On Improving the Robustness of Variational Optical Flow against Illumination Changes

Mahmoud A. Mohamed
GET lab.
Paderborn university
Pohlweg 47-49
33098 Paderborn Germany
mahmoud@get.upb.de

Hatem A. Rashwan
IRCV group
Rovira i Virgili university
Av. Paisos Catalans, 26
43007 Tarragona, Spain
hatem.abdellatif@urv.cat

Bärbel Mertsching
GET lab.
Paderborn university
Pohlweg 47-49
33098 Paderborn Germany
Mertsching@get.upb.de

Miguel Angel García
Autonomous university of Madrid
C. Francisco Tomás, 11
28049 Madrid,Spain
miguelangel.garcia@uam.es

Domenec Puig
IRCV group
Rovira i Virgili university
Av. Paisos Catalans, 26
43007 Tarragona, Spain
domenec.puig@urv.cat

ABSTRACT

The brightness constancy assumption is the base of estimating the flow fields in most differential optical flow approaches. However, the brightness constancy constraint easily violates with any variation in the lighting conditions in the scene. Thus, this work proposes a robust data term against illumination changes based on a rich descriptor. This descriptor extracts the textures features for each image in the two consecutive images using local edge responses. In addition, a weighted non-local term depending on the intensity similarity, the spatial distance and the occlusion state of pixels is integrated within the adapted duality total variational optical flow algorithm in order to obtain accurate flow fields. The proposed model yields state-of-the-art results on the the KITTI optical flow database and benchmark.

Categories and Subject Descriptors

I.2.10 [**Artificial Intelligence**]: Vision and Scene Understanding, motion; I.4.8 [**Image Processing**]: Scene Analysis time uarying imagery; I.4.9 [**Computing Methodologies**]: Image Processing and Computer Vision

General Terms

Algorithms, Measurement, Theory

Keywords

Motion Estimation, optical flow, illumination changes, large displacement, total variational

ARTEMIS'13, October 21, 2013, Barcelona, Spain.

http://dx.doi.org/10.1145/2510650.2510660 .

1. INTRODUCTION

Optical flow allows the estimation of the apparent motion of the scene. Motion estimation is a key task of a variety of applications, such as surveillance, robot-vision and driver-assistance. These applications require robust optical flow methods that are able to cope with different dramatically changing scenarios. The robustness of optical flow is badly affected by several surrounding environment factors such as fog, sunshine, clouds, shadow, shading, and lighting changes that yield brightness changes between two consecutive images.

Recently, most of the optical flow methods are concerned with estimating accurate flow fields under ideal conditions[1] rather than increasing the robustness in realistic scenes under various conditions. In addition, the brightness constancy assumption and the high-order constancy assumptions, such as gradient constancy were used by most variational optical flow approaches. Unfortunately, these assumptions are strongly become contaminated when applied to image sequences including illumination changes.

In the literature, many methods have presented different optical flow models that are robust against illumination changes. A robust energy function for solving the optical flow problem taking into account multiplicative and additive illumination factors was proposed by [5]. However, the simultaneously dealing of motion estimation and illumination variations in one energy function yields a more complex optimization problem. In addition, the accuracy of the optical flow estimation has adversely been affected, if the assumption of illumination factors is not accurate. Moreover, [6] proposed a photometric invariants of the dichromatic reflection model. However, this model is only applicable to color images with brightness variations

Furthermore, [8] proposed an illumination-invariant total variation with L1 norm (TV-L1) optical flow model by replacing the data term proposed by [14] by the Hamming distance of two Census Transform signatures. However, census signatures encode the local neighborhood intensity that is very sensitive to non-monotonic illumination variation and

[1] http://vision.middlebury.edu/flow/data/

random noise. In addition, the census transform discards most of the information casting from neighbors, and they cannot distinguish between dark and bright regions in a neighborhood.

In addition, the normalized cross correlation was proposed by [7] as a data term and leads to increasing the robustness against multiplicative illumination changes. In turn, [13] tackles the problems of poorly textured regions, occlusions and small scale image structure by incorporating a low level image segmentation process that has been used in a non-local total variation regularization term in a unified variational framework. In addition, [13] proposed the truncated normalized cross correlation that is robust against illumination changes in order to implement the data term.

Moreover, [15] presented a robust data term under outliers and varying illumination conditions based on constraint normalization, and an HSV color space with higher order constancy assumptions (gradient constancy assumption) as well as an anisotropic smoothness term is designed to work complementary to the data term. In addition, a weighted non-local term that depends on both the color similarity and the occlusion state of pixels and robustly integrates flow estimates over large spatial neighborhoods was proposed by [12]. However, the data term proposed by [12] still depends on the well-known brightness constraint.

The main contribution of this work is to replace the brightness constancy with a local texture feature which is more robust against illumination changes. Therefore, this work proposes the usage of the modified local directional pattern (MLDP) as a texture descriptor in order to extract texture features from two consecutive images. The extracted features are then utilized as a texture constancy assumption for the data term of the TV-L1 optical flow model proposed by [14]. In addition, this paper aims at compensating the loss of accuracy of the estimated flow field due to using an isotropic regularization term with an additional weighted non-local term that is similar to the one proposed by [12].

The rest of this paper is organized as follows: Section 2 discusses the local directional pattern and its modification version. In turn, the variational optical flow model containing a data term, a regularization term and a weighted non-local is summarized in section 3. Experimental results are shown and discussed in section 4, including a comparison with the state-of-the-art optical flow methods. Finally, conclusions and future work are given in section 5.

2. LOCAL DIRECTIONAL PATTERN

Census transform (CT) or local binary pattern is a widely used texture descriptor used as a facial descriptor. It can handle monotonic illumination changes; however, it depends on the intensity values that are more sensitive to noise and non-monotonic illumination changes. Furthermore, it cannot solve the problem of affine motion (rotatory motion), nor it can handle the problem of blocks with saturated center pixels, which appears when all neighbors are greater or smaller than the center pixel value, as shown figure 2.

In turn, the edge responses are more insensitive with illumination change and noise than intensity values that are used in the other descriptors. Thus, the LDP proposed by [4] is a gray-scale texture pattern and a robust descriptor that is recently used in a face recognition. Furthermore, The LDP feature describes the local primitives, including different types of curves, corners, and junctions, more stably and retains more information. An LDP operator computes the edges responses values (ER) in all eight directions (starting from east) at each pixel position by using compass Kirsch masks centered on its position as illustrated in [4] (see figure 1). LDP then generates an 8-bit binary code from the relative strength edge magnitude that is computed.

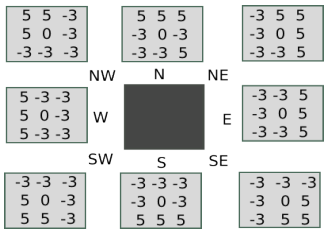

Figure 1: Kirsch masks used for extracting eight edge responses of a 3×3 neighborhood.

The presence of a corner or edge causes only high response values in some directions. Therefore, the LDP proposed by [4] is interested in the k ($k = 3$) most predominant directions to generate its code, also the top k directional bit responses are set to 1, as shown in figure 2.

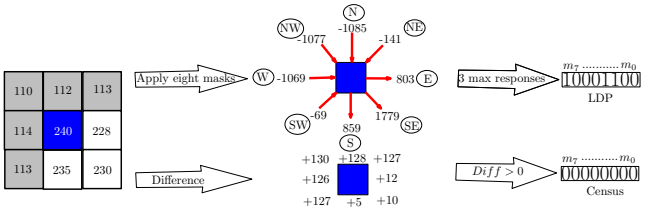

Figure 2: LDP vs. CT for a 3×3 neighborhood. (Up) A LDP descriptor, by applying the eight Kirsch masks to extract edge responses, and set then the corresponding bits to the three maximum responses to 1. (Down) A census descriptor, by calculating the difference between the center pixel and its neighbors, when the difference more than zero, the corresponding bit set to 0.

Another compass mask based on a Gaussian filter as formulated in (1) was proposed by [10]. This mask used for computing the edge responses is based on the derivative of a skewed Gaussian to create an asymmetric compass mask that is more robust to random noise. This mask can be formulated as:

$$M(x, y) = G_{\sigma}^{'}(x + k, y) * G_{\sigma}(x, y), \quad (1)$$

where $G_{\sigma}^{'}$ is the derivative of G_{σ} with respect to (x, y), $*$ is the convolution operation, and k is the offset of the Gaussian with respect to its center.

However, only relying on the k most prominent directions yields loosing a lot of structure information about a neighborhood. Thus, a new modification for the LDP descriptor called MLDP was proposed in this work. Likewise, eight edge responses are generated relating to eight masks operations. The 8-bit string is then generated by setting the corresponding bit of the positive edge responses to 1; in turn, the negative responses are set to 0.

$$s = \begin{cases} 1 & ER > 0 \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

where s is one bit in the MLDP signature.

Figure 3 shows that every pixel intensity is replaced by an 8-bit code. With MLDP, the extracted features (8-bit codes) only depend on the edge direction regardless of edge magnitude yielding a very robust descriptor against illumination. Therefore, in this paper, an 8-bit binary channels image is generated for each gray image with each channel corresponding to a directional mask operation.

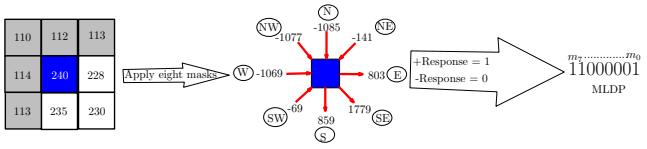

Figure 3: MLDP descriptor with setting the positive response to 1 and the negative response to 0.

In addition, if the features are extracted with a multi-scale approach that relies on rescaling the input image in a Gaussian pyramid, the features can be more insensitive to noise and illumination changes. On the one hand, the Kirsch masks [4] can be used for computing the eight edge responses for every pixel in every scale level. On the other hand, the shifted derivative of the Gaussian filter G'_σ shown in (1) can be utilized to generate edge responses without a convolution with the Gaussian filter G_σ, which is already applied on every scale level.

3. OPTICAL FLOW BASED ON MLDP

Given two consecutive frames $I_1(x, y)$ at time t and a frame $I_2(x+u, y+v)$ at time $t+1$, the duality of the TV-L1 optical flow model proposed by [14] is used to compute the vector flow field $w = (u, v)$ at a point $p = (x, y)$ in an image domain Ω.

$$\min_{u,v} \; E_I(u,v) = \sum_{\Omega} \left(\lambda\rho(x,y,u,v)^2 + \parallel \nabla u \parallel + \parallel \nabla v \parallel\right), \quad (3)$$

where ρ is the similarity function between two images; in turn, u and v are the horizontal and vertical optical flow components, and λ is the weight of the data term. Using a quadratic coupling term, the energy functional (3) is divided into two parts, which are then solved iteratively. The first part contains the data term:

$$\min_{u,v} \; E_{Id}(w) = \sum_{\Omega} \left(\lambda\rho(x,y,w)^2 + \frac{1}{2\theta}(w-\hat{w})^2\right), \quad (4)$$

where $w = (u, v)$ and $\hat{w} = (\hat{u}, \hat{v})$ is the auxiliary optical flow vector and θ is a threshold. The second part contains the regularization term:

$$\min_{\hat{u},\hat{v}} \; E_{Is}(\hat{w}) = \sum_{\Omega} \left(\frac{1}{2\theta}(w-\hat{w})^2 + \parallel \nabla\hat{u} \parallel + \parallel \nabla\hat{v} \parallel\right), \quad (5)$$

Assume $S_1(x, y)$ and $S_2(x+u, y+v)$ are the two descriptors extracted from the two images $I_1(x, y)$ and $I_2(x+u, y+v)$, respectively. Thus, the similarity function ρ used in between the two images can be rewritten as:

$$\rho(x,y,u,v) = S_2(x+u, y+v) - S_1(x,y), \quad (6)$$

Thus, $S_2(x+u, y+v)$ can be linearized around the starting value of w using a first order Taylor expansion:

$$S_2(x+u, y+v) = S_2(x,y) + \nabla^T S_2(x,y,\hat{w})(w-\hat{w}). \quad (7)$$

In turn, the derivative $\nabla^T S_2(x, y, \hat{w}) = \left[\frac{\partial S}{\partial x} = S_x, \frac{\partial S}{\partial y} = S_y\right]^T$ can be computed by applying a derivative mask to the binary image (i.e. Sobel) in the x and y directions. Thus, the similarity function will be:

$$\begin{aligned} \rho(x,y,w) &\approx \tilde{\rho}(x,y,w) \\ &= S_2(x,y) - S_1(x,y) + \nabla^T S_2(x,y,\hat{w})(w-\hat{w}). \end{aligned} \quad (8)$$

Now, (4) can be solved for (u, v) by doing:

$$\begin{aligned} \frac{\partial}{\partial u}(\lambda\tilde{\rho}(x,y,w)^2 + \frac{1}{2\theta}(u-\hat{u})^2) = 0, \\ \frac{\partial}{\partial v}(\lambda\tilde{\rho}(x,y,w)^2 + \frac{1}{2\theta}(v-\hat{v})^2) = 0. \end{aligned} \quad (9)$$

Both equations can be expressed in vector form as:

$$2\lambda\tilde{\rho}(x,y,w)\nabla S_2(x,y,\hat{w}) + \frac{1}{\theta}(w-\hat{w}) = 0. \quad (10)$$

Equation (10) is linear in (u, v) and can be solved as a linear system $Aw = b$.

In the proposed technique, the data term uses the similarity measure between two 8-bit channels descriptor extracted through MLDP descriptors. For every pixel, the similarity is calculated by counting the number of differences between the two descriptors. Assume $S_1(x, y)$ and $S_2(x+u, y+v)$ are the n binary channels descriptors extracted from the two images $I_1(x, y)$ and $I_2(x+u, y+v)$, respectively. In practice, the residual function between the two n-channels descriptors can be represented as:

$$\begin{aligned} \rho(x,y,u,v)^2 &= \sum_{i=1}^{n} \left(S_{2,i}(x+u, y+v) - S_{1,i}(x,y)\right)^2 \\ &= \sum_{i=1}^{n} \rho_i(x,y,u,v)^2, \end{aligned} \quad (11)$$

where n is the number of channels. Each ρ_i^2 will give 0 if both bits are the same and 1 if otherwise, $\rho_i = \{1, 0 \; or \; -1\}$. In practice, the summation over all ρ_i will measure the dissimilarity distance between the two descriptors.

Thus, the final data can be extended in order to be applicable to a multi channel descriptor as:

$$\min_{u,v} \; E_{Id}(w) = \sum_{\Omega} \left(\lambda \sum_{i=1}^{n} \rho(x,y,w)^2 + \frac{1}{2\theta}(w-\hat{w})^2\right), \quad (12)$$

Hence, A and b matrices of the linear system described in (10)can be written as:

$$A = \begin{bmatrix} \frac{1}{\theta} + 2\lambda\sum_{i=1}^{n} S_x^2 & 2\lambda\sum_{i=1}^{n} S_x S_y \\ 2\lambda\sum_{i=1}^{n} S_x S_y & \frac{1}{\theta} + 2\lambda\sum_{i=1}^{n} S_y^2 \end{bmatrix}. \quad (13)$$

and

$$b = \frac{1}{\theta}\begin{pmatrix}\hat{u}\\ \hat{v}\end{pmatrix} - 2\lambda\begin{pmatrix}\sum_{i=1}^{n} S_x\\ \sum_{i=1}^{n} S_y\end{pmatrix}(\sum_{i=1}^{n} S_t - (\sum_{i=1}^{n} S_x\hat{u} + \sum_{i=1}^{n} S_y\hat{v})). \quad (14)$$

Similarly, the smoothness term represents the isotropic total variation [9]. As a result, (5) can be decomposed into two equations and rewritten as:

$$E_u = \sum_{\Omega}\left[\frac{1}{2\theta}(u-\hat{u})^2 + \parallel \nabla\hat{u} \parallel\right], \quad (15)$$

$$E_v = \sum_{\Omega}\left[\frac{1}{2\theta}(v-\hat{v})^2 + \parallel \nabla\hat{v} \parallel\right]. \quad (16)$$

E_u and E_v have two unknowns, $\hat{u}$ and $\hat{v}$, while u, v are constants obtained after solving the data term.

For E_u, thus the Euler-Lagrange equation is:

$$-div[\frac{\nabla u}{\parallel \nabla u \parallel}] + \frac{1}{\theta}(u-\hat{u}) = 0 \quad (17)$$

Let $P_u = \nabla u / \parallel \nabla u \parallel$. Thus:

$$u = \lambda div(P_u) + \hat{u}, \quad (18)$$

By using (17) and (18), P_u can be rewritten as:

$$P_u^{h+1} = \frac{P_u^h + \tau\nabla(div(P_u^h) + \frac{\hat{u}}{\theta})}{1 + \tau \parallel \nabla(div(P_u^h) + \frac{\hat{u}}{\theta}) \parallel}, \quad (19)$$

where h is the iteration number, and $\tau \leq 1/8$ is the time step. The same can be applied to get P_v. That equation can be solved through a fixed-point iteration scheme as described in [9].

Furthermore, a coarse-to-fine scheme [1] is used for solving the energy function (3) in order to support both small and large displacements and improve the accuracy of flow fields. In each pyramid level, the scaled images are warped representations of the images based on the flow estimated at every preceding scale.

However, the only reliance on the isotropic L1 total variational in the regularization term causes the loss of some accuracy of the obtained flow fields for properly preserving edges and object boundaries as well as small details. In order to handle the motion discontinuity that is usually problematic due to occlusion and over-smoothing, the resulting flow field requires, with every pyramid level, a filtering stage to preserving edges and boundaries. Therefore, the motion boundary and edges are detected, and then dilated using a 5×5 mask in order to obtain flow boundary regions. For each pixel $p = (x, y)$ in the region, a weighted particular median filter proposed in [12] is performed:

$$E_w = \sum_{x,y}\sum_{(\acute{x},\acute{y})\in N_{x,y}} \varpi_{p,\acute{p}}(|u_{x,y} - u_{\acute{x},\acute{y}}| + |v_{x,y} - v_{\acute{x},\acute{y}}|). \quad (20)$$

where $(\acute{x}, \acute{y})$ is the spatial position of any pixel $\acute{p}$ belonging to a neighborhood of pixel p in a possibly large region $N_{x,y}$, and $\varpi_{p,\acute{p}}$ is the weighting function taking into account the occlusion state of pixels, $O(p)$ proposed in [11], in addition to color similarity and spatial distance. Thus $\varpi_{p,\acute{p}}$ can be formulated as shown in [12]:

$$\varpi_{p,\acute{p}} \propto exp\left(-\frac{(p-\acute{p})^2}{2\sigma_s^2} - \frac{(I(p) - I(\acute{p}))^2}{2\sigma_r^2}\right)\frac{O(\acute{p})}{O(p)}, \quad (21)$$

where $I(p)$ and $I(\acute{p})$ are the intensity values of points p and $\acute{p}$, respectively, and σ_s and σ_r are standard deviations with values 7.0 and 0.1, respectively. In motion boundary regions, the weighting in (21) is used in a 15×15 neighborhood. In the non-boundary regions, equal weights are used in a 5×5 neighborhood to compute the median as illustrated in [12].

4. EXPERIMENTS AND EVALUATION

In order to evaluate the robustness of the optical flow algorithm, synthetic illumination changes have been added to the GROVE2 sequence from the Middlebury datasets which has ground truth by changing the illumination of the second frame depending on:

$$I_o = uint8\left(255\left(\frac{mI_i + a}{255}\right)^{\gamma}\right), \quad (22)$$

where I_i and I_o are the input and output frames, respectively. $m > 0$ is a multiplicative factor, a is an additive change factor and $\gamma > 0$ is the gamma correction. The function $uint8$ is used for quantizing the values to an 8-bit unsigned integer format. Qualitative comparisons among MLDP and different feature descriptors have been done. Figure 4 shows the average end-point error (AEE) and the average angular error (AAE) between the flow fields obtained with MLDP, the census transform (CT) determined in a 3×3 neighborhood and the gradient constancy (GC). The effects of different values of m, a and γ have individually been assessed. As shown in figure 4, the gradient constancy is robust against small changes of m a and γ. In turn, MLDP increases the robustness against both small and large changes of m a and γ. In addition, the census transform yields adequate values for both AEE and AAE.

At the time of submission (July 2013), the results of the proposed model with (MLDP-OF) were evaluated with the KITTI[2] Vision Benchmark, which contains 195 testing sequences with ground truths, and it was ranked in the 9^{th} position against current state-of-the-art optical flow algorithms with an average of 8.90% incorrect pixels (percentage of pixels with AEE above 3 pixels) as shown in table 1. In turn, the baseline method [14] has 30.75% and [12] has 24.64%. Furthermore, the proposed methods were evaluated with different real image sequences that include illumination changes and large displacements of the training KITTI datasets. Table 2 shows the percentages of the incorrect pixels of four sequences that include illumination changes, as well as four sequences include large displacements calculated for the methods proposed in [15], [12], [2], [3] and [13], and the proposed method with MLDP and CT.

In addition, figures 5 and 6 show a visual example of the estimated flow field for sequence 44 and 181 of the KITTI datasets, which includes illumination changes and large displacement, respectively, as well as the error images and the error histograms. Among the evaluated approaches, the optical flow model based on MLDP yields the most accurate flow fields with respect to the state-of-the-art methods shown in table 2 for real images that include both illumination changes and large displacements.

5. CONCLUSION

This paper proposed an illumination robust texture constancy assumption to be used as a data term for variational

[2]http://www.cvlibs.net/datasets/kitti/

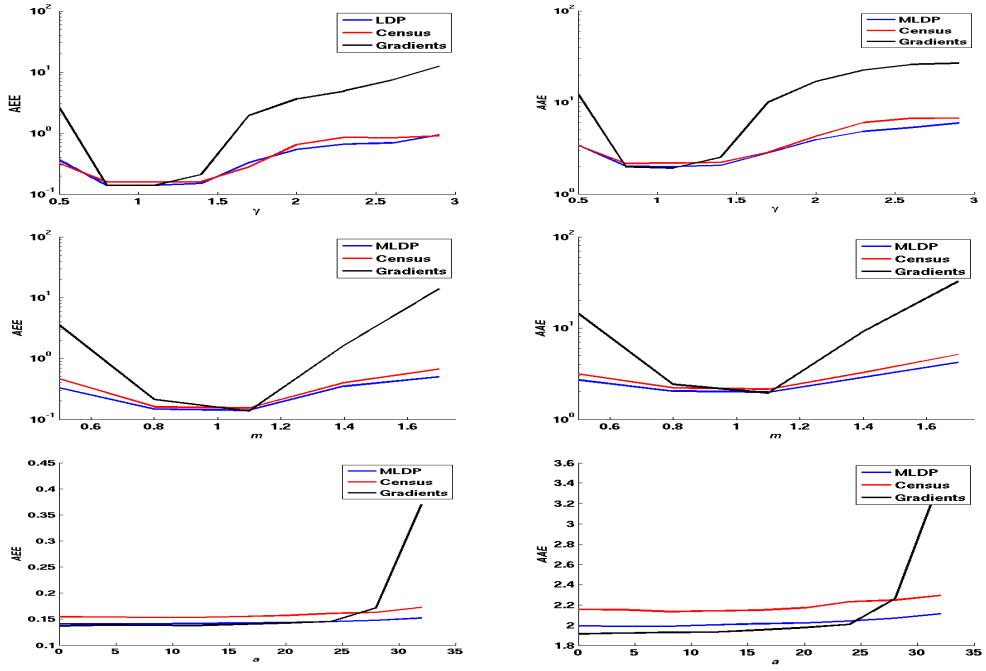

Figure 4: (Column 1) AEE and (column 2) AAE for MLDP, CT and GC for changing of γ, m and a respectively.

Table 1: The evaluation of the state-of-the-art method on the KITTI website July 2013.

Rank	Method	Out-Noc	Out-All	Avg-Noc	Avg-All
1	PR-Sf+E	4.08 %	7.79 %	0.9 px	1.7 px
2	PCBP-Flow	4.08 %	8.70 %	0.9 px	2.2 px
3	MotionSLIC	4.36 %	10.91 %	1.0 px	2.7 px
4	PR-Sceneflow	4.48 %	8.98 %	1.3 px	3.3 px
5	TGV2ADCSIFT	6.55 %	15.35 %	1.6 px	4.5 px
6	Data-Flow	8.22 %	15.78 %	2.3 px	5.7 px
7	TVL1-HOG	8.31 %	19.21 %	2.0 px	6.1 px
9	**MLDP-OF**	**8.91 %**	**18.95 %**	**2.5 px**	**6.7 px**
12	fSGM	11.03 %	22.90 %	3.2 px	12.2 px
13	TGV2CENSUS	11.14 %	18.42 %	2.9 px	6.6 px
14	C+NL-fast	12.42 %	22.27 %	3.2 px	7.8 px
25	DB-TV-L1	30.75 %	39.13 %	7.8 px	14.6 px

Table 2: Percentage of wrong pixels of state-of-the-art methods and the proposed method with eight sequences of the KITTI datasets: sequences 11, 15, 44 and 74, which include illumination changes ordered by difficulty , and sequences 117, 144, 147 and 181, which include large displacements ordered by difficulty.

Method	Seq44	Seq11	Seq15	Seq74	Seq147	Seq117	Seq144	Seq181
MLDP	**20.42%**	**29.67%**	**23.85%**	**56.01%**	**11.79%**	18.67%	**41.05%**	**59.40%**
CT (5 × 5)	35.23%	33.93%	29.04%	57.57%	13.98%	27.33%	47.68%	73.85%
CT (3 × 3)	29.55%	37.54%	33.74%	57.43%	14.76%	28.80%	48.97%	73.63%
GC	29.25%	35.72%	26.41%	59.20%	12.28%	17.70%	44.51%	67.63%
OFH 2011 [15]	23.22%	37.26%	32.20%	62.90%	15.04%	**16.26%**	42.04%	63.86%
SRB 2010 [12]	26.58%	40.61%	32.85%	62.94%	14.59%	24.71%	50.67%	67.11%
SRBF 2010[12]	31.83%	40.34%	35.13%	64.89%	14.79%	24.41%	50.66%	68.41%
BW 2005 [2]	32.44%	33.95%	47.70%	71.44%	16.98%	28.80%	46.98%	69.04%
HS 1981 [3]	42.96%	38.84%	58.08%	82.14%	24.84%	43.24%	51.89%	74.11%
WPB 2010 [13]	49.09%	49.99%	67.28%	88.67%	32.72%	46.80%	52.25%	76.00%

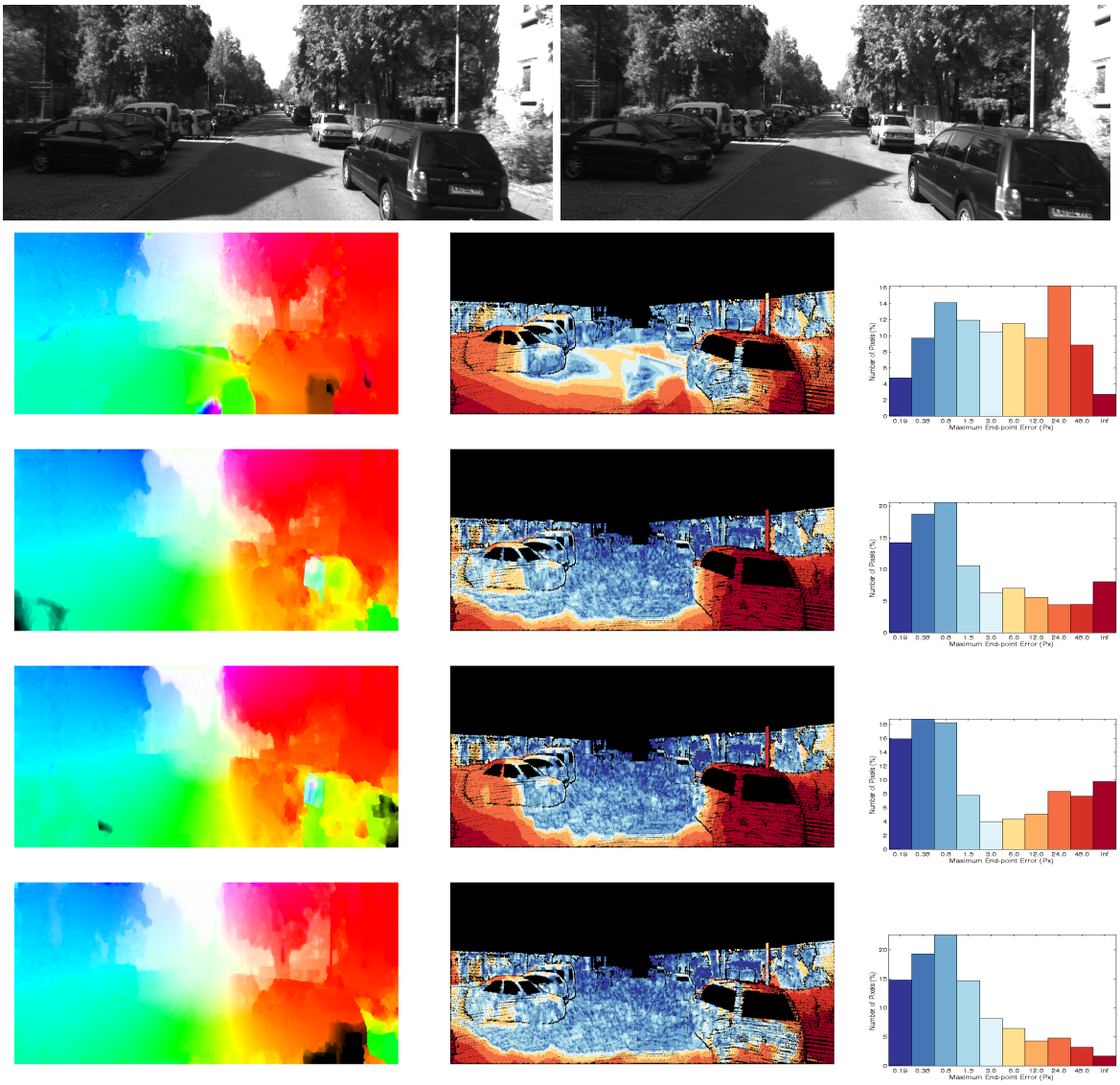

Figure 5: Row 1: Two original images for sequence 44 of the KITTI datasets. Resulting flow field, error image and error histogram for the proposed optical flow model with: Row 2: brightness constancy, Row 3: 3×3 census transform, Row 4: 5×5 census transform, Row 5: MLDP.

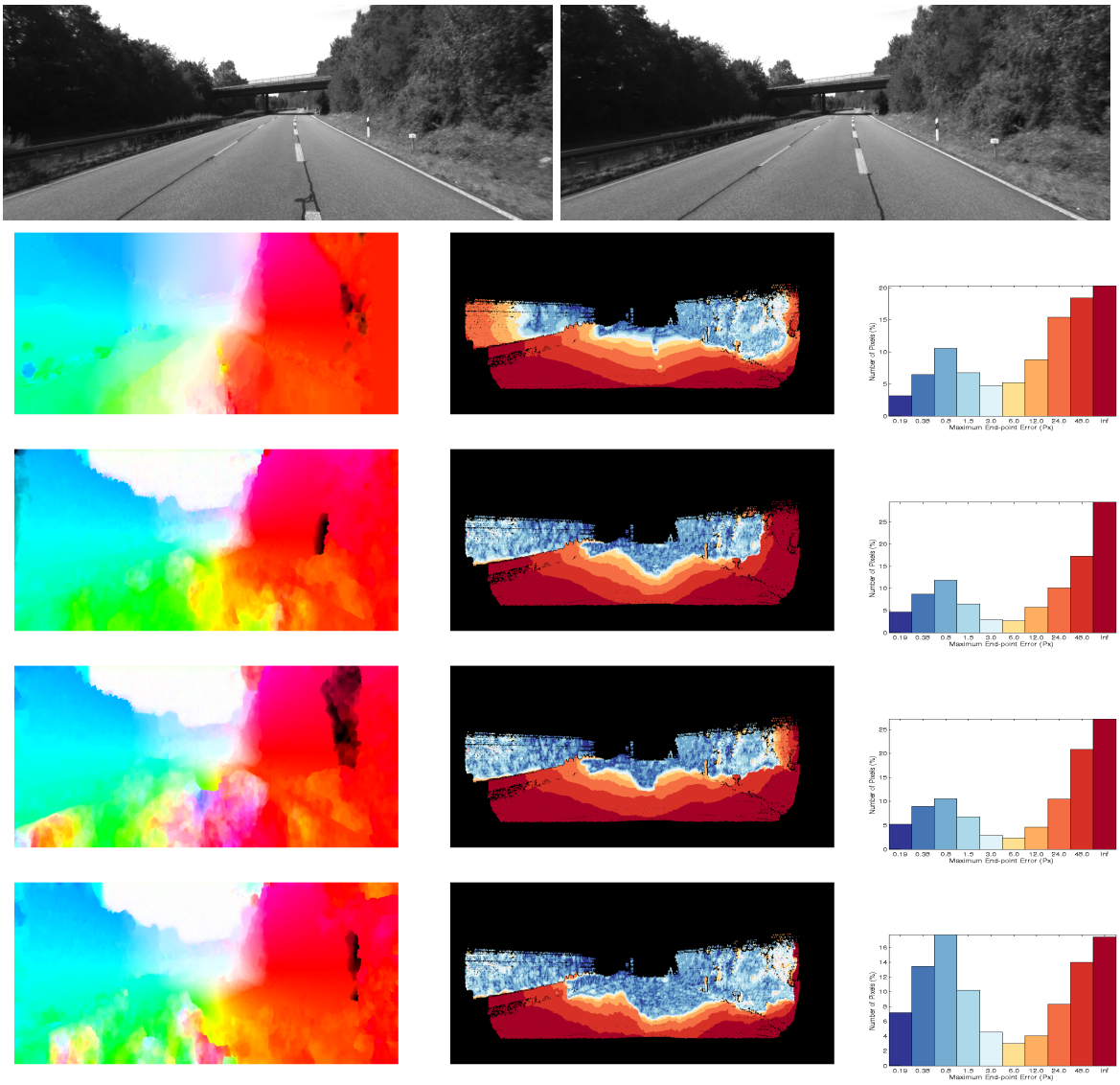

Figure 6: Row 1: Two original images for sequence 181 of the KITTI datasets. Resulting flow field, error image and error histogram for the proposed optical flow model with: Row 2: brightness constancy, Row 3: 3×3 census transform, Row 4: 5×5 census transform, Row 5: MLDP.

optical flow methods. The proposed method depends on local edge response in order to extract an 8-bit binary descriptor. In particular, the modified local directional pattern for each pixel within 3×3 local window was proposed in this paper as a texture descriptor in order to extract texture features from two consecutive images. In addition, the proposed method uses a duality of the TV-L1 optical flow model with a weighted non-local term in order to estimate accurate flow fields. The proposed algorithm assessed to different sequences of the KITTI datasets and it provided the more correct flow fields comparing to the sate-of-the-art methods. Ongoing work aims at including the epipolar geometry information for computing optical flow using stereo camera. In the future, the robust dense motion estimation will be applied in robot vision systems for outdoor tasks.

6. ACKNOWLEDGMENTS

Particular acknowledgment is due to DAAD (German Academic Exchange Service) and Egyptian Ministry of Higher Education for the financial support.

7. REFERENCES

[1] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In T. Pajdla and J. Matas, editors, *ECCV*, volume 3024 of *Lecture Notes in Computer Science*, pages 25–36. Springer, 2004.

[2] A. Bruhn and J. Weickert. Towards ultimate motion estimation: Combining highest accuracy with real-time performance. In *ICCV*, pages 749–755. IEEE Computer Society, 2005.

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

[4] T. Jabid, M. Kabir, and O. Chae. Local directional pattern (ldp) for face recognition. In *Consumer Electronics (ICCE), 2010 Digest of Technical Papers International Conference on*, pages 329–330. IEEE Computer Society, 2010.

[5] Y.-H. Kim, A. M. Martínez, and A. C. Kak. Robust motion estimation under varying illumination. *Image Vision Comput.*, 23(4):365–375, 2005.

[6] Y. Mileva, A. Bruhn, and J. Weickert. Illumination-robust variational optical flow with photometric invariants. In F. A. Hamprecht, C. Schnörr, and B. Jähne, editors, *DAGM-Symposium*, volume 4713 of *Lecture Notes in Computer Science*, pages 152–162. Springer, 2007.

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

[8] T. Mueller, C. Rabe, J. Rannacher, U. Franke, and R. Mester. Illumination-robust dense optical flow using census signatures. In R. Mester and M. Felsberg, editors, *DAGM-Symposium*, volume 6835 of *Lecture Notes in Computer Science*, pages 236–245. Springer, 2011.

[9] T. Pock, M. Urschler, C. Zach, R. Beichel, and H. Bischof. A duality based algorithm for tv-l1-optical-flow image registration. In *MICCAI*, volume 4792 of *Lecture Notes in Computer Science*, pages 511–518. Springer, 2007.

[10] A. Ramirez Rivera, J. Rojas Castillo, and O. Chae. Local directional number pattern for face analysis: Face and expression recognition. *IEEE Trans Image Process*, 2012.

[11] P. Sand and S. J. Teller. Particle video: Long-range motion estimation using point trajectories. *International Journal of Computer Vision*, 80(1):72–91, 2008.

[12] D. Sun, S. Roth, and M. J. Black. Secrets of optical flow estimation and their principles. In *CVPR*, pages 2432–2439. IEEE, 2010.

[13] M. Werlberger, T. Pock, and H. Bischof. Motion estimation with non-local total variation regularization. In *CVPR*, pages 2464–2471. IEEE, 2010.

[14] C. Zach, T. Pock, and H. Bischof. A duality based approach for realtime tv- l1 optical flow. In *DAGM*, pages 214–223, 2007.

[15] H. Zimmer, A. Bruhn, and J. Weickert. Optic flow in harmony. *International Journal of Computer Vision*, 93(3):368–388, 2011.