ROBUST LOCAL OPTICAL FLOW: LONG-RANGE MOTIONS AND VARYING ILLUMINATIONS

Tobias Senst, Jonas Geistert and Thomas Sikora

Technische Universität Berlin
Communication Systems Group
EN1, Einsteinufer 17, 10587 Berlin, Germany

ABSTRACT

Sparse motion estimation with local optical flow methods is fundamental for a wide range of computer vision application. Classical approaches like the pyramidal Lucas-Kanade method (PLK) or more sophisticated approaches like the Robust Local Optical Flow (RLOF) fail when it comes to environments with illumination changes and/or long-range motions. In this work we focus on these limitations and propose a novel local optical flow framework taking into account an illumination model to deal with varying illumination and a prediction step based on a perspective global motion model to deal with long-range motions. Experimental results on the Middlebury, KITTI and Sintel optical flow benchmarks demonstrate the superior performance of the proposed framework.

Index Terms— Optical Flow, RLOF, KLT, Feature Tracking, Global Motion Model, Illumination Model

1. INTRODUCTION

35 years after Lucas and Kanade published a gradient based local optical flow method [1] known as LK method and 16 years after Bouguet proposed the pyramidal Lucas Kanade (PLK) [2] which is most of the time used nowadays when speaking about the Lucas-Kanade method e.g. for the KLT tracker, the LK and PLK method play a fundamental role in many video-based computer vision applications. The PLK method belongs to the class of local optical flow methods and is competing with a wide range of global optical flow approaches [22]. In general global optical flow methods generate highly accurate and dense motion vector fields due to a global spatial coherence [4, 3] which pose the problem of motion estimation as the optimization of a global energy functional on the whole image data. The main disadvantage of global methods is that they are computational expensive.

For many applications in the field of video-based surveillance, medical imaging, video coding, robot navigation, augmented reality and video classification dense motion fields are not required instead only a small set of motion vectors are needed. In most of these systems computational efficiency, i.e. the real-time requirement, is a crucial aspect. For these applications global methods can become insufficient due to their high run-time and high computational complexity that is similar for small and large motion vector sets. In contrast, local methods are based on a local spatial coherence constrain and take into account only data of a limited region the so-called support region. In [5] it has been shown that for sparse motion fields local methods are competitive to state-of-the-art global solutions. These methods are due to the support region linear scalable with respect to the size of the motion vector set, i.e. linearly correlation between the computational complexity and the number of motion vectors to be estimated. This is an outstanding property of in general local optical flow methods which make these methods favorable for many practical video-processing solutions and motivates us in our research about a fast and accurate robust local optical flow method.

Recent research on local optical flow methods has been focused on improving the run-time performance, e.g. by parallelization through GPU implementations [6, 7] or more efficient numerical schemes [8, 5] and on improving the precision of the motion estimates. Therefore in [9, 10, 11] robust estimation frameworks dealing with statistical outliers have been introduced. Other work addresses the generalized aperture problem, see Black and Anandan [12], and propose to modify the support region, e.g. by Gaussian and Laplacian of Gaussian weighting functions [13], adapting region sizes [11] or adapting region shapes [14].

In this paper, we focus on the problem of estimating long-range motions in environments with varying illuminations which still is a limitation for local optical flow methods. In order to improve the accuracy of estimates with large motion, we propose to initialize the local method by predictions obtained from a global motion model. Therefore, we propose a perspective global motion model which is estimated from a sparse set of motion vectors. The peculiarity in this case is that this model allows to predict long-range motion scales from small motion estimates. In the following we provide a general framework to apply the predictions of the global motion model in order to improve the accuracy of local op-

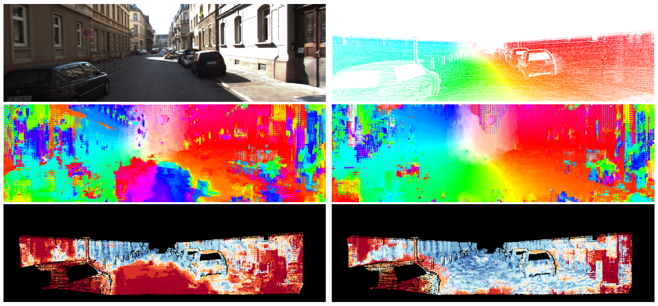

Fig. 1. Example from KITTI training set containing illumination changnes (sequence 15). Input and ground-truth (top), color-coded motion fields and error maps where dark blue denotes small and red large end-point error shows results of RLOF without (left) and with linear illumination model (right).

tical flow methods. An implementation will be given based on our previous work on Robust Local Optical Flow (RLOF) [11, 14]. The paper is organized as follows: In Section 2 we extend the baseline RLOF by implementing a linear illumination model proposed by [15] in order to handle scenes with varying illuminations. Section 3 describes a novel scheme which uses predictions form a perspective global motion model to initialize the local optical flow methods. Finally, the experimental results are given in Section 4.

2. LINEAR ILLUMINATION MODEL

The Intensity Constancy Assumption (ICA) [4] is the fundamental assumption for most optical flow methods although it almost never holds in real-life scenarios. For example in sequences that contain varying illuminations caused by shadows, moving light sources or changing weather conditions the ICA is violated. In the field of global optical flow several extension of the ICA have been proposed in order to deal with varying illuminations. For example Brox *et al.*[16] invented an additional gradient constancy constraint that is more robust to varying illuminations, Mileva *et al.*[17] combined Brox's approach with photometric invariances and in [18] a texture structure decomposition pre-processing step has been proposed. As stated by Kim *et al.*[10] for these methods parameters have to be tuned carefully in order to obtained desirable results.

In this work, we use an extended ICA based on a linear illumination model proposed by Gennert and Negahdaripour [15] which has been previously applied in [10] for local optical flow based on the Least Median of Squares (LMS) estimator. The ICA model proposed by Gennert and Negahdaripour is defined as follows:

$$I(\mathbf{x}, t) + m \cdot I(\mathbf{x}, t) + c = I(\mathbf{x} + \mathbf{d}, t + 1), \qquad (1)$$

with $I(\mathbf{x}, t)$ and $I(\mathbf{x} + \mathbf{d}, t + 1)$ being image intensity values of two consecutive images, $\mathbf{d}$ the motion vector and m and c the parameter of the illumination model at a position $\mathbf{x}$. Like ICA, Eq. (1) is solved by a first-order Taylor approximation. For the local optical flow the local motion constancy assumption has to be extended by assuming that the whole parameter vector $[\mathbf{d}\ m\ c]^T$ is constant in a small support region Ω. Let $\nabla I(\mathbf{x})$ be the spatial and $I_t(\mathbf{x})$ the temporal image gradient, $w(\mathbf{x})$ a weighting function and ρ a norm, then the resulting parameter vector is a vector that minimizes:

$$\min_{[\mathbf{d}\ m\ c]} \sum_{\mathbf{x}\in\Omega} w(\mathbf{x}) \cdot \rho \left(T(\mathbf{x}) \cdot \begin{bmatrix} \mathbf{d} \\ m \\ c \end{bmatrix} - I_t(\mathbf{x}), \boldsymbol{\sigma} \right) \qquad (2)$$

with $T(\mathbf{x}) = [\nabla I(\mathbf{x})\ \ -I(\mathbf{x})\ \ -1]^T$. This equation defines an M-Estimator solution which is derived from the generalized gradient-based optical flow equation [11]. If $\rho(x) = x^2$, $w(\mathbf{x}) = 1$ and $m = c = 0$, Eq. (2) denotes the Lucas Kanade formulation [1].

Please note that the following derivations are phrased in general terms to be applicable for a variety of local optical flow methods. For this paper we apply these derivations for the CB-RLOF where the weighting function $w(\mathbf{x})$ is used to implement the support region shape that correspond to the underlying color segment as proposed in [14]. To cope with small linearization errors, an iterative Newton-Raphson fashioned scheme proposed by Bouguet [2] for the Lucas Kanade method and described in [11] for the RLOF is applied to the solution of Eq. (2). Starting from the initial values $[\mathbf{d}_0\ m_0\ c_0]^T$, this scheme updates the parameters iteratively with:

$$[\mathbf{d}_{i+1}\ m_{i+1}\ c_{i+1}]^T = [\mathbf{d}_i\ m_i\ c_i]^T + [\Delta\mathbf{d}_i\ \Delta m_i\ \Delta c_i]^T. \qquad (3)$$

until convergence or a given maximal number of iteration has been reached. If the so called influence function $\psi = \dot{\rho}$ is a composite of linear functions, which holds for the RLOF and the PLK then Eq. (2) can be solved directly so that the new incremental motion and illumination parameters are given by:

$$\begin{aligned} [\Delta\mathbf{d}_i\ \Delta m_i\ \Delta c_i]^T &= \mathbf{G}_{IM}^{-1} \cdot \mathbf{b}_{IM,i} \\ \mathbf{b}_{IM,i} &= \sum_{\mathbf{x}\in\Omega} w(\mathbf{x}) \cdot T(\mathbf{x}) \cdot \psi\left(I_{t,i}(\mathbf{x})\right) \\ \mathbf{G}_{IM} &= \sum_{\mathbf{x}\in\Omega} w(\mathbf{x}) \cdot T(\mathbf{x}) \cdot \psi\left(T(\mathbf{x})^T\right) \end{aligned} \qquad (4)$$

with the iterative update of the temporal gradient:

$$I_{t,i}(\mathbf{x}) = I(\mathbf{x} + \mathbf{d}_i, t + 1) - I(\mathbf{x}, t) \cdot (1 - m_i) + c_i. \qquad (5)$$

In this work we will implement this scheme for the RLOF [14] by apply ρ being the shirnked Hampel norm. As stated in [11] the advantage of using the shrinked Hampel norm and the M-estimator is that the estimate is robust and less computational complex than the LMS proposed in [10]. Figure 1 gives an example of the improvements achieved for the linear illumination model in contrast to standard ICA.

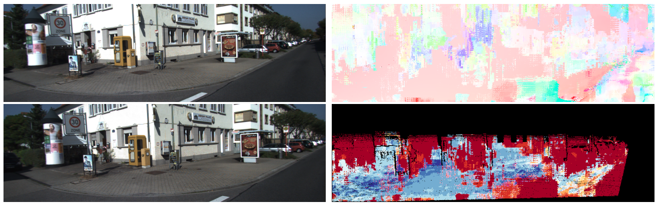

Fig. 2. Example from KITTI training set containing large displacements (sequence 147). Input images (left) color-coded motion field of the baseline RLOF (top right) and the corresponding error map (bottom right) where dark blue denotes small and red large end-point errors.

3. GLOBAL MOTION ESTIMATION

A major challenge for local optical flow methods is to deal with nonlinearities of the ICA or respectively the nonlinearities of the linear illumination model. Accurate motion estimates can be only achieved if the linearization by the first-order Taylor approximation is reasonable for the data. In real-world this holds on average for rather small displacements around a few pixels. As a result there is a subsequent difficulty in dealing with long-range motion.

The standard approach to deal with long-range motion is to embed the motion estimation in a coarse-to-fine scheme [12]. Therefore a pyramid of repeatedly low-pass filtered and down-sampled images is build. The optical flow estimation is started at the coarsest level and its result is used to initialize the iterative refinement, see Eq. (3) and Eq. (4), of the next level etc. until the finest pyramid level. This technique is effective for the PLK and RLOF but a drawback is that errors of the intermediate results are multiplied during the up-scaling to the next level. Erroneous intermediate results occur on the top level due to various reasons, e.g. image details can get lost and more homogenous areas can appear due to the repeated low-pass filtering which results into wrong estimates, see aperture problem. As a result the motion estimation fails, which is shown in Figure 2. The figure shows the result of the RLOF for sequence 147 of the KITTI benchmark [19]. The visualization of the error map has been provided by the KITTI website. Dark blue denotes locations with small end-point errors and dark red with large errors.

The idea of this paper is to use an initial guess predicted from the global motion of the scene instead of $\bar{0}$ to better deal with long-range motions. If a video sequence contains camera motion there can be found a parametric transform of the image that correspond to this global motion. The process of estimating the transform parameters is called global motion estimation (GME) and should not be confused with global optical flow estimation. There exists various practical 2-D parametric global motion models, a detailed evaluation can be found in [20]. The most general is the eight-parameter perspective model which can be derived from 3-D affine motions of objects under a perspective camera model. The eight perspective parameters $\mathbf{m} = [m_0, \ldots, m_7]$ can be estimated with at least four motion vectors. If the transform parameters $\mathbf{m}$ are known this models can be used to predict an optical flow field by the following function:

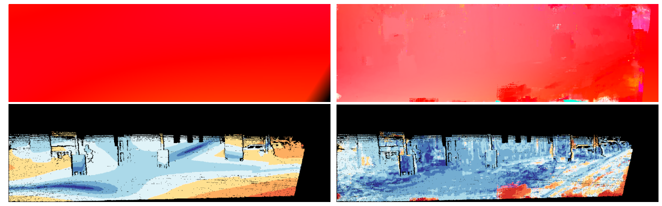

Fig. 3. Color-coded motion fields and error maps for sequence 147 of a motion field predicted with GME (left) and RLOF using the predicted field (right).

$$\mathbf{d}^{GM}(x, y) = f^{GM}(\mathbf{x}, \mathbf{m}) = \begin{bmatrix} \frac{m_0 \cdot x + m_1 \cdot y + m_2}{m_6 \cdot x + m_7 \cdot y + 1} - x \\ \frac{m_3 \cdot x + m_4 \cdot y + m_5}{m_6 \cdot x + m_7 \cdot y + 1} - y \end{bmatrix} \quad (6)$$

where $\mathbf{d}^{GM}(x, y)$ denotes a predicted motion vector at a position $\mathbf{x} = (x, y)$. The left row in Figure 3 shows an example of a predicted dense motion field using GME. Especially for scenes with a global motion that represent large and small motion components e.g. zooming, this approach can be used to predict long-range motion from a sparse set of short-range motion estimates. The predicted optical flow field can now be used to initialize the pyramidal iterative scheme on the coarsest level, i.e. $[\mathbf{d}_0 \ m_0 c_0] = [\mathbf{d}^{GM} \ 0 \ 0]$. The resulting motion field is shown in the right row of Figure 3. This examples show a tremendous improvement for the motion estimation for scenes with long-range motions.

To apply the GME prediction the following steps have to be performed. The goal is to estimate a set $S = \{\mathbf{d}_0, \ldots, \mathbf{d}_{n-1}\}$ of n motion vectors at defined locations $P = \{\mathbf{x}_0, \ldots, \mathbf{x}_{n-1}\}$ with a motion estimation function $f : P \rightarrow S$. In a first step a small set S^R of motion vectors located at a regular grid will be estimated with the PLK method. Note that PLK and RLOF integrate the linear illumination model. For each vector a forward-backward confidence [11] will be estimated. Erroneous vectors are than removed if the confidence distance is above one pixel. In the second step perspective global model parameters $\mathbf{m}$ are estimated using S_R and the well-known RANSAC [21]. Within these parameters a set of predicted motion vectors $S^{GM} = \{\mathbf{d}_0^{GM}, \ldots, \mathbf{d}_{n-1}^{GM}\}$ will be estimated following $f^{GM} : P \rightarrow S^{GM}$. The final motion vector set S is then computed with the RLOF and $[\mathbf{d}_0 \ m_0 c_0] = [\mathbf{d}^{GM} \ 0 \ 0]$ the initial guess values on the coarsest level.

4. EXPERIMENTAL RESULTS

In this section we evaluate the proposed global motion estimation (GME) and linear illumination model (IM) for local

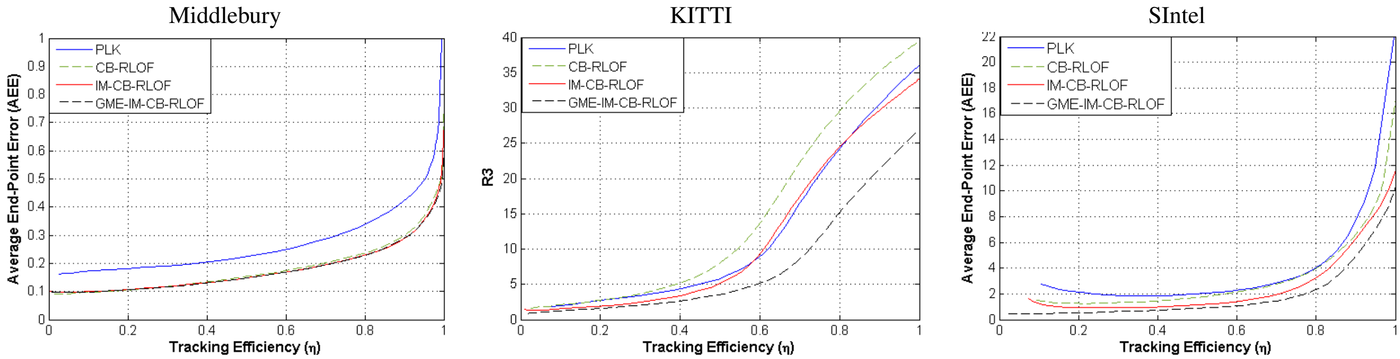

Fig. 4. Evaluation results on training sequences of Middlebury [22], KITTI [19] and Sintel [23] optical flow benchmarks. The plots show the average end-point error and for the KITTI dataset the R3 measure over η, the percentage of successfully estimated motion vectors, validated with forward-backward confidence.

	Middlebury		KITTI		SIntel	
	$\eta_{0.5}$	η_1	$\eta_{0.5}$	η_1	$\eta_{0.5}$	η_1
PLK	0.300	1.269	0.195	23.123	12.541	22.496
CB-RLOF	0.264	0.816	0.187	21.362	9.742	17.226
IM-CB-RLOF	0.232	0.706	**0.158**	19.729	3.081	11.640
GME-IM-CB-RLOF	**0.230**	**0.611**	0.158	**9.426**	**2.343**	**10.374**

Table 1. Average end-point error results of dense motion fields η_1 and sparse motion field with 50% outlier rejection $\eta_{0.5}$ for the training sequences of recent optical flow datasets.

optical flow methods. The proposed extensions have been integrated into the existing cross based Robust Local Optical Flow (CB-RLOF)[1] method [14]. Beside the CB-RLOF, the PLK[2] will be the benchmark for this evaluation. The experiments have been performed on the training sequences of the Middlebury [22], KITTI [19] and Sintel [23] optical flow datasets. For each method we use the same basic configurations, i.e. 3 pyramid levels, $\Omega = 19 \times 19$ support region size, 30 maximal number of iterations. For the cross based RLOF the minimal support region size is set to 9×9 and the color threshold $\tau = 35$.

The main field of application will be the estimation of sparse motion vector sets. Therefore we perform this evaluation based on a simple feature tracking framework which includes a motion vector validation (outlier filering) step to remove erroneous estimates. Following [11, 5, 14] forward-backward confidence is applied to validate the motion estimates.

Figure 4 shows the averaged tracking performance plot [11] for each of the datasets. This plot shows the AEE and for the KITTI the R3 measure of a subset of motion vectors validated by the forward-backward confidence measure where η denotes the percentage of not removed vectors, low AEE and high η are preferable. Table 1 shows the numerical results by the average end-point error for the three datasets. The table shows the mean AEE computed for the dense motion field (η_1) and the sparse motion fields. The sparse motion field consist of 50% of the best motion vectors evaluated with the forward-backward confidence ($\eta_{0.5}$). The results show that for sparse and dense motion vector fields there are tremendous improvements in terms of accuracy for the proposed GME-IM-CB-RLOF compared to the PLK and the baseline CB-RLOF. Especially for the challenging KITTI and Sintel datasets the improvements are significant and the results of the outlier filtered sparse field are competitive with state-of-the-art global methods [19, 23]. Our final approach requires 8 seconds for the dense field for Sintel's Cave 2 sequence (1024×436). The test are running on a Intel i7 3.40 GHz CPU. However the field of application will be the estimation of sparse motion. For a sparse motion vector set of 1.000 we achieved 0.3 seconds and for 10.000 motion vectors 0.5 seconds where 0.26 seconds are spend to estimate the GME.

[1] available at http://www.nue.tu-berlin.de/menue/forschung/projekte/rlof/
[2] available at http://www.opencv.org (v.2.4.9)

5. CONCLUSION AND FUTURE WORK

In this paper we propose two approaches to improve the accuracy of local optical flow methods. At first we implement an illumination model to deal with environments containing varying illuminations. Then we utilize predictions estimated with a perspective global motion model in order to initialize the iterative motion estimation scheme and deal with long-range motions. We have implemented both approaches for the RLOF method and demonstrated with e.g. an about 56% lower error for dense motion fields on the KITTI dataset and an about 76% lower error for sparse motion fields on the Sintel dataset, a significant improvement compared to state-of-the-art local methods.

It has been shown that the proposed local method and an outlier filtering step generates highly accurate sparse motion fields. In the future will further improve the run-time of the proposed approach e.g. by integrating the semi-direct iterative scheme [5] and provide an implementation based on the GPU.

6. REFERENCES

[1] Bruce D. Lucas and Takeo Kanade, "An iterative image registration technique with an application to stereo vision," in *International Joint Conference on Artificial Intelligence*, 1981, pp. 674–679.

[2] Jean-Yves Bouguet, "Pyramidal implementation of the lucas kanade feature tracker," Technical report, Intel Corporation Microprocessor Research Lab, 2000.

[3] Denis Fortun, Patrick Bouthemy, and Charles Kervrann, "Optical flow modeling and computation: A survey," *Computer Vision and Image Understanding*, vol. 134, pp. 1–21, 2015.

[4] Berthold K. P.Horn and Brian. G. Schunck, "Determining optical flow," *Artifical Intelligence*, vol. 17, pp. 185–203, 1981.

[5] Tobias Senst, Jonas Geistert, Ivo Keller, and Thomas Sikora, "Robust local optical flow estimation using bilinear equations for sparse motion estimation," in *20th IEEE International Conference on Image Processing*, Melbourne, Australia, Sept. 2013, pp. 2499–2503.

[6] Sudipta N. Sinha, Jan-Michael Frahm, Marc Pollefeys, and Yakup Genc, "Gpu-based video feature tracking and matching," Technical report 06-012, UNC Chapel Hill, 2006.

[7] Jan-Michael Frahm Christopher Zach, David Gallup, "Fast Gain-Adaptive KLT Tracking on the GPU," in *Computer Vision and Pattern Recognition Workshops (CVPRW 2008)*, 2008, pp. 1–7.

[8] A. Rav-Acha and S. Peleg, "Lucas-kanade without iterative warping," in *International Conference on Image Processing*, 2006, pp. 1097 –1100.

[9] J.M. Odobez and P. Bouthemy, "Robust multiresolution estimation of parametric motion models," *Journal of Visual Communication and Image Representation*, vol. 6, no. 4, pp. 348–365, December 1995.

[10] Y.H. Kim, A.M. Martinez, and A.C. Kak, "A local approach for robust optical flow estimation under varying illumination," in *British Machine Vision Conference*, 2004.

[11] Tobias Senst, Volker Eiselein, and Thomas Sikora, "Robust Local Optical Flow for Feature Tracking," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 22, no. 9, pp. 1377–1387, 2012.

[12] Michael J. Black and P. Anandan, "The robust estimation of multiple motion: parametric and piecewise-smooth flow fields," *Computer Vision and Image Understanding*, vol. 63, no. 1, pp. 75–104, 1996.

[13] Meghna Singh, Mrinal K. Mandal, and Anup Basu, "Gaussian and laplacian of gaussian weighting functions for robust feature based tracking," *Pattern Recognition Letters*, vol. 26, no. 13, pp. 1995–2005, 2005.

[14] Tobias Senst, Thilo Borgmann, and Thomas Sikora, "Cross based Robust Local Optical Flow," in *IEEE International Conference on Image Processing (ICIP)*, 2014, pp. 1967–1971.

[15] Michael A Gennert and Shahriar Negahdaripour, "Relaxing the brightness constancy assumption in computing optical flow," Tech. Rep., MIT AI Lab. Memo No.975, 1987.

[16] Thomas Brox, Andrés Bruhn, Nils Papenberg, and Joachim Weickert, "High accuracy optical flow estimation based on a theory for warping," in *European Conference on Computer Vision*, 2004, pp. 25–36.

[17] Yana Mileva, Andrés Bruhn, and Joachim Weickert, "Illumination-Robust Variational Optical Flow with Photometric Invariants," in *Annual Symposium of the German Association for Pattern Recognition (Dagm 2007)*, 2007, pp. 152–162.

[18] A Wedel, T Pock, C Zach, D Cremers, and H Bischof, "An Improved Algorithm for TV-L1 Optical Flow," in *Proceedings of the Dagstuhl Motion Workshop*. Sept. 2008, LNCS,, Springer.

[19] Andreas Geiger, Philip Lenz, and Raquel Urtasun, "Are we ready for autonomous driving? the kitti vision benchmark suite," in *Computer Vision and Pattern Recognition*, 2012, pp. 3354–3361.

[20] Yeping Su, Ming-Ting Sun, and V. Hsu, "Global motion estimation from coarsely sampled motion vector field and the applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 15, no. 2, pp. 232–242, 2005.

[21] Martin A Fischler and Robert C Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.

[22] Simon Baker, Daniel Scharstein, J P Lewis, Stefan Roth, Michael J Black, and Richard Szeliski, "A Database and Evaluation Methodology for Optical Flow," Technical report {MSR-TR-2009-179}, Microsoft Research, 2009.

[23] Daniel J Butler, Jonas Wulff, Garrett B Stanley, and Michael J Black, "A naturalistic open source movie for optical flow evaluation," in *European Conf. on Computer Vision*, 2012, pp. 611–625.