Feature- and Depth-Supported Modified Total Variation Optical Flow for 3D Motion Field Estimation in Real Scenes

Thomas Müller Jens Rannacher Clemens Rabe
Uwe Franke
Daimler Research, Sindelfingen, Germany
{thomas.an.mueller, jens.rannacher, clemens.rabe, uwe.franke}@daimler.com

Abstract

We propose and evaluate improvements in motion field estimation in order to cope with challenges in real world scenarios. To build a real-time stereo-based three-dimensional vision system which is able to handle illumination changes, textureless regions and fast moving objects observed by a moving platform, we introduce a new approach to support the variational optical flow computation scheme with stereo and feature information. The improved flow result is then used as input for a temporal integrated robust three-dimensional motion field estimation technique. We evaluate the results of our optical flow algorithm and the resulting three-dimensional motion field against approaches known from literature. Tests on both synthetic realistic and real stereo sequences show that our approach is superior to approaches known from literature with respect to density, accuracy and robustness.

1. Introduction

Real-time dense motion estimation from image sequences is a major technical challenge and has a wide range of applications, e.g. in the field of surveillance or in automotive driver assistance systems. Motion field estimation is directly related to optical flow, which is a fundamental technique in computer vision and was explored during the last three decades. In combination with real-time dense stereo approaches invented in recent years, dense three-dimensional motion information is now available [10].

However, optical flow estimation is still a major challenge for the computer vision community. Especially when dealing with real scenarios under non-ideal illumination conditions, when fast moving objects occur or the observer itself is moving (even in the case when the movement is roughly known from inertial sensors), real-time capable techniques for estimating dense flow fields are still a topic of research. Therefore, we propose an efficient way to support the state-of-the-art optical flow estimation technique by depth and feature information, having a grey-value stereo camera pair and inertial sensors available. The optical flow results presented in this paper are well-suited to be applied for robust three-dimensional motion field estimation in these non-ideal circumstances mentioned above, which is evaluated in the second part of this paper.

1.1. Outline

In the first section, we give a short overview of the state-of-the-art optical flow computation techniques with special regard to methods which handle challenges in real, non-ideal scenes.

In the central section of this paper, we introduce a modified total variation (MTV) regularization for the dense flow computation scheme. We describe how the MTV functional allows for the inclusion of stereo and featu
spondance information. We show how this inclusi
problems with weakly textured and fast moving i
gions in a very efficient way. We also show how
provements in optical flow estimation lead to robus
tion fields, using the framework presented in [10].

Dense, robust and accurate motion field estimation from stereo image sequences in

In the evaluation part, we compare our resulting optical flow fields to flow fields received from the TV-L^1 technique known from [17]. On synthetic sequences with known ground truth motion information, we show that our stereo- and feature-supported method outperforms this approach when we consider scenes with large displacements and weak-textured regions. We also demonstrate visually convincing results on typical real traffic scenes.

The last section concludes the paper by summarizing the achieved results and giving a short outlook on unresolved problems and further research.

1.2. Variational Optical Flow Estimation

Today's common approaches for computing the optical flow $\mathbf{u} \equiv (u_0, u_1)^\top : \Omega \to \mathbb{R}^2$, that is the two-dimensional apparent motion field between two images I_0 and I_1 on the

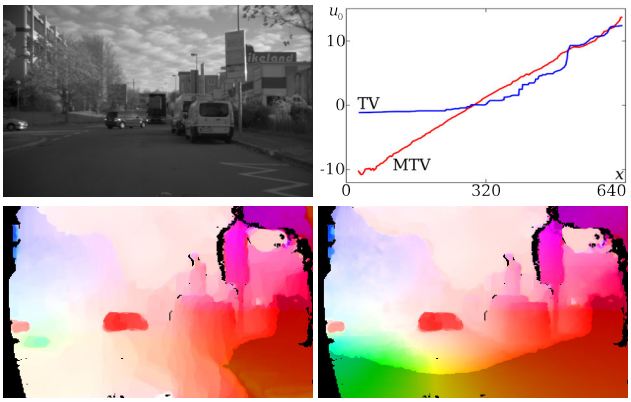

Figure 1. Top left: typical traffic scene. The observer is moving towards a turning car. Top Right: lateral optical flow component from the bottom part of the image, computed by the standard (TV, blue) and the modified total variational (MTV, red) approach. Bottom left: optical flow result of a typical traffic scene using standard total variation regularization. On the ground at the bottom of the image, the staircase-like structure is visible. Bottom right: optical flow result computed by the new MTV approach. The color encodes the direction, the saturation the magnitude of the optical flow. The discontinuity in the color gradient indicates a flow magnitude of $10\,\mathrm{px/frame}$. Stereo occlusions are colored in black.

image domain $\Omega \subset \mathbb{R}^2$, are based upon the classical framework established by Horn and Schunck [6]. In their work, a linearized version of the global energy functional

$$E_{\mathrm{HS}}[\mathbf{u}] = \sum_{\mathbf{x}\in\Omega} \left\{ \rho^2(\mathbf{x}, \mathbf{u}(\mathbf{x})) + \alpha^2 [\nabla \mathbf{u}(\mathbf{x})]^2 \right\} \qquad (1)$$

with the grey value residual $\rho(\mathbf{x}, \mathbf{u}) \equiv I_1(\mathbf{x}+\mathbf{u}) - I_0(\mathbf{x})$, the discretized gradient operator ∇ and a weighting parameter α^2 is minimized to receive the desired flow field $\mathbf{u}$. Modern approaches use the modified outlier-robust and flow-field-edge-preserving model by Zach and Pock [20],

$$E_{\mathrm{ZP}}[\mathbf{u}] = \int_\Omega \left\{ \lambda \left| \rho(\mathbf{x}, \mathbf{u}(\mathbf{x})) \right| + \sum_{i=0}^{1} |\nabla u_i(\mathbf{x})| \right\} \mathrm{d}\mathbf{x}\,, \qquad (2)$$

where $\lambda \in \mathbb{R}^+$ is a weighting parameter. The sum in (1) is replaced by an integral in (2) to emphasize the continuous, variational view on the problem.

1.3. Challenges in Real Scenes

When we try to use common variational techniques in real world applications, we encounter mainly three problems:

- Different illumination conditions in consecutive images: the grey value constancy constraint $\rho(\mathbf{x}, \mathbf{u})$ is directly affected by global or local changes in the grey value, caused by natural illumination changes like shadows or reflections, or by sensor noise or camera adjustments whose details are not known to the observer.

- Textureless regions: the regularizing smoothness term in Eq. (2) leads to flow results with preferably piecewise constant flow fields on the image domain. Especially in real, non-artificial scenarios with three-dimensional rigid motion (e.g. traffic scenes), piecewise constant image flow does not correspond to the desired solution. For example, the correct motion field on the ground (e.g. a street) perceived by a moving observer is by no means constant in the image; if the observer is rotating, it is not even linear. However, since the street is weakly textured, the flow result from Eq. (2) is a staircase-like structure, which can be observed in Fig. 1 on the left hand side.

- Large displacements: since generally the image functions $I_{\{0,1\}}$ are non-convex, the commonly used real-time iterative computation schemes lead to only locally optimal solutions of the flow field $\mathbf{u}$, which can at most differ from an initial flow field $\mathbf{u}_0$ by the size of the image details. Hence, when using the common initialization with the flow field $\mathbf{u}_0(\mathbf{x}) = \mathbf{0}\,\forall \mathbf{x} \in \Omega$, only small displacements can be estimated. However, when we e.g. consider moving cars in a common traffic scene at a usual frame rate, it is clear that special attention on large displacements is absolutely necessary to receive the correct optical flow.

1.4. Related Work

In recent years, several approaches have been proposed to overcome the typical problems mentioned in the section above. Since textureless regions and large displacements are related to our new method, we give a short overview about methods addressing these two problems.

1.4.1 Textureless Regions

To avoid total-variation-induced staircase effects, especially on weakly textured regions, the authors of [15] propose a regularization term which penalizes second derivatives. This leads to piecewise affine flow fields. While this works very well in some cases, it does not model general non-affine optical flow field patterns caused by the perspective projection of real three-dimensional rigid object motions the correct way. The reduction of the solution space of the optical flow field to typical flow patterns has been presented in [12], which has some relations to our approach. The authors propose to learn typical flow subspaces to estimate dense optical flow and to recover the camera motion.

We will re-address the extension of the model of Zach and Pock, Eq. (2), and handling of textureless regions in Section 2.4. There, we propose our solution for overcoming

the staircase effects (e.g. the flow field on the ground) in the stereo case when the observer itself is moving and this motion is roughly known by inertial sensors or by an image based ego-motion estimation technique.

1.4.2 Large Displacements

The standard approach for large displacement support is the implementation of a pyramid scheme, where input images are downsampled and optical flow is first computed on lower resolution images to establish the initialization for higher resolution pyramid levels. However, this technique is limited when fine structures become invisible at lower resolutions.

In [2], the authors proposed an extended model where differences between flow vectors of the flow field $\mathbf{u}$ and vectors computed by feature correspondences, $\mathbf{u}^* : \Omega \rightarrow \mathbb{R}^2$ (exemplary computed using SIFT [9]), are penalized in a variational scheme similar to

$$E_{\mathrm{B}}[\mathbf{u}] = E_{\mathrm{ZP}}[\mathbf{u}] + \lambda^* \int_{\Omega} m^*(\mathbf{x}) \left|\mathbf{u}(\mathbf{x}) - \mathbf{u}^*(\mathbf{x})\right| \mathrm{d}\mathbf{x}, \tag{3}$$

where $m^* : \Omega \rightarrow \{0, 1\}$ indicates the feature correspondences. Since information of the sparse feature-based flow field $\mathbf{u}^*$ is only local, a pyramid scheme is applied where at lower resolution levels the field $\mathbf{u}^*$ has a higher density and therefore has more influence on the resulting flow field $\mathbf{u}$. Within this approach, the feature correspondences themselves are not part of the variational optimization scheme, but are fixed during the iterations. This leads to a solution $\mathbf{u}$ which cannot fully correct wrong correspondences $\mathbf{u}^*$. In the case of an incorrect correspondence, the solution for $\mathbf{u}$ at this pixel will be a flow value somewhere between the one voted for by the grey value constancy constraint and the flow value from the feature correspondences.

Furthermore, Eq. (3) modifies the optical flow model itself. Please note, however, that for large displacement optical flow estimation, one does not need to modify the Zach-Pock model in Eq. (2), but has to find a way to solve it correctly, that means finding the global minumum rather than get stuck in local minima instead. We will come back to large displacement optical flow computation in Section 2.5.

1.5. Three-dimensional Motion Fields

Three-dimensional motion field estimation is a major task in computer vision. The direct differential straightforward approach of combining optical flow e.g. on the left part of a stereo image sequence with the disparity images (for example computed by semi-global matching, SGM [4]) usually leads to insufficient results due to, in practice, noisy disparity measurements. Variational approaches [16, 18] which try to estimate a three-dimensional motion field by a joint optimization over two stereo image pairs lead to far better results, but are either to expensive with respect to the computation time or simply not accurate enough. In [10], a Kalman filter based fusion of stereo and optical flow leads to the best results known so far. In the evaluation part, Section 3, we will take our improved optical flow results as input for this method and compare our results to the results of [10], where the optical flow technique from [17] based on Eq. (2) is used.

2. Modified Total Variation Regularization Optical Flow

2.1. Shortcomings of Total Variation Regularization

To preserve sharp flow edges, in Eq. (2) from [20], the total variation (TV) functional

$$\begin{aligned} \mathrm{TV} : \mathcal{C}^1(\Omega, \mathbb{R}^n) &\rightarrow \mathbb{R}^+, \quad n \in \mathbb{N} \\ u &\mapsto \int_{\Omega} |\nabla u(\mathbf{x})| \, \mathrm{d}\mathbf{x} \end{aligned} \tag{4}$$

is used to regularize the flow during the computation, where $\mathcal{C}^1(\mathbb{D}, \mathbb{W})$ is the space of differentiable functions $f : \mathbb{D} \rightarrow \mathbb{W}$. For a more general definition of the total variaion, we refer to [3]. The TV functional is convex, so a global minimum exists. However, *strict* convexity is not given, which results in a set of global minima rather than in one, and therefore in multiple optimal solutions to problems which involve the TV regularization. The left part of Fig. 2 illustrates this well-known property of the TV functional.

When computing optical flow in real scenes, we frequently have to cope with large regions with weak texture information (e.g. road surface due to motion blur). In these cases, the TV regularization fills in values from outside of this region, where the image data is discriminative and optical flow can be computed very robustly. Due to the abovementioned non-strict convexity of the TV functional, the result on textureless regions can be any flow field containing values between the flow values from the border of this region. So in these cases, the TV regularization usually leads to some staircase-like piecewise constant flow field, which is in general not the desired solution.

2.2. Support by Stereo and Ego-Motion Information

Please remind for a moment that the minimal energy representation of the optical flow model in Eq. (2) has its origin in the maximum a-posteriori probability problem for finding $\mathbf{u}$, given the data I_0 and I_1 and a prior model, which favors piecewise-constant flow fields. May this quite weak prior model be reasonable in the general case, in technical systems, however, where optical flow is applied for three-dimensional motion field estimation using stereo, much more information is known and should be used for regularization.

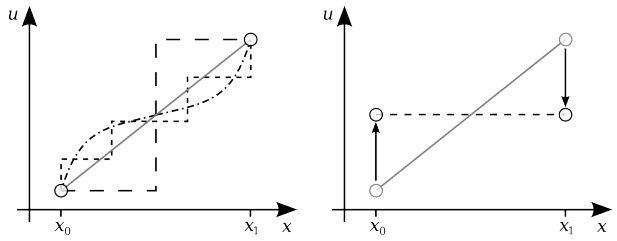

Figure 2. Left: Total Variation. Between strong data at the positions x_0 and x_1, every shown path (dashed) has the same total variation. The grey path corresponds to a flow field component u_0 on the ground while the observer is moving in forward direction. Due to the weak texture on the street, every shown path can be the result of a minimization process containing TV regularization. Right: Modified Total Variation. The compensation with the predicted shape of the flow field before the actual TV reduces the space of possible solutions. MTV smoothing leads to flow fields without staircasing artefacts, since between the strong data at x_0 and x_1, the direct path has now the uniquely smallest total variation.

In most cases, even the movement of the observer itself is (roughly) known from inertial sensors or can easily be computed using image-based ego-motion estimation techniques, see for example [1]. With known depth and ego-motion information, the optical flow $\mathbf{u}_{\text{static}}$ for the static part of the scene can be directly computed [8] by

$$\mathbf{u}_{\text{static}} = \frac{\mathrm{d}\mathbf{p}\left(\mathbf{X}\left(t\right)\right)}{\mathrm{d}t} = \nabla_3^\top \mathbf{p}\left(\mathbf{X}\right) \cdot \left(-\mathbf{t} - \omega \times \mathbf{X}\right) , \quad (5)$$

where $\mathbf{p} : \mathbb{R}^3 \rightarrow \mathbb{R}^2$ represents the perspective projection of a three-dimensional point $\mathbf{X}$ in the observer's coordinate system, $\nabla_3 \equiv \left(\frac{\partial}{\partial X_0}, \frac{\partial}{\partial X_1}, \frac{\partial}{\partial X_2}\right)^\top$ the three-dimensional gradient operator, and $\mathbf{t}, \omega \in \mathbb{R}^3$ the translational and rotational part of the observer's motion. The point $\mathbf{X}$ can easily be computed from the known depth field using triangulation.

Since this information is available, we propose a novel way to use it in a variational approach to regularize the optical flow field. An approach similar to Eq. (3), where the deviation from the model is penalized directly (in Eq. (3), the deviation from the feature correspondences $\mathbf{u}^*$ is penalized), would lead to the same shortcomings: One would only get correct estimates for $\mathbf{u}$ if both the model and the data vote for the same value. The data has little chance to over-vote the model completely. The next section introduces a novel straight forward and very effective way to include such additive information into our regularization.

2.3. Modified Total Variation Regularization

Following the reasoning of the last section, we now try to improve our model, combining the advantages of the TV regularization (edge-preserving and robustness) with the possibility to include additional information such as depth fields, ego-motion or feature corrensponcdence information. For this purpose, we propose to use a modified total variational functional to regularize the flow field,

$$\begin{aligned} \mathrm{MTV}_L : \mathcal{C}^1\left(\Omega, \mathbb{R}^n\right) &\rightarrow \mathbb{R}^+, \quad n \in \mathbb{N} \\ u &\mapsto \int_\Omega \left|\nabla L\left[u\right]\left(\mathbf{x}\right)\right| \mathrm{d}\mathbf{x}, \end{aligned} \quad (6)$$

where $L : \mathcal{C}^1\left(\Omega, \mathbb{R}^n\right) \rightarrow \mathcal{C}^1\left(\Omega, \mathbb{R}^n\right), u \mapsto \alpha u + \beta$ is a pointwise linear functional with $\alpha \in \mathbb{R}$ and $\beta \in \mathcal{C}^1\left(\Omega, \mathbb{R}^n\right)$. Our variational minimization scheme is now

$$\begin{aligned} E_{\text{MTV}}\left[\mathbf{u}\right] = \int_\Omega \Bigg\{ & \lambda \left|\rho\left(\mathbf{x}, \mathbf{u}\left(\mathbf{x}\right)\right)\right| \\ & + \sum_{i=0}^{1} \left|\nabla L_i\left[u_i\right]\left(\mathbf{x}\right)\right| \Bigg\} \mathrm{d}\mathbf{x} , \end{aligned} \quad (7)$$

where L_i and u_i are the i-th components of $\mathbf{L}$ or $\mathbf{u}$ respectively. The explicit choice of $\mathbf{L}$ now allows inclusion of additional information into the variational scheme.

2.4. Stereo Compensated Total Variation Optical Flow

If depth and ego-motion information are known, we can use the expected optical flow, assuming that our scene is static, computed with Eq. (5), to regularize our flow field. Therefore in Eq. (7), we set

$$\mathbf{L}\left[\mathbf{u}\right] \equiv \mathbf{u} - \mathbf{u}_{\text{static}} , \quad (8)$$

so we no longer assume a constant optical flow field to be the most probable solution. We now use the projected motion of the static scene as prior.

Note that this is still a weak regularization, since we only penalize the gradient of the deviation of the computed flow field from the expected static one. Note also that we do not depend on exact measurements of the ego-motion, since we still are in a variational approach, where strong data can over-vote the regularization. This also always happens in the standard model of Eq. (2) when the constant flow assumption is violated by the grey-value constancy term in the presence of motion.

A first result of such a stereo supported flow algorithm is shown in Fig. 1 on the right hand side. One can clearly see that the stereo and ego-motion (in this case received by a hardware SGM implementation [5] and by inertial sensors in our research vehicle) information helps significantly to receive reasonable flow fields without staircase effects. We see that especially on the ground surface, where large flow vectors have to be estimated on weakly textured areas, the new regularization with Eqs. (6) and (8) leads to correct

flow fields in contrast to the left hand side. The independently moving car is also covered correctly, since it is well-textured and the data is descriminative enough to outweigh the prior.

These effects can be explained by the model visualized at the top right of Fig. 2. While the standard TV functional allows for multiple optimal paths between fixed values (see again Fig. 2 on the left), the compensation of $\mathbf{u}$ by an expected flow field ($\mathbf{u}_{\text{static}}$ in this case) inside the TV functional reduces the space of optimal solutions with respect to the expected flow field. We evaluate the algorithm quantitatively in Section 3.

2.5. Feature Correspondence Support for Fast Moving Objects

It turns out that the MTV regularization works well on static scenes, even when the optical flow resulting from strong motion of the observer has large values. Furthermore, the MTV regularization does not spoil the flow field of independently moving objects if the motion difference to the ego-motion is not too big (see Fig. 1). However, even with the help of a pyramid approach, large motion relative to the observer is still challenging.

To cope with independently fast moving objects, we review the LDOF (large displacement optical flow) approach from [2], where deviations from feature correspondences are directly penalized in the energy functional to be minimized, see Eq. (3). We modify this approach in the following way:

1. As in [2], we downsample the flow features to lower resolution levels to gain density.

2. At each level, starting at the lowest resolution, we use the model of Eq. (3), now extended by the MTV regularization, to compute dense flow fields which are significantly dominated by the feature correspondences, supported by the grey-value constancy constraint. We upsample the flow field to the next higher level and repeat until we reach the second or first down-sampled level.

3. On this level, we have found an initialization for our optical flow field for which we hope it is near to the global optimum. We go on for the rest of the levels, using the model in Eqs. (11) and (8).

This way, we use the feature correspondence information as an initialization in the hope to find a global minimum of the Zach-Pock model Eq. (2) rather than trying to solve the Brox model in Eq. (3) locally. We evaluate this approach in Section 3 using census features [14].

2.6. Numerical Implementation

In [20], the authors split the energy functional in Eq. (2) into two parts and couple them by an additive term to establish an iterative numerical solution scheme. On the one hand at every iteration step,

$$\arg\min_{\mathbf{v}} \int_{\Omega} \left\{ \lambda \left| \rho\left(\mathbf{x}, \mathbf{v}\left(\mathbf{x}\right)\right)\right| + \frac{1}{2\theta}\left(\mathbf{u}\left(\mathbf{x}\right) - \mathbf{v}\left(\mathbf{x}\right)\right)^2 \right\} \mathrm{d}\mathbf{x} \tag{9}$$

is solved for $\mathbf{v} : \Omega \to \mathbb{R}^2$ with fixed $\mathbf{u} : \Omega \to \mathbb{R}^2$ from the last iteration of Eq. (10), independently pixelwise for every $\mathbf{x} \in \Omega$. On the other hand, for fixed $\mathbf{v}$ the functional

$$\arg\min_{u_i} \left\{ \int_{\Omega} \frac{1}{2\theta}\left(u_i\left(\mathbf{x}\right) - v_i\left(\mathbf{x}\right)\right)^2 \mathrm{d}\mathbf{x} + \mathrm{TV}\left[u_i\right] \right\} \tag{10}$$

is minimized independently in all i components with respect to the flow $\mathbf{u}$ for fixed $\mathbf{v}$ from the previous step. Eq. (10) is also well-known as the Rudin-Osher-Fatemi (ROF) model [13], which frequently occurs in image processing tasks.

The modification with the linear functional L proposed in the last sections combines two advantages: First, it allows the inclusion of additional information such as depth and ego-motion information, and in addition, the regularization part which has to be minimized instead of Eq. (10), namely

$$\arg\min_{u_i} \left\{ \int_{\Omega} \frac{1}{2\theta}\left(u_i\left(\mathbf{x}\right) - v_i\left(\mathbf{x}\right)\right)^2 \mathrm{d}\mathbf{x} + \mathrm{MTV}_L\left[u_i\right] \right\}, \tag{11}$$

can be reduced to the well-understood and efficiently solvable [3] classical ROF model by writing Eq. (11) (with $L_i\left[u_i\right] \equiv \alpha u_i + \beta_i$) as

$$\frac{1}{\alpha} \arg\min_{u_i'} \left\{ \frac{1}{2\alpha^2\theta}\left(u_i' - \alpha^2\left(v_i + \beta_i\right)\right)^2 + \mathrm{TV}\left[u_i'\right] \right\} - \beta_i\,. \tag{12}$$

The implementation of Eq. (11) is straight-forward and we do not have to change the structure of our computation algorithm: in our iterative approach, when entering the regularizer step, we have to replace the old flow component v_i by $\alpha^2\left(v_i + \beta_i\right)$ and θ by $\alpha^2\theta$. After solving the classical ROF model for u_i', we compute the actual solution for this iteration step by $\mathbf{u}_{\text{min}} = \frac{1}{\alpha}\mathbf{u}'_{\text{min}} - \beta$.

2.7. Robust 3D Motion Fields with MTV Optical Flow

2.7.1 Variational Scene Flow

In Section (2.3), we have introduced the MTV regularization to include depth and ego-motion information. This idea can directly be extended to variational scene flow [17]. Therefore, the disparity change between two consecutive stereo image pairs is no longer regularized by total variation, but also by an MTV functional, which uses the ex-

Figure 3. Three-dimensional motion field computed by the Dense6D algorithm presented in [10]. Left: original left image. Right: three-dimensional motion field. The color encodes the motion magnitude, green=static, red=moving.

pected disparity change (using ego-motion and depth information while assuming a static world) for compensation before smoothing. When the observer is moving, this leads to better results than known from the standard approach presented in [17].

2.7.2 Dense6D

In [10], the authors show that a Kalman filter [7] based approach to three-dimensional motion field estimation from stereo image sequences (*Dense6D*, see Fig. 3) leads to the most robust results and allows for the real-time application in non-artificial scenarios. Although the combined optical flow and depth measurements are filtered to gain robustness and outliers are detected by a 3σ test, the quality of the resulting three-dimensional motion field significantly depends on the quality of the input data.

Since the authors in [10] use a standard optical flow approach (Eq. (2) based on [20]), their system is not able to cope with fast ego- or object motion, which would, however, be necessary to use the system for e.g. driver assistance systems in real traffic scenarios. We show in this section how the application of our new MTV regularized optical flow within the Dense6D framework leads to better results regarding accuracy, having depth and ego-motion information available as described in [10].

3. Evaluation and Results

In this section, the performance of the new depth-, ego-motion-, and feature-supported MTV optical flow approach is evaluated against the state-of-the-art real-time variational optical flow algorithm known from literature [17]. It is evaluated on synthetic stereo image sequences and on real scenes, both with strong camera motion commonly occuring in traffic scenes and fast independently moving objects. The system runs at 12 Hz on an Intel Core i7 and the main part of the algorithm runs on an nVidia GTX 480 graphics card. The whole system, including the stereo camera and the hardware SGM stereo module [5] is implemented in our research vehicle.

RMS	Dense6D			Crossing		
[px]	TV	MTV	LDMTV	TV	MTV	LDMTV
u_0	0.643	0.918	0.456	9.12	3.91	4.05
u_1	1.58	2.66	0.304	6.97	1.06	1.01

Table 1. Comparison of the RMS error results for the optical flow for three different optical flow approaches, computed over the whole synthetic image sequences.

3.1. Synthetic Sequences with Ground Truth Data

Firstly, the new MTV regularized approach for optical flow estimation is evaluated on two different synthetic stereo sequences ($640\times480\,\mathrm{px}\times12\,\mathrm{bit}$) with known ground truth motion, generated by POV-Ray. We focus on usual traffic situations, where the camera is mounted on a vehicle. In addition to the sequence presented in [10] (published on [11], *Dense6D*), we simulate a scene where the ego-vehicle is moving along a road while other cars are crossing very quickly in front of the observer (*crossing*). Note that all the optical flow algorithms used in this paper are running on highpass images, since we have to cope with illumination changes occuring in real scenes, which are the focus of this paper.

3.1.1 Optical Flow

Three different optical flow approaches are considered in our evaluation:

1. Standard TV-L^1 optical flow from Eq. (2), based on [17] (TV)

2. Optical flow with MTV regularization, using ego-motion and stereo information from SGM [4] (MTV)

3. MTV optical flow from above, with feature support as described in Section 2.5 (LDMTV)

For each of the three approaches, the root mean squared error (RMS) of the flow field $\mathbf{u}$ is computed over the whole image sequence and is presented in Tab. 1. In Fig. 4, sample images of the computed optical flow and the difference to the ground-truth are shown. One can clearly see from the RMS values and the sample pictures that the standard optical flow algorithm from [17] (TV) is not able to estimate the large flow caused by the strong camera motion correctly. It can also be seen in the sample pictures that both TV and MTV do not cover the large displacements ($\gtrsim 10\,\mathrm{px/frame}$). In contrast, the new LDMTV approach combines the advantage of the MTV approach with large displacement support. Due to the blurring effects around the moving vehicles (note that there is almost no texture on the ground), this is not reflected in the RMS values in Tab. 1 for the *crossing* sequence.

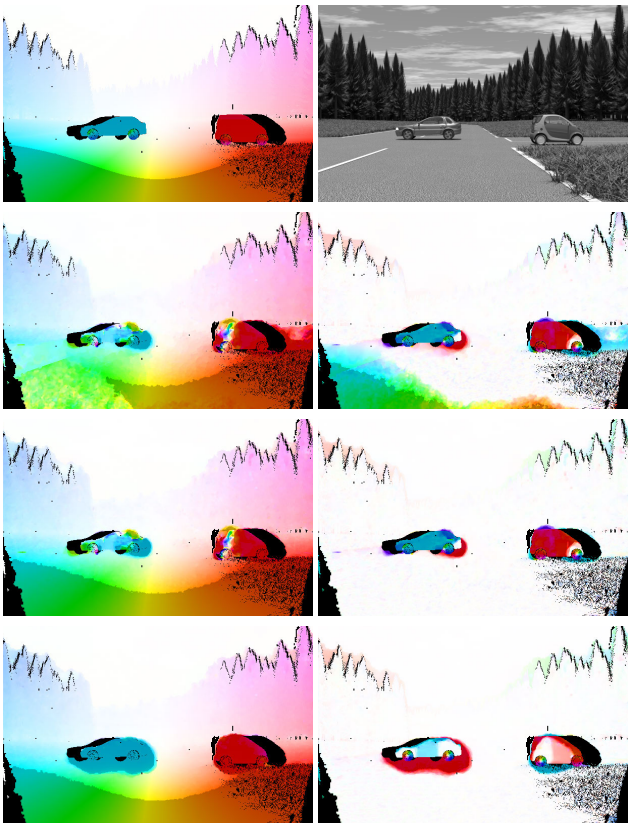

Figure 4. Sample pictures of the *crossing* ground truth scene. Top: ground truth optical flow (left) and left input image (right). Rows from second row to bottom: TV optical flow, MTV optical flow, LDMTV optical flow; left: optical flow $\mathbf{u}$, right: difference to ground truth optical flow ($\mathbf{u}_{\mathrm{GT}} - \mathbf{u}$). Color encoding as in Fig. 1. Flow and stereo occlusions are colored black.

RMS	Dense6D		Crossing	
[m/s]	TV	LDMTV	TV	LDMTV
v_x	0.487	0.500	0.812	0.764
v_y	0.366	0.368	0.242	0.118
v_z	3.11	3.12	5.27	1.09

Table 2. Comparison of the RMS error results for the three-dimensional motion field, having used two different optical flow approaches as input for the Dense6D system.

3.1.2 Dense6D

Since the LDMTV algorithm has shown its advantages, it is now applied to the Dense6D system instead of TV as in the original implementation [10]. We expect to receive robust three-dimensional motion fields even in the case of large ego-motion and fast independently moving objects. In Tab. 2, the RMS errors of the corresponding motion fields are shown for two different synthetic stereo sequences: the *crossing* and the *Dense6D* sequence already used in the original work for evaluation.

As expected, in challenging scenes, the new LDMTV approach leads to improvements in the flow field, regarding accuracy and also density (note that less tracked features are rejected by the 3σ outlier test). However, we cannot achieve better results than known from [10] in the *Dense6D* sequence, since both the observer and the independent objects are moving slowly enough for the standard optical flow computation scheme.

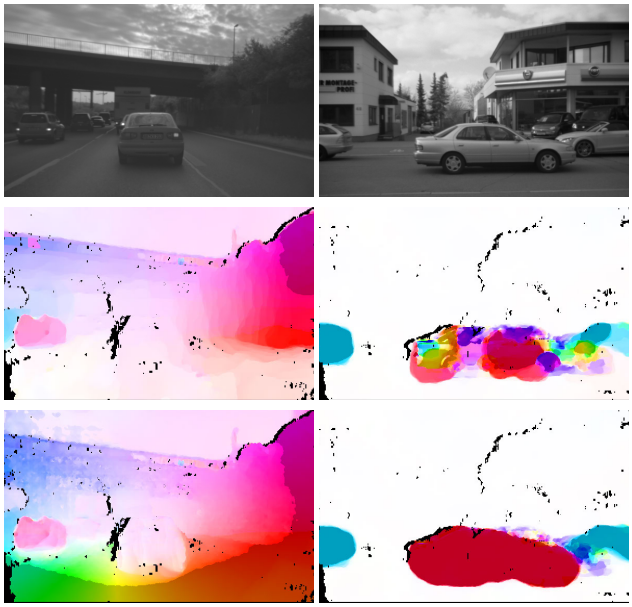

Figure 5. Real Scenes. Left: the fast ego-vehicle is overtaken by another car. Right: a T intersection with fast crossing vehicles in front of a moving observer. Top: original left images. Middle: standard TV optical flow. Bottom: LDMTV optical flow. Color encoding as in Fig. 1.

3.2. Real Scenarios

The actual scope of this paper are real scenes. In Fig. 5, two challenging situations and the computed optical flow (standard TV and LDMTV) are shown. Obviously, only the ego-motion-, depth- and feature-supported LDMTV optical flow algorithm is able to cope with these challenging sequences.

4. Conclusion

We have introduced a way to include additional information into a variational scheme to estimate optical flow, especially in real, non-ideal situations. We have shown how additional information can improve the optical flow results, so that they can be used for real-time three-dimensional motion field estimation from a fast moving observer platform. Our evaluation has demonstrated that the inclusion of ego-motion, as well as depth information from stereo or motion information from feature correspondences in a modified total variation regularization term efficiently supports the computation of optical flow. This leads to results which

are not known in that quality from real-time approaches on real scenes so far.

We are convinced that MTV regularized optical flow will be applied to perceive three-dimensional motion fields from a moving platform. This has a variety of important applications, e.g. in driver assistance. Further research will focus on a more sophisticated densification scheme for the feature information and the combination with anisotropic approaches (see e.g. [19]) to avoid the blurring effects in Fig. 4. Many challenges in real scenarios, such as bad weather conditions, reflections or shadows are not addressed in this work, but are of strong interest for future research.

References

[1] Hernàn Badino. A robust approach for ego-motion estimation using a mobile stereo platform. In *International Workshop on Compact Modeling*, Günzburg, Germany, October 2004. Springer.

[2] Thomas Brox, Christoph Bregler, and Jitendra Malik. Large displacement optical flow. In *IEEE International Conference on Computer Vision and Pattern Recognition*, Miami Beach, Florida, USA, 2009.

[3] Antonin Chambolle. An algorithm for total variation minimization and applications. *Journal of Mathematical Imaging and Vision*, 20(1–2):89–97, 2004.

[4] Heiko Hirschmüller. Accurate and efficient stereo processing by semi-global matching and mutual information. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2005.

[5] Heiko Hirschmüller and Stefan Gehrig. Stereo matching in the presence of sub-pixel calibration errors. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[6] Berthold K. P. Horn and Brian G. Schunck. Determining optical flow. In *Artificial Intelligence*, volume 17, pages 185–203, 1981.

[7] Rudolf Emil Kalman. A new approach to linear filtering and prediction problems. *ASME-Journal of Basic Engineering*, 82(D):35–45, 1960.

[8] Hugh Christopher Longuet-Higgins and Kvetoslav Prazdny. The interpretation of a moving retinal image. In *Proceedings of the Royal Society of London*, volume 208 B, pages 385–397, 1980.

[9] David G. Lowe. Object recognition from local scale-invariant features. In *International Conference on Computer Vision*, 1999.

[10] Clemens Rabe, Thomas Müller, Andreas Wedel, and Uwe Franke. Dense, robust and accurate motion field estimation from stereo image sequences in real-time. In *European Conference on Computer Vision*, pages 582–595, Heraklion, Greece, September 2010.

[11] Clemens Rabe and Toby Vaudrey. University of Auckland. enpeda. image sequence analysis test site (EISATS). `http://www.mi.auckland.ac.nz/EISATS`, June 2010.

[12] Richard Roberts, Christian Potthast, and Frank Dellaert. Learning general optical flow subspaces for egomotion estimation and detection of motion anomalies. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[13] Leonid I. Rudin, Stanley Osher, and Emad Fatemi. Nonlinear total variation based noise removal algorithms. *Physica D*, 60:259–268, 1992.

[14] Fridtjof Stein. Efficient computation of optical flow using the census transform. In *DAGM-Symposium*, 2004.

[15] Werner Trobin, Thomas Pock, Daniel Cremers, and Horst Bischof. An unbiased second-order prior for high-accuracy motion estimation. In *DAGM-Symposium*, pages 396–405, 2008.

[16] Sundar Vedula, Simon Baker, Peter Rander, Robert Collins, and Takeo Kanade. Three-dimensional scene flow. In *International Conference on Computer Vision*, 1999.

[17] Andreas Wedel, Thomas Pock, Christopher Zach, Horst Bischof, and Daniel Cremers. *Statistical and Geometrical Approaches to Visual Motion Analysis*, chapter An Improved Algorithm for TV-L1 Optical Flow, pages 23–45. Springer-Verlag Berlin, 2009.

[18] Andreas Wedel, Clemens Rabe, Toby Vaudrey, Thomas Brox, Uwe Franke, and Daniel Cremers. Efficient dense scene flow from sparse or dense stereo data. In *European Conference on Computer Vision*, 2008.

[19] Manuel Werlberger, Werner Trobin, Thomas Pock, Andreas Wedel, Daniel Cremers, and Horst Bischof. Anisotropic Huber-L1 optical flow. In *British Machine Vision Conference*, 2009.

[20] Christopher Zach, Thomas Pock, and Horst Bischof. A duality based approach for realtime TV-L1 optical flow. In *DAGM-Symposium*, 2007.