Experts of Probabilistic Flow Subspaces for Robust Monocular Odometry in Urban Areas

Christian Herdtweck
Max Planck Institute for Biological Cybernetics
Tübingen, Germany
Email: christian.herdtweck@tuebingen.mpg.de

Cristóbal Curio
Max Planck Institute for Biological Cybernetics
Tübingen, Germany
Email: cristobal.curio@tuebingen.mpg.de

***Abstract*— Visual odometry has been promoted as a fundamental component for intelligent vehicles. Relying solely on monocular image cues would be desirable. Nevertheless, this is a challenge especially in dynamically varying urban areas due to scale ambiguities, independent motions, and measurement noise. We propose to use probabilistic learning with auxiliar depth cues. Specifically, we developed an expert model that specializes monocular egomotion estimation units on typical scene structures, i.e. statistical variations of scene depth layouts. The framework adaptively selects the best fitting expert. For on-line estimation of egomotion, we adopted a probabilistic subspace flow estimation method. Learning in our framework consists of two components: 1) Partitioning of datasets of video and ground truth odometry data based on unsupervised clustering of dense stereo depth profiles and 2) training a cascade of subspace flow expert models. A probabilistic quality measure from the estimates of the experts provides a selection rule overall leading to improvements of egomotion estimation for long test sequences.**

I. Introduction

Visual odometry aims to recover an agent's motion through the world by means of cameras, which is known to be ill-posed due to the lack of depth knowledge. Although sensors like laser scanners or time-of-flight cameras can bypass large portions of this problem, they have drawbacks like cost, weight, range limits, limited resolution and reflection. GPS would be a viable alternative for pure position and orientation estimation if it was more precise and always available. Inertial measurement units (IMUs) offer a good alternative but are prone to drift. Fusion of these sensors has been an active area of research, but still, many purposeful applications need camera images for other tasks like object detection or further high-level image interpretation.

Visual odometry has recently advanced with new outlier rejection methods and the incorporation of stereo cues with time constraints ([1], [2], [3]).

Along this line, an interesting learning based and versatile approach for outlier rejection has been proposed by [4]. It is based on the observation that if the depth profile, as seen by the camera, is roughly constant over time, the flow field created by a moving agent lies in a very low-dimensional subspace of all possible flow fields [5]. In fact, this has the following interpretation: Flow fields observed by a forward-facing camera in a car moving through an environment with roughly constant depth profile can be represented as a linear combination of a looming flow field and a flow field caused by a rotation (c.f. Fig 3). This approach makes the reasonable assumption that the scene flow has several interdependent causes, like deviations from the depth profile, independently moving objects and erroneous measurements. Their approach sets these causes apart from inlier flow vectors with a robust variant of probabilistic PCA (pPCA, [6]) that is trained using an Expectation-Maximization (EM) scheme. Inferred coefficients of the low-dimensional subspace flow representation are linearly related to the movement of the observing platform. The approach has the additional advantage that arbitrary camera configurations can be used without calibration, including multiple-camera systems, fish-eye lenses and catadioptric setups.

This and other monocular approaches rely on the depth constancy assumption which is likely to be violated in structured urban areas. Our approach to on-line motion estimation from camera images directly addresses the ill-posed nature of camera data. We use unsupervised learning methods at different stages to incorporate knowledge about the scene structure and apply this knowledge adaptively. Instead of using further specialized sensors, we are motivated to approach the underlying estimation problem using a minimum number of cues with the intention that a more versatile image interpretation can emerge from the solution.

In this paper we present a framework consisting of two contributions: 1) we assume different amounts of variation of flow within the image; preliminary tests showed that flow vectors in the lower left and right image corners of a forward-facing camera stem from close-by objects and therefore have greater amplitude and variability; 2) we further address the constant-depth assumption by training several estimators, each specialized on a characteristic scene depth profile, and each yielding a motion estimate with an associated quality measure.

We acquire context knowledge with the help of the K.I.T. datasets of video & ground truth odometry data ([7], [8]) by unsupervised clustering of stereo depth profiles, and train a cascade of subspace flow expert models. This way we implicitly make use of depth cues in the optical flow without explicitly calculating them on-line. We show results that can compete with a state-of-the-art visual odometry system.

The remainder of this paper is structured as follows: after putting our method in context in section II, we describe our framework in more detail in III. In section IV we evaluate

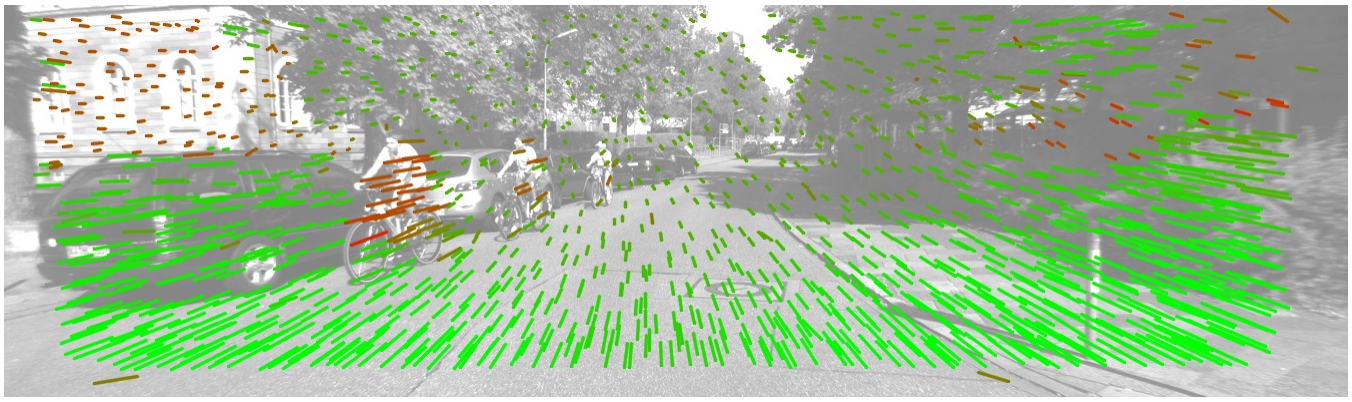


Fig. 1. Example frame from a test sequence captured from a forward-facing camera in a car, with optical flow overlaid. Our model assigns high probabilities to the flow vectors being produced by self-motion (green) and low probabilities for other causes (red). Causes could be erroneous correspondence (e.g. on the right), independently moving objects (e.g. bicycles on the left) or deviations from the constant depth profile (e.g. the building in the background on the left).

the performance of our method including comparisons to other visual odometry approaches. We summarize and point to further research directions in section V.

II. Related Work

Our approach for probabilistic and monocular visual odometry borrows ideas from ensemble learning ([9]) and transfer learning ([10]). Ensemble learning has been applied in machine vision for detecting human poses from monocular images [11], as well as pedestrian detection and view classification ([12], [13], [14], [15]) for single and sequences of images. Romer and Rosales [11] estimate an ensemble of regression experts, each specializing on smaller domains of feature input spaces and a mapping through regression to corresponding pose hypotheses. In a similar context, classification experts are trained to enhance the overall pedestrian classification rate with view-depending classification experts ([12], [13]). A more compact model has been presented, that exploits mixture-of-expert models for pedestrian classification considering shape information and view-points ([14], [15]). In our paper we do not exploit stereo cues directly like [2]. Instead we adopt the approach of Roberts *et al*. [4] for monocular egomotion estimation. This approach, like any other monocular algorithm, is prone to violations of depth constancy, but models outliers caused by e.g. independently moving objects in a probabilistic subspace model. In particular, we use the approach to learn experts of egomotion estimators adapted to different depth profiles. As such, we exploit the implicit knowledge of depth cues by sorting the training data for our expert models according to clusterings of scene depth profiles that we obtain with the publicly available stereo code for rectified images from K.I.T. (*libelas*) [3]. Our approach is new in that we use characteristic scene depth layouts, obtained with unsupervised learning, just to define the domains of training samples that the experts should be responsible for. They are trained on these domains with the corresponding flow as feature input.

III. Methods

In this section we describe our approach in detail. We give an overview of the work flow of our framework in Fig 2.

A. Optical Flow Extraction

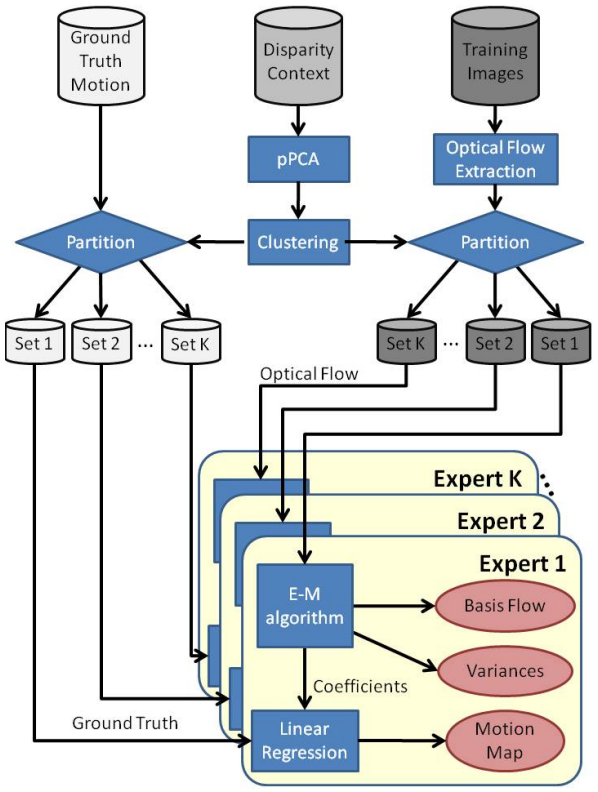
For optical flow extraction we use the publicly available library *libviso2* from K.I.T. [16]. We divide the image into a grid of non-overlapping cells of size 20×20 pixels. For each cell *libviso2* provides at random one flow vector, which overall results in a flow field $\mathbf{f}_t$. We chose this grid cell size trading off large training efforts with possibly many missing values from small grid cells against high variance of flow vectors as consequence of larger cells.

B. Flow Subspace Representation

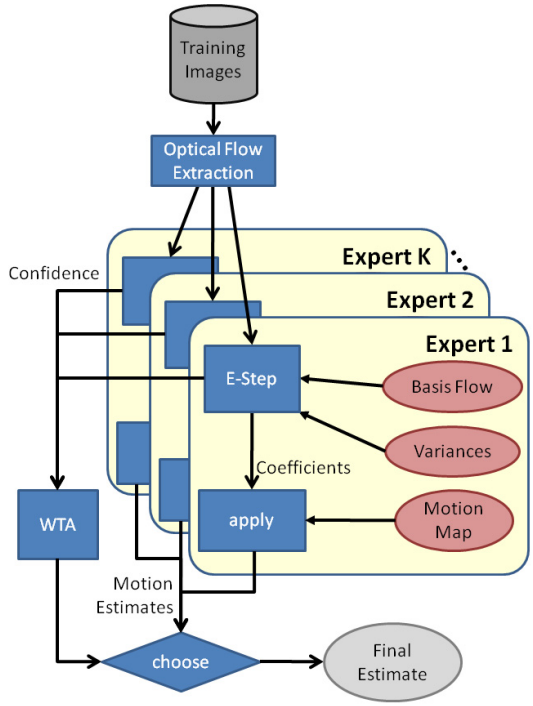
Our framework is based on the learnt robust decomposition of optical flow vectors, $\mathbf{f}_t$, into outlier and inlier flow vector components [4]. In this probabilistic approach inlier flow is modeled as a linear combination of basis flow vector fields $\mathbf{B} = (\mathbf{B}_1, \cdots, \mathbf{B}_J)$ and added noise, while outlier flow is modelled as pure noise

$$\mathbf{f}_{ti} = \begin{cases} (\mathbf{B}\mathbf{x}_t)_i + \mu_i + {\epsilon^{\text{in}}}_{ti} & \text{if } \mathbf{z}_{ti} = 1 \\ {\epsilon^{\text{out}}}_{ti} & \text{if } \mathbf{z}_{ti} = 0 \end{cases} . \tag{1}$$

Flow vector components are indexed by $i = 1, \ldots, I$, and frame numbers are denoted by $t = 1, \ldots, T$ throughout the paper. The binary latent variables $\mathbf{z}_{ti}$ determine whether a measured flow vector component is regarded as an inlier ($\mathbf{z}_{ti} = 1$) or not. Outliers are caused by false flow measurements, independently moving objects or deviations from the expected depth profile (c.f. Fig. 1). They are modeled by a zero-mean Gaussian error ${\epsilon^{\text{out}}}_{ti}$ with mean 0 and variance σ^2_{out}. Inlier flow is modeled with error $\epsilon^{\text{in}}_{ti} \sim \mathcal{N}(0, \sigma_{\text{in},i})$ with either a different variance $\sigma^2_{\text{in},i}$ for each flow vector component (inhomogeneous case) or with same variance $\sigma^2_{\text{in},i} = \sigma^2_{\text{in}}$ for all components i (original, homogeneous case described in [4]).



(a) Training procedure of our system



(b) Performing inference with our system

Fig. 2. System overview for training (a) and inference (b).

During training (c.f. Fig. 2(a)), basis flow fields $\mathbf{B}_1, \ldots, \mathbf{B}_J$, coefficients $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_t)$, inlier assignments $\mathbf{z} = (\mathbf{z}_1, \ldots, \mathbf{z}_t)$, as well as error variances $\sigma_{\text{in},i}, \sigma_{\text{out}}$ have to be optimized together from a set of observed flow fields. To perform this optimization, Roberts *et al.* [4] propose an Expectation-Maximization (EM) scheme based on probabilistic PCA (pPCA) [6]. The probabilistic model underlying pPCA is extended in two ways: it is modified to handle missing values and outliers using equation 1. The authors formulate the optimization as a maximum-likelihood problem, from which they derive update equations for the EM algorithm. Our implementation follows their formulae which we do not repeat here. We just specify here the modified formula for our extension of training individual variances σ_i for each flow vector component i

$$\sigma_{\text{in},i}^2 = \left(\sum_{t \in S_i} \mathrm{E}(\mathbf{z}_{ti}) \right)^{-1} \sum_{t \in S_i} \mathrm{E}(\mathbf{z}_{ti}) \left(y_{ti}^2 - 2 y_{ti} \mathbf{B}_i \mathrm{E}(\mathbf{x}_t) + \mathrm{trace}(\mathrm{Cov}(\mathbf{x}_t) \mathbf{B}_i^T \mathbf{B}_i) \right) \quad . \quad (2)$$

Here, S_i is the set of flow vectors for which flow vector component i is observed (i.e., not a missing value), $y_{ti} = f_{ti} - (\sum_{j=1}^{d} \mathbf{B}_{ji} \mathbf{x}_{tj} + \mu_i)$ is the deviation of the observed flow from a flow field synthesized from basis flow fields, and $\mathrm{E}(\mathbf{x}_t)$ and $\mathrm{Cov}(\mathbf{x}_t)$ are mean and covariance of the coefficients $\mathbf{x}_t$. The expectation $\mathrm{E}(\mathbf{z}_{ti})$ can be viewed as the probability for flow vector component $\mathbf{f}_{ti}$ being an inlier, and is referred to as the *inlier probability*.

We thus arrive at an EM scheme that alternates between updating the assignments $\mathbf{z}$ and coefficients $\mathbf{x}$ in the E-step and optimizing the remaining model parameters in the M-step. After convergence of the EM procedure we orthogonalize and normalize the obtained basis flow vectors to make the subspace representation more stable. During inference (see also Fig. 2(b)), only the E-step has to be performed, which requires a few iterations because subspace coefficients and inlier assignments are co-dependent. Figure 1 visualizes inlier probabilities assigned to a sample frame.

C. Expert System Training

A general drawback of this monocular approach is, that it requires flow originating from a scene profile that stays roughly constant over time for training and on-line inference. Static objects and geometry can alter the depth profile, resulting in different flow; objects rendering own motion further alter the ego-motion induced flow fields, especially if they get close. The statistical approach described in the last section can deal with these cases as long as those outliers form a minority of the flow in the image or do not conform to the same wrong motion estimate. If, however, the system encounters a large field of consistent outlier flow, caused e.g. by a crossing bus occupying more than half of the image, the system may be fooled to take those flow vectors as inliers, resulting in a wrong motion estimate.

Our working hypothesis is to increase robustness of performance through flexibility without losing specificity.

Thus, trained estimators are specialized on different situations. Each of these estimators is specialized implicitly in a prototypical scene depth profile and is referred to as an *expert*. For defining these 'situations' during training, we exploit additional data in the form of disparity maps that we calculate from stereo image pairs using open library *libelas* [8]. Since we only need a very rough idea about the statistical distribution of depth profiles, we save computational effort by subsampling the disparity images of every tenth frame by a factor of three. We further reduce their dimensionality using the probabilistic missing-value algorithm pPCA by Verbeek[1]) with ten principal components (c.f. Fig. 2(a)). The resulting projections $\mathbf{d}_t, t = 1, \ldots, D$ of the disparity images onto the lower-dimensional PCA space are then clustered by a Mixture of Gaussians (MoG) model ([17]) with parameterization Θ consisting of K multidimensional Gaussian distributions with means $\mathbf{m}_1, \ldots, \mathbf{m}_K$ and general covariance matrices $\mathbf{C}_1, \ldots, \mathbf{C}_K$:

$$p(\mathbf{d}_t|\Theta) = \sum_{k=1}^{K} \pi_k N(\mathbf{d}_t; \mathbf{m}_k, \mathbf{C}_k) \quad , \tag{3}$$

where $\sum_{k=1}^{K} \pi_k = 1$. We assign images based on their corresponding embedded disparity profiles $\mathbf{d}_1, \ldots, \mathbf{d}_T$ to one of the resulting 'situation', i.e., mixture components $(\mathbf{m}_k, \mathbf{C}_k), k = 1, \ldots, K$, except for a few cases where the assignment to a component is not clear (i.e., the entropy over the component membership exceeding a threshold). Subsequently, the resulting K subsets of training frames are used to train egomotion estimators as described above, resulting in K estimators, each with their own set of basis flow vectors, inlier variance and motion map.

For the testing phase we select for each time step, t, the estimate from the expert k_t, that can best explain the measured flow. This is determined by calculating on-line the mean inlier confidence of each expert. We deliberately chose a winner-takes-all (WTA) selection criterion

$$k_t = \text{argmax}_{k=1,\ldots,K} \frac{1}{I} \sum_{i=1}^{I} \text{E}(\mathbf{z}_{tik}) \quad , \tag{4}$$

where $\text{E}(\mathbf{z}_{tik})$ is the probability that expert k considers flow vector component i of frame t as an inlier i.e., caused by self-motion.Note that the choice of expert is done purely on the confidences returned by the experts from monocular input, no disparity or other additional data is needed during on-line inference.

D. Mapping to Motion

The final step in our system pipeline is the mapping from inferred subspace flow coefficients to incremental platform motion. To train this linear mapping we use iteratively reweighted least squares with a bisquare weighting function. Since each of our experts works on a different flow subspace, the motion mapping differs for all of them.

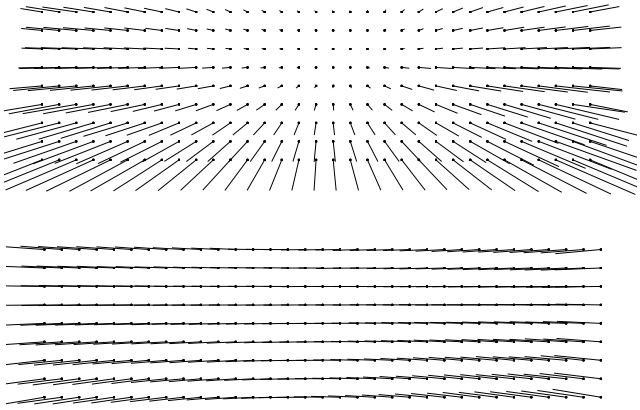


Fig. 3. Basis flow fields of an estimator trained with inhomogeneous inlier variance. **Top**: first basis component corresponding to a strong forward motion. **Bottom**: second basis flow field, corresponding to a rotation with slight backward motion.

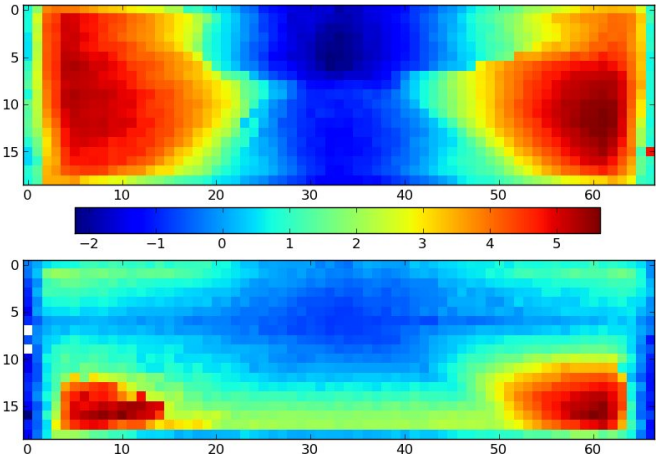


Fig. 4. Estimated inlier variances in horizontal (**top**) and vertical (**bottom**) flow vector components of an estimator trained with inhomogeneous inlier variance $\sigma^2_{\text{in},i}$. As expected the variance is much higher in the lower image corners than in the middle. Colors are on a logarithmic scale.

IV. Evaluation

For evaluating our system we performed training and inference on a set of rectified images captured by a car in an urban environment [7], [8] at approximately 10 frames per second ($\rightarrow$ dt $\approx$ 0.1s). The car was equipped with GPS and an IMU to obtain ground truth motion and position information, as well as a stereo camera rig for estimating disparities that we used for defining our experts. To keep the data consistent we worked on five sequences of the first day (sequences 2009_09_08_drive_0010, 15, 16, 19, and 21), dividing each sequence into a first half for training and a second half for testing. Overall, we arrived at a training and testing sequence (3201 frames). As ground truth, we preferred the IMU data over the GPS since that seemed to be more consistent with motion estimates in a preliminary test.

[1] http://lear.inrialpes.fr/~verbeek

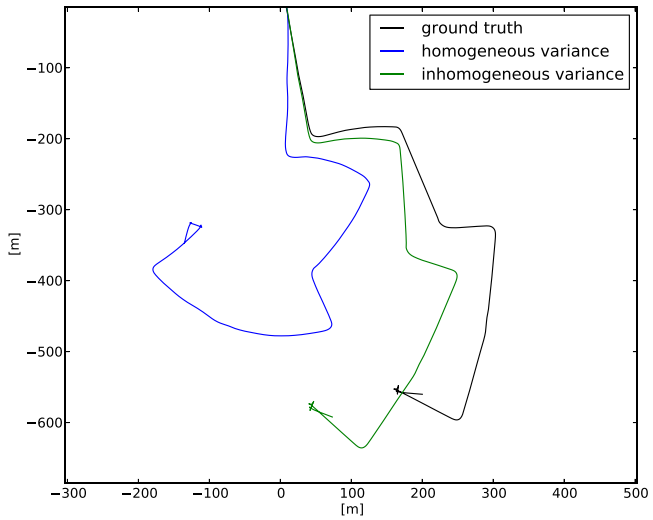


Fig. 5. Estimated trajectories of single estimators trained with homogeneous (blue) and inhomogeneous (green) inlier variances, ground truth (black).

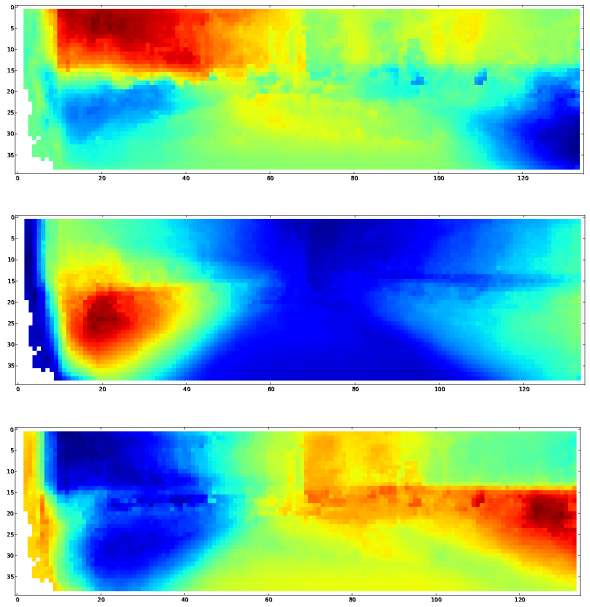


Fig. 6. Means of disparity clusters shown as deviations from the overall disparity mean. Blue corresponds to low disparities i.e., high distance, red to close objects. Clusters 2 and 3 correspond to the typical road depth layout with the difference that an expert trained on cluster 2 (**middle**) expects closer objects in the lower left than usual (e.g. approaching or parked cars) while the expert trained on cluster 3 (**bottom**) represents the opposite: lower disparities on the left correspond to a relatively wide road, higher disparities on the right require parked cars or other objects there.

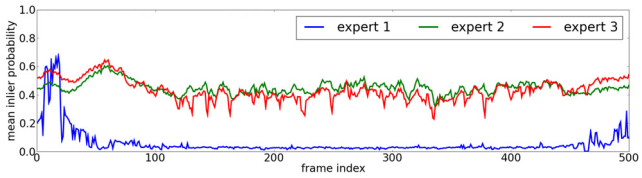


Fig. 7. Expert confidences of a 3-expert system trained on depth profiles in Fig. 6 on the first 500 frames of the test set.

We smoothed x,y,z and yaw data and calculated speed as the euclidean distance between x,y,z values of successive frames. During inference we measured the motion error as the root mean squared difference between our system's estimate and the ground truth for rotation and forward motion.

A. Single estimators

As first evaluation we trained only a single estimator with homogeneous and inhomogeneous inlier variance on our training set and did an analysis of performance on the test set. The flow basis fields learnt by the estimator with homogeneous inlier variance are depicted in Fig. 3, maps of the inhomogeneous inlier variance of the other estimator are shown in Fig. 4. Horizontal and Vertical flow vector components are shown at all flow grid cells. Figure 5 shows the trajectories resulting from inference on the test set. The root mean squared error per frame of the estimator with homogeneous inlier variance was $0.050\,\frac{\mathrm{m}}{\mathrm{dt}}$ in speed and $1.59 \times 10^{-3}\,\frac{\mathrm{rad}}{\mathrm{dt}}$ in angular velocity. Training an estimator with inhomogeneous inlier variance in the same way reduced those errors to $0.021\,\frac{\mathrm{m}}{\mathrm{dt}}$ and $5.29 \times 10^{-4}\,\frac{\mathrm{rad}}{\mathrm{dt}}$, respectively, confirming that the introduction of inhomogeneous inlier variance paid off.

B. Expert System

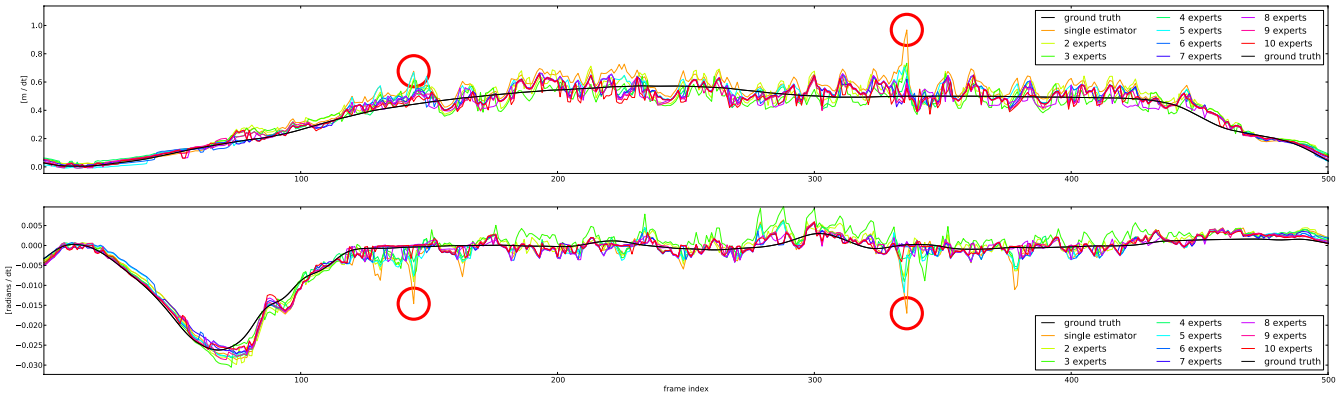
Next, we put our expert system to the test. As a first evaluation we inspected the means of the Gaussian mixture components, back-projected into the disparity space as shown in Fig. 6 for the example of a 3-expert system. The experts seem to have specialized in clearly distinct depth profiles. Fig. 7 shows the confidences of these three experts on the first 500 frames of the test set.

Fig. 8 illustrates two interesting failure cases of a single estimator that can be resolved using an expert system.

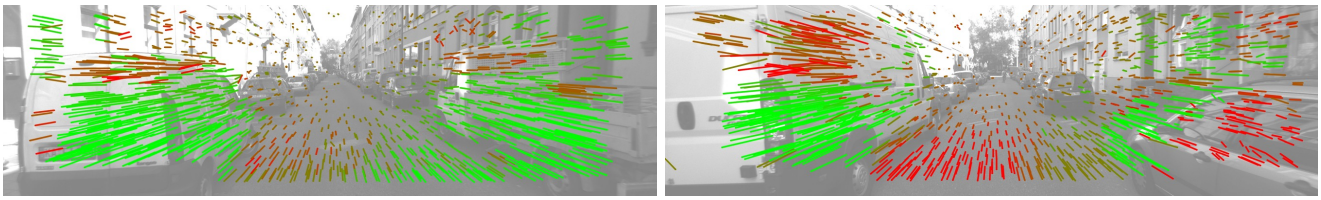
In expert systems the choice of number of experts i.e., the model complexity, is critical. Therefore, we tested the effect of expert number K on mean squared motion error in forward and rotation motion and compared the results to the ground truth and to a single estimator.

Fig. 9 shows an overview. In the homogeneous case, the trend seems to be that more experts perform better for rotation. For speed, the error shows clear improvement only for 2 and 3 experts. For higher model complexities, it stays constant but is still below the single-estimator error. In the inhomogeneous case, there seems to be an optimum of 3 experts, after which the error raises with number of experts and even goes considerably above the single-estimator error. This effect might have to do with the drastically higher number of variables in the inhomogeneous case. This might have lead to an overfitting for clusters with smaller sample numbers and the training data available.

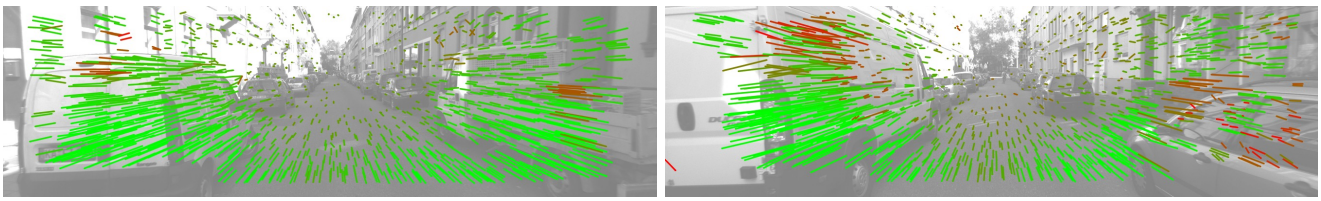
To show that our expert system captures the relevant depth information, we compared the choice of expert made by our system based on the flow data alone with the choice that would result from the full disparity data. Experts trained



(a) Estimates of speed (**top**) and rotation (**bottom**) of single-estimator (orange), expert systems of different numbers of experts, and ground truth (black) on part of the test data. In this example, expert systems and single estimator were trained with homogeneous inlier variance.The red circles show failure cases of the single estimator that are detailed in the following.



(b) Flow fields for failure cases in frames 144 (**left**) and 336 (**right**) indicated by red circles above. Flow vector color corresponds to inlier probability assigned by a single estimator: green = caused by egomotion, red=outlier. In both examples the flow vectors on the left and right cars are much longer than one would expect for the observer's true speed since the cars violate the expected depth profile. The estimator therefore misinterprets the flow as caused by a quick right turn and discards the flow on the street as outliers.



(c) Same flow fields as above but with inlier probability assigned by our 10-expert system which is able to cope with this situation since one of the experts was trained on depth profiles that roughly correspond to this geometric layout. That expert is therefore able to correctly interpret the observed flow and give a very confident and more accurate motion estimate.

Fig. 8. Time course of motion estimates (a) and details for two failure cases of the single-estimator system (b) that are resolved by an expert system (c)

with inhomogeneous inlier variance were correctly chosen for 73.60% of frames of our testing sequence when using 3 Experts. For higher numbers of experts that number declined to e.g. 28.12% with 10 experts. Using experts with homogeneous inlier variance showed similar results. This shows that our monocular system is able to 'hallucinate' depth — or the relevant aspect of depth — to a certain extent.

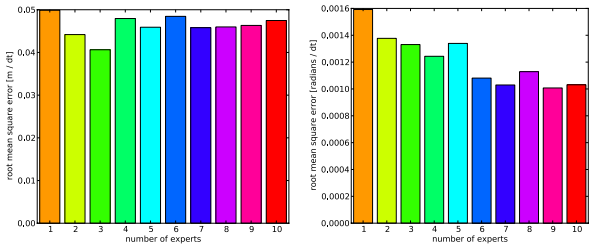
C. Comparison to Other Approaches

To show that our approach can compete with current visual odometry approaches, we compared the trajectories estimated by our system with those estimated with the system by Geiger *et al.* [16]. Fig. 10 shows the resulting trajectories for homogenous (10(a)) and inhomogeneous (10(b)) estimators. We compared a single estimator and our system with model complexity of $K = 3$ and $K = 7$ experts with the Geiger approach using mono and stereo information as well as the ground truth. As expected, the stereo algorithm outperforms our system, since it makes excessive use of stereo information data in the estimation. The mono version seems to yield results with larger errors. However we have to mention that this code is labelled 'experimental'. For comparison, we only show the x and y component of the full 6 degree-of-freedom results that [16] produces.

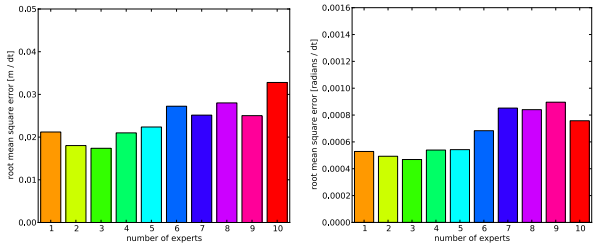
V. Conclusion and Outlook

To summarize, we have presented a system for monocular odometry that uses a system of experts, which allows for a robust interpretation of flow by decomposing optical flow into basis flow fields and linearly mapping those to platform motion. We have detailed our expert system as well as our extension to the probabilistic PCA and outlier model from [4]. Evaluating our system shows that our extensions lead to lower errors in estimated platform motion, comparable to state-of-the-art self-motion estimation approaches.

We plan to extend our system in several respects. The selection process of experts can surely be improved upon. Using more training data reduces overtraining problems e.g.



(a) Results of experts trained with homogeneous inlier variance.



(b) Results of experts trained with inhomogeneous inlier variance.

Fig. 9. Effect of model complexity (number of experts) on the system's root mean squard error for forward translation (**left**) and rotation (**Right**), evaluated on the test set. '1 expert' are the results of a single estimator.

in cases of inhomogeneous inlier variance with many experts. Stochastic temporal modeling and model selection might be an interesting direction to investigate on top of our approach. Constructing high-level cognitive loops comprising knowledge about outlier flow and detected objects might pay off since these two sources of information might be beneficial to both, object detection and identification of outlier flow.

REFERENCES

[1] A. Wedel, C. Rabe, T. Vaudrey, T. Brox, U. Franke, and D. Cremers, "Efficient Dense Scene Flow from Sparse or Dense Stereo Data," in *ECCV*, ser. LNCS, D. Forsyth, P. Torr, and A. Zisserman, Eds., vol. 5302, no. 1, University of Auckland. Computer Science Department, The University of Auckland, New Zealand, 2008, pp. 739–751.

[2] H. Badino, U. Franke, and D. Pfeiffer, "The Stixel World - A Compact Medium Level Representation of the 3D-World," in *Pattern Recognition*, ser. Lecture Notes in Computer Science, J. Denzler, G. Notni, and H. Süße, Eds. Springer Berlin / Heidelberg, 2009, vol. 5748, pp. 51–60.

[3] A. Geiger, M. Roser, and R. Urtasun, "Efficient Large-Scale Stereo Matching," in *Asian Conference on Computer Vision*, ser. Lecture Notes in Computer Science, R. Kimmel, R. Klette, and A. Sugimoto, Eds., vol. 6492. Berlin, Heidelberg: Springer Berlin Heidelberg, 2011, pp. 25–38.

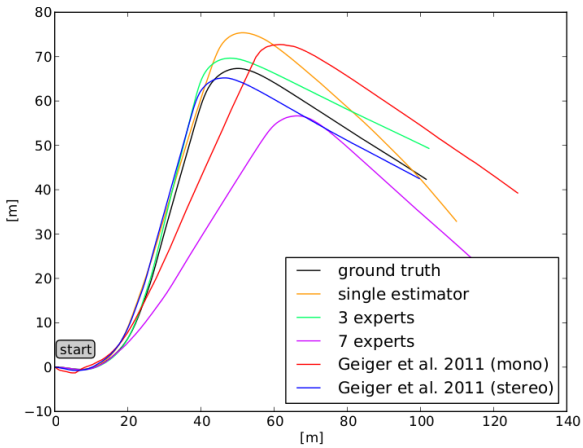
[4] R. Roberts, C. Potthast, and F. Dellaert, "Learning general optical flow subspaces for egomotion estimation and detection of motion anomalies," in *IEEE Conference on Computer Vision and Pattern Recognition*. Miami: IEEE, Jun. 2009, pp. 57–64.

[5] M. Irani, "Multi-frame optical flow estimation using subspace constraints," in *Proceedings ICCV*, vol. 1. IEEE, 1999, pp. 626–633.

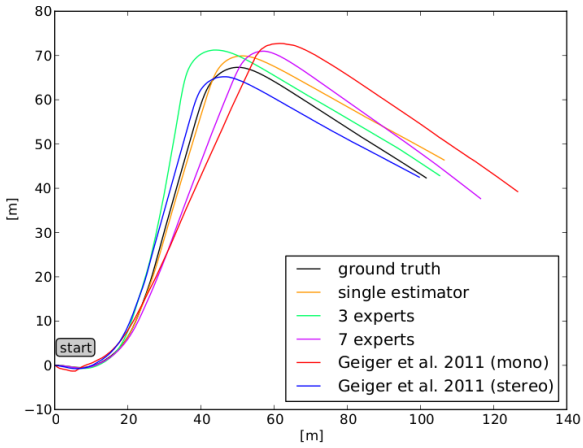
[6] M. E. Tipping and C. M. Bishop, "Probabilistic principal component analysis for metabolomic data." *Journal of the Royal Statistical Society, Series B (Statistical Methodology)*, vol. 61, no. 3, pp. 611–622, Jan. 1999.

[7] B. Kitt and A. Geiger, "Visual odometry based on stereo image sequences with RANSAC-based outlier rejection scheme," *Intelligent Vehicles Symposium*, 2010.

[8] A. Geiger, M. Roser, and R. Urtasun, "Efficient Large-Scale Stereo Matching," in *Asian Conference on Computer Vision 2010*, ser. Lecture Notes in Computer Science, R. Kimmel, R. Klette, and A. Sugimoto, Eds., vol. 6492. Berlin, Heidelberg: Springer, 2011, pp. 25–38.



(a) Comparing our homogeneous-variance system with [16].



(b) Comparing our inhomogeneous-variance system with [16].

Fig. 10. Trajectories estimated by our system using estimators with homogeneous (**top**) and inhomogeneous (**bottom**) inlier variances. The shown trajectories are for a single estimator (orange), our system with 3 (mint) and 7 (purple) experts, the system by Geiger *et al*. using mono (red) or stereo information (blue) and ground truth (black). Test sequence is the second half of 2009_09_08_drive_0016.

[9] S. J. Pan and Q. Yang, "A Survey on Transfer Learning," *IEEE Transactions on Knowledge and Data Engineering*, vol. 22, no. 10, pp. 1345–1359, 2010.

[10] G. Brown, "Ensemble learning," in *Encyclopedia of Machine Learning*, 1st ed., C. Sammut and G. Webb, Eds. Springer, 2010, ch. 400.

[11] R. Rosales and S. Sclaroff, "Learning body pose via specialized maps," in *NIPS*, 2001, pp. 1263–1270.

[12] C. Nakajima, M. Pontil, B. Heisele, and T. Poggio, "Full-body person recognition system," *Pattern Recognition*, vol. 36, no. 9, pp. 1997–2006, 2003.

[13] H. Shimizu and T. Poggio, "Direction estimation of pedestrian from multiple still images," in *In Proc. IEEE Intelligent Vehicles Symposium 2004*, 2004, pp. 596–600.

[14] D. M. Gavrila and S. Munder, "Multi-cue pedestrian detection and tracking from a moving vehicle," *International Journal of Computer Vision*, vol. 73, no. 1, pp. 41–59, 2007.

[15] M. Enzweiler and D. M. Gavrila, "Integrated pedestrian classification and orientation estimation," in *CVPR*, 2010, pp. 982–989.

[16] A. Geiger, J. Ziegler, and C. Stiller, "StereoScan: Dense 3d Reconstruction in Real-time," in *IEEE Intelligent Vehicles Symposium*, 2011.

[17] C. M. Bishop, *Pattern recognition and machine learning*, M. Jordan, J. Kleinberg, and B. Schölkopf, Eds. Springer New York, 2006, vol. 4.