

# Segmentation of independently moving objects using a maximum-likelihood principle

Martin Clauss, Pierre Bayerl, and Heiko Neumann

Dept. of Neural Information Processing, University of Ulm, 89081 Ulm, Germany,  
{martin.clauss,pierre.bayerl,heiko.neumann}@uni-ulm.de

**Abstract.** Detection of independently moving objects (IMOs) is a demanding task especially in situations, where the observer is moving himself. In such situations detection of IMOs as well as estimation of egomotion depend on each other and thus have to be handled simultaneously. We present an algorithm based on the Expectation/Maximization algorithm, which is capable of sharply separating background and independently moving objects, whilst the observer itself is moving. Furthermore it incorporates temporal integration of extracted information to improve estimation.

## 1 Introduction

When an observer (camera) is moving freely through a dominantly rigid scene, the detection of independently moving objects (IMOs) is a difficult task. In this situation, optical flow may result from either self-motion or from IMOs. On the one hand, detection of IMOs requires knowledge of the observer's egomotion to eliminate those flow components induced by the observer's own motion. On the other hand, estimation of the observer's motion is based on the global optical flow pattern, which is disturbed by the presence of IMOs in the current scene. Since both problems depend on each other, they have to be dealt with simultaneously in order to achieve a robust solution.

Unlike other previous proposals, we propose an approach that employs a single camera to segment global flow patterns (due to self-motion) from motion that is induced by other moving objects in the scene. The proposed approach may be utilized to feed further applications like collision warning, autonomous robot navigation, guidance, etc.

We will outline some previously proposed approaches for the problem scenario in section 2. Subsequently, we present our new method to simultaneously solve scene segmentation and egomotion estimation in section 3, together with a brief outline of the expectation/maximization algorithm and a short description of the underlying data representation. Some results of our algorithm will be presented in section 4, followed by a brief conclusion in section 5.

## 2 Related work

Approaches have been proposed to detect independent motion utilizing multiple cameras or object shape constants (e.g. [11]). Algorithms that utilize a single camera to detect IMOs during observer motion can be basically categorized into two groups: The

first group are algorithms based on motion similarity. Smith and Brady [12] presented an approach that groups similar flow vectors returned by a feature tracking algorithm. The flow magnitude needs to contrast the background estimates in order to be considered as candidate. Their approach uses information extracted over time to learn the contour and the motion parameters of detected IMOs. The method does not employ knowledge of the observer’s egomotion.

The other group are algorithms which detect IMOs based on knowledge of egomotion. Pauwels and van Hulle [10] iteratively estimate the observer’s motion and remove data points not matching the observer’s current motion-estimate. The method does not use any knowledge obtained from earlier image frames, but only uses flow information of the current frame. In order to circumvent an estimation of egomotion, several approaches (e.g. [1]) employ additional sensors to recover egomotion.

Woelk and Koch [14] utilize a particle filter [4] for sampling of optical flow estimation. The focus of expansion (FOE) and the translational component of the observer’s motion are estimated, while relying on an inertial sensor to determine the rotational component. After removal of the rotational component, the flow vectors are classified depending on their deviation from the radial flow direction, pointing outwards from the FOE.

MacLean et al. [8] apply a subspace method to segregate rotation from translation. The constraints on the translation vectors, which are obtained from the subspace method, are then associated to a dynamical number of processes using the EM-algorithm. The number of processes is estimated according to the total fit of the constraint vectors to the estimated translation vectors.

Our approach combines previous proposals ([8][10][12]) to develop a framework for robust segmentation of background and IMOs during egomotion.

### 3 Detecting IMOs using the EM-Algorithm

Optical flow calculated from an image stream that is captured by a moving camera represents a superposition of four components:  $F = T + R + I + \xi$ . Translational motion  $T$  and rotational motion  $R$  of the observer, motion  $I$  induced by independently moving objects and noise  $\xi$  resulting from either camera measurement or the optical flow algorithm. Being able to only observe  $F$  by means of a single camera, these four components cannot be easily split.

The EM algorithm provides an iterative framework for finding the corresponding unobservable data association. After introducing our underlying data representation, we briefly introduce the EM algorithm and how its principle can be utilized in our scenario, leading to our new approach.

#### 3.1 Data representation

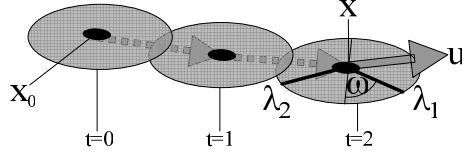
We utilize a spatially sparse multimodal token-based data representation. This data representation is based on the biologically motivated work of Krüger et al. [6], which elaborates on the primal sketch introduced by Marr [9]. Different local feature attributes are stored in a common symbolic token  $T$ :

- position  $\mathbf{x}$  of the token within the input image,
- optical flow  $\mathbf{u}$  (Lucas-Kanade, [7]),
- greyscale structure  $\lambda_1, \lambda_2, \omega$  (corner-/edgeness and orientation, [13]),
- status information (initial position  $x_0$ , age  $t$ ).

In sum, a token  $T$  can be stored as vector

$$\mathbf{T} = (\text{position}(x), \text{flow}(u), \text{greyscale}(\lambda_1, \lambda_2, \omega), \text{status}(x_0, t))$$

This token-based representation reduces the memory requirements, whilst preserving relevant information.



**Fig. 1.** Token-based symbolic data representation. Shown is one token at three successive time steps. Each token holds information on its position  $\mathbf{x}$ , its movement  $\mathbf{u}$  (from optical flow), underlying greyscale information (from structure tensor) given by  $\lambda_1, \lambda_2, \omega$  and token status information. All values are coupled with confidence values.

### 3.2 EM-Algorithm

When observing data samples  $y$ , with each sample originating from exactly one out of  $n$  models  $M_n$ , the maximum-likelihood (ml) parameters are those parameters for the generating models  $M_n$ , that maximize the probability of observing the data samples  $y$ . In general, however, the association of the data samples  $y$  to the models  $M_n$  cannot be observed, that is, the complete data  $x$  is unavailable. The Expectation/Maximization (EM) algorithm [3] seeks to iteratively find an optimized solution for model estimation from such incomplete observed data in an ml-fashion. The algorithm exploits, that the ml-parameters as well as the association of data samples to the models can be computed depending on each other. Thus, the EM algorithm iteratively solves the problem as follows: Be  $y$  the observation and  $x$  the corresponding complete data. The probability density function is  $f(x|\theta)$  with  $\theta$  denoting the parameters of the density. In the beginning, the parameters  $\theta$  are initialized randomly.

In the **Expectation** step, the algorithm estimates the probability that the data sample originates from a model, for each model and every data sample, given the current model parameter estimates:

$$Q(\theta|\theta_k) = E[\log f(x|\theta)|y, \theta_k]$$

Where  $\theta_k$  is the parameter set after the  $k$ -th iteration of the algorithm.

In the **Maximization** step, the model parameters are optimized in a maximum-likelihood fashion given the current data association estimate from the Expectation step:

$$\theta_{k+1} = \arg \max_{\theta} Q(\theta|\theta_k)$$

Expectation and Maximization step are alternately executed until precision is regarded sufficient, e.g. if  $\|\theta_k - \theta_{k+1}\| < \epsilon$  for suitable  $\epsilon$  and  $\|\cdot\|$ . The EM algorithm is guaranteed to converge [3] with respect to a local maximum of the likelihood function. One drawback of the EM algorithm is, that the number of underlying models  $n$  has to be known in advance. In our scenario, the number of IMOs is not known a priori and, therefore, the number of models is unknown as well, which is one of the major issues solved by our new approach.

### 3.3 Method

The EM-principle can be utilized for detecting IMOs via an indirect approach based on the following reasoning: If we could estimate the translational and rotational flow field that is induced by the observer's self-motion, then any deviating flow must arise from independent motions of an unknown number of objects. As the number of models (that is, the number of IMOs+1) is not known in advance for our scenario, we propose a simplified version of the EM algorithm in which only one model needs to be estimated. This model is the one associated with the observer's motion.

**Expectation Step:** During the expectation step, the fit of data samples is determined only for the model corresponding to the observer. First, the rotational component of the flow field is removed. As the rotational component only depends on the actual motion parameters and not on the scene geometry, the rotational parameters are sufficient to achieve this. These parameters are calculated by the subspace method employed during the maximization step. Under ideal conditions in a rigid environment without presence of IMOs, a flow field radially expanding from the focus of expansion (FOE) is obtained. The location  $f_0$  of the FOE depends on the observer's translational parameters. The angular deviation (Fig. 2) from the expected radial flow field is then calculated for each flow vector (compare [14][10]):

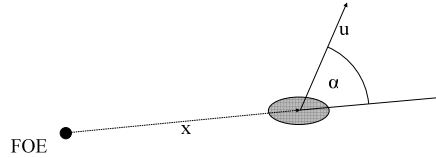
$$\alpha = \arccos \frac{u \cdot x}{\|u\| \cdot \|x\|}$$

where  $x$  denotes the token's position relative to the focus of expansion  $f_0$  and  $u$  is the optical flow component stored in the corresponding token. The deviation represents the grade of fit to the observer's motion model for a given flow vector.

**Maximization Step:** In the maximization step, the motion parameters are estimated using the subspace method presented by Heeger and Jepson [5]. By transforming to a subspace the rotational component vanishes and constraints for translation are obtained. Translational parameters are then computed from these constraints in a least-squares sense. As this subspace method does not support a continuous weighting of the input data samples per se, the model association of data points is first transformed to a binary membership using a fixed threshold.

Instead of initializing the model parameters randomly, we determine an initial guess by starting with the maximization step. For the first input frame, all data samples are assumed originating from the observer's motion model. For subsequent frames, model associations obtained from previously analysed video frames are used. This temporal feedback results in a better initial guess as well as a reduction in computational cost. In order to ensure that the correct model is estimated, we need to assume that more than

50% of the tokens represent the scene background that yield to estimates corresponding with the observer’s egomotion. After convergence of the proposed algorithm, the segmentation of the image into background and IMOs is given by the association of data points to the observer’s motion model obtained in the final expectation step.



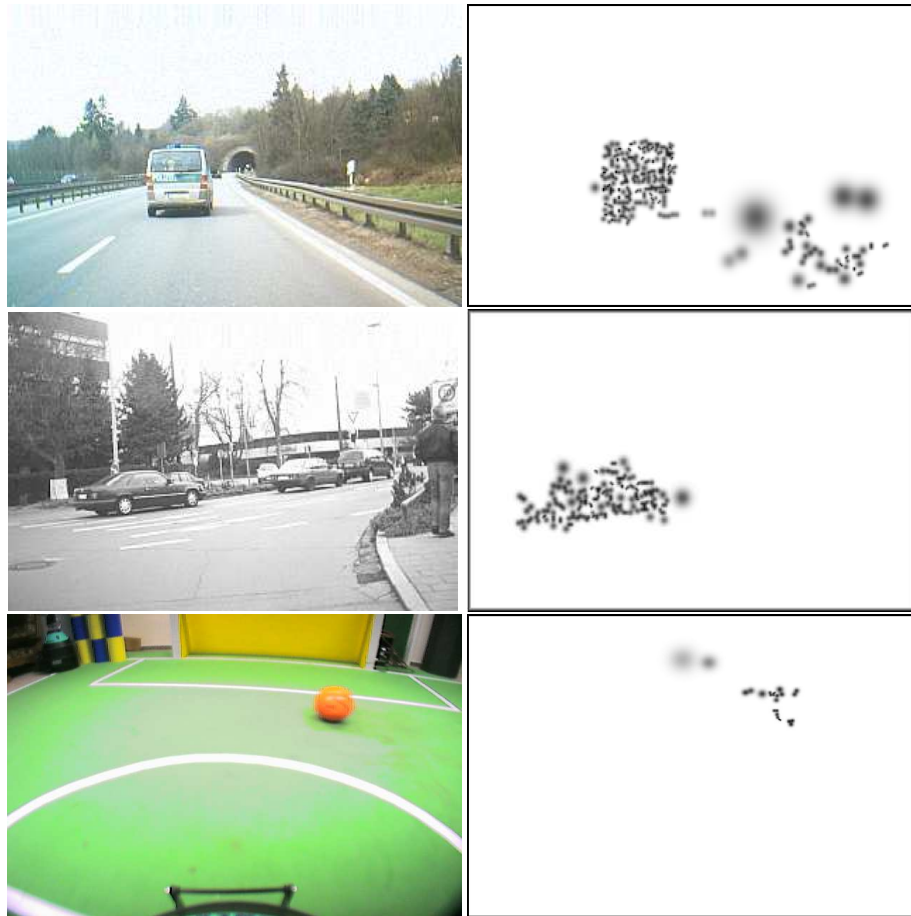
**Fig. 2.** Deviation of flow from to the expected motion pattern. Vector  $\mathbf{x}$  points from the focus of expansion to the position of the token. Vector  $\mathbf{u}$  is the movement of the token, i.e. the optical flow corresponding to the token. The angular error  $\alpha$  employed in the segmentation algorithm is calculated as the angle between vector  $\mathbf{x}$  and vector  $\mathbf{u}$

## 4 Results

We show results of our proposed algorithm from three scenarios. The input data originates from a moving car on a freeway and in an inner city environment, as well as from an autonomous mobile soccer robot (Fig. 3 left). The model association of the flow vectors obtained during the final expectation step are plotted. These associations represent the probabilities that a flow vector originates from an IMO. The sparse data representation is transformed into pixel-maps, which are shown on the right hand side (Fig. 3, right). In the top row a van is overtaking the observer’s vehicle. The middle row is taken from an observer approaching an intersection. In this sequence, the observer moves while another vehicle is moving to the right. All other vehicles are waiting at traffic signs and were therefore not detected as IMOs. In the bottom row a soccer ball crosses the robot’s path. The plots on the right hand side show that the IMOs (cars, ball) are detected correctly by the algorithm. There are some false positives, especially in the freeway and the robot sequence. These yield from incorrectly determined optical flow, which partly occurs because tracked features moved out of the image (e.g. freeway). Partly this is also due to the aperture problem (e.g. robot), which can be solved by advanced motion algorithms [2].

## 5 Conclusion

We presented an algorithmic approach for separating IMOs from background using techniques based on the expectation/maximization principle. The algorithm can handle arbitrary motion including rotation of the observer by iterative estimation of egomotion from optical flow. It features a spatially sparse data representation together with usage of feedback processing, which yields in a reduction of the amount of data points



**Fig. 3.** Results of the proposed algorithm. Left: Input images taken from moving observers (car/robot). Right: Probability of IMO presence. We transform the sparse data by adding gaussians in a pixel-map at the corresponding locations. The height of these gaussians is proportional to the probability of the optical flow of originating from an IMO. The width is proportional to the flow vector's distance to its nearest neighbour.

to be processed over time as well as an enhancement regarding the estimation of the observer's egomotion. The approach only assumes that the majority of estimated flow vectors is induced by the static background.

## 6 Acknowledgements

This work has been supported by a grant from the Ministry of Science, Research and the Arts of Baden-Württemberg (Az: 23-7532.24-12-19) to Martin Claus and Heiko Neumann and a scholarship funded by the University of Ulm granted to Martin Claus.

## References

1. D. Baehring, S. Simon, W. Niehsen and C. Stiller. Detection of close cut-in and overtaking vehicles for driver assistance based on planar parallax. *Proc. Intelligent Vehicles*, pages 289–294, 2005.
2. P. Bayerl and H. Neumann. Disambiguating visual motion through contextual feedback modulation. *Neural Computation*, 16:2041–2066, 2004.
3. A. Dempster, N. Laird and D. Rubin. Maximum likelihood from incomplete data via the em algorithm. *Journal of the Royal Statistical Society, Ser. B* 39:1–38, 1977.
4. M. Isard and A. Blake. Condensation – conditional density propagation for visual tracking. *Intl. Journal of Computer Vision*, 29:5–28, 1998.
5. A. D. Jepson and D. J. Heeger. Linear subspace methods for recovering translational direction. In *Proceedings of the 1991 York conference on spatial vision in humans and robots*, pages 39–62, New York, NY, USA, 1993. Cambridge University Press.
6. N. Krüger, M. Lappe and F. Wörgötter. Biologically motivated multi-modal processing of visual primitives. *Journal of Artificial Intelligence and Simulation of Behaviour*, 1:417–428, 2004.
7. B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Int. Joint Conf. on Artificial Intelligence*, pages 674–679, 1981.
8. J. MacLean, A. Jepson and R. Frecker. Recovery of egomotion and segmentation of independent object motion using the em-algorithm. In *British Machine Vision Conference*, pages 175–184, 1994.
9. D. Marr. *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information*. W. H. Freeman, New York, 1982.
10. K. Pauwels and M. van Hulle. Segmenting independently moving objects from egomotion flow fields. *Early Cognitive Vision Workshop*, 2004.
11. H. S. Sawhney, Y. Guo and R. Kumar. Independent motion detection in 3d scenes. *IEEE Trans. Pattern Anal. Mach. Intell.*, 22(10):1191–1199, 2000.
12. S. M. Smith and J. M. Brady. Asset-2: Real-time motion segmentation and shape tracking. *IEEE Trans. Pattern Anal. Mach. Intell.*, 17:814–820, 1995.
13. E. Trucco and A. Verri. *Introductory Techniques for 3-D Computer Vision*. Prentice Hall PTR, Upper Saddle River, NJ, USA, 1998.
14. F. Woelk and R. Koch. Fast monocular bayesian detection of independently moving objects by a moving observer. In *DAGM-Symposium*, volume 3175 of *Lecture Notes in Computer Science*, pages 27–35. Springer, 2004.