1}{\sqrt{n+1}}\Big(\sum_{i=0}^{n}\Big(\sum_{j=0}^{n}\alpha_{i,j}I(\mathbf{x}_i)\Big)^2\Big)^{\frac{1}{2}} \tag{68}$$

where $\alpha_{i,j}$ is defined by (66). Referring to (63), it can be seen that in (67), the parameters γ, L, τ, and η are set to 1, while in (68), $\gamma = \frac{1}{\sqrt{n+1}}$, $L = n$, $\tau = 2$, and $\eta = \frac{1}{2}$.

5. From the Generalized Formulations to New Descriptors

As shown before, the existing descriptors given in Section 3.2 are particular cases of the two generalized formulations given in Section 4. Using these two formulations, one can design new illumination-invariant descriptors. Note that the generalized formulations are the starting points for the construction of illumination invariant descriptors. The main issue now lies in the choice of appropriate functions and parameters in (58) for the sign invariance descriptors and in (60) for the illumination model parameter elimination approach.

This section introduces, as examples, two novel descriptors derived from the second general form (60)-(63). Similarly, other descriptors could be derived from the first general form (58). The assessment of the performance of the proposed descriptors is given in Section 6.

As visible in (17), a linear model is used to represent the illumination changes. This model leads to high accuracy if it is applied on small enough neighborhoods. Thus, to define the descriptor for pixel $\mathbf{x}_0$ in image I, we consider a 3×3 patch $P_I(\mathbf{x}_0) = [I(\mathbf{x}_0), I(\mathbf{x}_1), \ldots, I(\mathbf{x}_8)]^T$ centered at $\mathbf{x}_0$ (see Fig. 3).

5.1. First Proposed Descriptor

The first descriptor ($\mathbf{D}_1$) based on the elimination of the illumination change model parameters $a_{\mathbf{x}}$ and $b_{\mathbf{x}}$ is designed as follows:

$$\mathbf{D}_1(P_I(\mathbf{x}_0)) = \frac{\mathcal{A}_{\mathbf{x}_0}}{\|\mathcal{A}_{\mathbf{x}_0}\|_2}, \tag{69}$$

where $\mathcal{A}_{\mathbf{x}_0} = [M_1 \otimes P_I(\mathbf{x}_0), \ldots, M_8 \otimes P_I(\mathbf{x}_0)]^T$ is a vector in $\mathbb{R}^8$ and $M_1, M_2, \ldots, M_8$ are eight Robinson compass kernels (see Fig. 3).

Note that the sum of the elements in the Robinson compass kernels is 0. Thus, a Robinson kernel corresponds to a sequence of coefficients $\{\alpha_j\}_{j=1}^{9}$ satisfying condition (53). The descriptor given in (69) relates to local images structures since each kernel gives a response in one of eight directions around pixel $\mathbf{x}_0$.

Referring to (60), one can see that the i-th component in $\mathbf{D}_1$ is given by $\mathbf{D}_1^i(P_I(\mathbf{x}_0)) = \Psi\left(\frac{g_{1,i}(P_I(\mathbf{x}_0))}{g_{2,i}(P_I(\mathbf{x}_0))}\right)$ in which

$$\begin{cases}\Psi(x) = x\\ g_{1,i}(P_I(\mathbf{x}_0)) = M_i \otimes P_I(\mathbf{x}_0) = \sum_{j=0}^{9}\alpha_{i,j}I(\mathbf{x}_i)\\ g_{2,i}(P_I(\mathbf{x}_0)) = \left(\sum_{i=0}^{8}\left(\sum_{j=0}^{9}\alpha_{i,j}I(\mathbf{x}_j)\right)^2\right)^{\frac{1}{2}}.\end{cases} \tag{70}$$

Functions $g_{1,i}$, $g_{2,i}$ are defined using (63), and $\{\alpha_{i,j}\}_{j=1}^{9}$ is the sequence of the elements of matrix M_i satisfying (53).

Descriptor $\mathbf{D}_1$ simulates a star-shaped structure which gives normalized grey-level variations in eight directions, along horizontal, vertical and diagonal line segments originating all from the patch center (pixel $\mathbf{x}_0$). It captures a local 2D intensity variation information which relates to the shape and sharpness of the textures overlapped by the patch. For homologous points in two images, the descriptor response is similar when the local grey-level distributions are similar. The similarity measurement with this descriptor is robust since intensity differences attenuate the effect of additive intensity changes (elimination of $b_{\mathbf{x}_0}$), whereas the normalization with $\|A_{\mathbf{x}_0}\|$ limits the effect of multiplicative intensity changes (elimination of $a_{\mathbf{x}_0}$).

5.2. Second Proposed Descriptor

In this example, the functions in the general form (60) are defined as:

$$\begin{cases} \Psi(x) = \exp(x) \\ g_{1,i}(P_I(\mathbf{x}_0)) = I(\mathbf{x}_i) - \min(P_I(\mathbf{x}_0)) \\ g_{2,i}(P_I(\mathbf{x}_0)) = \max(P_I(\mathbf{x}_0) - \min(P_I(\mathbf{x}_0)). \end{cases} \tag{71}$$

This leads to descriptor $\mathbf{D}_2$:

$$\mathbf{D}_2(P_I(\mathbf{x}_0)) = \begin{bmatrix} \exp \frac{I(\mathbf{x}_0)-\min(P_I(\mathbf{x}_0))}{\max(P_I(\mathbf{x}_0)-\min(P_I(\mathbf{x}_0))} \\ \exp \frac{I(\mathbf{x}_1)-\min(P_I(\mathbf{x}_0))}{\max(P_I(\mathbf{x}_0)-\min(P_I(\mathbf{x}_0))} \\ \vdots \\ \exp \frac{I(\mathbf{x}_n)-\min(P_I(\mathbf{x}_0))}{\max(P_I(\mathbf{x}_0)-\min(P_I(\mathbf{x}_0))} \end{bmatrix}. \tag{72}$$

The idea behind descriptor $\mathbf{D}_2$ is similar to that which motivated the design of $\mathbf{D}_1$. The main difference is that the structure corresponding to $\mathbf{D}_2$ is not necessarily "star-shaped" because the structure center from which the line segments originate is not $\mathbf{x}_0$ but the pixel with the smallest intensity $\min(P_I(\mathbf{x}_0))$ in patch $P_I(\mathbf{x}_0)$. With this minimum, the intensity changes relating to textures are maximized. Effects of additives and multiplicative illumination terms are attenuated by subtracting $\min(P_I(\mathbf{x}_0))$ from the patch intensities and by normalization with ($\max(P_I(\mathbf{x}_0))$ - $\min(P_I(\mathbf{x}_0))$).

The next section compares the performance of all descriptors discussed in this paper (the descriptors of the literature and the two proposed ones).

6. Descriptor Performance Comparison

As mentioned in Section 1, the aim of this work is not to propose a competitive OF method. The main objective of this contribution is to propose a theoretical study on the illumination-invariant descriptors used in variational OF models. For this reason, this paper focuses on the evaluation and comparison of the performance of descriptors by estimating the OF in scenes with and without strong illumination changes.

The proposed descriptors are compared to the descriptors presented in Section 3.2: Census [39], CRT [43], LDP [52], MLDP [15], Corr [14], and NND [16]. For a fair evaluation, all descriptors are placed in the same variational OF model as presented in Section 2. The configuration of the variational model is presented in subsection 6.1.

The performance of the descriptors were evaluated on five datasets with known ground truth OF and including images corresponding to five scene conditions:

1. Weak illumination changes combined with small displacements (Section 6.2).
2. Weak illumination changes combined with large displacements (Section 6.3).
3. Strong illumination changes associated with small displacements (Section 6.4).
4. Strong illumination changes associated with large displacements (Section 6.5).
5. All possible illumination change and displacement combinations (weak or strong illumination changes combined with small, or large displacements, Section 6.6).

Moreover, this section also proposes a subjective comparison using real endoscopic images. In addition, the effects of some parameters on the performance of each descriptor are reported in this section.

6.1. Experimental configuration

In this performance study, all descriptors are placed in exactly the same optimization scheme such that the only factors influencing the differences in the OF field results are the descriptors themselves. Thus, to evaluate the descriptors, the same regularization term (see (9)) and the same data-term (see (8)) with a quadratic penalty function $\Psi(\mathbf{v}) = \|\mathbf{v}\|^2$ were used as energy in the minimization process (the quadratic penalty function facilitates the optimization problem solving).

Similarly to [14], the minimization in (14) is performed with the projected-proximal-point algorithm [51] and the classical coarse-to-fine warping strategy is used to cope with large displacements. This pyramidal approach uses a classical bilinear interpolation for building the images at different levels, and for

Table 1. Default values of parameters σ_1, σ_2, Py_s and λ obtained with the Middlebury and KITTI training datasets.

Desc.	Middlebury				KITTI			
	σ_1	σ_2	Py_s	λ	σ_1	σ_2	Py_s	λ
$\mathbf{D}_1$	3	5	0.8	50	1	5	0.9	50
$\mathbf{D}_2$	3	5	0.7	15	1	5	0.9	30
Census	3	5	0.8	20	1	5	0.9	30
CRT	5	7	0.5	0.8	1	5	0.9	1.1
LDP	5	7	0.8	17	1	5	0.7	25
MLDP	3	5	0.5	9	1	5	0.8	7
Corr	3	5	0.5	12	1	5	0.9	10
NND	3	5	0.7	100	1	5	0.9	75

up-sampling the flow field to the finer level. At each pyramid level, 5 warps and 40 iterations per warp are used to optimize the energy. As recommended in [18], a median filter of size 5×5 is applied to the intermediate flow results after every warping iteration.

In the data-term, the default size of the descriptor patch $P_I(\mathbf{x}_0)$ is 3×3 pixels, except for the patch in the NND descriptor where this number is 5×5 (a size lower than 5×5 is not possible for this descriptor). In the regularization-term, the size of neighborhood $\mathcal{N}_x$ in (9) is systematically set to 5×5.

The remaining parameters consisting of σ_1 and σ_2 in (10), λ in (14), and the pyramid scale factor Py_s (parameter Py_s corresponds classically to the ratio of both the image width and height when passing from level n to level $n+1$) are specifically adjusted for each descriptor. The optimal parameter quadruplet (σ_1, σ_2, Py_s, λ) is obtained for each descriptor by computing the OF for all combinations of following parameter values: σ_1 and $\sigma_2 \in \{1, 3, 5, 7\}$, $Py_s \in \{0.5, 0.6, 0.7, 0.8, 0.9\}$, and $\lambda \in (0, 140]$. The effects of parameters Py_s and λ will be studied because they are the most crucial for ensuring robust and accurate flow field estimation.

6.2. *Weak illumination changes associated with small displacements*

The Middlebury training benchmark [9] was used in this case. This dataset consists of eight training image pairs with known ground truth. The training image pairs include synthetic images, hidden textures in real scenes, indoor and outdoor scenes, etc. Besides weak illumination changes and small displacements, these images also include rather pronounced textures leading to optimal OF computation conditions. The Average End-point Error (AEE) and the Average Angular Error (AAE) are the two reference quality criteria used to measure the accuracy of the estimated flow fields on the eight training image pairs.

Different combinations of σ_1, σ_2, Py_s and λ values (see at the end of Section 6.1) were systematically tested to find the optimal settings for these parameters. The optimal values correspond to the parameter combinations for which the mean AEE and the mean AAE computed for the eight image pairs of the training Middlebury dataset are the lowest. Table 1 gives for each descriptor the optimal parameter values. The strategy of this experimental adjustment method was to consider large interval values for all parameters to be sure to find an optimal combination for the $\{\sigma_1, \sigma_2, Py_s, \lambda\}$ quadruplet.

The OF results in terms of AEE and AAE values are given in Table 2 for the different descriptors. The lowest and the second lowest errors obtained for each image pair correspond to bold numbers and underlined numbers, respectively. As visible, most of the best results were obtained by one of the two proposed descriptors ($\mathbf{D}_1$). It is noticeable that, whatever the descriptor, for weak illumination changes there is no significant difference in terms of AEE and AAE for some images (e.g. Grove2 or Hydrangea), while for some other images some descriptors, like LDP, may lead to high AEE (e.g. for Urban3) or AAE (e.g. for Venus) errors. For all images, the AEE and AAE values remain among the smallest for the two proposed descriptors ($\mathbf{D}_1$ and $\mathbf{D}_2$). This result is confirmed in Table 3 which ranks the descriptors according to two criteria (average AEE and average AAE computed for the eight training image pairs). In terms of accuracy (AEE and AAE), MLDP is the only descriptor of the literature which is placed on one of the first three places of this ranking while descriptor $\mathbf{D}_1$ is the most accurate.

Table 2. AEE and AAE values obtained for the Middlebury training database. Parameters σ_1, σ_2, λ and Py_s were set on the default values given in Table 1.

Des.	Dimetrodon		Grove2		Grove3		Hydrangea		RubberWhale		Urban2		Urban3		Venus	
	AEE	AAE	AEE	AAE	AEE	AAE	AEE	AAE	AEE	AAE	AEE	AAE	AEE	AAE	AEE	AAE
D_1	**0.11**	**2.09**	**0.13**	1.80	**0.46**	**4.79**	**0.17**	2.07	**0.08**	2.68	0.35	3.30	**0.48**	**3.55**	**0.25**	3.88
D_2	0.20	4.37	**0.13**	**1.76**	0.48	5.08	0.18	2.16	**0.08**	2.68	0.35	3.48	0.71	5.06	**0.25**	3.75
Census	0.26	5.51	0.15	2.01	0.52	5.34	0.19	2.21	0.09	2.73	0.37	3.50	0.68	4.79	0.91	7.48
CRT	0.22	4.82	0.17	2.24	0.56	5.83	0.19	2.17	0.23	2.99	0.44	3.98	0.71	5.02	0.30	4.06
LDP	0.14	2.83	0.26	3.32	1.03	9.31	0.22	2.49	0.11	3.36	0.64	4.94	2.8	11.32	7.25	18.48
MLDP	0.13	2.35	**0.13**	1.84	0.48	5.02	**0.17**	2.06	**0.08**	2.71	**0.33**	**3.18**	0.57	4.18	0.26	**3.68**
Corr	0.23	4.86	0.20	2.53	0.49	5.21	**0.17**	**2.01**	**0.08**	**2.60**	0.34	3.65	0.75	5.67	0.28	4.26
NND	0.22	4.62	0.19	2.77	0.63	6.50	0.18	2.29	0.09	3.06	0.52	3.93	0.59	4.28	0.34	5.3

Table 3. Descriptor ranking according to the accuracy (the average errors) on the eight image pairs of Table 2.

Rank	Average AEE		Average AAE	
1	**D_1**	0.255	**D_1**	3.02
2	**MLDP**	0.268	**MLDP**	3.12
3	**D_2**	0.297	**D_2**	3.54
4	**Corr**	0.320	**Corr**	3.85
5	**NND**	0.345	**CRT**	3.89
6	**CRT**	0.352	**NND**	4.09
7	**Census**	0.396	**Census**	4.19
8	**LDP**	1.556	**LDP**	7.00

6.3. Weak illumination changes associated with large displacements

The KITTI 2012 [11] and KITTI 2015 [12] datasets are composed of real-world images taken from a driving platform and consist of 194 and 200 image pairs, respectively. All sequences are with known ground truth and, even if the illumination changes are rather weak in average, these changes are varying along the image sequences since they are arising in real scenes. However, even for the large displacements between images, the illumination changes remain moderate.

The AEE and BP3 (the percentage of bad-pixels which have an end-point error above 3 pixels) are the two quality criteria used for an objective comparison. Four sequences in the KITTI 2012 training set, namely sequences 11, 15, 44, and 74, were chosen to determine the optimal (default) values of parameters σ_1, σ_2, Py_s and λ for each descriptor. The adjusted parameter combinations are those leading to the lowest mean BP3 value in the non-occluded areas. The default values for the parameters are shown in the KITTI column of Table 1 for all descriptors.

Table 4 shows a comparison of the proposed descriptors $\mathbf{D}_1$ and $\mathbf{D}_2$ with the existing descriptors. It is noticeable that the Corr descriptor is ranked at the first position according to all the quality criteria. Descriptors $\mathbf{D}_1$ and $\mathbf{D}_2$ take the remaining two positions in the top three. The 4th and 5th places were obtained by NND and MLDP. However, there is no significant difference in the quality indexes between the descriptors of the top five. Globally, the average errors of descriptors Census, CRT and LDP are significantly higher than those of the five first places.

6.4. Strong illumination changes associated with small displacements

In order to compare descriptor performance in this case, we report experimental results on image pairs with simulated illumination changes. The RubberWhale image pair of the Middlebury training dataset was used in these experiments. As visible in Fig. 4(d) and Fig. 4(e), this image pair exhibits few illumination changes and small displacements. Illumination variations were simulated and applied on this pair to obtain images with strong illumination changes. The ground truth OF of these image pairs being available, the AEE and AAE criteria are again used to quantify the descriptor performance.

Table 4. Results for the KITTI 2012 [11] and the KITTI 2015 [12] training datasets. The reported error measures are the mean of average end-point errors (AEE), and the percentage of erroneous pixels over a threshold of 3 pixels (BP3). Noc and Occ stand for non-occluded and occluded, respectively.

Descriptors	KITTI 2012				KITTI 2015			
	Noc		Occ		Noc		Occ	
	AEE	BP3(%)	AEE	BP3(%)	AEE	BP3(%)	AEE	BP3(%)
Corr	**1.95**	**8.81**	**4.97**	**18.33**	**6.07**	**21.74**	**12.66**	**30.11**
$\mathbf{D}_2$	2.10	9.33	5.28	18.74	6.40	22.65	13.26	30.78
$\mathbf{D}_1$	2.12	9.39	5.38	19.05	6.46	23.20	13.14	31.35
NND	2.12	9.57	5.36	19.52	6.40	23.12	13.02	31.40
MLDP	2.19	9.57	5.44	19.27	6.57	23.32	13.33	31.47
Census	3.79	14.35	10.78	24.91	9.24	28.05	19.18	36.23
CRT	4.25	15.36	12.04	25.83	9.79	28.67	20.46	36.80
LDP	10.03	24.06	20.95	33.39	14.71	34.69	27.49	41.93

The illumination changes between the source and target images are controlled using:

$$I_{out}(x,y) = \begin{cases} 0 \text{ if } \mathcal{M}(x,y).I_{in}(x,y) + C \leq 0 \\ 255 \text{ if } \mathcal{M}(x,y).I_{in}(x,y) + C \geq 255 \\ [\mathcal{M}(x,y).I_{in}(x,y) + C] \text{ otherwise} \end{cases} \quad (73)$$

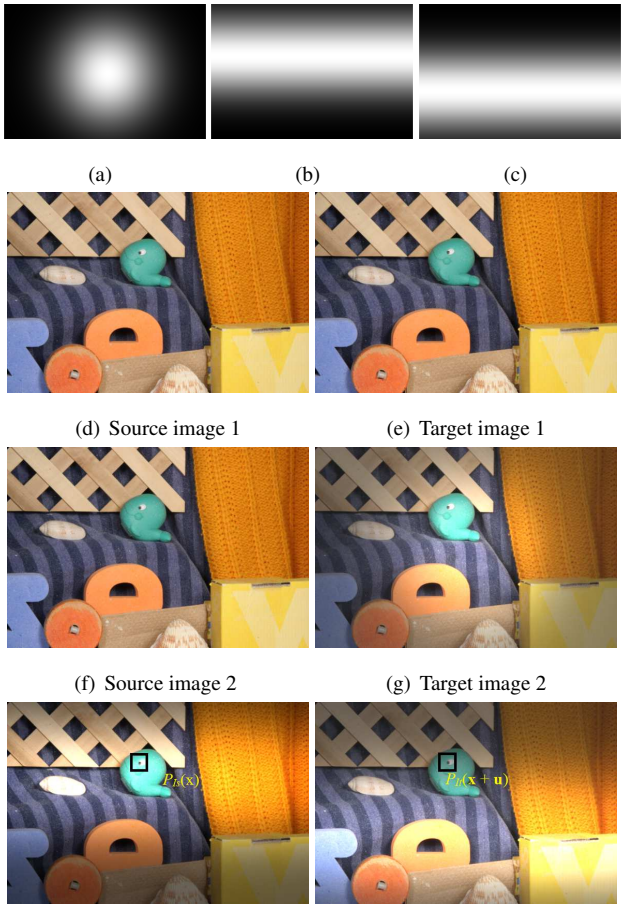
where I_{in} and I_{out} are the (original) input and (illumination modified) output images respectively, $\mathcal{M}$ is a multiplicative mask, C is an additive constant, and $[.]$ is the rounding symbol of $\mathcal{M}(x,y).I_{in}(x,y) + C$ to its closest integer.

Three cases of illuminations changes are considered:

- Case 1: The source-target image pair corresponds to the original RubberWhale image pair (Figs. 4(d) and 4(e)). In this case, the illumination variation between two images is very small.
- Case 2: The source-target image pair is in Fig. 4(f) - Fig. 4(g), where the illumination of the target image is changed by the multiplier $\mathcal{M}$ illustrated in Fig. 4(a), and by additive factor $C = 20$.
- Case 3: The source-target image pair is in Fig. 4(h) - Fig. 4(i). In this case, the illumination in both source and target images are changed by the multiplicative masks given in Fig. 4(b) and Fig. 4(c) respectively, and with $C = 0$ in the source image and $C = 20$ in the target image.

The small squares in Figs. 4(h) and 4(i) indicate two 3×3 neighborhoods of homologous pixels of the source (I_s) and target (I_t) images. In Fig. 4(j), the intensities (computed from the RGB colors) of these homologous regions are given by the left and right 3×3 grids, while the central grid with dashed lines gives the intensity ratios I_s/I_t of corresponding pixels. One can notice that in the central grid the nine intensity ratios are slightly different. In fact, small ratio differences indicate that the illumination variation between the 3×3 homologous patches is not exactly affine. These test conditions allow to simulate in a realistic way a real situation in which the intensity changes can only deviate a little bit from a perfect affine illumination model in small 3×3 neighborhoods.

For each descriptor, the optimal values of parameters σ_1, σ_2, Py_s and λ are determined by testing all their combinations on the three images pairs in Fig. 4. The optimal values of σ_1 and σ_2 of the regularization term (see (9) and (10)) were the same for all descriptors, namely $\sigma_1 = 3$ and $\sigma_2 = 5$. On the contrary, scale parameter Py_s and weight λ have different values for each descriptor: $(Py_s, \lambda) = (0.8, 50)$ for descriptor $\mathbf{D}_1$, $(0.7, 15)$ for $\mathbf{D}_2$, $(0.8, 20)$ for Census, $(0.6, 0.8)$ for CRT, $(0.8, 17)$ for LDP, $(0.6, 5)$ for MLDP, $(0.6, 12)$ for Corr, and $(0.7, 100)$ for NND. It is noticeable that, compared to the optimal values obtained for the weak illumination changes given in Table 1, the opti-



219	251	255
216	237	240
143	126	123

3x3 patch $P_{I_s}(\mathbf{x})$

1.7381	1.7310	1.6558
1.7008	1.8231	1.7647
1.6824	1.6154	1.5185

Ratio $P_{I_s}(\mathbf{x})\,/P_{I_t}(\mathbf{x}+\mathbf{u})$

126	145	154
127	130	136
85	78	81

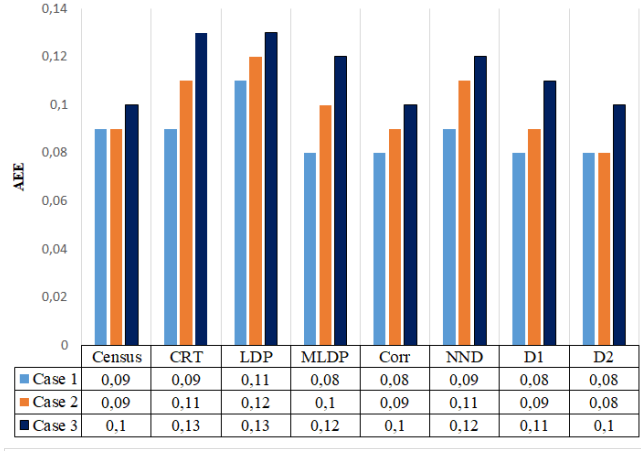
3x3 patch $P_{I_t}(\mathbf{x}+\mathbf{u})$

(j) Intensity values and ratios

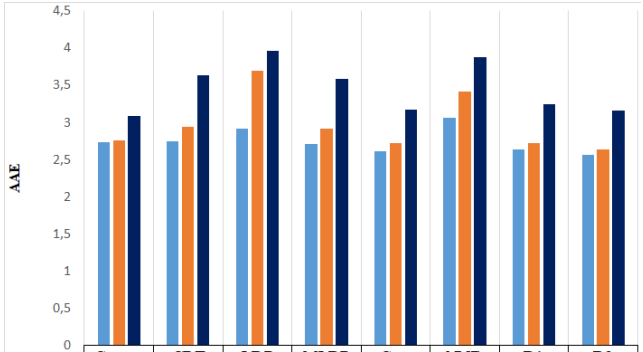
Fig. 4. Experiments with simulated illumination changes. Images (a), (b) and (c) represent the multiplicative masks used to generate the synthetic illuminations in (g), (h) and (i), respectively. Sub-figures (d) to (i) present the source-target image pairs used to evaluate the performance of the descriptors and their stability against illumination changes in terms of OF accuracy. (d) and (e) : image pair simulating weak illumination changes. (f) and (g) : strong illumination change in the target image center. (h) and (i) strong vertical illumination intensity gradient between the source and target images. (j) Intensity values of the two 3×3 homologous regions marked by squares in Figs. 4(h) and 4(i). The central grid gives the intensity ratios of homologous pixels of these regions.

mal values of the pyramid scale parameter Py_s corresponding to descriptors CRT, MLDP and Corr increased from 0.5 to 0.6. The impact of the parameter pair (Py_s, λ) on the performance of each descriptor will be discussed in detail in Section 6.6.1.

The results of these experiments with strong illumination



	Census	CRT	LDP	MLDP	Corr	NND	D1	D2
Case 1	0,09	0,09	0,11	0,08	0,08	0,09	0,08	0,08
Case 2	0,09	0,11	0,12	0,1	0,09	0,11	0,09	0,08
Case 3	0,1	0,13	0,13	0,12	0,1	0,12	0,11	0,1

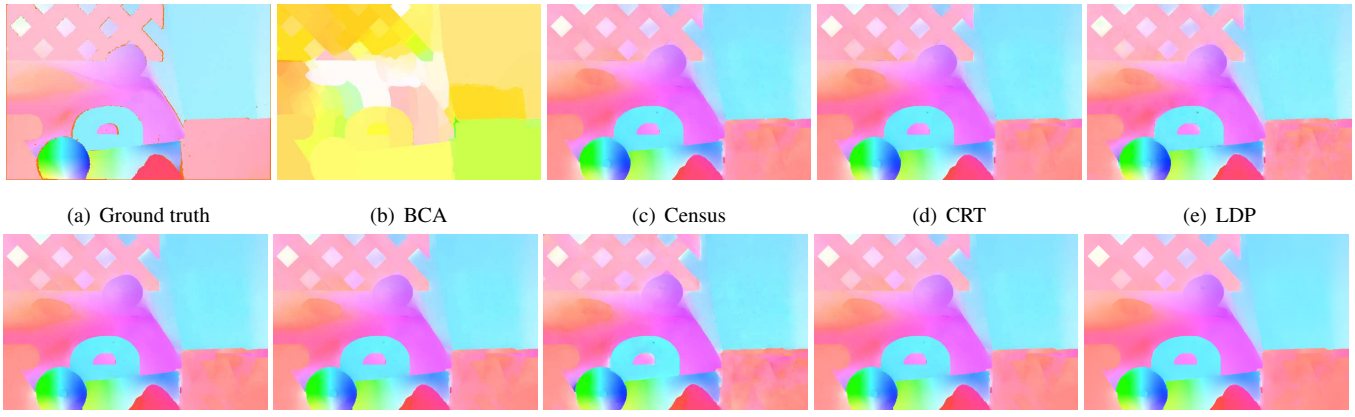


	Census	CRT	LDP	MLDP	Corr	NND	D1	D2
Case 1	2,73	2,75	2,92	2,71	2,61	3,06	2,64	2,57
Case 2	2,76	2,94	3,69	2,92	2,72	3,42	2,72	2,64
Case 3	3,09	3,63	3,96	3,59	3,17	3,88	3,25	3,16

Fig. 5. Performances and stability in terms of AEE and AAE values obtained for different descriptors with the three different illumination change simulations represented in Fig. 4.

changes are given in Fig. 5 for the six descriptors of the literature and the two proposed descriptors. AEE and AAE are the two reference quality criteria used to measure the accuracy of the estimated flow fields. In the following, the terms "stable" or "stability" refer to the ability of a descriptor to lead to similar OF accuracy under weak and strong illumination changes (i.e. the accuracy remains constant when passing from weak to strong illumination changes). It is noticeable that all descriptors demonstrate stability with respect to illumination variations. This confirms the relevance of the proposed illumination-invariant criterion in (18). We can observe that, in this experiment, the quantitative results of the four descriptors $\mathbf{D}_1$, $\mathbf{D}_2$, Corr and Census are slightly better compared to those of the remaining four (CRT, LDP, MLDP and NND).

This result in terms of accuracy stability under strong illumination change conditions is visually illustrated in Fig. 6 which



(a) Ground truth (b) BCA (c) Census (d) CRT (e) LDP

(f) MLDP (g) Corr (h) NND (i) $\mathbf{D}_1$ (j) $\mathbf{D}_2$

Fig. 6. OF results for the RubberWhale sequence with simulated illumination variations. The source-target image pair is given in Figs. 4(h)-4(i).

shows different OF images computed for Figs. 4(h) and 4(i) with the vertical gradient illumination change (case 3). It visible in Fig. 6 that the colors of the OF image obtained with the brightness constancy assumption (BCA, see Fig. 6(b)) is quite different from those of the ground truth given in Fig. 6(a). Unsurprisingly, the worst OF is that obtained with the BCA assumption. The differences with the ground truth is by far less perceptible for the illumination-invariant descriptors whose OF are shown in Figs. 6(c)-6(j).

6.5. *Strong illumination changes associated with large displacements*

Both the clean pass and the final pass of the Bamboo-2 and Shaman-3 sequences (each sequence consists of 50 images) of the MPI Sintel training dataset [10] are used in these experiments because they include strong illumination changes and large displacements between two consecutive images (see Fig. 7). Similarly to the parameter adjustment done in Section 6.4, the best combination of (σ_1, σ_2, Py_s, λ) quadruplets consists of constant σ_1 and σ_2 parameters (their values are 3 and 5, respectively) and descriptor dependent Py_s pyramid scale and λ weight values.

The experimental results are reported in Table 5. By considering only this table, one can conclude that four descriptors ($\mathbf{D}_1$, $\mathbf{D}_2$, Corr and MLDP) outperform the remaining four descriptors which are almost systematically among the least accurate for all images. One can notice that Corr notably exhibits high accuracy in this table.

6.6. *Descriptor comparison on a more general dataset*

In subsections 6.2 to 6.5, the descriptor performances were evaluated on images corresponding to four distinct situations obtained by using images with either small or large displacements combined with either weak or strong illumination changes. The aim of this subsection is to assess the descriptor performances on a dataset including simultaneously all the four situations tested separately in previous subsections.

This "more general" dataset consists of nine image pairs involving two benchmarks, namely Middlebury and MPI Sintel. The first three image pairs are those given in Fig. 4 (modified Middleburry images) and including either weak or strong illumination changes combined with small displacements. The other six image pairs are extracted from two final pass sequences of the MPI Sintel training dataset [10]. More specifically, four consecutive images from frame-0039 to frame-0042 of the Bamboo-2 sequence (see Figs. 7(a)-7(d)) represent three image pairs, and four other consecutive images from frame-0001 to frame-0004 of the Shaman-3 sequence (see Figs. 7(e)-7(h)) correspond to the last 3 image pairs. These MPI-Sintel data were chosen since they include both small and strong illumination changes associated to large displacements.

The AEE and AAE metrics are used to objectively evaluate the accuracy of the OF field. Besides the OF accuracy tests, this custom-made dataset is also used to evaluate the effects of some

(a) Bamboo-2, frame 0039 (b) Bamboo-2, frame 0040 (c) Bamboo-2, frame 0041 (d) Bamboo-2, frame 0042

(e) Shaman-3, frame 0001 (f) Shaman-3, frame 0002 (g) Shaman-3, frame 0003 (h) Shaman-3, frame 0004

Fig. 7. Four consecutive images in the final pass of the Bamboo-2 and Shaman-3 sequences of the MPI Sintel training dataset [10]. The first row includes the Bamboo images and the second row shows the Shaman images.

Table 5. Results for the Bamboo 2 and Shaman 3 sequences of the MPI Sintel [10] training dataset. These sequences include illumination changes.

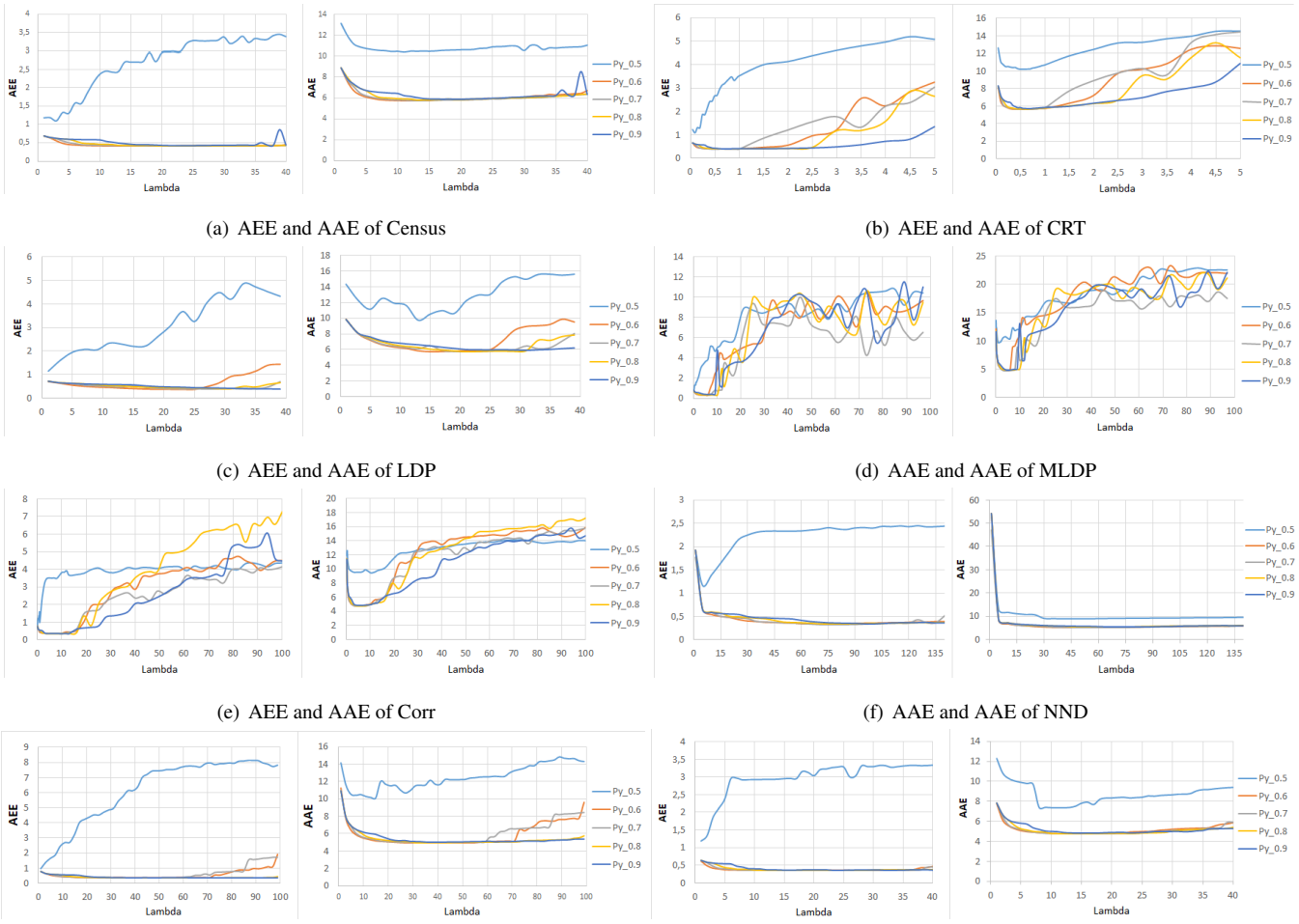
Clean								**Final**							
Bamboo 2				Shaman 3				Bamboo 2				Shaman 3			
Avg. AEE		Avg. AAE		Avg. AEE		Avg. AAE		Avg. AEE		Avg. AAE		Avg. AEE		Avg. AAE	
Corr	1.73	$\mathbf{D}_1$	7.53	$\mathbf{D}_1$	0.16	$\mathbf{D}_2$	2.32	**Corr**	1.62	**Corr**	7.43	**MLDP**	0.31	**MLDP**	4.31
$\mathbf{D}_1$	1.85	**Corr**	7.66	**Corr**	0.16	$\mathbf{D}_1$	2.43	$\mathbf{D}_1$	1.68	$\mathbf{D}_1$	7.48	$\mathbf{D}_1$	0.32	$\mathbf{D}_1$	4.41
MLDP	1.88	$\mathbf{D}_2$	7.72	$\mathbf{D}_2$	0.17	**Corr**	2.53	**MLDP**	1.70	**MLDP**	7.52	**Corr**	0.33	$\mathbf{D}_2$	4.59
$\mathbf{D}_2$	2.01	**MLDP**	7.95	**MLDP**	0.17	**MLDP**	2.59	$\mathbf{D}_2$	1.86	$\mathbf{D}_2$	7.65	**NND**	0.35	**Corr**	4.66
NND	2.04	**NND**	7.97	**NND**	0.17	**NND**	2.63	**NND**	2.00	**NND**	7.66	$\mathbf{D}_2$	0.36	**NND**	4.98
LDP	2.60	**LDP**	9.22	**CRT**	0.21	**Census**	2.82	**CRT**	2.56	**LDP**	9.85	**LDP**	0.41	**LDP**	5.46
CRT	2.65	**CRT**	10.29	**Census**	0.21	**CRT**	2.89	**Census**	2.64	**CRT**	10.26	**CRT**	0.46	**CRT**	6.13
Census	2.80	**Census**	10.83	**LDP**	0.25	**LDP**	3.72	**LDP**	2.80	**Census**	10.67	**Census**	0.47	**Census**	6.30

important parameters on the performance of the descriptors.

6.6.1. Influence of parameters

This subsection investigates the impact of two crucial OF scheme parameters on the descriptor performances. These parameters are the pyramid scale parameter Py_s and the trade-off parameter λ in (14). It is usually difficult to adjust these two parameters under different displacements between image and for different illumination changes. Therefore, a descriptor should not only be accurate, but *(i)* its accuracy should as less as possible be affected by scene variations and *(ii)* the crucial OF scheme parameters should be easily adjustable.

The experiments in Sections 6.2, 6.3, 6.4 and 6.5 have shown that setting the values of σ_1 and σ_2 in (10) to 3 and 5 respectively is appropriate for the four different scene conditions. These values were also chosen in these experiments. The results in this section were obtained by testing in a systematic way all combinations of Py_s and λ. The pyramid scale parameter Py_s took values 0.5, 0.6, 0.7, 0.8 and 0.9 successively, while the trade-off parameter λ in (14) varied from 0.05 to 140. However, to show more clearly the optimal value range of λ, we only present here, for each descriptor, the experimental results with λ-values around its optimal range. For each descriptor, the AEE and AAE curves according to λ are given for every values of Py_s. Thus, for the Census descriptor λ belongs to [1:1:40], where [1:1:40] means that λ takes the values ranging from 1 to 40 and are changing with a step of 1. Similarly, $\lambda \in$ [0.05:0.05:5] for CRT, $\lambda \in$ [1:1:40] for LDP, $\lambda \in$ [0.05:0.05:1] $\cup$ [2:1:100] for MLDP and Corr, $\lambda \in$ [1:1:140] for NND, $\lambda \in$ [1:1:100] for $\mathbf{D}_1$, and $\lambda \in$ [1:1:40] for $\mathbf{D}_2$.



(a) AEE and AAE of Census

(b) AEE and AAE of CRT

(c) AEE and AAE of LDP

(d) AAE and AAE of MLDP

(e) AEE and AAE of Corr

(f) AAE and AAE of NND

(g) AEE and AAE of $\mathbf{D}_1$

(h) AEE and AAE of $\mathbf{D}_2$

Fig. 8. OF accuracy according the values of parameters Py_s and λ. The mean AAE and AEE values are computed for the dataset described at the beginning of Section 6.6 and plotted according λ for different values of parameter Py_s.

The OF of every image pair in the "general" dataset was computed for each descriptor and for all (Py_s, λ) pairs. The two mean values of the nine AEE and AAE criteria obtained for each (Py_s, λ) pair act as descriptor accuracy measure. Fig. 8 shows, for each descriptor and for each value of Py_s, the behavior of the mean AEE and AAE errors according to the λ values.

As visible in Fig. 8 for all descriptors, the mean AEE and AAE curves with $Py_s = 0.5$ (light blue plots) are always above the curves corresponding to $Py_s = 0.6$, 0.7, 0.8 and 0.9. These non constant and rather large AEE and AAE mean values according to λ show that $Py_s = 0.5$ is inappropriate for scenes with strong variations in terms of illumination conditions and displacement magnitudes. In fact, as shown by the other curves, it appears that parameter Py_s should at least have a minimal value of 0.6.

Curve parts with constantly small AEE and AAE values (curve plateau with minimal error, CPME) in the plots of Fig. 8 indicate that a descriptor is both accurate and robust towards changes of illumination conditions and displacement magnitudes. Except for CRT (see Fig. 8(b)), MLDP (see Fig. 8(d)) and LDP (see Fig. 8(c)), rather large CPME exist for the descriptors. It is noticeable for the five descriptors with a CPME in their plots (Census in Fig. 8(a)), LPD in Fig. 8(c), NND in Fig. 8(f), $\mathbf{D}_1$ in Fig. 8(g) and $\mathbf{D}_2$ in Fig. 8(h)) that the plateaus are preserved when Py_s increases from 0.6 to 0.9. One can also notice that an increase of the Py_s parameter value also increases the number of levels in the pyramid scheme. For this reason, a solution to minimize the computation time while preserving the CPME in the AEE and AAE plots is to set Py_s to 0.6.

For a given descriptor, the optimal value of λ should be in

an interval in which the AEE and AAE values are constantly weak (i.e. in the λ interval defined by the common λ values of the CPME in the AEE and AAE curves of a descriptor). The optimal value ranges of λ for descriptors are approximately

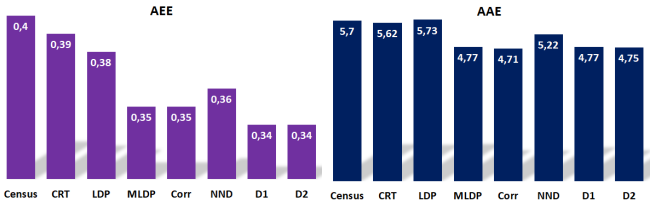
- Census : $[15, 35]$ (interval width : 20, see Fig. 8(a)),
- CRT : $[0.5, 1]$ (interval width : 0.5, see Fig. 8(b)),
- LDP : $[17, 25]$ (interval width : 8, see Fig. 8(c)),
- MLDP : $[3, 9]$ (interval width : 6, see Fig. 8(d)),
- Corr : $[3, 12]$ (interval width : 9, see Fig. 8(e)),
- NND : $[75, 120]$ (interval width : 45, see Fig. 8(f)),
- $\mathbf{D}_1$: $[20, 60]$ (interval width : 40, see Fig. 8(g)), and
- $\mathbf{D}_2$: $[12, 31]$ (interval width : 19, see Fig. 8(h)).

It is noticeable that, for $Py_s = 0.6$, descriptors Census, NND, $\mathbf{D}_1$, and $\mathbf{D}_2$ lead to constantly high accuracy for large ranges of λ-values. In practice it means that the optimal λ value is easily adjustable and that these descriptors are the less sensitive to illumination changes. Descriptor LDP has also high accuracy in a large range of λ-values when $Py_s = 0.9$ (Dark blue curve in Fig. 8(c)). Descriptors CRT, MLDP, and Corr have small λ intervals with high accuracy and seem less appropriate when the illumination and displacement magnitudes change from one scene type to another.

6.6.2. *Descriptor performance with optimal parameter settings*

The mean AAE and AEE values over the nine image pairs of the "more general" dataset (as described at the beginning of Section 6.6) are used to compare the performance of the eight descriptors. These values are computed using the parameter set optimized in Subsection 6.6.1 (i.e., $\sigma_1 = 3$, $\sigma_1 = 5$, $Py_s = 0.6$ for all 8 descriptors, while $\lambda = 20$, 0.8, 25, 8, 12, 100, 50, and 15 for descriptors Census, CRT, LDP, MLDP, Corr, NND, $\mathbf{D}_1$, and $\mathbf{D}_2$, respectively).

As visible in Fig. 9, even if the differences between the eight descriptors are very small, the values of the objective criteria



(a) AEE comparison (b) AAE comparison

Fig. 9. Objective comparison between the eight descriptors on the synthetic dataset.

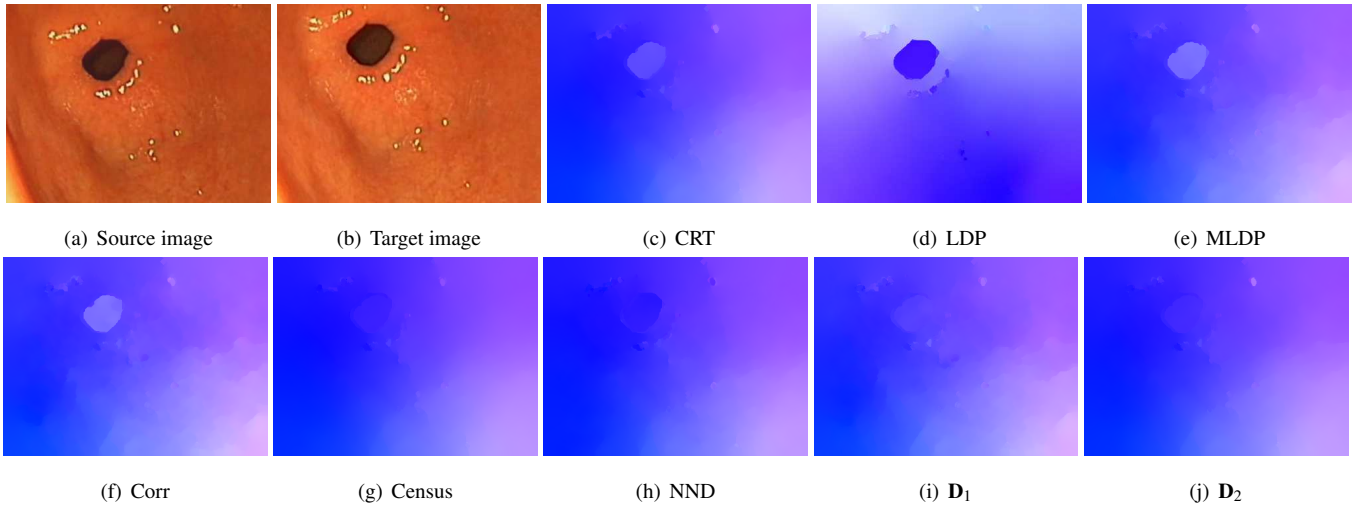
(AAE and AEE) of descriptors MLDP, Corr and of the two proposed descriptors ($\mathbf{D}_1$ and $\mathbf{D}_2$) indicate lower OF errors than those of the other methods (Census, CRT, LDP and NND). By comparing MLDP, Corr, $\mathbf{D}_1$ and $\mathbf{D}_2$ is it visible that the AAE values are nearly the same for the four descriptors, while the AEE values of the proposed descriptors are slightly better.

By considering globally the results of section 6.6, one can conclude that the eight descriptors can be separated into four groups as follows:

- CRT and LDP are among the group of four descriptors with the least accuracy when the parameters are optimally tuned and their OF is rather less λ independent (scene condition independent).
- Census and NDD are the other two descriptors of the group with the least accuracy when the parameters are optimally tuned, but their OF is more λ independent.
- MLDP and Corr are among the group of four descriptors with the best accuracy when the parameters are optimally tuned. However their OF is rather less λ independent.
- $\mathbf{D}_1$ and $\mathbf{D}_2$ are both in the descriptor group with the best accuracy when the parameters are optimally tuned and in the group of descriptors with a λ independent OF.

6.7. *Descriptor comparison in very complex scenes*

The aim of this section is to test the descriptors on a scene type with extrem conditions (very inhomogeneous scene illumination, strong illumination changes between images, lack of textures and large camera displacements). A test on such scenes



(a) Source image (b) Target image (c) CRT (d) LDP (e) MLDP

(f) Corr (g) Census (h) NND (i) $\mathbf{D}_1$ (j) $\mathbf{D}_2$

Fig. 10. Experimental result on a gastroscopic image pair. (a) and (b) are two endoscopic images of the pyloric antrum region (without textures and with illumination changes) used to compute OF fields. These images were provided by Pr. Dominique Lamarque (Ambroise Paré Hospital, Boulogne-Billancourt, France). Figures from (c) to (j) give the OF results for each descriptor using the classical color code.

is of particular interest since dense OF remains a useful technique in many applications which need to precisely register images under complicated conditions. For instance, the mosaicing of image sequences acquired in gastroscopy [53] or in white light [54] or fluorescence [55] cystoscopy involves such complex conditions.

As visible in Figs. 10(a)-10(b) and Figs. 11-12, endoscopic images often include regions without textures, inhomogeneous scene illumination, varying lighting conditions according the viewpoint and/or specular reflections. In order to demonstrate the potential and robustness of the proposed descriptors in such medical scenes, this section gives OF results for an endoscopic image pair of the pyloric antrum region (see the source and target images in Fig. 10). Objective quality criteria such as AAE or AEE cannot be estimated since no ground truth is available for these images. Thus, to present OF results, we use the classical color code representation [9] for which the hue and the saturation correspond to the vector orientation and modulus respectively. In regions with constant colors the OF is also constant.

The homologous point displacements between the two images in Figs. 10(a)-10(b) are caused both by inner stomach movements and the endoscopic camera movements. In practice, due to the varying depth of the surface, the modulus of the flow vectors should be larger for tissue which is close to the camera (e.g. pixels in the down left image corner in Fig. 10(a)), and smaller for tissue which is farther from the camera (e.g. pixels in the image center). 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. As noticeable in Fig. 10, the OF results of Census, NND, $\mathbf{D}_1$ and $\mathbf{D}_2$ globally correspond to realistic flow fields since *i)* their OF is smooth, and *ii)* the flow field is constant onto and around the duodenum black disc (in Figs. 10(g)-10(j) the absence of the disc shows that flow field is effectively constant in this region). These results are globally less coherent for the CRT, LDP, MLDP and Corr descriptors. In particular, the light and dark blue discs in Figs. 10(c)-10(f) indicate a flow discontinuity which does not correspond to a real displacement continuity.

6.8. Computation time

This subsection evaluates the run-times requested by each descriptor to determine the descriptor vectors at all pixels in an image (descriptor time in millisecond, *ms*), as well as the total time for computing the OF using that descriptor (OF time in second, *s*). The evaluation was performed on the Middle-

bury Urban2 image pair with a size of 640 × 480 pixels and using MATLAB R2016a running under Windows 10, 64bits on a HP Desktop Core i7-4910MQ (2.90 GHz, 16 GB RAM). Descriptor vectors are completely computed with the Matlab code, while the core of the algorithm optimizing the OF (see (14)) is available as a *.mex* file (it is first written with the C++ language, and then compiled into a *.mex* code to be called in the Matlab environment). To compute the OF time, parameters σ_1, σ_2, Py_s, and λ are set based on the results in Section 6.6.1. Specifically, the three parameters σ_1, σ_2, Py_s are set to 3, 5 and 0.6, respectively. The λ parameters of descriptors $\mathbf{D}_1$, $\mathbf{D}_2$, Census, CRT, LDP, MLDP, Corr, and NND are set to 50, 15, 20, 0.8, 25, 8, 12, 100, respectively.

The descriptors are ordered according to increasing computation time:

- Descriptor time : MLDP (23.5 *ms*), Census (42 *ms*), $\mathbf{D}_1$ (44.5 *ms*), Corr (57 *ms*), $\mathbf{D}_2$ (60 *ms*), LDP (81 *ms*), NND (104 *ms*), and CRT (119 *ms*).
- OF time : $\mathbf{D}_2$ (87.22 *s*), $\mathbf{D}_1$ (87.67 *s*), Census (89.12 *s*), MLDP (91.68 *s*), Corr (92.14 *s*), NND (94.22 *s*), LDP (94.98 *s*), and CRT (96.72 *s*).

6.9. Global discussion

The experimental results have shown that when a descriptor fulfills condition (18) it is indeed illumination invariant. With the variational OF model presented in this work, all the eight descriptors can reach their best OF accuracy with the pyramid scale parameter $Py_s \geq 0.6$.

Considering globally Sections 6.2, 6.3, 6.4, 6.5 and 6.6, descriptors Census, CRT and LDP led to the less accurate OF accuracy in comparison to the other descriptors. The NND descriptor accuracy is nearly equivalent to that of the proposed descriptors ($\mathbf{D}_1$ and $\mathbf{D}_2$) on the KITTI training datasets (see Table 4), whereas its ranking was not too high on the Middlebury training set (see Table 3). Moreover, the computation time of NND is higher (104 *ms*) than that of the proposed descriptors. Descriptors MLDP, Corr, $\mathbf{D}_1$ and $\mathbf{D}_2$ can be considered as being globally the most accurate for particular illumination change conditions (Subsections 6.2 to 6.5). However, when considering the criterion relating to OF accuracy independence towards the λ parameter, the two proposed descriptors can be considered as being more robust than MLDP and Corr. Moreover, in the experiment on gastroscopic images shown in Fig. 10, the proposed descriptors gave subjectively more coherent results than Corr and MLDP.

The ability to deal both with textureless scenes and with scenes with various textures is illustrated and confirmed in the next section on different real-data scenes which are not represented in the standard benchmarks (Middlebury, KITTI 2012, KITTI 2015 and MPI-Sintel).

7. Applications to Image Mosaicing in Gastroscopy and Other Scenes

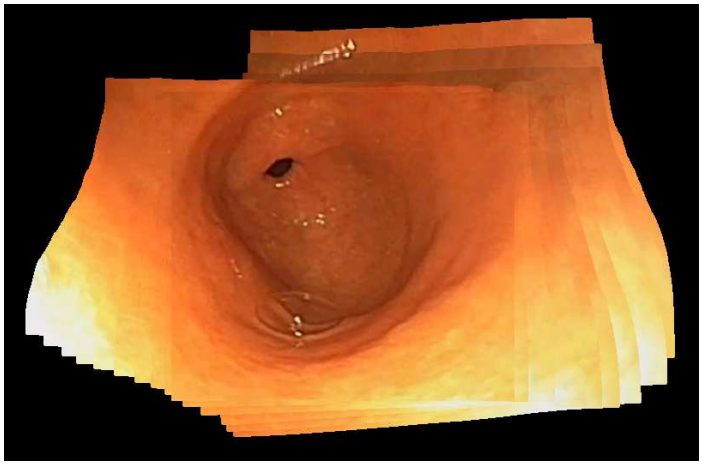
A major challenge in an image mosaicing application is to register accurately and robustly a long sequence of images with poor quality. This subsection shows the mosaicing results for two image sequences (two patients) of the pyloric antrum region (see Figs. 11-12). The video-sequences were provided by Pr. Dominique Lamarque (Ambroise Paré Hospital, Boulogne-Billancourt, France). The image registration is based on the flow field obtained with descriptor $\mathbf{D}_1$ (i.e. the IIOF-NLDP scheme) using the parameter values as in Table 1: $(\sigma_1, \sigma_2, Py_s, \lambda) = (3, 5, 0.8, 70)$.

Fig. 11(f) shows a panoramic image computed with 21 images, while Fig. 12(e) shows the mosaicing result computed for a sequence of 45 images. The precise mosaics (without structure discontinuities) of Fig. 11 and Fig. 12 confirm the potential and robustness of the proposed decriptor[2]. In these mosaicing examples, the illumination discontinuities were intentionally not corrected to show the image superimposition. However, these discontinuities can be corrected with an approach as described in [56].

Fig. 13 gives another mosaicing result for textureless images. This panorama was built with 161 images and represents

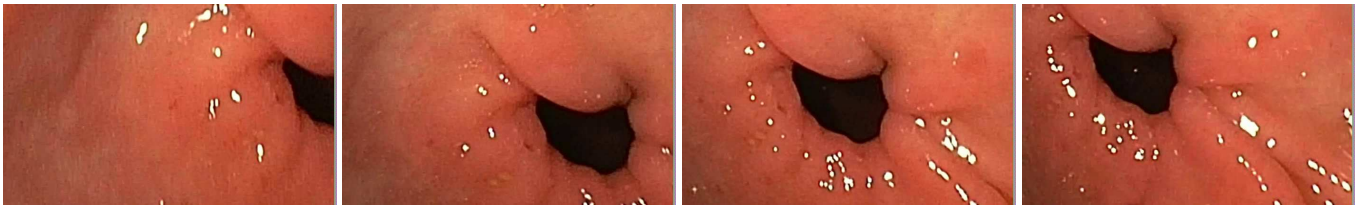
[2]Two videos illustrating the image mosaicing process from which Figs. 11 and 12 have been built are provided as supplementary material.

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



(f) Mosaiced image

Fig. 11. Mosaic built with 21 images of a gastroscopic sequence (5 images of the sequence are also given in the first row). The OF between the images was directly used to place the pixels of the 21 frames into the coordinate system of the first image which acts as mosaic reference.



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

(e) Mosaiced image

Fig. 12. Mosaic built with 45 images of a gastroscopic sequence. The first row gives 4 images of the sequence while (e) shows the mosaiced image.

the landing site of the NASA's curiosity rover on planet Mars. The field of view was significantly increased in comparison to the Mars landscape parts perceptible in a single image. While the lack of textures is the only common point of gastroscopic scenes and the red planet landscape, the mosaics in Figs. 11, 12 and 13 were all obtained with the same descriptor and OF

Fig. 13. Mosaic of a landscape of the red planet. This mosaic was built with 161 images almost weak textures.

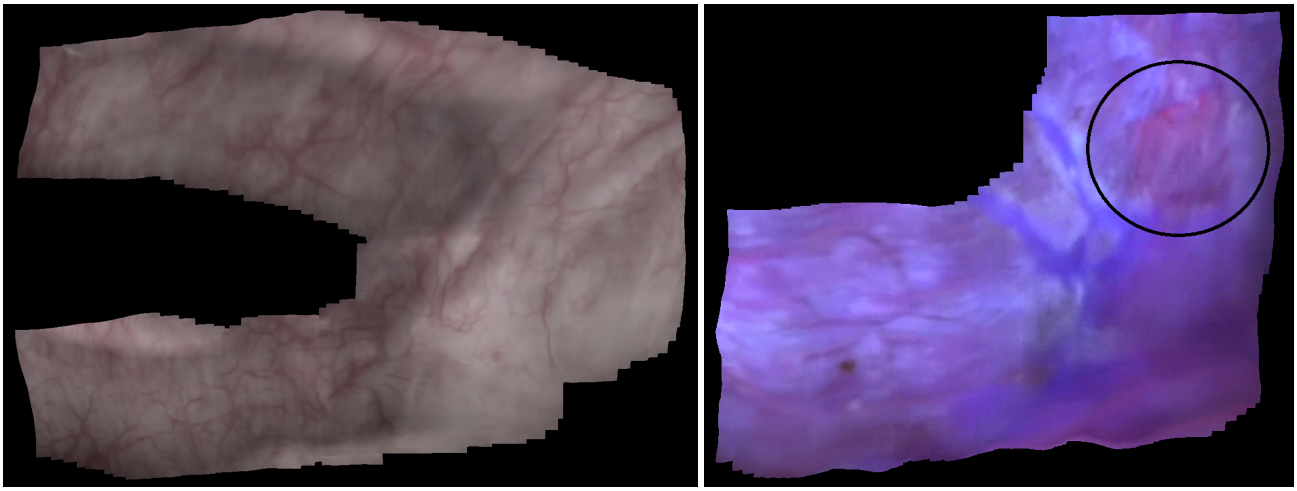
scheme parameters. This illustrates the variety of scenes with can targeted by the proposed variational OF method.

In the cystoscopic data of Fig. 14, the endoscopic bladder images include textures whose aspect and contrast vary inside and between the sequences. White light cystoscopy (see Fig. 14(a)) is the standard examination for cancerous bladder lesion diagnosis and patient follow-up. Wide field of view mosaics have several advantages in cystoscopy: the organ inspection is facilitated since the endoscope and potential lesions can be located with respect to anatomical landmarks (this is most often not the case when simply observing video-sequences), two mosaics built at a some week or month interval facilitate the lesion evolution assessment (urologists are unable to assess the evolution by comparing two video-sequences), and the mosaics can be used by different specialists as a medium of concertation (i.e., urologists, surgeons, radio-physicians and oncologists can use a mosaic as a basis of discussion). In the white light modality, although more or less contrasted textures are present, the images are affected by strong illumination changes between different viewpoints and the endoscopic displacement can be large and saccadic, as in gastroenterology. Fig. 14(a) shows that the proposed variational OF scheme can be used to significantly increase the field of view in white light cystoscopy (the dark and bright regions in this mosaic are due to illumination changes). Some cystoscopic set-ups also enable to switch between the white light modality and the fluorescence modality. In the latter (see Fig. 14(b)), the natural aspect of the bladder colours and textures is lost, but the detection of cancerous lesions located under the epithelium becomes possible. In urology, white light and fluorescence endoscopy are two complementary image modalities. The mark delineated by a circle in the mosaic of Fig. 14(b) is due to a transurethral resection of a lesion. The mosaic is used to check whether the whole tumoral tissue was removed or not during the surgical intervention. It is worth noticing that the same descriptor and OF scheme parameters (that used for the gastroscopic data and given at section beginning) were again used to build both the white light and fluorescence mosaics.

8. Conclusion

This contribution shows how illumination invariant data-terms can be constructed in the frame of OF algorithms. The illumination changes between images are modelled with affine transformations which locally link the intensities in small images regions (in this paper the affine transformation parameters are constant in a 3×3 neighborhood). These local affine transformations enable to deal with complicated (i.e. strong and/or inhomogeneous) illumination changes between images. It was also shown how it can be verified whether a descriptor used to construct a data-term is illumination invariant or not with respect to the proposed model. Until now, this invariance was most often experimentally highlighted (i.e. without mathematical proof).

The major contribution of this paper lies in the development of two general formulations of illumination invariant descriptors, one formulation being based on a descriptor sign invariance, while in the other formulation the idea is to eliminate the parameters of the affine transformation. These two formulations help to understand the appropriate way to construct illumination invariant descriptors and act as a starting point for constructing accurate and robust data-terms for complex scenes.



(a) (b)

Fig. 14. Two mosaics of the inner bladder wall epithelium. (a) Wide field of view map constructed with 200 images acquired in the white light modality. In this reference modality in urology, a clinician (urologist or surgeon) can, due to the natural colors and textures, interpret the scene content in order to (mentally) localize the endoscope with respect to anatomical landmarks. This modality allows for the diagnosis of multifocal cancerous lesions which pierce the epithelium. (b) Panoramic mosaic constructed with 95 images acquired in the fluorescence modality. This complementary modality facilitates the detection of tumoral lesions potentially located under the epithelium and which are not visible in the white light modality.

Two novel illumination invariant descriptors were then proposed in this contribution to illustrate the potential and interest of the proposed general formulations. The performances of existing and the novel descriptors were evaluated using datasets with and without strong illumination changes and with small and large displacements. One of the main practical results lies in the fact that the two proposed descriptors led to data-terms allowing to compute an OF with high and constant accuracy, even for changing illumination conditions (the strong simulated illumination changes did not significantly affect the OF accuracy). When comparing the overall results of the two proposed descriptors it can be observed that $\mathbf{D}_1$ has a little bit higher performance than $\mathbf{D}_2$. The difference between the two descriptors is that $\mathbf{D}_1$ uses six pixels to determine the vector components with Robinson kernels, while $\mathbf{D}_2$ is only based on a difference of two pixel values. $\mathbf{D}_1$ is probably able to capture more accurately weak texture structures or is less sensitive to noise (Robinson kernels compute a difference of mean values). Tests on complex endoscopic data (with strong illumination changes and almost no texture) confirmed the robustness of the proposed descriptors to illumination changes.

More generally, all descriptors can be classified in two groups. On the one hand, the group of sign invariance descriptors (LDP, Census, CRT, MLDP, see Section 4.1) gathers the vectors those components correspond to binary information obtained by thresholding iconic data based values. On the other hand, the group of descriptors based on the illumination model parameter elimination described in section 4.2 (Corr, NND, $\mathbf{D}_1$, $\mathbf{D}_2$) consists of descriptor vectors with real value components. The components of the vectors of both groups are all related to a description of the texture overlapped by a patch. However, the thresholding of real values in the binary descriptors generally leads to a loss of valuable information, especially when the textures are weakly contrasted. For this reason, the real-value descriptor performance is, in average, higher for images with few or poorly contrasted textures. This observation must be taken into account when developing a descriptor.

The proposed method is quite appropriate for scenes without textures or with very weakly contrasted textures. However, this method is not limited to such scenes since it can also be applied to textured scenes. For textured scenes, the proposed OF method can be improved in terms of computation speed and accuracy by combining it with matching-based approaches.

As a perspective for applications in endoscopy, the results

of this paper will be used to optimize the descriptor design for the registration of endoscopic images with the aim to build wide-field of view 2D mosaics of the internal bladder wall (cystoscopy) and the internal stomach epithelium (gastroscopy). The OF will be used to determine the non-linear geometrical transformations between homologous image points with the aim to superimpose the images of the video-sequences.

A natural and further extension of this work is to build 3D mosaics (internal bladder or stomach wall surfaces superimposed by the image textures/colors). To do so, the homologous image points of image pairs will be associated to structure from motion (SfM) or simultaneous localization and mapping (SLAM) techniques. The feasibility of such approaches was shown in [57, 58] for the bladder. However, these approaches use feature extraction methods (as SIFT) in the point correspondence step. Such correspondences can often not be robustly determined due to a lack of pronounced textures in the cystoscopic and gastroscopic images [59]. The proposed OF method not only determines homologous points in a more robust way, even in the stomach images, but provides also a more dense correspondence field as that given by a SIFT approach (for few and weakly contrasted textures a SIFT method provides few matches). It is noticeable that SfM or SLAM techniques are only based on images. This avoids the use of active vision methods [60] which require the modification of standard endoscopes [61] or the use of non conventional devices [57].

Acknowledgments

This work is funded by the Agence nationale de la recherche (ANR) in the framework of the EMMIE (Endoscopie Multi-Modale pour les lésions Inflammatoires de l'Estomac) project (project number: ANR-15-CE17-0015).

References

[1] A. Dosovitskiy, P. Fischer, E. Ilg, P. Hausser, C. Haz?rbas, V. Golkov, Flownet: Learning optical flow with convolutional networks, in: ICCV, 2015, pp. 2758–2766.

[2] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, T. Brox, Flownet2.0: Evolution of optical flow estimation with deep networks, in: CVPR, 2017, pp. 2462–2470.

[3] D. Sun, X. Yang, M.-Y. Liu, J. Kautz, Pwc-net: Cnns for optical flow using pyramid, warping, and cost volume, in: CVPR, 2018.

[4] T.-W. Hui, X. Tang, C. C. Loy, Liteflownet: A lightweight convolutional neural network for optical flow estimation, in: CVPR, 2018.

[5] P. Weinzaepfel, J. Revaud, Z. Harchaoui, C. Schmid, Deepflow: Large displacement optical flow with deep matching, in: Proceedings of IEEE International Conference on Computer Vision (ICCV), Sydney, Australia, 2013, pp. 1385–1392.

[6] C. Bailer, B. Taetz, D. Stricker, Flow fields: Dense correspondence fields for highly accurate large displacement optical flow estimation, in: ICCV, 2015, pp. 4015–4023.

[7] J. Revaud, P. Weinzaepfel, Z. Harchaoui, C. Schmid, Epicflow: Edge-preserving interpolation of correspondences for optical flow, in: Proceedings of IEEE conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 1164–1172.

[8] Y. Hu, R. Song, Y. Li, Efficient coarse-to-fine patchmatch for large displacement optical flow, in: Proceedings of IEEE conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, IEEE, 2016.

[9] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, R. Szeliski, A database and evaluation methodology for optical flow, International Journal of Computer Vision 92 (1) (2011) 1–31.

[10] D. J. Butler, J. Wulff, G. B. Stanley, M. J. Black, A naturalistic open source movie for optical flow evaluation, in: A. Fitzgibbon et al. (Eds.) (Ed.), Proceedings of European Conference on Computer Vision (ECCV), Part IV, LNCS 7577, Springer-Verlag, 2012, pp. 611–625.

[11] A. Geiger, P. Lenz, R. Urtasun, Are we ready for autonomous driving? the KITTI vision benchmark suite, in: Proceedings of IEEE conference on Computer Vision and Pattern Recognition (CVPR), 2012.

[12] M. Menze, A. Geiger, Object scene flow for autonomous vehicles, in: Proceedings of IEEE conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015.

[13] L. Xu, J. Jia, Y. Matsushita, Motion detail preserving optical flow estimation, IEEE Transactions on Pattern Analysis and Machine Intelligence 34 (9) (2012) 1744–1757.

[14] M. Drulea, S. Nedevschi, Motion estimation using the correlation transform, IEEE Transactions on Image Processing 22 (8) (2013) 3260–3270.

[15] M. A. Mohamed, H. A. Rashwan, B. Mertsching, M. A. García, D. Puig, Illumination-robust optical flow using a local directional pattern, IEEE Transactions on Circuits and Systems for Video Technology 24 (9) (2014) 1499–1508.

[16] S. Ali, C. Daul, E. Galbrun, W. Blondel, Illumination invariant optical flow using neighborhood descriptors, Computer Vision and Image Understanding 145 (2016) 95–110.

[17] B. K. P. Horn, B. G. Schunck, Determining optical flow, Artificial Intelligence 17 (1-3) (1981) 185–203.

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

[19] D. Fortun, P. Bouthemy, C. Kervrann, Optical flow modeling and computation: A survey, Computer Vision and Image Understanding 134 (2015) 1–21.

[20] H. Nagel, W. Enkelmann, An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences, IEEE Transactions on Pattern Analysis and Machine Intelligence 8 (5) (1986) 565–593.

[21] C. Zach, T. Pock, H. Bischof, A duality based approach for realtime TV-L^1 optical flow, in: Pattern Recognition, 29th DAGM Symposium, Heidelberg, Germany, 2007, pp. 214–223.

[22] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, H. Bischof, Anisotropic Huber-L1 optical flow, in: Proceedings of British Machine Vision Conference (BMVC), London, UK, 2009, pp. 1–11.

[23] T. Brox, A. Bruhn, N. Papenberg, J. Weickert, High accuracy optical flow estimation based on a theory for warping, in: Proceedings of European Conference on Computer Vision (ECCV), Prague, Czech Republic,, 2004, pp. 25–36.

[24] J. Yang, H. Li, Dense, accurate optical flow estimation with piecewise parametric model, in: Proceedings of IEEE conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, USA, 2015, pp. 1019–1027.

[25] C. Zhang, Z. Chen, M. Wang, M. Li, S. Jiang, Robust non-local TV- L1 optical flow estimation with occlusion detection, IEEE Trans. on Image Processing 26 (2017) 4055 – 4067.

[26] D. Hafner, O. Demetz, J. Weickert, Why is the census transform good for robust optic flow computation?, in: The 4th International Conference on Scale Space and Variational Methods in Computer Vision, Leibnitz, Austria, 2013, pp. 210–221.

[27] A. Kumar, F. Tung, A. Wong, D. A. Clausi, A decoupled approach to illumination-robust optical flow estimation, IEEE Transactions on Image Processing 22 (10) (2013) 4136–4147.

[28] N. Papenberg, A. Bruhn, T. Brox, S. Didas, J. Weickert, Highly accurate optic flow computation with theoretically justified warping, International Journal of Computer Vision 67 (2) (2006) 141–158.

[29] A. Wedel, T. Brox, C. Zach, H. Bischof, D. Cremers, An improved algorithm for TV-L1 optical flow, in: Dagstuhl Motion Workshop, 2008.

[30] P. Golland, A. M. Bruckstein, Montion from color, Computer Vision and Image Understanding 68 (3) (1997) 346–362.

[31] J. V. de Weijer, T. Gevers, Robust optical flow from photometric invariants, in: Proceedings of IEEE International Conference on Image Processing (ICIP), Vol. 3, 2004, pp. 1835–1838.

[32] Y. Mileva, A. Bruhn, J. Weickert, Illumination-robust variational optical flow with photometric invariants, in: Pattern Recognition, 29th DAGM Symposium, Heidelberg, Germany,, 2007, pp. 152–162.

[33] T. E. Zickler, S. P. Mallick, D. J. Kriegman, P. N. Belhumeur, Color subspaces as photometric invariants, International Journal of Computer Vision 79 (1) (2008) 13–30.

[34] H. Zimmer, A. Bruhn, J. Weickert, Optical flow in harmony, International Journal of Computer Vision 93 (2011) 368–388.

[35] S. Park, N. Kwak, Illumination robust optical flow estimation by illumination-chromaticity decoupling, in: Proceedings of IEEE International Conference on Image Processing (ICIP), Australia, Quebec City, Canada, 2015, pp. 1910–1914.

[36] H. W. Haussecker, D. J. Fleet, Computing optical flow with physical model of brightness variation, IEEE Transactions on Pattern Analysis and Machine Intelligence 23 (6) (2001) 661–673.

[37] S. Negahdaripour, Revised definition of optical flow: Integration of radiometric and geometric cues for dynamic scene analysis, IEEE Transactions on Pattern Analysis and Machine Intelligence 20 (9) (1998) 961–979.

[38] Y. Kim, A. M. Martínez, A. C. Kak, Robust motion estimation under varying illumination, Image and Vision Computing 23 (4) (2005) 365–375.

[39] R. Zabih, J. Woodfill, Non-parametric local transforms for computing visual correspondence, in: Proceedings of European Conference on Computer Vision (ECCV), Stockholm, Sweden, 1994, pp. 151–158.

[40] F. Stein, Efficient computation of optical flow using the census transform, in: Pattern Recognition, 26th DAGM Symposium, Tübingen, Germany, 2004, pp. 79–86.

[41] T. Müller, C. Rabe, J. Rannacher, U. Franke, R. Mester, Illumination-robust dense optical flow using census signatures, in: Pattern Recognition - 33rd DAGM Symposium, Frankfurt/Main, Germany, 2011, pp. 236–245.

[42] R. Ranftl, S. Gehrig, T. Pock, H. Bischof, Pushing the limits of stereo using variational stereo estimation, in: IEEE Intell. Vehicles Symposium, Alcal de Henares, Madrid, Spain, 2012, pp. 401–407.

[43] O. Demetz, D. Hafner, J. Weickert, The complete rank transform: A tool for accurate and morphologically invariant matching of structures, in: Proceedings of British Machine Vision Conference (BMVC), Bristol, UK, 2013.

[44] J. Molnár, D. Chetverikov, S. Fazekas, Illumination-robust variational optical flow using cross-correlation, Computer Vision and Image Understanding 114 (10) (2010) 1104–1114.

[45] M. Werlberger, T. Pock, H. Bischof, Motion estimation with non-local total variation regularization, in: Proceedings of IEEE conf. on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 2010, pp. 2464–2471.

[46] D. Fortun, P. Bouthemy, P. Paul-Gilloteaux, C. Kervrann, Aggregation of patch-based estimations for illumination-invariant optical flow in live cell imaging, in: IEEE 10th International Symposium on Biomedical Imaging, San Francisco, CA, USA, 2013, pp. 660–663.

[47] D.-H. Trinh, W. Blondel, D. Lamarque, C. Daul, Illumination-invariant optical flow: Application to endoscopic image mosaicing, in: XXVIe Colloque GRETSI Traitement du Signal et des Images, GRETSI 2017,

Juan-Les-Pins, France, 2017.

[48] D.-H. Trinh, W. Blondel, C. Daul, A general form of illumination-invariant descriptors in variational optical flow estimation, in: Proceedings of IEEE International Conference on Image Processing (ICIP), Beijing, China, 2017.

[49] D. Sun, S. Roth, M. J. Black, Secrets of optical flow estimation and their principles, in: Proceedings of IEEE conf. on Computer Vision and Pattern Recognition (CVPR), San Francisco, CA, USA, 2010, pp. 2432–2439.

[50] P. Krähenbühl, V. Koltun, Efficient nonlocal regularization for optical flow, in: Proceedings of European Conference on Computer Vision (ECCV), Florence, Italy, 2012, pp. 356–369.

[51] A. Chambolle, T. Pock, A first-order primal-dual algorithm for convex problems with applications to imaging, Journal of Mathematical Imaging and Vision 40 (1) (2011) 120–145.

[52] M. H. Kabir, T. Jabid, O. Chae, A local directional pattern variance (LDPv) based face descriptor for human facial expression recognition, in: Proceedings of IEEE International Conference on Advanced Video and Signal Based Surveillance, 2010, pp. 526–532.

[53] S. Ali, C. Daul, E. Galbrun, F. Guillemin, W. Blondel, Anisotropic motion estimation on edge preserving riesz wavelets for robust video mosaicing, Pattern Recognition 51 (2016) 425–442.

[54] R. Miranda-Luna, C. Daul, W. Blondel, Y. Hernandez-Mier, D. Wolf, F. Guillemin, Mosaicing of bladder endoscopic image sequences: Distortion calibration and registration algorithm, IEEE Transactions on Biomedical Engineering 55 (2) (2008) 541–553.

[55] Y. Hernández-Mier, W. Blondel, C. Daul, D. Wolf, G. Bourg-Heckly, 2D panoramas from cystoscopic image sequences and potential application to fluorescence imaging, in: 6th IFAC Symposium on Modelling and Control in Biomedical Systems, Reims, France, 2006, pp. 291–296.

[56] T. Weibel, C. Daul, D. Wolf, R. Rösch, Contrast-enhancing seam detection and blending using graph cuts, in: Proceedings of the 21st International Conf. on Pattern Recognition (ICPR), Japan, 2012, pp. 2732–2735.

[57] T. D. Soper, M. P. Porter, E. J. Seibel, Surface mosaics of the bladder reconstructed from endoscopic video for automated surveillance, IEEE Transactions on Biomedical Engineering 59 (6) (2012) 1670–1680.

[58] K. L. Lurie, R. Angst, D. V. Zlatev, J. C. Liao, A. K. E. Bowden, 3D reconstruction of cystoscopy videos for comprehensive bladder records, Biomedical Optics Express 8 (4) (2017) 2106–2123.

[59] S. Ali, C. Daul, T. Weibel, W. Blondel, Fast mosaicing of cystoscopic images from dense correspondence: Combined SURF and TV-L1 optical flow method, in: Proceedings of IEEE International Conference on Image Processing (ICIP), Australia, 2013, pp. 1291–1295.

[60] A. Ben-Hamadou, C. Soussen, C. Daul, W. Blondel, D. Wolf, Flexible calibration of structured-light systems projecting point patterns, Computer Vision and Image Understanding 117 (10) (2013) 1468–1481.

[61] A. Ben-Hamadou, C. Daul, C. Soussen, Construction of extended 3D field of views of the internal bladder wall surface: a proof of concept, 3D Research 7 (3) (2016) 95:1–95:23.