Detection and Segmentation of Independently Moving Objects from Dense Scene Flow

Andreas Wedel, Annemarie Meißner, Clemens Rabe,
Uwe Franke, and Daniel Cremers

Daimler Group Research, Sindelfingen, Germany
University of Applied Sciences, Stuttgart, Germany
Department of Computer Science, University of Bonn, Germany

Abstract. We present an approach for identifying and segmenting independently moving objects from dense scene flow information, using a moving stereo camera system. The detection and segmentation is challenging due to camera movement and non-rigid object motion. The disparity, change in disparity, and the optical flow are estimated in the image domain and the three-dimensional motion is inferred from the binocular triangulation of the translation vector. Using error propagation and scene flow reliability measures, we assign dense motion likelihoods to every pixel of a reference frame. These likelihoods are then used for the segmentation of independently moving objects in the reference image. In our results we systematically demonstrate the improvement using reliability measures for the scene flow variables. Furthermore, we compare the binocular segmentation of independently moving objects with a monocular version, using solely the optical flow component of the scene flow.

1 Introduction and Related Work

In this paper we present the segmentation of independently moving objects from stereo camera sequences, obtained from a moving platform. Classically, moving objects are separated from the stationary background by *change detection* (e. g. [1]). But if the camera is also moving in a dynamic scene, motion fields become rather complex. Thus, the classic change detection approach is not suitable as it can be seen in Fig. 1. Our goal is to derive a segmentation of moving objects for this general dynamic setting.

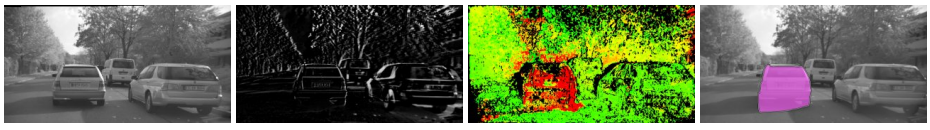

Fig. 1. From left to right: input image, difference image between two consecutive frames, motion likelihood, and segmentation result. With the motion likelihood derived from the scene flow, the segmentation of the moving object becomes possible although the camera itself is moving.

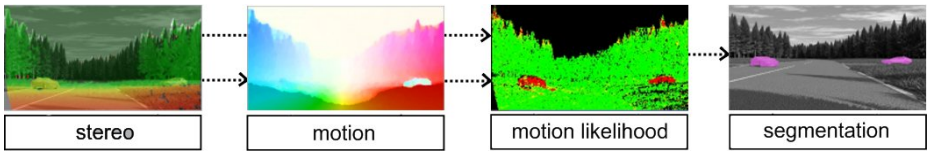

Fig. 2. The segmentation pipeline. Firstly, disparity and scene flow are computed; secondly, motion likelihoods are derived, and thirdly the image is segmented using graph cut.

We do not constrain the motion of the camera itself nor imply assumptions on the structure of the scene, such as rigid body motion. Rigid objects constrain the motion onto sub-spaces which yield efficient means to segment dynamic scenes using two views [2]. Another approach is used in [3], where the segmentation process is solved efficiently by incorporating a shape prior. If nothing about object appearance is known, the segmentation clearly becomes more challenging.

High-dynamic scenes with a variety of different conceivable motion patterns are especially challenging and reach the limits of many state-of-the-art motion segmentation approaches (e. g. [4, 5]). This is a pity because the detection of moving objects implies certain scene dynamics. Although we do not constraint the camera motion, we assume that it is approximately known. In particular, we compute the fundamental matrix together with the scene flow, as proposed for the optical flow setting in [6, 7]. From this, the motion of the camera is derived where the free scale parameter is fixed using the velocity sensor of the moving platform.

In [8] the authors use dense optical flow fields over multiple frames and estimate the camera motion and the segmentation of a moving object by bundle adjustment. The necessity of rather long input sequences however limits its practicability; furthermore, the moving object has to cover a large part of the image in order to detect its motion. The closest work related to our work is the work presented in [9]. It presents a monocular and a binocular approach to moving object detection and segmentation in high-dynamic situations using sparsely tracked features over multiple frames. In this paper we focus on moving object detection using only two consecutive stereo pairs, we use a dense scene flow fiels, and we show how per-pixel motion confidences are derived.

Fig. 2 illustrates the segmentation pipeline. The segmentation is performed in the image of a reference frame (left frame at time t) employing the graph cut segmentation algorithm [10]. The motion cues we use are derived from dense scene flow and calculated from the two stereo image pairs at time t-1 and t. Furthermore, we consider individual reliability measures for the variances of the flow vectors and the disparities at each image pixel. To our knowledge, the direct use of dense scene flow estimates for the detection and segmentation of moving objects is novel.

Paper Outline

In Section 2 we present the core graph cut segmentation algorithm. It minimizes an energy consisting of a motion likelihood for every pixel and a length term, favoring segmentation boundaries along intensity gradients.

The employed motion likelihoods are derived from dense scene flow in Section 3. Scene flow consists of the optical flow, the disparity, and the change of disparity over time. In the monocular setting, only the optical flow component of the scene flow is used. Compensating for the camera motion is a prerequisite step to detecting moving objects; additionally, one has to deal with inaccuracies in the estimates. We show how inaccuracies in the images can be modelled with reliability measures for the disparity and scene flow variables, and use error propagation to derive the motion likelihoods.

In Section 4 we compare the monocular method and the binocular method for the segmentation of independently moving objects in different scenarios. We systematically demonstrate that the consideration of inaccuracies, when computing the motion likelihoods for every pixel, yields increased robustness for the segmentation. Furthermore, we demonstrate the limits of the monocular and binocular segmentation methods and provide ideas for further research to overcome these limitations.

2 Segmentation Algorithm

The segmentation of the reference frame into moving and stationary parts can be expressed by a binary labelling of the pixels,

$$\mathcal{L}(\mathbf{x}) = \begin{cases} 1 & \text{if the pixel } \mathbf{x} \text{ is part of a moving object} \\ 0 & \text{otherwise.} \end{cases} \tag{1}$$

The goal is now to determine an optimal assignment of each pixel to *moving* or *non moving*. There are two competing constraints. Firstly, a point should be labelled *moving* if it has a high motion likelihood ξ_{motion} derived from the scene flow information and vice versa. Secondly, points should favour a labelling which matches that of their neighbors. Both constraints enter a joint energy of the form

$$E(\mathcal{L}) = E_{\mathrm{data}}(\mathcal{L}) + \lambda E_{\mathrm{reg}}(\mathcal{L})\,, \tag{2}$$

where λ weighs the influence of the regularization force. The data term is given by

$$E_{\mathrm{data}} = -\sum_{\Omega} \left\{ \mathcal{L}(\mathrm{x})\, \xi_{\mathrm{motion}}(\mathbf{x}) + \Big(1 - \mathcal{L}(\mathrm{x})\Big)\ \xi_{\mathrm{static}}(\mathbf{x}) \right\} \tag{3}$$

on the image plane Ω, where ξ_{static} is a fixed prior likelihood of a point to be static. The regularity term favors labellings of neighboring pixels to be identical. This regularity is imposed more strongly for pixels with similar brightness:

$$E_{\mathrm{reg}} = \sum_{\Omega} \left\{ \sum_{\hat{\mathbf{x}} \in \mathcal{N}_4(\mathbf{x})} g\left(I(\mathbf{x}) - I(\hat{\mathbf{x}})\right)\ \left|\mathcal{L}(\hat{\mathbf{x}}) - \mathcal{L}(\mathbf{x})\right| \right\}\,, \tag{4}$$

where $\mathcal{N}_4$ is the 4 neighborhood (upper, lower, left, right) of a pixel and $g(\cdot)$ is a positive, monotonically decreasing function of the brightness difference between neighboring pixels. Here, we set $g(z) = \frac{1}{z+\alpha}$ with a positive constant α.

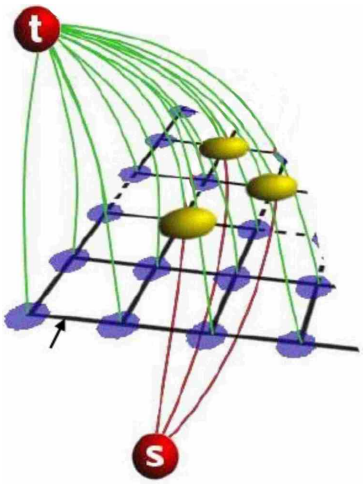

Fig. 3. Illustration of the graph mapping. Red connections illustrate graph edges from the source node s to the nodes, green connections illustrate graph edges from nodes to the target node t. Note, that the ξ_{motion} likelihood may be sparse due to occlusion. In the illustration only pixels with yellow spheres contribute to this motion likelihood. Black connections (indicated by the arrow) illustrate edges between neighboring pixels.

Graph Mapping.

Summarizing the above equations, this yields for the energy (Equation 2)

$$\sum_{\Omega} \left\{ -\mathcal{L}(\mathrm{x})\, \xi_{\text{motion}}(\mathbf{x}) - \Big(1 - \mathcal{L}(\mathrm{x})\Big)\, \xi_{\text{static}}(\mathbf{x}) + \lambda \sum_{\hat{\mathbf{x}} \in \mathcal{N}_4(\mathbf{x})} \frac{|\mathcal{L}(\hat{\mathbf{x}}) - \mathcal{L}(\mathbf{x})|}{|I(\mathbf{x}) - I(\hat{\mathbf{x}})| + \alpha} \right\}. \quad (5)$$

Due to the combinatorial nature, finding the minimum of this energy is equivalent to finding the s-t-separating cut with minimum costs of a particular graph $\mathcal{G}(v, s, t, e)$, consisting of nodes $v(\mathbf{x})$ for every pixel $\mathbf{x}$ in the reference image and two distinct nodes: the source node s and the target node t [11]. The edges e in this graph connect each node with the source, target, and its $\mathcal{N}_4$ neighbors. The individual edge costs are defined as follows:

edge	edge cost
source link: $s \rightarrow v(\mathbf{x})$	$-\xi_{\text{motion}}(\mathbf{x})$
target link: $v(\mathbf{x}) \rightarrow t$	$-\xi_{\text{static}}(\mathbf{x})$
$\mathcal{N}_4$ neighborhood: $v(\hat{\mathbf{x}}) \leftrightarrow v(\mathbf{x})$	$\lambda \frac{1}{\lvert I(\mathbf{x}) - I(\hat{\mathbf{x}})\rvert + \alpha}$

The cost of a cut in the graph is computed by summing up the costs of the cut (removed) edges. Removing the edges of an s-t-separating cut from the graph yields a

graph where every node v is connected to exactly one terminal node, either the source s or the target t. If we define nodes that are connected to the source as static and those connected to the target as moving, it turns out that the cost of an s-t-separating cut is equal to the energy in Equation (5) with the corresponding labelling, and vice versa. Thus, the minimum s-t-separating cut yields the labeling that minimizes Equation (5). The minimum cut is found using the graph cut algorithm in [10].

In the next Section, we will derive the likelihoods $\xi_{\text{motion}}(\mathbf{x})$ from the disparity and scene flow estimates.

3 Motion Likelihoods

Independently moving objects can only be detected from an image sequences if at least two consecutive images are evaluated. In this paper we constraint ourselves to the minimum case of only two consecutive images. If more images are available, the detection task essentially becomes a tracking task because previously detected objects influence the current segmentation.

We analyze a monocular and a binocular camera setting, and derive likelihoods that pixels of a reference frame depict moving objects. In the monocular case, these constraints have been proposed in [12]. We will review the constraints and derive a Mahalanobis distance for every pixel in the image space which corresponds to the likelihood that the depicted object is moving. In the binocular case, the three-dimensional position for every pixel and its three-dimensional motion vector are reconstructed. Then the Mahalanobis distance of the three-dimensional translation vector yields a likelihood that the depicted object is moving.

3.1 Scene Flow Computation.

The input for the motion likelihood is given by dense disparity and scene flow estimates $[d, u, v, p]$ for every pixel in the reference frame. The image position, $\mathbf{x} = [x, y]$, and the disparity, d, encode the three-dimensional position of a point. The optical flow (change of image position in between two frames), $[u, v]$, and the change in disparity, p, encode the scene flow motion information. Note, that for the monocular setting only the optical flow information, $[u, v]$, is used.

A variational approach to estimating this flow field was first proposed in [13]. The authors imposed regularity over all four variables and estimated all variables by minimizing a resulting single functional. Here we use the approach proposed in [14], where the authors split the position and motion estimation steps into two separate problems,

$$\text{(A)} \quad \Omega \to \mathbb{R}\,, \qquad [x, y] \mapsto d \tag{6}$$

$$\text{and (B)} \quad \Omega \times \mathbb{R} \to \mathbb{R}^3\,, \quad [x, y] \times d \mapsto [u, v, p]\,. \tag{7}$$

While (A) is the well-known disparity estimation step, (B) implies minimizing a scene flow energy, consisting of a data term and a smoothness term,

$$SF(u, v, p, d) = SF_{\text{data}}(u, v, p, d) + SF_{\text{smooth}}(u, v, p)\,. \tag{8}$$

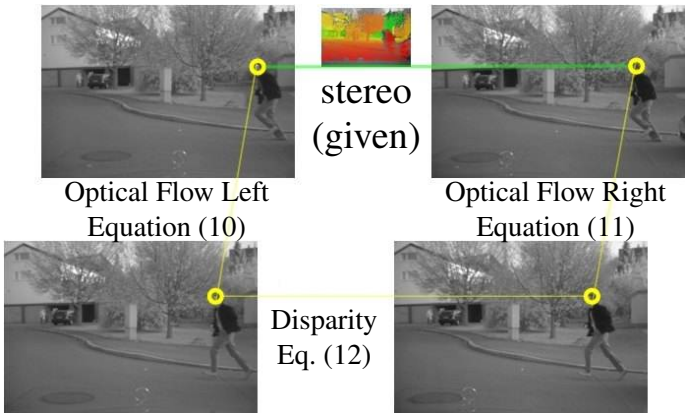

Fig. 4. Scene flow computation from two stereo image pairs. The stereo at the last time instance, t-1, is given by the semi-global matching algorithm. The data terms and smoothness term are described in the text in Equations (10 - 13).

The implicit dependancy of the variables u, v, p, and d on $[x, y]$ (e.g. $u(x, y)$) is left out in the notation to keep the notation uncluttered. Note, that the coupling between position and motion in such an approach is taken care of implicitly as the motion estimation step in (B) depends on the position estimation, which is the previously computed disparity map in (A).

The data term evaluates the gray value constancy of the scene flow,

$$SF_{\text{data}}(u, v, p, d) = \int_{\Omega} \left\{ E_{\text{sf-data-left}} + E_{\text{sf-data-right}} + E_{\text{sf-data-disp}} \right\} dx\, dy\,. \quad (9)$$

It evaluates the gray value constancy assumption for the optical flow field in the left image pair (I_L) and the right image pair (I_R):

$$E_{\text{sf-data-left}} = |I_L(x, y, t-1) - I_L(x+u, y+v, t)| \quad (10)$$
$$E_{\text{sf-data-right}} = |I_R(x+d, y, t-1) - I_R(x+d+u+p, y+v, t)|\,. \quad (11)$$

Additionally, the gray value constancy assumption for the stereo disparity field at time t is evaluated:

$$E_{\text{sf-data-disp}} = |I_L(x+u, y+v, t) - I_R(x+d+u+p, y+v, t)|\,. \quad (12)$$

The smoothness term minimizes the fluctuation in the scene flow field by penalyzing the flow field derivatives,

$$SF_{\text{smooth}}(u, v, p) = \int_{\Omega} E_{\text{sf-reg}}\, dx\, dy \quad \text{with} \quad E_{\text{sf-reg}} = |\nabla u| + |\nabla v| + |\nabla p|\,. \quad (13)$$

The resulting energy can be solved by calculus of variation. For the numerical solution scheme we refer to [14].

3.2 Variances for Disparity and Scene Flow.

Computing the Mahalanobis distance implies that variances for the image position (monocular setting) or three-dimensional translation vector (binocular setting) need to be known. Although constant variances for the whole image may be used, our experiments show that individual variances yield more reliable segmentation results. Therefore, we derive such variances for the disparity and scene flow estimates for every pixel, depending on the corresponding underlying energy functional.

Disparity Reliability. The scene flow algorithm in [14] uses the semi-global matching algorithm [15] for the disparity estimation and a variational framework for the scene flow estimates. The core semi-global matching algorithm is pixel-accurate.

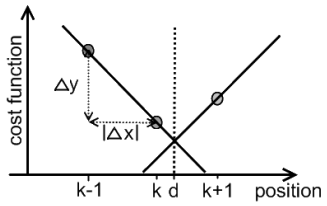

Fig. 5. The slope of the disparity cost function serves as a quality measure for the disparity estimate.

Let k be the disparity estimate of the core SGM method for a certain pixel in the left image. The SGM method in [15] is formulated as an energy minimization problem. Hence, changing the disparity by ± 1 yields an increase in costs (yielding an increased energy). The minimum, however, may be located in between pixels, motivating a subsequent sub-pixel estimation step. Sub-pixel accuracy is achieved by a subsequent fit of a symmetric equiangular function (see [16]) in the cost volume. The basic idea of this step is illustrated in Figure 5. The costs for the three disparity assumptions k-1 px, k px, and k+1 px are taken and a symmetric first order function is fitted to the costs. This fit is unique and yields a specific sub-pixel minimum, located at the minimum of the function. Note, that this might not be the minimum of the underlying energy but is a close approximation, evaluating the energy only at pixel position.

The slope of this fitting function (the larger of the two relative cost differences between the current estimate and neighboring costs, Δy) serves as a quality measure for the goodness-of-fit. If the slope is low, the disparity estimate is not accurate in the sense that other disparity values could also be valid. If on the other hand the slope is large, the sub-pixel position of the disparity is expected to be quite accurate as deviation from this position increases the energy. Hence, the larger the slope, the better is the expected quality of the disparity estimate. Note that the costs mentioned here are accumulated costs that also incorporate smoothness terms.

Based on this observation an uncertainty measure is derived for the expected variance of the disparity estimate:

$$U_D(x, y, d) = \frac{1}{\Delta y} \; . \tag{14}$$

Scene Flow Reliability. For variational optic flow methods the idea of using the incline of the cost function or energy function as uncertainty measure becomes more complex than in the disparity setting. This is due to the higher dimensionality of the input and solution space. An alternative, energy-based confidence measure was proposed in [17]. The novel idea is that the reliability is inversely proportional to the local energy contribution in the energy functional, used to compute the optical flow. A large contribution

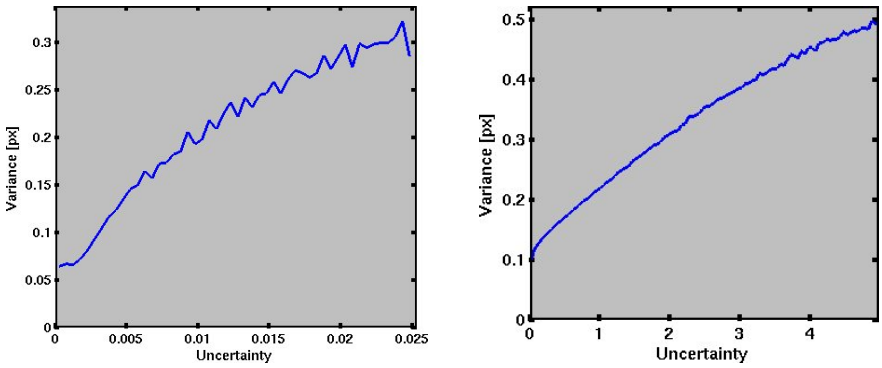

Fig. 6. Plots of the proposed reliability measures and corresponding variances for the disparity ($\mathrm{VAR}(d)$ vs. U_D, *left*) and for the scene flow u-component ($\mathrm{VAR}(u)$ vs. U_{SF}, *right*). The plots reveal that the proposed reliability measures are approx. proportional to the observed variances.

to the total energy implies low expected accuracy while the accuracy is expected to be good if the energy contribution is small. The authors show that this energy-based measure yields a better approximation of the *optimal* confidence for optic flow estimates than an image-gradient-based measure. The same idea is now applied to the scene flow case, yielding an expected variance of the scene flow estimate:

$$U_{SF}(x, y, d, u, v, p) = E_{\text{sf-data-left}} + E_{\text{sf-data-right}} + E_{\text{sf-data-disp}} + \lambda E_{\text{sf-reg}}\,. \quad (15)$$

Comparing Variances and Reliability Measures. To evaluate the reliability measures for the disparity and scene flow estimates, we plot the derived uncertainty measures against the observed error in Fig. 6 (for the disparity d and the u-component of the optical flow). To generate the plots a 400 frames long evaluation sequence, rendered with Povray and available in [18] together with the ground truth flow, is used.

The plots illustrate, that the proposed reliability measures are correlated to the true variances of the errors. Furthermore, the variance σ_z (for a scene flow component $z \in \{d, u, v, p\}$) can be approximated by a linear function of the reliability measure, denoted by γ_z, with fixed parameters a_z and b_z: $\sigma_z^2(\mathbf{x}) = a_z + b_z \gamma_z(\mathbf{x})$.

3.3 Monocular Motion Likelihood.

For the monocular case we use the motion likelihood proposed for sparse data in [12]. There is a fundamental weakness of monocular three-dimensional reconstruction when compared to stereo methods – moving points cannot be correctly reconstructed by monocular vision. This is due to the camera movement between the two sequential images. Thus, optical flow vectors are triangulated, assuming that every point belongs to a static object. Such triangulation is only possible, if the displacement vector itself does not violate the fundamental matrix constraint. Needless to say that every track violating the fundamental matrix constraint belongs to a moving object and the distance to the fundamental rays directly serves as a motion likelihood.

However, even if flow vectors are aligned with the epipolar lines, they may belong to moving objects. This is due to the fact that the triangulated point may be located behind one of the two cameras or below the ground surface (for this constraint we make a planar road assumption). Certainly such constellations are only virtually possible, assuming that the point is stationary. In reality such constellations are prohibited by the law of physics. Therefore, such points must be located on moving objects.

In summary, a point is detected as moving if its 3D reconstruction is identified as erroneous. For calculating the distance $d_{\text{valid}}(\mathbf{x})$ between the observed optical flow vector and the closest optical flow vector fulfilling above constraints, we refer to [12] where above verbal descriptions are expressed in mathematical formulations. We calculate the Mahalanobis distance to this closest optical flow vector by weighing the distance with the variance of the optical flow vector, yielding

$$\xi_{\text{motion}}(\mathbf{x}) = \sqrt{d_{\text{valid}}\left(\mathbf{x}\right)^2 \, \sigma_{u,v}\left(\mathbf{x}\right)^2}\,. \tag{16}$$

Note, that due to the coupling in the variational framework, the variances σ_u and σ_v are assumed to be equal.

3.4 Binocular Motion Likelihood.

In the stereo setting, the full disparity and scene flow information is available. A point is transformed from the image coordinates (x, y, d) into world coordinates (X, Y, Z) according to $X = (x - x_0)\frac{b}{d}$, $Y = (y - y_0)\frac{b}{d}\frac{f_x}{f_y}$, and $Z = \frac{f_x\, b}{d}$, where b is the basis length of the stereo camera system, $f_x\, f_y$ are the focal lengths of the camera in pixels, and (x_0, y_0) its principal point. As a simplification, we assume the focal lengths $f_x\, f_y$ to be equal. Transforming the points (x, y, d) and $(x + u, y + v, d + p)$ into world coordinates and compensating the camera rotation $\mathbf{R}$ and translation $\mathbf{T}$ yields the three-dimensional residual translation (or motion) vector $\mathbf{M}$ with

$$\mathbf{M} = \frac{b}{d}\,\mathbf{R}\begin{bmatrix} x - x_0 \\ y - y_0 \\ f_x \end{bmatrix} - \frac{b}{d+p}\begin{bmatrix} x + u - x_0 \\ y + v - y_0 \\ f_x \end{bmatrix} + \mathbf{T} \tag{17}$$

Using error propagation we calculate the Mahalanobis length of the translation vector. Essentially, this incorporates the variances of the disparity, scene flow estimates, and the camera rotation and translation. Here, we assume the variances of the camera rotation to be negligible. Although this is certainly not true, such procedure is possible because the estimation of the fundamental matrix from the complete optical flow field does yield vanishing variances for the rotational parts. We do however use fixed variances in the camera translation because the translation information from the velocity sensor of the ego-vehicle is rather inaccurate. With the variances σ_u^2, σ_v^2, σ_p^2, and σ_d^2 for the scene flow and $\sigma_{\mathbf{T}}^2$ for the translation this yields the Mahalanobis distance

$$\xi_{\text{motion}}(\mathbf{x}) = \sqrt{\mathbf{M}^\top \left(\mathbf{J}^\top \text{diag}(\sigma_u, \sigma_v, \sigma_p, \sigma_d, \sigma_{\mathbf{T}})\,\mathbf{J}\right)^{-1}\mathbf{M}}\,, \tag{18}$$

where $\mathbf{J}$ is the Jacobian of Equation (17).

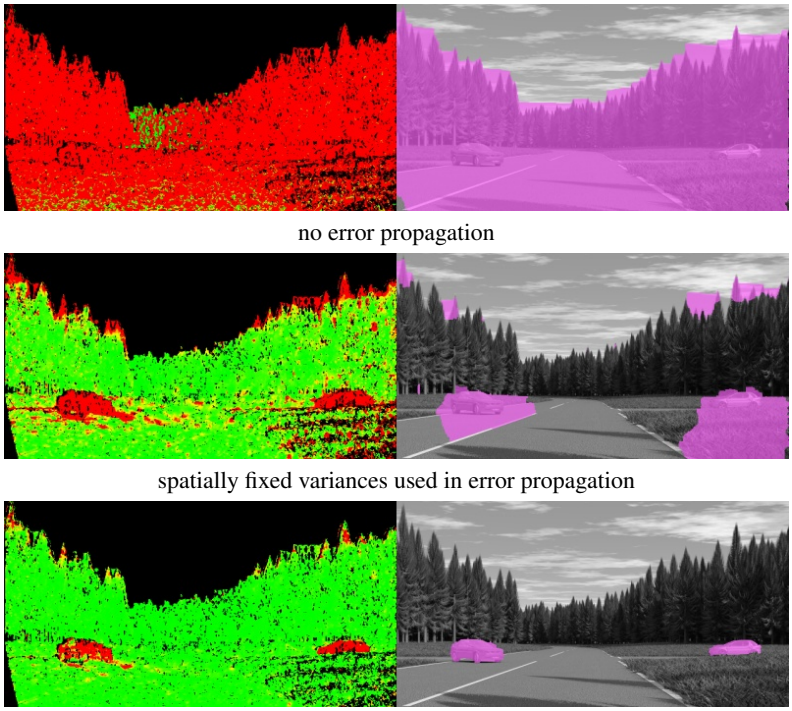

Fig. 7. Results for different error propagation methods. The *left* images show the motion likelihoods and the *right* images the segmentation results.

4 Experimental Results and Discussion

In this section we present results which demonstrate the accurate segmentation of moving objects using scene flow. In the first part, we show that the presented reliability measures greatly improve the segmentation results when compared to a fixed variance for the disparity and scene flow variables. In the second part, we compare the segmentation results using the monocular and binocular motion segmentation approaches.

4.1 Robust Segmentation

Figure 7 illustrates the importance of using the reliability measures to derive individual variances for the scene flow variables. If the propagation of uncertainties is not used at all, the segmentation of moving objects is not possible (top row). Using the same variance for every image pixel the segmentation is more meaningful; but still outliers are present in both, the motion likelihoods and the segmentation results (middle row). Only when the reliability measures are used to derive individual variances for the pixels, is the segmentation accurate and outlier influence is minimized (bottom row).

4.2 Monocular versus Binocular Segmentation of Independently Moving Object

A binocular camera system will always outperform a monocular system, simply because more information is available. However, in many situations a monocular system is able to detect independent motion and segment the moving objects in the scene. In this section we demonstrate the segmentation of independently moving objects using a monocular and a binocular camera system and discuss the results.

In a monocular setting, motion which is aligned with the epipolar lines cannot be detected without prior knowledge about the scene. Amongst other motion patterns, this includes objects moving parallel to the camera motion. For a camera moving in depth this includes all (directly) preceding objects and (directly) approaching objects. The *PreceedingCar* and *HillSide* sequences in Figure 8 show such constellations.

Using the ground plane assumption in the monocular setting (no virtually triangulated point is allowed to be located below the road surface) facilitates the detection of preceding objects. This can be seen in the *PreceedingCar* experiment, where lower parts of the car become visible. If compared to the stereo settings, which does not use any information about scene structure, the motion likelihood for the lower part of the preceding car is more discriminative. However, if parts of the scene are truly located below the ground plane, as the landscape at the right in the *HillSide* experiment, these will always be detected as moving, too. Additionally, this does not help to detect approaching objects. Both situations are solved using a binocular camera.

If objects do not move parallel to the camera motion, they are essentially *detectable* in the monocular setting (*Bushes* and *Running* sequences in Figure 9). However, the motion likelihood using a binocular system is more discriminative. This is due to the fact that the three-dimensional position of an image point is known from the stereo disparity. Thus, the complete viewing ray for a pixel does not need to be tested for apparent motion in the images, as in the monocular setting. In the unconstrained setting (not considering the ground plane assumption), the stereo motion likelihood therefore is more restrictive than the monocular motion likelihood. Note, that non-rigid objects (as in the *Running* sequence in Figure 9) are detected as well as rigid objects and do not limit the detection and segmentation at any stage.

5 Conclusion

Building up on a recent variational approach to scene flow estimation, we proposed in this paper an energy minimization method to detect and segment independently moving objects filmed in two video cameras installed in a driving car. The central idea is to assign, to each pixel in the image plane, a motion likelihood which specifies whether, based on 3D structure and motion, the point is likely to be part of an independently moving object. Subsequently, these local likelihoods are fused in an MRF framework and a globally optimal spatially coherent labelling is computed using the min cut max flow duality. In challenging real world scenarios where traditional background subtraction techniques would not work (because everything is moving), we are able to accurately localize independently moving objects. The results of our algorithm could directly be employed for automatic driver assistance.

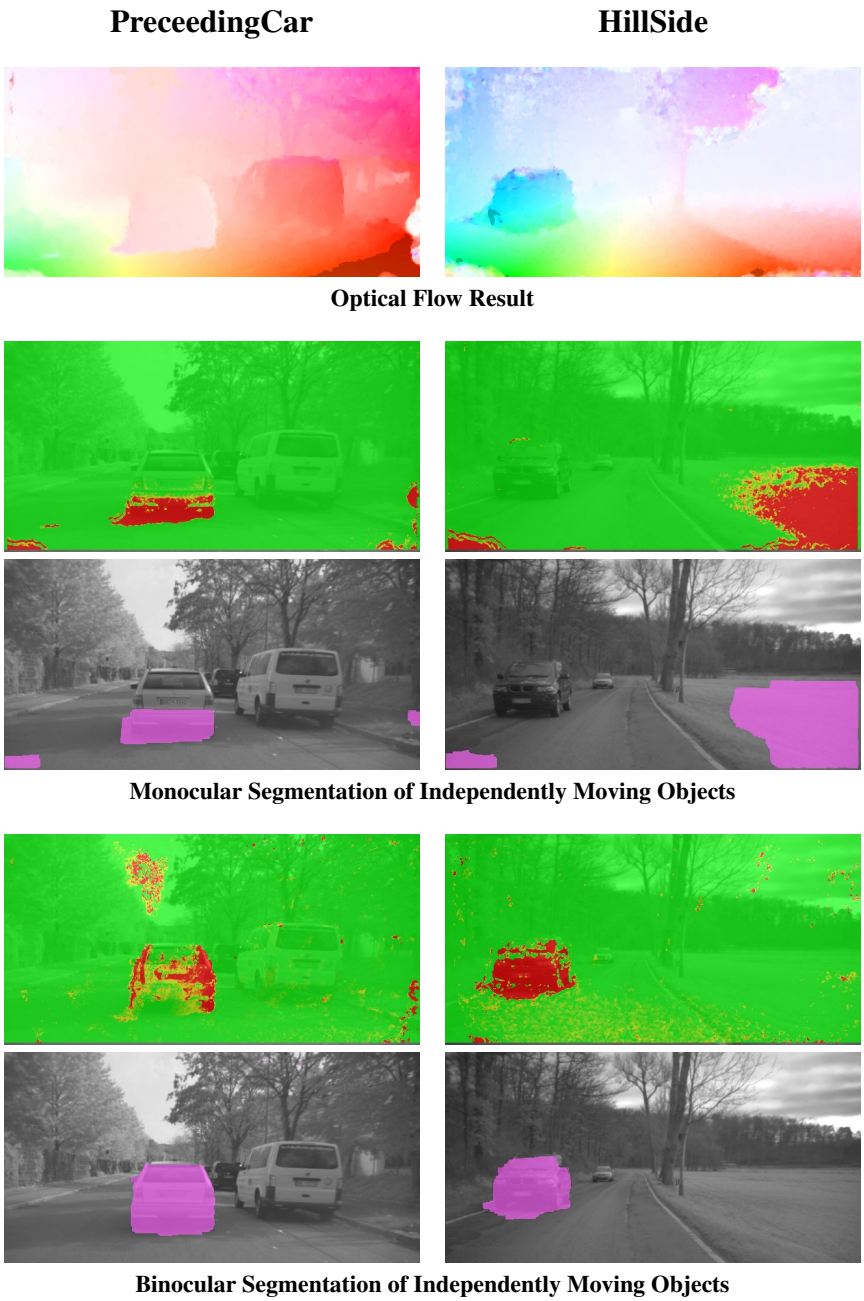

Fig. 8. The figure shows the energy images and the segmentation results for objects moving parallel to the camera movement. This movement cannot be detected monocularly without additional constraints, such as a planar ground assumption. Moreover if this assumption is violated, this yields errors (as in the *HillSide* sequence). In a stereo setting prior knowledge is not needed to solve the segmentation task in these two scenes.

Fig. 9. The figure shows the energy images and the segmentation results for objects which move not parallel to the camera motion. In such constallations a monocular as well as a binocular segmentation approach is successfull. However, one can see in the energy images and in the more accurate segmentation results (the head of the person in the *Running* sequence) that stereo is more discriminative. Note, that the also non-rigid independently moving objects are segmented.

References

1. Sun, J., Zhang, W., Tang, X., Shum, H.: Background Cut. In: Proc. European Conference on Computer Vision. Volume 3952., Springer (2006) 628
2. Vidal, R., Sastry, S.: Optimal segmentation of dynamic scenes from two perspective views. In: 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2003. Proceedings. Volume 2. (2003)
3. Brox, T., Rosenhahn, B., Cremers, D., Seidel, H.P.: High accuracy optical flow serves 3-D pose tracking: exploiting contour and flow based constraints. In Leonardis, A., Bischof, H., Pinz, A., eds.: Proc. European Conference on Computer Vision. Volume 3952 of LNCS., Graz, Austria, Springer (May 2006) 98–111
4. Cremers, D., Soatto, S.: Motion competition: A variational framework for piecewise parametric motion segmentation. International Journal of Computer Vision **62**(3) (May 2005) 249–265
5. Kolmogorov, V., Criminisi, A., Blake, A., Cross, G., Rother, C.: Bi-layer segmentation of binocular stereo video. In: Proc. International Conference on Computer Vision and Pattern Recognition. Volume 2. (2005)
6. Wedel, A., Pock, T., Braun, J., Franke, U., Cremers, D.: Duality TV-L1 flow with fundamental matrix prior. In: Proc. Image and Vision Computing New Zealand, Christchurch, New Zealand (November 2008)
7. Valgaerts, L., Bruhn, A., Weickert, J.: A variational approach for the joint recovery of the optical flow and the fundamental matrix. In: Pattern Recognition (Proc. DAGM), Munich, Germany (June 2008) 314–324
8. Zhang, G., Jia, J., Xiong, W., Wong, T., Heng, P., Bao, H.: Moving object extraction with a hand-held camera. In: Proc. International Conference on Computer Vision. (2006) 1–8
9. Vaudrey, T., Wedel, A., Rabe, C., Klappstein, J., Klette, R.: Evaluation of moving object segmentation comparing 6D-Vision and monocular motion constraints. In: Proc. Image and Vision Computing New Zealand, Christchurch, New Zealand (November 2008)
10. Boykov, Y., Kolmogorov, V.: An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision. IEEE Transactions on Pattern Analysis and Machine Intelligence (2004) 1124–1137
11. Kolmogorov, V., Zabih, R.: What energy functions can be minimized via graph cuts? In: Proc. European Conference on Computer Vision. (2002) 65–81
12. Klappstein, J., Stein, F., Franke, U.: Detectability of Moving Objects Using Correspondences over Two and Three Frames. In: Pattern Recognition (Proc. DAGM). Volume 4713 of LNCS., Springer (2007) 112
13. Huguet, F., Devernay, F.: A variational method for scene flow estimation from stereo sequences. In: IEEE Eleventh International Conference on Computer Vision, ICCV 07, Rio de Janeiro, Brazil. (October 2007)
14. Wedel, A., Rabe, C., Vaudrey, T., Brox, T., Franke, U., Cremers, D.: Efficient dense scene flow from sparse or dense stereo data. In: Proc. European Conference on Computer Vision. LNCS, Marseille, France, Springer (October 2008) 739–751
15. Hirschmüller, H.: Stereo vision in structured environments by consistent semi-global matching. In: Proc. International Conference on Computer Vision and Pattern Recognition. (2006) 2386–2393
16. Shimizu, M., Okutomi, M.: Precise sub-pixel estimation on area-based matching. In: Proc. International Conference on Computer Vision. (2001) 90–97
17. Bruhn, A., Weickert, J.: A confidence measure for variational optic flow methods. Geometric Properties for Incomplete Data (March 2006)
18. University of Auckland: *.enpeda..* Image Sequence Analysis Test Site (EISATS) (2008) http://www.mi.auckland.ac.nz/EISATS/.