

Differential Optical Flow Estimation Under Monocular Epipolar Line Constraint

Mahmoud A. Mohamed^(✉), M. Hossein Mirabdollah,
and Bärbel Mertsching

GET Lab, University of Paderborn,
33098 Paderborn, Germany
{mahmoud,mirabdollah,mertsching}@get.upb.de
<http://getwww.upb.de>

Abstract. In this paper, a new method is presented to use the epipolar constraint for the estimations of optical flows. We derive the necessary formulation to add the epipolar constraint in terms of optical flow components and force the components to transform points from the first frame to the next consecutive frame such that the points lie on their correspondent epipolar lines. In this work, no smoothness term is utilized and the performance of the proposed method is evaluated based only on data terms. We conducted different evaluations using two different point matching methods (SIFT and Lucas-Kanade) and used them in two different fundamental matrix estimation methods required to calculate epipolar line coefficients. It is demonstrated that epipolar constraint yields noticeable improvements almost in all of the cases.

1 Introduction

Optical flow is defined as the apparent motion of pixels between two consecutive frames. Thus, a flow field describes the dynamics of a scene and involves two independent motions; the ego-motion of the camera and motions of objects. In literature, many differential optical flow methods, which are mainly dependent on the brightness constancy assumption have been proposed. In these methods, it is assumed that the intensities of correspondent pixels do not change if objects or cameras move. Unfortunately, not every object motion yields changes in gray values and not every changes in gray values are generated by body motion. Challenging outdoor scenes with poorly textured regions, illumination changes, shadows, reflections, glare, and the inherent noise of the camera image may yield gray value changes while the depicted object remains stationary. Especially in poor illumination conditions, the number of photons collected by a pixel, may vary over time.

Recently, many methods have presented different optical flow models to tackle with the problem of illumination changes. A robust energy is proposed in [1], which take into account multiplicative and additive illumination factors. Nevertheless, dealing with motion estimations and illumination variations in one energy function gives rise to a more complex optimization problem. Additionally, the accuracy

of the optical flow estimation is adversely affected, if the assumption of illumination factors is not accurate. Moreover, [2] proposed a photometric invariant of the dichromatic reflection model. This model is only applicable to color images with brightness variations. Furthermore, [3] proposed an illumination-invariant total variation with L1 norm (TV-L1) optical flow model by replacing the data term proposed in [4] with the Hamming distance of two census transform signatures. Census signatures encode local neighborhood intensity variations, which are very sensitive to non-monotonic illumination variation and random noise. In addition, the census transform discards most of the information casting from neighbors, and cannot distinguish between dark and bright regions in a neighborhood which called saturated center.

In addition, the normalized cross correlation was utilized in [5] as a data term and led to increasing the robustness of the estimated optical flow. In turn, [6] tackles the problem of poorly textured regions, occlusions and small scale image structures by incorporating a low level image segmentation process that has been used in a non-local total variational regularization term in a unified variational framework. In addition, an optical flow estimation method based on the zero-mean normalized cross-correlation transform was introduced in [7].

Dense descriptors such as census [3], histogram of oriented gradient (HOG) [8] and local directional pattern (LDP) [9] and [10] have been incorporated directly into the classical energy minimization framework in order to gain robustness for the estimated optical flow in real-world vision systems. Nevertheless, the algorithms fail in the case of poorly textured regions.

In case that optical flows are mostly induced by camera motion, i.e. the objects are stationary, the epipolar geometry can be used to obtain one more extra constraint for the flow of pixels. In this regard, the fundamental matrix related to the motion of the camera between two frames is firstly estimated; consequently, given each point in the first frame, the place of the correspondent point in the next frame will be constrained over a line known as the epipolar line. Fundamental matrices can be estimated using the 8-point [11] or the 7-point [12] methods. In [13], by searching over epipolar lines and using a semi-global block matching technique, the correspondent points are found. The method has two main shortcomings: block matching methods are slow and they use calibration information and also an approximation for the rotation matrix which is valid for rotation angles of less than 10° . In this paper, we introduce a differential method which works with uncalibrated cameras and does not make any assumptions concerning rotation matrices. Our contribution in this paper is to utilize epipolar constraint in a differential multi resolution scheme to gain much more accurate optical flow estimations even for low textured scenes.

The paper is organized as follows: in Sect. 2, the sparse optical flow calculation based on brightness consistency is reviewed, while the optical flow model is discussed in Sect. 3. The inclusion of the epipolar constraint in the calculation of optical flow is discussed in Sect. 4. Evaluation of the proposed method is conducted in Sect. 5. Section 6 concludes this paper.

2 Brightness Constancy Assumption (BCA)

Most sub-pixel optical flow estimation are based on the (linearized) brightness constraint, which assumes that the intensity of a pixel stays constant if objects or cameras move. Given two gray level images such as $I(x, y)$ and $I'(x, y)$, we are interested to map each pixel, namely (x, y) , in the image I to a pixel, namely (x', y') , in the image I' using a translation vector such as $[u \ v]^T$. The brightness constraint states that:

$$x'_i = x_i + u_i \quad (1)$$

$$y'_i = y_i + v_i \quad (2)$$

such that $I(x, y) = I'(x', y')$. In a differential optical flow calculation, the intensity consistency constraint can be approximated as follows:

$$I_x(x, y)u + I_y(x, y)v = -I_t(x, y) \quad (3)$$

where $I_x(x, y) = \frac{\partial I(x, y)}{\partial x}$, $I_y(x, y) = \frac{\partial I(x, y)}{\partial y}$ and $I_t(x, y) = I'(x, y) - I(x, y)$.

Nevertheless, the brightness constraint in Eq. 3 has one inherent problem: it yields only one constraint to solve for two variables. It is well known that such an under-determined equation system gives an infinite number of solutions. For every fixed u a valid v can be found fulfilling the constraint and vice versa.

3 Optical Flow Model

Lucas-Kanade method [14] assumed a constant (or affine) optical flow field in a small neighborhood of a pixel and calculate the flow based on the least square method. Such a neighborhood typically consists of $N = n \times n$ pixels with n smaller than 15. The brightness constraint is then evaluated with respect to all pixels within this neighborhood window N . The brightness constraint will usually not be perfectly fulfilled for all pixels as the assumption of equal flow vectors within the window might be violated.

Assuming that the optical flow changes smoothly in a neighborhood or remains constant in a small neighborhood, more equations in terms of u and v can be obtained. Nevertheless, it is also necessary that the intensities of pixels vary at least in two different directions, otherwise the equation system becomes singular and will have infinite solutions. The singularities occur in low textured images or for the pixels located at the edges (aperture problem). Typically, the equation system is formed with more than two equations and it should be solved using the least squared method. Consequently, weighting equations differently yields different solutions. Thus, given a set of neighbor pixels such as $\{(x_i, y_i) | i = 0 \dots n\}$, the equation system will be:

$$\begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & & 0 \\ \vdots & \dots & \ddots & 0 \\ 0 & & & w_N \end{bmatrix} \begin{bmatrix} I_x(x_1, y_1) & I_y(x_1, y_1) \\ I_x(x_2, y_2) & I_y(x_2, y_2) \\ \vdots \\ I_x(x_N, y_N) & I_y(x_N, y_N) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} w_1 & 0 & \dots & 0 \\ 0 & w_2 & & 0 \\ \vdots & \dots & \ddots & 0 \\ 0 & & & w_N \end{bmatrix} \begin{bmatrix} I_t(x_1, y_1) \\ I_t(x_2, y_2) \\ \vdots \\ I_t(x_N, y_N) \end{bmatrix} \quad (4)$$

where (x_0, y_0) is the center point for which the optical flow should be calculated and (w_1, w_2, \dots, w_N) are the weight of each equation. The weights are also assumed to be normalized: $\sum_{i=1}^N w_i = 1$.

The computed flow vector is sub-pixel accurate. But due to the Taylor approximation, the method is only valid for small displacement vectors. Larger displacements are found by embedding the method into a pyramid approach, solving for low frequency structures in low resolution images first and refining the search on higher resolved images. While the maximum track length depends on the image content, generally speaking, flow vectors with large displacements are less likely to be found than those within few pixel displacements.

4 Epipolar Constraint

As stated before, Eq. 4 can be singular in low textured scenes, i.e., all equations will be dependent. Nevertheless, if the optical flow field between two images is mostly induced by the camera motion, another constraint known as the epipolar constraint can be leveraged to calculate optical flow more robustly even for low textured images. Given two matched points: (x, y) and (x', y') between two images captured at two different camera positions, the following equation holds:

$$\mathbf{q}^T F \mathbf{p} = 0 \quad (5)$$

where $\mathbf{p} = [x \ y \ 1]^T$, $\mathbf{q} = [x' \ y' \ 1]^T$ and F is a 3×3 matrix known as the fundamental matrix. The fundamental matrix can be determined using the 8-point method [11] or the 7-point method [12]. As a result, given a point in the image I the location of the correspondent point in the image I' is constrained with a line equation (epipolar line). In this regard, Eq. 5 can be rewritten as follows:

$$ax'_i + by'_i + c = 0 \quad (6)$$

where $[a \ b \ c]^T = \frac{1}{\eta} F \mathbf{p}$, where η is a normalization factor such that $a^2 + b^2 = 1$. By substituting Eq. 2 in Eq. 6, an equation in terms of u and v is obtained:

$$au + bv = -ax - by - c \quad (7)$$

Equation 7 gives one more linear equation of u and v which can be inserted in Eq. 4:

$$\begin{bmatrix} w_1 I_x(x_0, y_0) & w_0 I_y(x_0, y_0) \\ \vdots & \vdots \\ w_n I_x(x_n, y_n) & w_n I_y(x_n, y_n) \\ w_{n+1} a & w_{n+1} b \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -w_1 I_t(x_0, y_0) \\ \vdots \\ -w_n I_t(x_n, y_n) \\ -w_{n+1} (ax_0 + by_0 + c) \end{bmatrix} \quad (8)$$

Since Eq. 2 is based on a linear approximation, usually the optical flow is calculated in an iterative context in which the optical flow is enhanced in each iteration. Therefore, the equation system should be reformed based on the enhanced variables, namely δu^k and δv^k . As a result, the optical flow components are iteratively modified as follows:

$$u^{k+1} = u^k + \delta u^k \tag{9}$$

$$v^{k+1} = v^k + \delta v^k \tag{10}$$

In this case, to guarantee that the matched point in the second image lies on the epipolar line, the following equation should hold in each iteration:

$$a(u^k + \delta u^k) + b(v^k + \delta v^k) = -ax_0 - by_0 - c \tag{11}$$

Thus

$$a\delta u^k + b\delta v^k = -a(x_0 + u^k) - b(y_0 + v^k) - c \tag{12}$$

Therefore, we obtain the following equation system in an iterative context:

$$\begin{bmatrix} w_1 I_x(x_0, y_0) & w_1 I_y(x_0, y_0) \\ \vdots & \vdots \\ w_n I_x(x_n, y_n) & w_n I_y(x_n, y_n) \\ w_{n+1} a & w_{n+1} b \end{bmatrix} \begin{bmatrix} \delta u^k \\ \delta v^k \end{bmatrix} = \begin{bmatrix} -w_1 I_t(x_0 + u^k, y_0 + v^k) \\ \vdots \\ -w_n I_t(x_n + u^k, y_n + v^k) \\ -w_{n+1}(a(x_0 + u^k) + b(y_0 + v^k) + c) \end{bmatrix} \tag{13}$$

So far we have focused on the calculation of small optical flow if the flows are about a few pixels. Obviously, in case of a large displacement optical flows, using the discussed method causes the iterations to get stuck in local minima. An efficient well-known solution to this problem is the pyramid analysis in which an optical flow is calculated step by step from coarse to fine levels. In this case, using the epipolar line is a bit tricky since the line equations should be changed depending on the level of the pyramid in which the optical flow is calculated. Assuming the level of the pyramid denoted as $l < 1$ and the scale factor of the pyramid to be s , it can be verified that in the epipolar line equation c will be changed as follows:

$$c_l = s^l c_0 \tag{14}$$

where $c_0 = c$.

The weight of the neighborhood pixels is chosen to follow a Gaussian distribution. The weight of the epipolar constraint determines the importance of the constraint. Experimentally, we found that the weight 1.5 gives the best results.

Figure 2 shows comparisons of the AEE and the AAE for each image from 194 training images and presents significant improvements of the accuracy for the

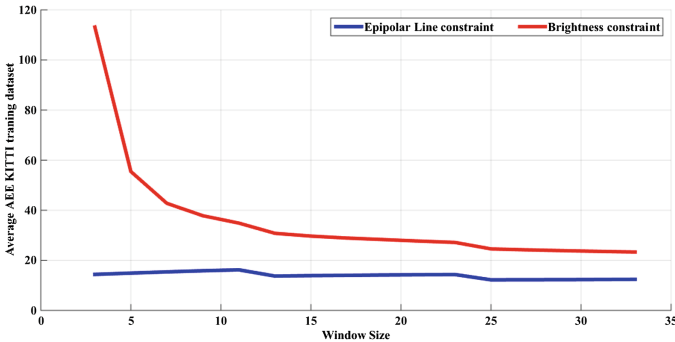
new data term as shown in Fig. 3. However, in some scenarios such as sequence 150, the average errors based on epipolar constraint were larger for the epipolar constraint. In such scenarios such as Fig. 4, the camera had obviously side translations and relatively high rotations which typically gave rise to high errors in the estimation of the fundamental matrix for even a small amount of measurement noise.

5 Experimental Results

For a quantitative evaluation, we tested our method on the well-known KITTI dataset [15]. The dataset contained images with a resolution of (1240×376) pixels. The KITTI dataset provided a very challenging testbed for the evaluation of optical flow algorithms. Pixel displacements in the data set are generally large, exceeding 250 pixels. Furthermore, the images exhibit less texture regions, strongly varying lighting conditions, and many non-Lambertian surfaces, especially translucent windows and specular glass, and metal surfaces. Moreover, the

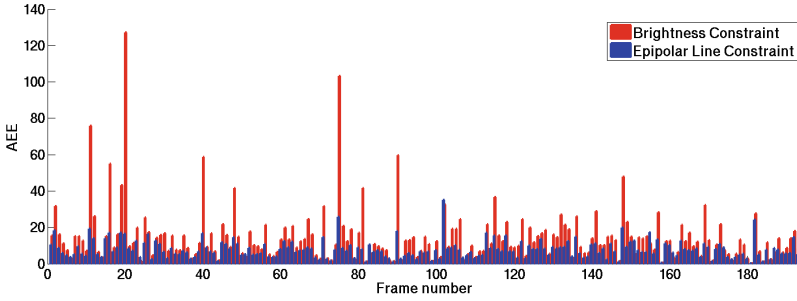
Table 1. Average end point and average angular error of the estimated optical flow for the 194 training sequences from KITTI data set.

Method		Average End Point Error AEE			Average Angular Error AAE		
		(3×3)	(7×7)	(15×15)	(3×3)	(7×7)	(15×15)
Epipolar Line Constraint	LK 7 Points	17.74	10.01	7.83	9.93	7.93	7.29
	LK 8 Points	19.11	11.55	9.33	13.42	11.24	10.47
	SIFT 7 Points	17.87	10.25	8.19	10.35	8.12	7.92
	SIFT 8 Points	18.65	11.15	9.01	12.43	10.51	9.84
Brightness	Constraint	127.85	29.69	14.56	36.88	17.23	11.07

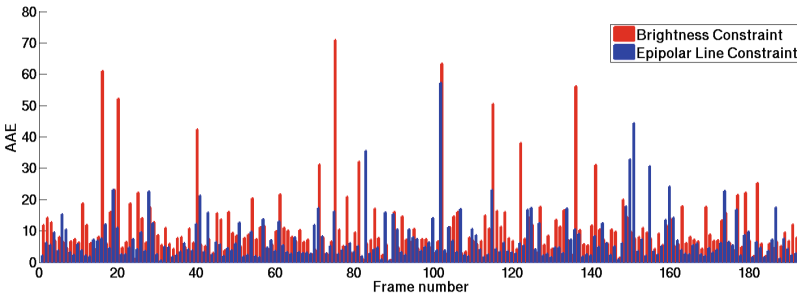


(a)

Fig. 1. The average error for each sequence of the 194 sequences of the KITTI training data set. (a) AEE. (b) AAE.



(a)



(b)

Fig. 2. The average error for each sequence of the 194 sequences of the KITTI training data set. (a) AEE. (b) AAE.

high speed of the forward motion creates large regions on the image boundaries that move out of the field of view between frames, such that no correspondence can be established.

For the estimation of optical flow, there are two concerns: which feature matching method should be applied and which method for the estimation of the fundamental matrix is more appropriate. Hence, we applied two different feature matching methods: SIFT [16] and the pyramid Lucas-Kanade optical flow [17] (both are implemented in openCV). Additionally, we applied 7-point and 8-point methods in the context of a RANSAC algorithm for the fundamental matrix estimation. For the 8-point method, we used 9 matched points and for the 7-point method we used 8 matched points. A problem concerning the 7-point method was that it might yield up to three distinct valid solutions, while only one solution was needed. As eight points were used in the 7-point method, we assumed that the solution should not deviate from the 8-point solution; therefore, we selected the solution which was generated based on the smallest root of the third order polynomial obtained in the 7-point method.

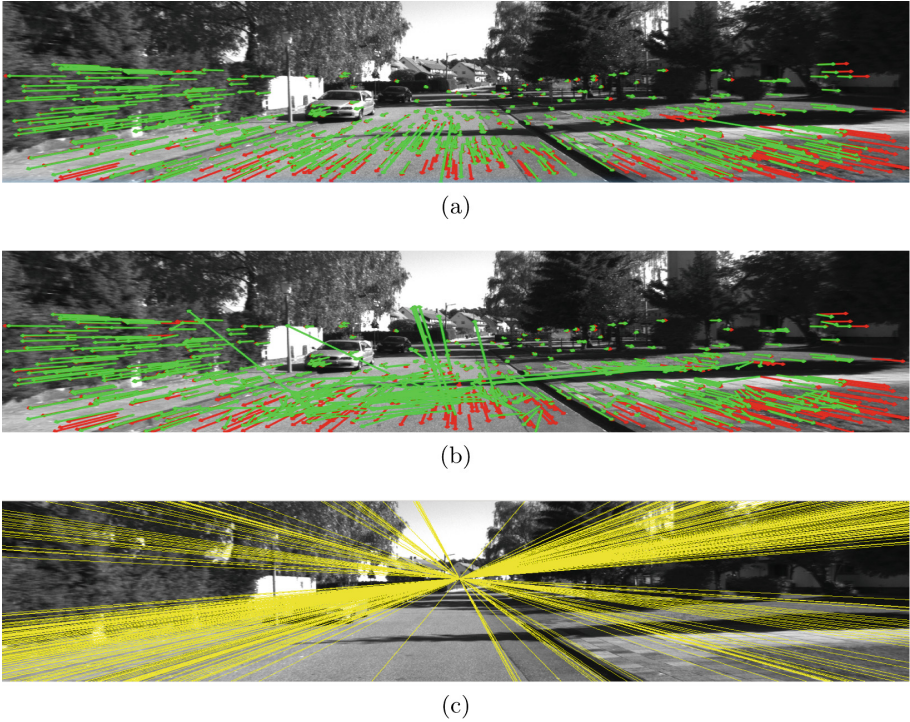
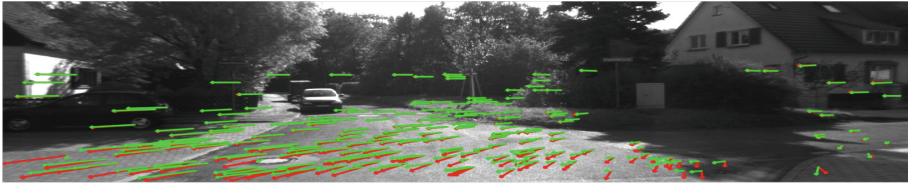
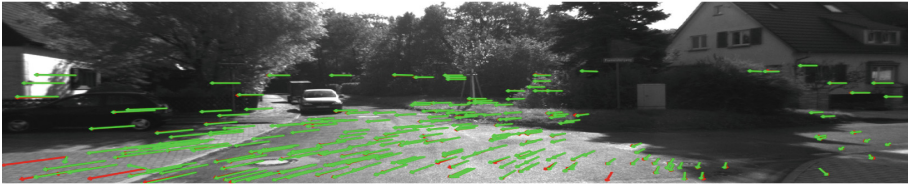


Fig. 3. Comparison between ground truth (red) and estimated optical flow (green) for sequence 24 of the KITTI training data set for some feature points. (a) Using epipolar constraint based on 7-points method. (b) Using brightness constrain. (c) Estimated epipolar lines (Color figure online).

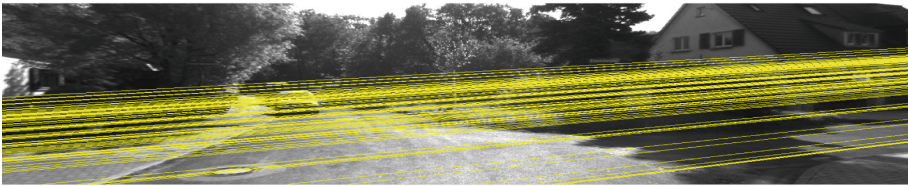
In the first experiment, we calculated the average end-point error (AEE) and the average angular error (AAE) of the estimated optical flow using the data term based on the epipolar constraint and the data term of brightness constraint based on [14]. Table 1 shows the average errors using different window sizes and demonstrates that the epipolar constraint has lower average errors for 194 training images. As it can be seen, the best performance was achieved using the LK tracker and the 7-point method. The reason lies in the fact that SIFT works based on blob features and these features are not sub-pixel or even pixel accurate enough. The 7-point method also performs better than the 8-point method as it needs less points, which make it robust against outliers in the RANSAC algorithm. In Fig. 1, the effect of increasing the size of surrounding window is depicted. As can be seen, brightness constraint has very poor performance for the small window sizes, whereas using the epipolar constraint yielded good results even for the small windows.



(a)



(b)



(c)

Fig. 4. Comparison between ground truth (red) and estimated optical flow (green) for sequence 150 of the KITTI training data set for some feature points. (a) Using epipolar constraint based on 7-points method. (b) Using brightness constrain. (c) Estimated epipolar lines (Color figure online).

6 Conclusion

We derived the necessary formulation to augment epipolar constraint for an uncalibrated camera in a differential method for the calculation of the optical flow. The proposed algorithm was evaluated with different sequences of the KITTI datasets and provided more correct flow fields and increased the robustness. For future work, applying this method to dense flow estimations should be considered.

References

1. Kim, Y.H., Martinez, A.M., Kak, A.C.: Robust motion estimation under varying illumination. *Image Vision Comput.* **23**, 365–375 (2005)
2. Mileva, Y., Bruhn, A., Weickert, J.: Illumination-robust variational optical flow with photometric invariants. In: Hamprecht, F.A., Schnörr, C., Jähne, B. (eds.) *DAGM 2007*. LNCS, vol. 4713, pp. 152–162. Springer, Heidelberg (2007)

3. Mueller, T., Rabe, C., Rannacher, J., Franke, U., Mester, R.: Illumination-robust dense optical flow using census signatures. In: Mester, R., Felsberg, M. (eds.) DAGM 2011. LNCS, vol. 6835, pp. 236–245. Springer, Heidelberg (2011)
4. Zach, C., Pock, T., Bischof, H.: A duality based approach for realtime TV- $L1$ optical flow. In: Hamprecht, F.A., Schnörr, C., Jähne, B. (eds.) DAGM 2007. LNCS, vol. 4713, pp. 214–223. Springer, Heidelberg (2007)
5. Molnar, J., Chetverikov, D., Fazekas, S.: Illumination-robust variational optical flow using cross-correlation. *Comput. Vis. Image Underst.* **114**, 1104–1114 (2010)
6. Werlberger, M., Pock, T., Bischof, H.: Motion estimation with non-local total variation regularization. In: CVPR, pp. 2464–2471. IEEE (2010)
7. Drulea, M., Nedeveschi, S.: Motion estimation using the correlation transform. *IEEE Trans. Image Process.* **22**, 3260–3270 (2013)
8. Rashwan, H.A., Mohamed, M.A., García, M.A., Mertsching, B., Puig, D.: Illumination robust optical flow model based on histogram of oriented gradients. In: Weickert, J., Hein, M., Schiele, B. (eds.) GCPR 2013. LNCS, vol. 8142, pp. 354–363. Springer, Heidelberg (2013)
9. Mohamed, M.A., Rashwan, H.A., Mertsching, B., Garcia, M.A., Puig, D.: Illumination-robust optical flow using local directional pattern. *IEEE Trans. Circuits Syst. Video Technol.* **24**, 1–9 (2014)
10. Mohamed, M.A., Rashwan, H.A., Mertsching, B., Garcia, M.A., Puig, D.: On improving the robustness of variational optical flow against illumination changes. In: Proceedings of the 4th ACM/IEEE International Workshop on Analysis and Retrieval of Tracked Events and Motion in Imagery Stream, pp. 1–8. ACM (2013)
11. Hartley, R.I.: In defense of the eight-point algorithm. *IEEE Trans. Pattern Anal. Mach. Intell.* **19**, 580–593 (1997)
12. Hartley, R.I., Zisserman, A.: *Multiple View Geometry in Computer Vision*. Cambridge University, Cambridge (2004)
13. Yamaguchi, K., McAllester, D.A., Urtasun, R.: Robust monocular epipolar flow estimation. In: CVPR, pp. 1862–1869. IEEE (2013)
14. Lukas, B., Kanade, T.: An iterative image registration technique with an application to stereo vision. In: *Image Understanding Workshop* (1981)
15. Geiger, A., Lenz, P., Stiller, C., Urtasun, R.: Vision meets robotics: the KITTI dataset. *Int. J. Robot. Res.* **32**(11), 1231–1237 (2013)
16. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vision* **60**(2), 91–110 (2004)
17. Bouguet, J.Y.: Pyramidal implementation of the Lucas Kanade feature tracker. Intel Corporation, Microprocessor Research Labs (2000)