

Motion estimation using the correlation transform

Marius Drulea, Sergiu Nedevschi, *Member, IEEE*

Abstract—The zero-mean normalized cross-correlation was shown to improve the accuracy of optical flow, but its analytical form is quite complicated for the variational framework. This work addresses this issue and presents a new direct approach to this matching measure. Our approach uses the correlation transform to define very discriminative descriptors that are pre-computed and that have to be matched in the target frame. It is equivalent to the computation of the optical flow for the correlation transforms of the images. The smoothness energy is non-local and uses a robust penalty in order to preserve motion discontinuities. The model is associated with a fast and parallelizable minimization procedure based on the projected-proximal point algorithm. The experiments confirm the strength of this model and implicitly demonstrate the correctness of our solution. The results demonstrate that the involved data term is very robust with respect to changes in illumination, especially where large illumination exists.

Index Terms—correlation transform, correlation flow, correlation-based descriptors, non-local flow propagation, parallelizable numerical scheme, changes in illumination

I. INTRODUCTION

DENSE optical flow estimation gained a lot of interest in the last years due to its importance in environment perception. The increasing power of computation units facilitated the development of increasingly accurate models. Work on dense optical flow spans over three decades. In [1], Horn and Schunck (HS) have put the basis of the variational framework for motion estimation. The variational problems consist of two data terms: one is the data matching or data fidelity and the other is the smoothness or regularization term. The data fidelity term measures the degree of similarity between two pixels (or regions) in terms of intensity. It is effective at objects' boundaries because they present important features and the best matching is unique here or there are few possible matches. For large homogeneous areas, there are multiple possible matches and the true flow here inherits the flow from boundaries. The variational methods are integrated into coarse-to-fine warping strategies in order to detect large motion because they rely on the Taylor expansion of the data matching term, which is valid only for small displacements.

Manuscript received May 29, 2012, accepted May 02, 2013.

Copyright (c) 2013 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

This paper was supported in part by the Romanian Agency of Scientific Research under Contract PN-II-ID-PCE-2011-3-1086.

M. Drulea and S. Nedevschi are with the Computer Science Department, Technical University of Cluj-Napoca, Memorandumului Str. 28, 400114 Cluj-Napoca (e-mail: {marius.drulea, sergiu.nedevschi}@cs.utcluj.ro).

A lot of effort went into improving the original variational framework. One of the highly investigated directions regards the norms used for penalizing the errors. The use of robust penalty functions [2-17] defines a non-smooth optimization problem, enabling non-smooth solutions and therefore preserving the motion boundaries, in contrast to the HS model which is smooth and allows only smooth transitions in the solution. To overcome the non-differentiability of the robust L1 norm, the square root approximation $|x| \approx \sqrt{x^2 + \varepsilon}$ is used in [2-5, 10, 14, 15] and most often in combination with a successive-over-relaxation scheme, sacrificing the parallelism. The TV-L1 model [6] does not use this approximation and incorporates the parallelizable convex dual algorithm presented in [18].

Nagel and Enkelmann [19] have investigated the smoothness constraint of the HS model and have proposed an anisotropic propagation, decreasing the importance of the smoothing in regions with high gradient and favoring the propagation inside regions and not across boundaries. Recently, Sun [13] and Werlberger [7] have proposed a patch based smoothing, known as non-local regularization, by extending the traditional derivatives in constraints. Such a constraint enhances the propagation of the flow.

The matching criteria used for the data term in dense optical flow estimation do not vary too much. The simple point-to-point equality of the intensities [1, 6, 8, 13, 19, 20] supported or not by the simple point-to-point equality of gradients [2, 3, 12, 14, 15] is used in many methods. More criteria might be desirable because we want to increase the accuracy and the robustness. A common problem that has to be solved is the change in illumination, which occurs quite frequently in real-world applications. One option is to preprocess the images to remove the illumination [8, 13, 20]. We can avoid this step if we can use a matching function that is invariant to this type of perturbations.

One of the first attempts to improve the data matching term is in [4], where the local method Lucas-Kanade (LK) is integrated into the HS model. The combined model is equivalent to using a SSD intensity matching and it was shown to increase the robustness against noise [4, 21]. In terms of accuracy, however, the performance of this model is limited [21, 22]. In [9], the local window of the combined model is weighted by means of bilateral filtering, in order to reduce the over-smoothing of boundaries.

The literature provides sophisticated matching techniques. The SIFT detector [23] computes the scale of each feature in a first step and then it uses a histogram of oriented gradients to build strong descriptors for the involved features. SURF [24]

uses sums of wavelet responses to characterize the features. The Census transform was proven to be a very good descriptor in the context of stereo-vision. In template matching several measures applies: SSD (sum of squared differences), SAD (sum of absolute differences), ZNCC (zero-mean normalized cross-correlation). To summarize, there are two main steps in establishing the similarity of two pixels/regions/features: the first one is the building of the descriptors and the second one is the computation of the “distance” between two descriptors. For variational optical flow there are some restrictions however. For any matching criteria the overall energy to be minimized is non-convex. To find the solution quickly we should approximate the energy such that it is convex. This usually implies the use of a Taylor linearization and it also requires the use of convex “distance” measures. Due to these limitations, the integration of the above matching functions into the variational framework is not a trivial task. In this paper we solve this problem for the ZNCC data term.

The zero-mean normalized cross-correlation is a strong matching function because it is very discriminative, providing accuracy. It also offers robustness with respect to illumination changes. Later in this paper we explain why ZNCC is more discriminative than patch intensity matching. Due to these properties, ZNCC is a good candidate in enhancing the data matching term. Unfortunately, ZNCC has a more complicated form and the direct use of the classical Taylor linearization for the input images does not solve the difficulties: from a highly non-linear system we would reach another highly non-linear system of equations.

This paper contributes with a fast and reliable optical flow system that uses zero-mean normalized cross-correlation as a matching measure. We have observed that ZNCC of two signals is in fact the SSD distance between the correlation transforms of the given signals. Using this interesting property, we transform the ZNCC problem into the problem of flow estimation between the correlation transforms of the images. After this transformation we were able to follow the traditional linearization framework. Our approach shows accurate results on the online Middlebury evaluation website [25].

The paper is organized as follows. Section 2 presents the related work. Section 3 introduces the basic notations, the mathematical model and our solution to this model. Section 4 presents the derivation of the corresponding numerical scheme and Section 5 discusses related implementation issues. Section 6 presents evaluations with regard to accuracy and robustness with respect to changes in illumination. Finally, Section 7 concludes the paper.

II. RELATED WORK

Four works share similarities with the proposed approach. All these works use ZNCC as a matching measure.

Molnar et al. [26] simplifies the mathematical analysis by assuming that the standard deviation of the warped image remains constant over the iterative process. However, such an assumption is not very realistic, modifies the objective and can lead to inaccurate results. We do not make this assumption and we solve the exact model.

In [27], some variants for data fidelity matching are presented, including zero-mean normalized cross-correlation. The solutions are computed using a search mechanism guided by the local flow propagation. The search approach is very slow. This model was designed to compare more complicated (but useful) matching functions, regardless of the running time. Although inaccurate and not reliable, ZNCC has provided interesting results on some situations. In contrast to this work, our approach provides accurate results and we have found a fast, parallelizable optimization scheme.

Fazekas et al. [28] introduce a method able to handle any data term. They define the matching term as a function of two variables and analyze how it varies if the flow (the second variable) is changed locally. The numerical scheme is not completely stable and they use a “step tolerance” to avoid instabilities [29]. For some pixels, especially in boundaries, the flow has a high value and is completely wrong. To avoid these anomalies they clip the flow vectors that fall outside the step tolerance window. Without this trick the algorithm produces a lot of errors because the errors from the boundaries are propagated in all directions. This approach is also slow due to the large number of iterations required to converge. The method was shown to be quite robust to changes in illumination, but not to provide very accurate results.

Werlberger et al. [7] use the same idea as in [28] to solve the truncated ZNCC problem. They further apply a second-order approximation of the data term in order to meet convexity conditions. This approach also needs the step tolerance window to solve the instabilities. This method is fast and provides very accurate results on the Middlebury evaluations. As seen in the online Middlebury evaluation website [25], our results are even better.

In contrast to the above two works [7, 28], we directly solve the ZNCC problem. Instead of analyzing the explicit form of ZNCC we exploit the correlation transform of each image separately. Therefore, we are able to solve the optical flow system using the classical first-order framework. Our approach is numerically stable for all values of the involved parameters.

III. THE PROPOSED APPROACH

In the followings we present the zero-mean normalized cross-correlation matching measure, define an optical flow model that uses ZNCC and we derive a solution for this model.

A. Definitions and notations

Let f and T be two patches having the same dimension, as the ones in Fig. 1. A common metric for measuring the similarity of f and T is SSD:

$$SSD(f, T) := \sum_{s \in \mathcal{N}} (f_s - T_s)^2, \quad (1)$$

where \mathcal{N} contains the indexes s of each location of the patch.

The zero-mean normalized cross correlation ZNCC is defined as:

$$ZNCC(f, T) = \frac{1}{|\mathcal{N}|} \frac{\langle f - \mu_f, T - \mu_T \rangle}{\sigma_f \cdot \sigma_T}, \quad (2)$$

where $|\mathcal{N}|$ represents the size of the patch, μ_x and σ_x are the mean and standard deviation of $X \in \{f, T\}$ and $\langle \cdot, \cdot \rangle$ represents the standard dot product. If f and T are the same, the value of this measure is 1. The best match minimizes the expression $\{1 - \text{ZNCC}(f, T)\}$.

B. ZNCC data energy for optical flow

Let $I_1, I_2 : \Omega(\subseteq \mathbb{N}^2) \rightarrow \mathbb{R}$ be two consecutive images of a sequence and $w = (u, v) : \Omega \rightarrow \mathbb{R}^2$ the unknown flow between them. $\Omega = \{(x, y) | 1 \leq x \leq N, 1 \leq y \leq M\}$ is the standard two dimensional Cartesian grid and (N, M) represents the size of the images. Each pixel has a two-dimensional index $i \in \Omega$ and a neighborhood denoted as \mathcal{N}_i , containing the indices around i . The images can take real data as arguments, but the values at these locations are computed using interpolation. We denote with w_i or $w(i)$ the flow at location i , and the same notations are used for all the other variables.

We assume that at a given moment an initial flow estimate $w_0 = (u_0, v_0)$ is available, and we denote $\tilde{I}_2(i) = I_2(i + w_0(i))$ the warped image. We have to further refine it with the displacement $dw = w - w_0$ in order to match the fixed image I_1 in the sense of ZNCC distance between blocks:

$$E_d(w) := \sum_{i \in \Omega} \left\{ 1 - \frac{1}{|\mathcal{N}_i|} \sum_{s \in \mathcal{N}_i} \frac{(\tilde{I}_2(s + dw_i) - \tilde{\mu}_2(i + dw_i))(I_1(s) - \mu_1(i))}{\tilde{\sigma}_2(i + dw_i) \cdot \sigma_1(i)} \right\} \quad (3)$$

In the above formula $\mu_1(i)$ and $\sigma_1(i)$ are the mean and the standard deviation of the patch \mathcal{N}_i around location i in image I_1 . The definitions of $\tilde{\mu}_2$ and $\tilde{\sigma}_2$ are similar. $|\mathcal{N}_i|$ represents the size of the patch i (the sizes are equal). Each term of the global sum represents the ZNCC error between the patch of the refined image $\tilde{I}_2(i + dw_i)$ and the patch of $I_1(i)$ at the given location. We have assumed that the local displacement dw_i is constant over the patch around location i .

The above energy is quite complicated for the traditional variational approaches because we have to linearize it and have to reach a convex function. Our idea is to express this energy as the squared distance between the correlation transforms of the images, as described below.

C. ZNCC as SSD

In order to solve (3), we use an interesting relationship between ZNCC, SSD and the correlation transform. With the notations used in section A, the correlation transform of a patch f is given by:

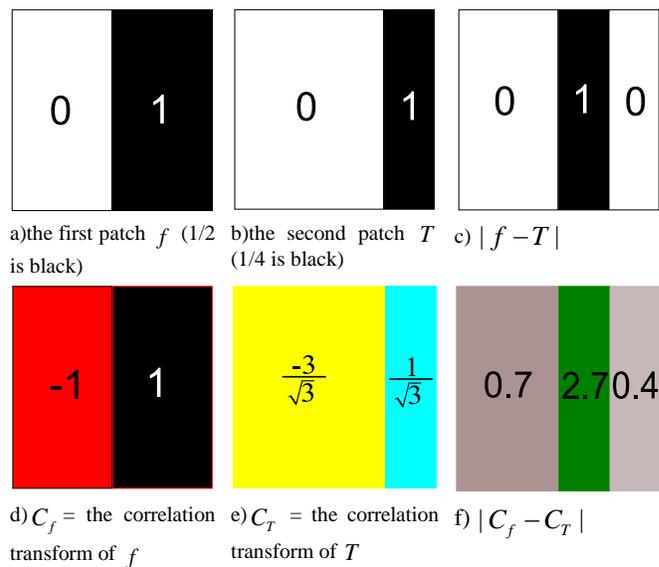


Fig. 1. The first row shows two patches denoted as f and T . The difference between f and T , used by SSD and SAD, appears only in the black part of c). The second row shows the correlation transforms of the patches and the point-to-point distance f) between these transforms in the last column. Now, the differences appear all over the patch. Therefore, ZNCC strongly penalizes deviations from the correct match.

$$\left(\frac{f_s - \mu_f}{\sigma_f} \right), s \in \mathcal{N} \quad (4)$$

We have to compute the mean and the standard deviation of the signal f and to replace each f_s with the expression (4). It is easy to check that the mean of the new patch is zero and its standard deviation is one. Therefore, the correlation transform maps each patch into a patch with a normal distribution $(N(0,1))$.

$$\left\{ \left(\frac{f_s - \mu_f}{\sigma_f} \right), s \in \mathcal{N} \right\} \approx N(0,1) \quad (5)$$

Simple computations show that *the correlation measure is in fact the mean SSD between the correlation transforms*:

$$\frac{1}{|\mathcal{N}|} \sum_{s \in \mathcal{N}} \left(\frac{f_s - \mu_f}{\sigma_f} - \frac{T_s - \mu_T}{\sigma_T} \right)^2 = 2(1 - \text{ZNCC}(f, T)) \quad (6)$$

Fig.1 shows why ZNCC is more discriminative than SSD or SAD. We consider two patches f and T , shown in Fig. 1. a) and b). SSD and SAD use the point-to-point difference between f and T , which is shown in Fig. 1. c). In this case, the differences appear only in the center, where the intensities do not match. In the second row, d) and e) represent the correlation transforms of f and T , denoted as C_f and C_T . The mean and the standard deviation of the patches are different and therefore C_f and C_T are different in every point. Fig. 1. f) shows the point-to-point distance between C_f and C_T . The mismatches appear all over the patch, not only in

the center. In conclusion, ZNCC penalizes more severe the deviations from the correct match, in comparison with SSD or SAD.

D. The correlation's optical flow

The relation (6) is useful in defining a special descriptor. For each point i of an image I it consists of the values of the correlation transform of the surrounding patch:

$$C(i) := \left(\frac{I(s) - \mu(i)}{\sigma(i)} \right), s \in \mathcal{N}_i \quad (7)$$

Two examples are given in the bottom of Fig 2. Given the image I_1 and the warped second image \tilde{I}_2 , instead of matching the intensities we have to match their descriptors C_1, \tilde{C}_2 , point-to-point. For each i the refinement flow $dw = w - w_0$ should satisfy the vector equality:

$$\tilde{C}_2(i + dw(i)) = C_1(i) \quad (8)$$

Based on this equality, we can define the total matching error (or data error) by summing the distances between descriptors:

$$E_d = \sum_{i \in \Omega} \frac{1}{|\mathcal{N}_i|} \|\tilde{C}_2(i + dw_i) - C_1(i)\|^2 \quad (9)$$

$$E_d = \sum_{i \in \Omega} \left\{ \frac{1}{|\mathcal{N}_i|} \sum_k (\tilde{C}_2(i + dw_i, k) - C_1(i, k))^2 \right\} \quad (10)$$

where k indexes the components of the descriptor. The above matching error is the same as the ZNCC error (3), as deduced from the relation (6). The involved descriptors are pre-computed after each warp and we can ignore their analytical form. In this case, we can view the correlation transforms as two multichannel images and we have to calculate the flow of these channels: the correlations' flow. The first channel $C_1(\Omega, 1)$ ($k = 1$) for "Teddy" appears in Fig. 2. It collects the first value of the descriptor for each point. It has the size of the original image, only a part of it is depicted on Fig.2. The data matching energy (10) is non-convex. To meet the convexity conditions it is necessary to apply a first order expansion, as in the traditional optical flow systems.

$$\tilde{C}_2(i + dw_i, k) - C_1(i, k) = 0 \quad (11)$$

The above equality takes place for each position i and each index k of the descriptor. The expansion takes place for each component of the vector.

$$\tilde{C}_2(i, k) + \nabla \tilde{C}_2(i, k) \cdot dw_i - C_1(i, k) = 0 \quad (12)$$

The term $C_i(i, k) := \tilde{C}_2(i, k) - C_1(i, k)$ represents the temporal gradient. A common practice in implementations is to blend the derivatives [20]: $\nabla C(i, k) := (\nabla \tilde{C}_2(i, k) + \nabla C_1(i, k)) / 2$ is used instead of $\nabla \tilde{C}_2(i, k)$. Other settings for this blending are also possible. With these notations the equality now reads:

$$C_i(i, k) + \nabla C(i, k) \cdot dw_i = 0 \quad (13)$$

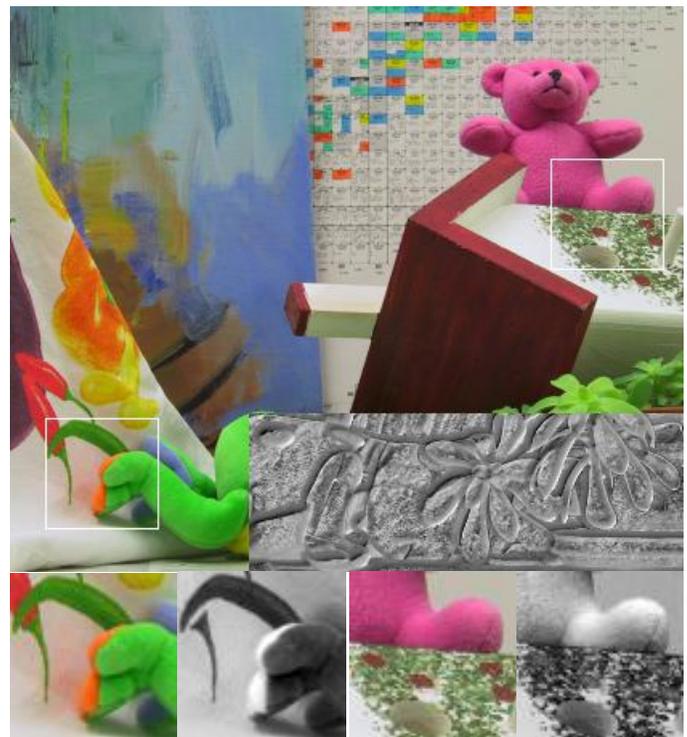


Fig. 2. Visualization of the correlation transform. Top: window positions in the original image. Bottom: close-up view of the 71x71 patches and their correlation transforms. Top image, right bottom: a part of the first channel of the 3x3 correlation transform. The output of the correlation transform was remapped from the original range into [0 1].

That last equation expresses the residual at location i and index k . After the linearization the total matching energy error is a convex function:

$$E_d(w) = \sum_{i \in \Omega} \sum_k (C_i(i, k) + \nabla C(i, k) \cdot (w_i - w_{0,i}))^2 \quad (14)$$

The second part of a variational optical flow system is the regularization. The regularization term is responsible for the propagation of the optical flow from boundaries, where the descriptors are effective, to homogenous regions. This propagation relies on the spatial coherence of natural images. The pixels belonging to the same object have almost the same flow; their flow is almost constant, at least on sub-regions. The deviations from this constancy assumption are penalized by the following smoothness error:

$$\begin{aligned} E_{smooth}(i) &= \sum_{s \in \mathcal{N}_i} bf_{i,s} \|w_s - w_i\|_1 \\ &= \sum_{s \in \mathcal{N}_i} \{bf_{i,s} |u_s - u_i| + bf_{i,s} |v_s - v_i|\} \end{aligned} \quad (15)$$

Here $bf_{i,s}$ measures how likely the pixels i and s belong to the same object. If s is in the neighborhood of i and if their colors are the same then probably i and s belong to the same object and $bf_{i,s}$ is close to 1. Otherwise this measure is close to zero, ignoring that pixel.

The last error (15) is zero in the ideal case and says the following: if s belongs to the same surface as i , then their

flow should be the same ($w_s - w_i = 0$). The total smoothness error is:

$$E_s(w) = \sum_{i \in \Omega} \sum_{s \in \mathcal{N}_i} bf_{i,s} \|w_s - w_i\|_1 \quad (16)$$

The bilateral filter [30, 31] formalizes the similarity between two locations. It was used in image de-noising, stereo-vision applications [32], and recently, in the context of optical flow [7, 9, 13, 33].

$$bf_{i,s} = e^{-\left(\frac{\Delta_c(i,s)}{2\sigma_c^2} + \frac{\Delta_d(i,s)}{2\sigma_d^2}\right)}, \quad (17)$$

where $\Delta_c(i,s)$ represents the distance between the colors of i and s computed in the Lab color space, $\Delta_d(i,s) = \|s - i\|^2$ is the distance between the positions. The parameters σ_c and σ_d control the similarity measure. The bilateral filter is the product of two types of Gaussian kernels: a color (or intensity) kernel and a distance kernel. An illustrative example is provided in Fig. 4. The weight of a pixel decreases as the distance from the center increases. In addition, the weight decreases if the intensity of the pixel is different from the center's intensity.

The discrete smoothing energy error defined in (15) is known in the literature as non-local regularization [34, 35]. The latter was inspired by the neighborhood filters [31, 36, 37], which were used in the context of image denoising and deblurring. Recently, the idea of non-local regularization was also used for optical flow in [7, 13]. It uses the L_1 robust penalty function in order to allow sharp solutions and to preserve the motion's discontinuities.

In order to recover the optical flow we have to minimize both the matching error (E_d) and the smoothness error (E_s), and this is usually expressed in the literature as a combined energy:

$$E(w) = \lambda \cdot E_d(w) + E_s(w), \quad (18)$$

where λ controls the amount of regularization.

The next section covers the minimization procedure of the resulted energy.

IV. THE OPTIMIZATION PROCEDURE

The functional E_s in (16) is non-smooth, but it is a convex function. Let us now consider a higher dimensional space $\mathbb{R}^{2 \times |\Omega| \times |\mathcal{N}|}$, where $|\mathcal{N}|$ is the size of the neighborhoods \mathcal{N}_i , and a linear operator $K : \mathbb{R}^{2 \times |\Omega|} \rightarrow \mathbb{R}^{2 \times |\Omega| \times |\mathcal{N}|}$

$$Kw := \begin{pmatrix} w_s - w_{(1,1)} & \cdots & w_s - w_{(1,1)} \\ \vdots & \ddots & \vdots \\ w_s - w_{(N,M)} & \cdots & w_s - w_{(N,M)} \end{pmatrix} \begin{matrix} \rightarrow s \in \mathcal{N}_{(1,1)} \\ \dots \\ \rightarrow s \in \mathcal{N}_{(N,M)} \end{matrix} \quad (19)$$

For each row representing the location i , s spans the values of the neighborhood \mathcal{N}_i (as in (15)). Using this notation the smoothness term can be represented in a more compact form:

$$E_s = F(Kw), \quad (20)$$

where $F : \mathbb{R}^{2 \times |\Omega| \times |\mathcal{N}|} \rightarrow \mathbb{R}$ is the function defined by

$$F(y) := \|bf \cdot y\|_1 \quad (21)$$

The argument y has two components: $y = (yu, yv)$. The bilateral filter bf and the components yu , yv are matrices in $\mathbb{R}^{|\Omega| \times |\mathcal{N}|}$ and the multiplication in the function takes place element-by-element: $bf_{i,s} \cdot y_{i,s} = (bf_{i,s} \cdot yu_{i,s}, bf_{i,s} \cdot yv_{i,s})$. The expanded form of F is:

$$F(y) = \sum_{i \in \Omega} \sum_{s \in \mathcal{N}_i} (|bf_{i,s} \cdot yu_{i,s}| + |bf_{i,s} \cdot yv_{i,s}|) \quad (22)$$

Substituting $y_{i,s} = (Kw)_{i,s} = w_s - w_i = (u_s - u_i, v_s - v_i)$

in the above formula, we obtain (20).

Using these notations, the optical flow energy error to be minimized now reads:

$$\min_w \{\lambda \cdot E_d(w) + F(Kw)\} \quad (23)$$

In the following we present a minimization scheme based on the projected-proximal-point algorithm (PPA). It can be found in [38-41], but we also present its principle for completeness and for clarity.

Projected proximal point strategy

The discretized optical flow problem (23) is convex in w . The data term $E_d(w)$ is differentiable and the only issue is the smoothness term. The function F is quite simple, but the operator K introduces the complications and the idea is to eliminate it, by means of duality as below:

$$\min_w \left\{ \lambda \cdot E_d(w) + \left[\max_q \left\{ \langle Kw, q \rangle - F^*(q) \right\} \right] \right\}, \quad (24)$$

where $\langle \cdot, \cdot \rangle$ is the standard dot product of the space $\mathbb{R}^{2 \times |\Omega| \times |\mathcal{N}|}$

and F^* is the convex conjugate of F , given by:

$$F^*(q) = \begin{cases} 0, & q \in Q \\ \infty, & \text{otherwise} \end{cases} \quad (25)$$

The convex set Q is given by

$$Q = \left\{ q \in \mathbb{R}^{2 \times |\Omega| \times |\mathcal{N}|} : \begin{matrix} q_{i,s} = (qu_{i,s}, qv_{i,s}), \\ \|qu_{i,s}\| \leq bf_{i,s}, \|qv_{i,s}\| \leq bf_{i,s} \forall i \in \Omega, \forall s \in \mathcal{N}_i \end{matrix} \right\} \quad (26)$$

This dual-optimization problem (min, max) can be combined into a single one with multiple variables.

$$\min_w \max_q \left\{ \lambda \cdot E_d(w) + \langle Kw, q \rangle - F^*(q) \right\} \quad (27)$$

Min and max can be swapped because the min-max theorem [42] holds here. The problem is convex in the primal variable w and is concave in the dual variable q .

The proximal point strategy [38-41] implies the minimaximization of the following function, denoted as $\Phi(w, q)$:

$$\lambda \cdot E_d(w) + \langle Kw, q \rangle - F^*(q) + \frac{1}{2\tau} \|w - w^k\|^2 - \frac{1}{2\eta} \|q - q^k\|^2, \quad (28)$$

where w^k, q^k are the results from the previous estimate. The idea is to find a point w^{k+1}, q^{k+1} which is as close as possible to the optimum and not too far from the previous estimate. The parameters τ, η control the proximity to the previous estimations w^k, q^k . This objective is also convex-concave, as the original. The new estimates are the solutions of the problems:

$$\begin{cases} a) w^{k+1} = \arg \min_w \Phi(w, q^k) \\ b) q^{k+1} = \arg \max_q \Phi(w^{k+1}, q) \end{cases} \quad (29)$$

As can be seen in the above equalities, in order to find one variable the other is kept fixed. (29) b) is a point-wise problem in q . Also, (29) a) is a point-wise problem in w due to the property $\langle Kw, q^k \rangle = \langle w, K^T q^k \rangle$, where K^T is the adjoint operator of K . The exact expression of $K^T q^k$ can be easily deduced from this property. Since Φ is differentiable, the gradients of the functions in (29) are zero in the desired points:

$$\begin{cases} a) \frac{\partial \Phi}{\partial w}(w, q^k) = 0 = \lambda \cdot \frac{\partial E_d}{\partial w} + K^T q^k + \tau^{-1}(w - w^k) \\ b) \frac{\partial \Phi}{\partial q}(w^{k+1}, q) = 0 = Kw^{k+1} - \eta^{-1}(q - q^k) \end{cases} \quad (30)$$

There results a linear system of equations (a) in every $w_i = (u_i, v_i)$, which is very simple to solve. Their explicit form is not included, in order to save space. The only unsolved issue is how to keep the variable q in the proper set \mathcal{Q} , because the solution derived from the gradient does not guarantee this. In fact, we have to find a point in the set \mathcal{Q} that maximizes the objective (in q). The desired point is the projection of the gradient solution onto \mathcal{Q} . The following relation on $z = (zu, zv)$ defines this projection:

$$\text{Pr}(z) = \left(\frac{zu}{\max(|zu|, bf)} \cdot bf, \frac{zv}{\max(|zv|, bf)} \cdot bf \right) \quad (31)$$

All operations are element-wise. If z is in \mathcal{Q} , the projection leaves it unchanged, otherwise z is normalized such that $|zu_{i,s}| \leq bf_{i,s}$, $|zv_{i,s}| \leq bf_{i,s}$, in accordance with the definition of \mathcal{Q} (26).

Very simple computations of the gradient equations (30) and the use of projection onto \mathcal{Q} lead to the final numerical scheme that solves the considered optical flow model:

$$\begin{cases} a) w^{k+1} = \text{sol. of linear system}(w^k, K^T q^k, \lambda, \tau) \\ b) q^{k+1} = \text{Pr}(q^k + \sigma Kw^{k+1}) \end{cases} \quad (32)$$

- 2) Set up **pyramids** I_1, I_2 .
- 3) For each level starting with the coarsest (below, each referred variable is at the current level)
 - a) If the coarsest level, then initialize w, q to zero.
 - b) If not, upsample w, q from the previous level to the current level.
 - c) Computes the **bilateral coefficients** using the first image I_1 in the Lab color space.
 - d) Compute the correlation transform of the first image I_1
 - e) for $i = 1$ to warps (apply the warping technique)
 - i) Apply a **median filter** to u, v in order to avoid strong outliers.
 - ii) **Warp** the second image I_2 and compute the **correlation transforms** necessary for the temporal and spatial derivatives (12), (13)
 - iii) For $k=1$ to eq_iterations (find the minimum of (23))
 - A. **Primal step**: update w using (32), a)
 - B. **Dual step**: update q using (32), c)

Fig. 3 The sequence of the performed operations

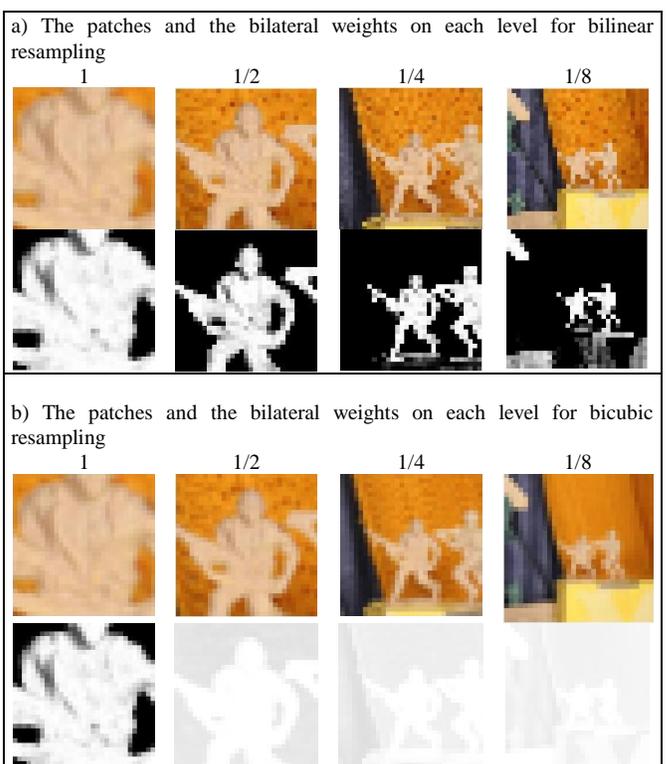


Fig. 4. The neighborhood and the coefficients of the bilateral filter for the left soldier in "Rubber Whale", after bilinear resampling (the first two rows of images) and after bicubic resampling (the next two rows of images), for each level of the pyramid. The bicubic resampling leads to an oversmoothing of the flow in the lower levels. The bilinear resampling preserves the object's boundaries. For these images we have considered only the distance between colors in the bilateral filter.

An approximate interpretation is that the algorithm performs a gradient descent operation in the primal variable w (convex) and a gradient ascent operation in the dual variable q (concave), but the scheme provided (32) is better because it is semi-implicit (q^{k+1} depends on w^{k+1}).

V. IMPLEMENTATION

Like all variational approaches, the presented model relies on the Taylor approximation of the data matching term. It is integrated into the well-known coarse-to-fine warping strategy in order to cope with large displacements. The sequence in Fig. 3 shows the main steps performed.

We use bilinear interpolation to build the pyramids and bicubic interpolation for upsampling the flow. We have also tried the bicubic downsampling, but it affects the bilateral coefficients. It considerably mixes the intensities and disturbs the boundaries and the local properties of an object. An illustrative example is given in Fig. 4. For bicubic resampling Fig. 4. b), the coefficients on the coarser levels (1/2, 1/4, 1/8) do not delimitate the object from background, the differences are hardly seen, causing an oversmoothing of the flow on these levels. The bilinear resampling Fig. 4. a), separates the object from other pixels. The use of antialiasing when downsampling also mixes the intensities and causes oversmoothing in some situations. Finally, we have chosen bilinear downsampling for building the pyramids and an additional pyramid for computing the bilateral weights. This additional pyramid still uses bilinear downsampling, but with no antialiasing on fine levels. On coarse levels, antialiasing is employed. This setting depends on the number of levels. For a pyramid factor of 0.5, we don't use antialiasing for the level 2 of the pyramid.

For the computation of the derivatives we have used simple finite differences. At location $i = (x, y)$ and index k in the descriptor we have:

$$\nabla \tilde{C}_2((x, y), k) = \begin{pmatrix} \frac{\tilde{C}_2((x+1, y), k) - \tilde{C}_2((x-1, y), k)}{2} \\ \frac{\tilde{C}_2((x, y+1), k) - \tilde{C}_2((x, y-1), k)}{2} \end{pmatrix}^T \quad (33)$$

The quantity $\{\tilde{C}_2((x+1, y)) : \forall (x, y) \in \Omega\}$

represents the correlation transform of the image

$\{I_2(x+u_0(x, y)+1, y+v_0(x, y)) : \forall (x, y) \in \Omega\}$ which is computed by means of interpolation. We do not warp the correlation transform of the second image. Instead, we warp

the second image and we recompute the correlation transforms. Therefore, five interpolations and transforms take place at each warp: 4 for space derivatives in equation (33) and one for the temporal derivative in equation (13). The cubic interpolation was used to warp the second image, as recommended in [13]. We use a median filter of 3x3 before each warp.

The developed numerical scheme is highly parallelizable and the implementation on CUDA architectures is straightforward.

VI. RESULTS

For ground truth evaluation we turn to the Middlebury datasets [22]. Two types of evaluations are provided in order to measure the accuracy: the average endpoint error (AEPE) and the average angular error (AAE). We have used a pyramid factor of 0.5, 5 warps per level and 30 iterations per warp. We have set to 5x5 the size of the neighborhood for non-local propagation. Based on visual inspections of the bilateral coefficients for various points (an example is in Fig. 4. a)) we have set $\sigma_c = 7, \sigma_d = 7$. With these values, the bilateral coefficients look very natural (see Fig. 4. a)). A larger neighborhood might be beneficial, as shown in [7, 13], but we restrict this dimension in order to keep the running time in a reasonable range. In this paper we are more interested in the evaluation of the cross-correlation as matching measure. We have to determine the optimal window size for ZNCC and the optimal λ . The Middlebury training dataset contains eight pairs of images and we compute the mean AEPE and AAE over all these frames. For ZNCC window size we have tested three configurations: 3x3, 5x5 and 7x7. For each configuration we vary the values of λ from 3 to 40. Fig. 6 plots the evolution of the AEPE and AAE. The plots reveal that 3x3 is the optimal configuration and the best value for λ is between 11 and 13. We have chosen $\lambda = 12$ as the optimal weighting between the data term and the smoothness term. Fig. 5 shows the color encoded output of the optimal configuration for these semi-synthetic images. With these settings, the running time of our Matlab implementation for "Rubber Whale" was 150

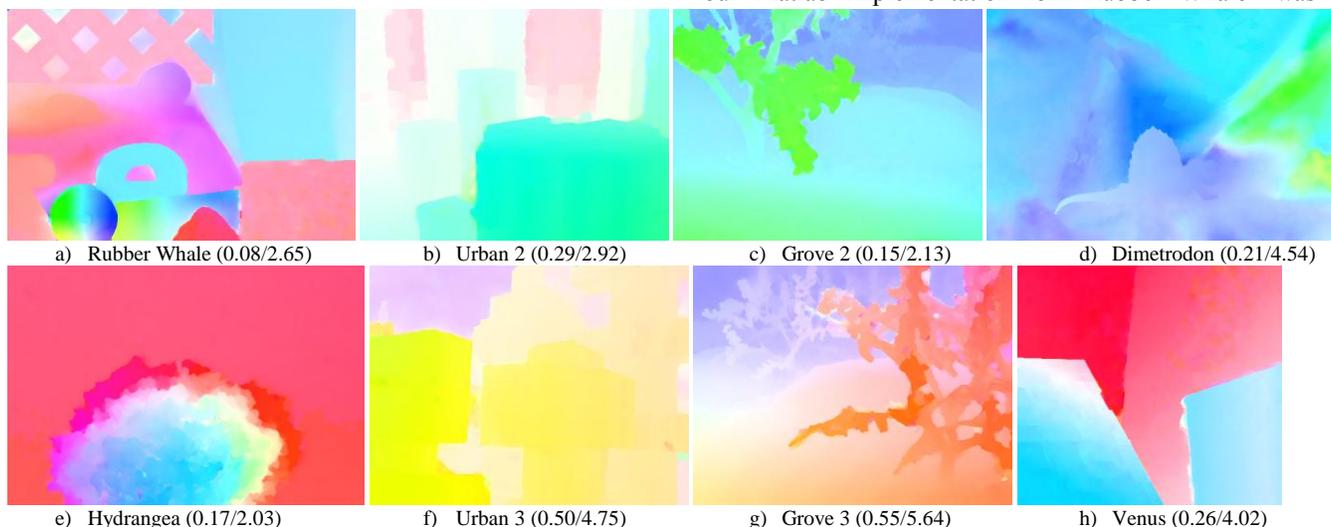


Fig. 5. Flow results on the training sequences of the Middlebury dataset. The corresponding AEPE and AAE are shown in brackets (AEPE/AAE).

seconds. Remarkable for the ZNCC data-term are the outputs of “Rubber Whale” (the hole in the central circle, the shadowed region between the wheel and the blue figure on the left) and Venus (the central bottom part).

Our method, denoted as “Correlation Flow”, is present on the Middlebury evaluation website. At the time of submission, “Correlation flow” was ranked 8 for the endpoint error. There were a total of 79 methods. The related method “NL-TV-NCC” [7] was ranked 28. The other related methods [26-29] were not listed there. Fig. 7 shows the results of our method for the Middlebury datasets with hidden ground-truth and Fig.8 shows a snapshot of the online Middlebury list. We have obtained good results, especially for “Wooden” (rank 2), “Teddy” (rank 4), “Mequon” (rank 4) and “Urban” (rank 10).

The real-world images are more complex than the semi-synthetic images provided by Middlebury. A common problem is represented by the changes in illumination. This occurs when the system’s camera changes its relative orientation to the sun (it takes a left or a right turn) or when it enters a shadowed/spotted region. In addition, the cameras might change their internal settings by modifying the gain or the exposure time. The first two scenarios in Fig. 9 depict such situations. In order to further reveal the advantages of ZNCC against other data terms we have applied non-uniform illuminations to the Middlebury pairs RubberWhale (third column) and Grove3 (last column). For RubberWhale we have modified the intensities of the second image using the formula:

$$I_2(i) = 255 \cdot I_1(i) \cdot (1 + 2 \cdot \rho(i)), \quad \text{where } \rho(i) = \exp(-\|i - i_0\|^2 / 6000)$$

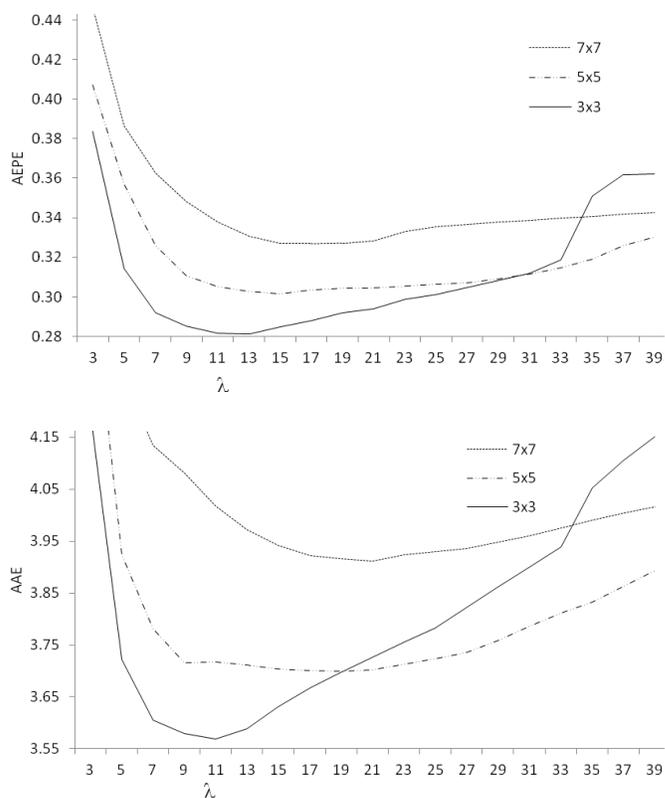


Fig. 6. The evolution of AEPE and AAE. Top: AEPE. Bottom: AAE.



Fig. 7. Flow result on the Middlebury sequences with hidden ground-truth.

Average endpoint error	avg. rank	Army (Hidden texture)			Mequon (Hidden texture)			Schefflera (Hidden texture)			Wooden (Hidden texture)			Grove (Synthetic)			Urban (Synthetic)			Yosemite (Synthetic)			Teddy (Stereo)					
		GT	im0	im1	GT	im0	im1	GT	im0	im1	GT	im0	im1	GT	im0	im1	GT	im0	im1	GT	im0	im1	GT	im0	im1			
		all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext	all	disc	untext			
MDP-Flow2 [70]	5.4	0.08	0.21	0.07	0.15	0.48	0.11	0.20	0.40	0.14	0.15	0.80	0.19	0.08	0.63	0.93	0.43	0.26	0.76	0.23	0.11	0.12	0.17	0.10	0.38	0.79	0.44	
NN-Field [73]	6.1	0.08	0.22	0.05	0.17	0.55	0.13	0.19	0.39	0.15	0.09	0.48	0.05	0.20	0.41	0.61	0.20	0.52	0.33	0.64	0.26	0.13	0.13	0.20	0.19	0.35	0.83	0.21
ADF [67]	11.8	0.08	0.22	0.06	0.18	0.62	0.14	0.29	0.17	0.22	0.17	0.16	0.21	0.91	0.31	0.07	0.69	0.19	1.03	0.47	0.43	0.13	0.91	0.28	0.12	0.15	0.12	0.20
Layers++ [37]	12.1	0.08	0.21	0.07	0.19	0.56	0.17	0.20	0.20	0.42	0.18	0.13	0.58	0.07	0.48	0.70	0.33	0.47	0.22	1.01	0.33	0.15	0.14	0.36	0.24	0.33	0.46	1.0
LME [72]	12.1	0.08	0.22	0.06	0.15	0.49	0.11	0.30	0.20	0.64	0.31	0.15	0.78	0.17	0.66	0.96	0.53	0.33	1.18	0.21	0.28	0.12	0.15	0.12	0.18	0.13	0.44	0.91
IROF++ [58]	12.5	0.08	0.23	0.07	0.21	0.68	0.17	0.20	0.28	0.63	0.19	0.15	0.73	0.09	0.60	0.89	0.42	0.43	1.08	0.31	0.16	0.10	0.12	0.12	0.44	0.12	0.98	
nLayers [57]	12.7	0.07	0.19	0.06	0.22	0.59	0.19	0.33	0.25	0.54	0.20	0.15	0.84	0.08	0.53	0.78	0.34	0.44	0.84	0.30	0.13	0.13	0.13	0.21	0.20	0.19	0.47	
Correlation Flow [79]	14.4	0.09	0.23	0.07	0.17	0.58	0.11	0.43	0.34	0.99	0.15	0.11	0.47	0.08	0.75	1.08	0.56	0.41	0.92	0.30	0.13	0.14	0.13	0.21	0.27	0.46	0.40	
FC-Layers-FF [77]	14.5	0.08	0.21	0.07	0.21	0.70	0.17	0.20	0.21	0.40	0.18	0.15	0.76	0.08	0.53	0.77	0.37	0.49	0.27	1.02	0.33	0.16	0.15	0.13	0.21	0.29	0.44	
FESL [75]	17.0	0.08	0.21	0.07	0.25	0.75	0.19	0.33	0.27	0.61	0.18	0.14	0.68	0.08	0.61	0.89	0.44	0.47	0.22	1.03	0.32	0.14	0.15	0.15	0.25	0.38	0.50	
ALD-Flow [68]	17.1	0.07	0.21	0.06	0.19	0.64	0.13	0.30	0.20	0.73	0.15	0.17	0.26	0.92	0.35	0.07	0.78	0.27	1.14	0.27	0.59	0.33	0.13	0.29	0.21	0.12	0.15	
COFM [59]	17.2	0.08	0.26	0.06	0.18	0.62	0.14	0.30	0.20	0.74	0.19	0.15	0.86	0.07	0.79	1.14	0.74	0.35	0.87	0.28	0.14	0.13	0.12	0.28	0.50	0.49		
SCR [74]	17.2	0.08	0.23	0.07	0.22	0.71	0.17	0.27	0.10	0.60	0.19	0.14	0.5	0.73	0.63	1.02	0.44	0.51	0.31	1.08	0.33	0.15	0.14	0.13	0.21	0.29	0.47	
TC-Flow [46]	17.3	0.07	0.21	0.06	0.15	0.59	0.11	0.31	0.25	0.78	0.14	0.16	0.21	0.86	0.75	1.11	0.54	0.42	1.2	1.40	0.25	0.11	0.12	0.29	0.55	0.62		
Sparse-NonSparse [56]	17.7	0.08	0.23	0.07	0.22	0.73	0.18	0.28	0.14	0.64	0.19	0.14	0.5	0.71	0.67	1.08	0.48	0.49	0.27	1.06	0.32	0.14	0.13	0.11	0.28	0.49		
Efficient-NL [60]	17.9	0.08	0.22	0.06	0.21	0.67	0.17	0.31	0.25	0.73	0.18	0.14	0.5	0.71	0.59	0.88	0.39	1.30	0.59	1.35	0.32	0.14	0.13	0.21	0.26	0.45		

Fig. 8. Snapshot of the Middlebury top for end-point error. Our method, “Correlation Flow” is ranked 8.

and $i_0 = (280, 250)$ (i_0 corresponds to the central circle). We have also normalized the result to keep the intensities in the range $[0, 1]$. This modification makes lighter the region around location i_0 , while the rest of the image becomes darker. It is similar to using two light sources: the intensity of the main light is reduced in order to have a darker image, while the second source of light is focused around location i_0 and has a higher intensity. We have applied similar formulas for the red and for the blue channels of the Grove image, producing a green image with two spots of light, as seen in Fig. 9 last column, second row. Four other methods are selected for comparison. The first method is obtained from our framework, by replacing ZNCC with SSD. SSD fails on all selected scenarios. The second method, Classic+NL[13] uses an intensity based data term and a preprocessing step to remove the illumination changes. LDOF[14] uses the brightness constraint assumption (BCA) and also the gradient constraint assumption (GCA). GCA is useful because it is invariant to additive multiplication changes. In addition to these two data terms, LDOF uses a large-displacement detector, based on HOG-like descriptors, which is also invariant to changes in illumination. As seen in Fig. 9, ZNCC is more robust than LDOF w.r.t. changes in illumination. MDP-Flow2[12] also uses BCA and GCA, but it selects the best between these constraints. At each level SIFT vectors are employed to help the initialization of optical flow. From these four selected methods MDP-Flow2 produces the best results (Fig. 9, Table I). However, the differences between the original errors and the errors after applying the illumination are higher than in the case of ZNCC. As seen in Table I, ZNCC is almost invariant to the considered illumination change, the errors have increased only a bit.

TABLE I

ERRORS (AEPE, AAE) BEFORE AND AFTER THE CHANGE IN ILLUMINATION

Method	RubberWhale	RubberWhale + illumination	Grove3	Grove3 +illumination
Correlation Flow	(0.08, 2.65)	(0.08, 2.81)	(0.55, 5.64)	(0.56, 5.64)
SSD	(0.13, 4.37)	(63.4, 68.2)	(0.73, 6.71)	(52.3, 78.6)
Classic+NL[13]	(0.07, 2.40)	(29.6, 66.3)	(0.46, 4.92)	(9.25, 31.7)
LDOF [14]	(0.12, 4.23)	(0.62, 17.2)	(0.69, 6.38)	(2.60, 25.5)
MDP-Flow2 [12]	(0.08, 2.51)	(0.14, 4.29)	(0.46, 4.87)	(0.55, 5.79)

VII. CONCLUSIONS

We have presented a direct approach to optical flow estimation with zero-mean normalized cross-correlation (ZNCC) as matching measure. The approach exploits the equivalence between the ZNCC and the SSD distance between the correlation transforms. We have shown that ZNCC is better than patch intensity matching (SSD or SAD). The deviations from the correct match between two regions are penalized stronger by ZNCC. The model employs a non-local propagation mechanism that uses a robust penalty in order to preserve motion boundaries. The bilateral filter guides the propagation process. The involved optimization problem is associated with a suitable numerical scheme derived using the projected proximal-point strategy. The numerical scheme is fast and highly parallelizable and directly fits to the programming capabilities of today’s graphic processing units. The experiments showed a very good robustness of ZNCC even in cases where severe illumination changes are present. The discriminative power of ZNCC and its invariance to illumination changes have led to highly accurate results on the Middlebury evaluations, both for the test and the training datasets.

REFERENCES

- [1] B. Horn and B. Schunck, "Determining optical flow," *Artificial Intelligence*, vol. 17, pp. 185-203, 1981.
- [2] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert, "High accuracy optical flow estimation based on a theory for warping," in *European Conference and Computer Vision*, 2004, pp. 25-36.
- [3] A. Bruhn and J. Weickert, "Towards Ultimate Motion Estimation: Combining Highest Accuracy with Real-Time Performance," presented at the Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1 - Volume 01, 2005.
- [4] A. Bruhn, J. Weickert, and C. Schnörr, "Lucas/Kanade Meets Horn/Schunck: Combining Local and Global Optical Flow Methods," *International Journal of Computer Vision*, vol. 61, pp. 211-231, 2005.
- [5] J. Weickert, A. Bruhn, N. Papenberg, and T. Brox, "Variational Optical Flow Computation: From Continuous Models to Algorithms," in *International Workshop on Computer Vision and Image Analysis*, 2003.
- [6] C. Zach, T. Pock, and H. Bischof, "A duality based approach for realtime TV-L1 optical flow," in *Proceedings of the 29th DAGM conference on Pattern recognition*, Heidelberg, Germany, 2007, pp. 214-223.

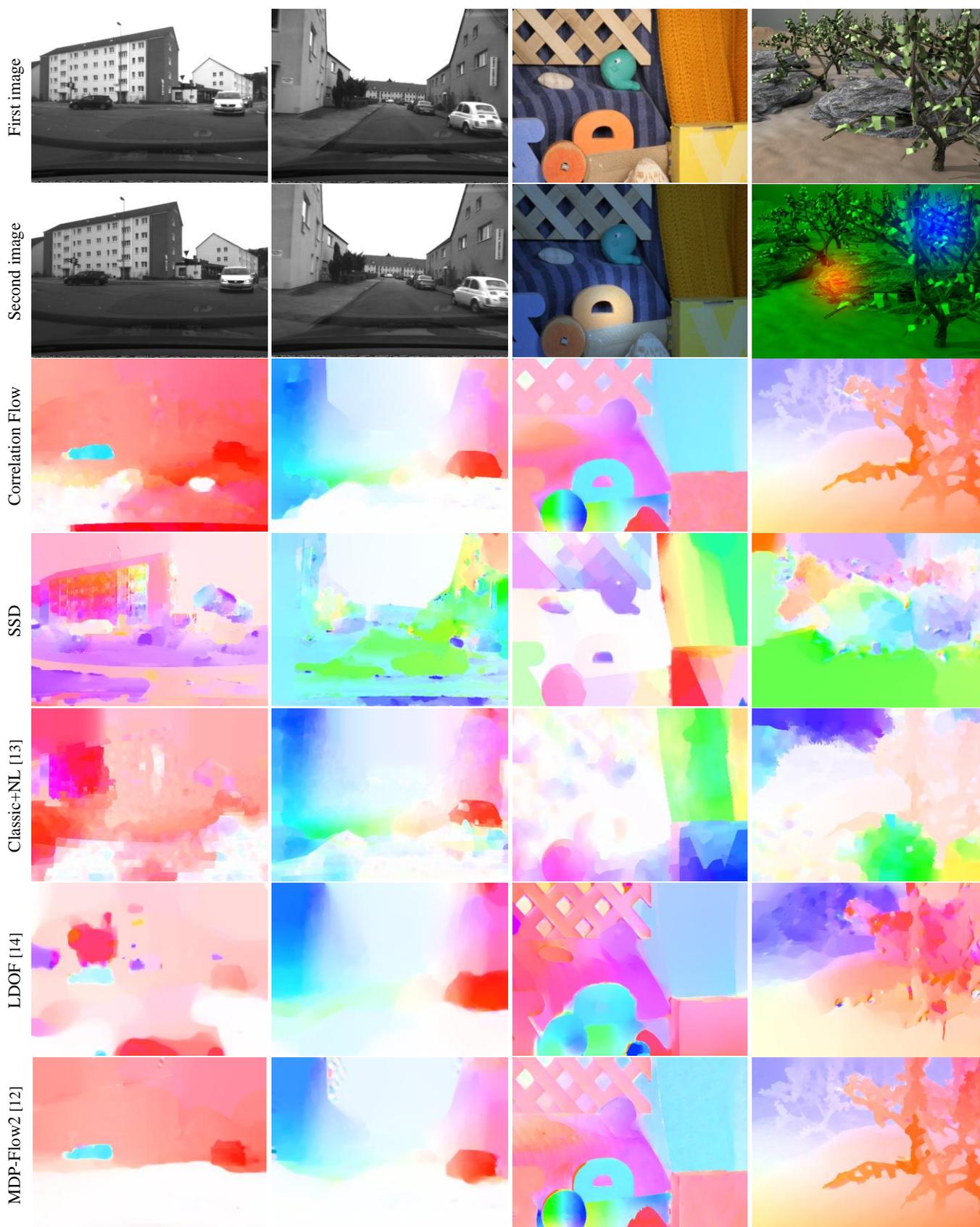


Fig. 9. The first two rows show the input images. The first two columns show real-world scenarios. The last two columns shows the RubberWhale and Grove3 pairs, where we have added an artificial illumination to the second image. The other rows show the output flow of each selected method. Our approach, “Correlation Flow” is the most robust with respect to changes in illumination.

- [7] M. Werlberger, T. Pock, and H. Bischof, "Motion estimation with non-local total variation regularization," *Computer Vision and Pattern Recognition, IEEE Computer Society Conference on*, vol. 0, pp. 2464-2471, 2010.
- [8] M. Werlberger, W. Trobin, T. Pock, A. Wedel, D. Cremers, and H. Bischof, "Anisotropic Huber-L1 Optical Flow," in *Proceedings of British Machine Vision Conference (BMVC)*, 2009.
- [9] M. Drulea and S. Nedevschi, "Total variation regularization of local-global optical flow," in *Intelligent Transportation Systems (ITSC), 2011 14th International IEEE Conference on*, 2011, pp. 318-323.
- [10] M. Drulea, I. R. Peter, and S. Nedevschi, "Optical flow. A combined local-global approach using L1 norm," in *Proceedings of the 2010 IEEE 6th International Conference on Intelligent Computer Communication and Processing*, 2010, pp. 217-222.
- [11] L. Xu, J. Chen, and J. Jia, "A Segmentation Based Variational Model for Accurate Optical Flow Estimation," in *Proceedings of the 10th European Conference on Computer Vision: Part I*, Marseille, France, 2008, pp. 671-684.
- [12] L. Xu, J. Jia, and Y. Matsushita, "Motion detail preserving optical flow estimation," in *Computer Vision and Pattern Recognition*, San Francisco, CA, USA, 2010, pp. 1293-1300.
- [13] D. Sun, S. Roth, and M. Black, "Secrets of optical flow estimation and their principles," in *In Proc. of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*, San Francisco, CA, USA, 2010, pp. 2432-2439.
- [14] T. Brox and J. Malik, "Large Displacement Optical Flow: Descriptor Matching in Variational Motion Estimation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 500-513, March, 2011 2011.
- [15] T. Brox, C. Begler, and J. Malik, "Large displacement optical flow," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 41-48.
- [16] M. A. Mohamed and B. Mertsching, "TV-L1 Optical Flow Estimation with Image Details Recovering Based on Modified Census Transform," in *Advances in Visual Computing*, ed: Springer, 2012, pp. 482-491.
- [17] H. A. Rashwan, D. Puig, and M. A. Garcia, "Improving the robustness of variational optical flow through tensor voting," *Computer Vision and Image Understanding*, 2012.
- [18] A. Chambolle, "An Algorithm for Total Variation Minimization and Applications," *J. Math. Imaging Vis.*, vol. 20, pp. 89-97, 2004.
- [19] H. H. Nagel and W. Enkelmann, "An investigation of smoothness constraints for the estimation of displacement vector fields from image sequences," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, pp. 565-593, 1986.
- [20] A. Wedel, T. Pock, C. Zach, H. Bischof, and D. Cremers, "An Improved Algorithm for TV-L1 Optical Flow," in *Statistical and Geometrical Approaches to Visual Motion Analysis*, ed: Springer-Verlag, 2009, pp. 23-45.
- [21] H. Zimmer, A. Bruhn, and J. Weickert, "Optic flow in harmony," *International Journal of Computer Vision*, vol. 93, pp. 368-388, 2011.
- [22] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A Database and Evaluation Methodology for Optical Flow," *Int. J. Comput. Vision*, vol. 92, pp. 1-31, 2011.
- [23] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *Int. J. Comput. Vision*, vol. 60, pp. 91-110, 2004.
- [24] H. Bay, A. Ess, T. Tuytelaars, and L. V. Gool, "Speeded-Up Robust Features (SURF)," *Comput. Vis. Image Underst.*, vol. 110, pp. 346-359, 2008.
- [25] <http://vision.middlebury.edu/flow/>.
- [26] J. Molnár and D. Chetverikov. (2010). Illumination-robust variational optical flow based on cross-correlation. *Proc. 33rd Workshop of the Austrian Association for Pattern Recognition*, 119-128.
- [27] F. Steinbruecker, T. Pock, and D. Cremers, "Advanced data terms for variational optic flow estimation," presented at the Vision, Modeling, and Visualization Workshop, Braunschweig, Germany, 2009.
- [28] S. Fazekas, D. Chetverikov, and J. Molnár, "An implicit non-linear numerical scheme for illumination-robust variational optical flow," in *British Machine Vision Conference*, 2009.
- [29] J. Molnár, D. Chetverikov, and S. Fazekas, "Illumination-robust variational optical flow using cross-correlation," *Computer Vision and Image Understanding*, vol. 114, pp. 1104-1114, 2010.
- [30] F. Durand and J. Dorsey, "Fast bilateral filtering for the display of high-dynamic-range images," in *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, San Antonio, Texas, 2002, pp. 257-266.
- [31] C. Tomasi and R. Manduchi, "Bilateral Filtering for Gray and Color Images," in *Proceedings of the Sixth International Conference on Computer Vision*, 1998.
- [32] K.-J. Yoon and I. S. Kweon, "Adaptive Support-Weight Approach for Correspondence Search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 28, p. 650, 2006.
- [33] J. Xiao, H. Cheng, H. Sawhney, C. Rao, and M. Isnardi, "Bilateral Filtering-Based Optical Flow Estimation with Occlusion Detection," in *European Conference on Computer Vision*, 2006, pp. 211-224.
- [34] G. Gilboa and S. Osher, "Nonlocal operators with applications to image processing," *Multiscale Model. Simul.*, vol. 7, pp. 1005-1028, 2008.
- [35] S. Kindermann, S. Osher, and P. W. Jones, "Deblurring and denoising of images by nonlocal functionals," *Multiscale Modeling and simulation*, vol. 4, pp. 1091-1115, 2006.
- [36] L. P. Yaroslavsky, *Digital Picture Processing: An Introduction*: Springer-Verlag, 1985.
- [37] A. Buades, B. Coll, and J.-M. Morel, "A Non-Local Algorithm for Image Denoising," presented at the Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05) - Volume 2 - Volume 02, 2005.
- [38] A. Chambolle and T. Pock, "A First-Order Primal-Dual Algorithm for Convex Problems with Applications to Imaging," *Journal of Mathematical Imaging and Vision*, vol. 40, pp. 120-145, 2011.
- [39] R. T. Rockafellar, "Monotone Operators and the Proximal Point Algorithm," *SIAM Journal on Control and Optimization*, vol. 14, pp. 877-898, 1976.
- [40] B. Martinet, "Régularisation d'inéquations variationnelles par approximations successives," in *RIRO* vol. 4, ed, 1970, pp. 154-159.
- [41] B. S. He and X. M. Yuan, "Convergence analysis of primal-dual algorithms for a saddle-point problem: From contraction perspective," *SIAM Journal on Imaging Science*, vol. 5, pp. 119-449, 2012 2012.
- [42] R. T. Rockafellar, *Convex Analysis (Princeton Mathematical Series)*: Princeton University Press.



Marius Drulea received his M.S. and B.S. degrees in computer science from the Technical University of Cluj-Napoca, Romania in 2011 and 2009, respectively. He is currently working toward the PhD. degree in computer science, focusing on environment perception for intelligent vehicles. His research interests include image processing, pattern recognition, motion estimation, convex optimization, stereo-vision and high performance computing on multi-core architectures.



Sergiu Nedevschi (M'99) received the M.S. and PhD degrees in Electrical Engineering from the Technical University of Cluj-Napoca (TUCN), Cluj-Napoca, Romania, in 1975 and 1993, respectively. From 1976 to 1983, he was with the Research Institute for Computer Technologies, Cluj-Napoca, as researcher. In 1998, he was appointed Professor in computer science and founded the Image Processing and Pattern Recognition Research Laboratory at the TUCN. He is currently vice-rector at TUCN. He has published more than 200 scientific papers and has edited over ten volumes, including books and conference proceedings. His research interests include Image Processing, Pattern Recognition, Computer Vision, Intelligent Vehicles, Signal Processing, and Computer Architecture.