Deliverable D5.2
Task optimized representation and the center of interest

1 Introduction

The extraction of information from the motion field estimated from camera sequences can be split in several modi or tasks, estimate the parameters of self motion, identify obstacles, track the center of interest, detect independent moving objects, determine the structure of the environment, etc. Motion fields estimated from camera sequences are often very noisy and affected by the aperture problem and to extract optic flow information requires the processing of the motion field. Depending on the information to be extracted from the motion field, different processing modi are required. For determining the global parameter of ego motion noise can be strongly decreased by bundling motion information from large domains of the visual field. By this method local information is several reduced but the global motion pattern is preserved. For identifying possible obstacles or for detecting independent moving objects local motion information is very important and the prior processing of the motion field can only take into account small domains of the visual field for a weaker noise decreasing, but stronger preservation of the local properties of the motion field. Here we focus on the first task, global motion estimation.
The present space variant filter model is optimized to extract the global motion pattern from motion fields estimated from camera sequences. We use the task of heading detection from optic flow as a way to estimate improvement of the flow field. Our filter model adopts properties of a particular motion sensitive area of the brain which averages the incoming motion signals over receptive fields whose size increases with the distance from the center of the projection. This representation is optimized for motion fields elicited by forward ego- motion as in car driving situations and in combination with tracking movements that stabilize the center of interest in the field of view. The tests are conducted with two different sets of flow fields. The first set is calculated with a standard flow algorithm from image sequences recorded by a camera installed in a moving car. The second set is derived from a data base containing 3D data and reflectance information from natural scenes with full control of camera motion and ground truth of the flow field and the heading. We test the filtering method by comparing heading estimation results between filtered and raw flow for both sets of flow fields. Due to noise and the aperture problem the results for the raw flows are often unreliable. Estimated heading differs widely for different sub-sampled calculations. In contrast, the results obtained from the space- variant filtered flows are much less variable and therefore more confidential. We suggest extensions to this scheme that takes other properties of the motion representation in this area into account to further improve the flow representation.

2 Global motion patterns

The patterns of optical flow fields elicited on optical detectors of visual systems by self motion encode much information about the direction of motion, the velocity, and the direction and magnitude of camera rotation, and the position and relative velocity of independent moving objects. Biological systems use such information for path planning, obstacle avoidance, ego-motion control and foreground-background segregation [11, 20]. One wishes to make vision based technical applications

like driver assistance systems or autonomous robots also capable to use optic flow information to solve these tasks. For instance, the fast and correct assessment of the current driving situation by a driver assistance system could help to make appropriate decisions and to avoid accidents. Such a driver assistance system capable to estimate heading could warn the driver about any sideward drifting and unintended lane changing on slippery roads. Furthermore, the correct estimation of the parameters of self-motion is required for the task of identification of independent moving objects like other cars and pedestrians from optic flow ([17]).

Optical flow fields obtained from flow algorithms applied to camera image sequences are usually very noisy [8, 4]. Optic flow estimation is plagued by ambiguities due to the aperture problem, the correspondence problem, noise inherent in the algorithms due to resolution and quantization effects, lack of signal in homogeneous image areas, and ambiguities at depth discontinuities. Therefore, estimates of self motion from optic flow are error prone. Since motion estimation in biological systems also starts from spatiotemporal variations of image intensity it is likely to face the same problematic optic flow on the input stage. But biological systems estimate self motion rather exact (overview in [13]). Somehow, therefore, the brain must have developed methods to remedy the shortcomings of the early flow detectors. We searched for features which enable biological vision systems to handle noisy flow fields successfully.

In the visual system of primates, the raw flow field, i.e., the collection of early image motion signals, is measured by motion sensitive cells in the primary visual cortex (V1). Their output is further processed by motion sensitive cells in the medial temporal area (MT). From area MT the signals are transferred to the medial superior temporal area (MST), which is thought to analyse the entire optic flow pattern and to extract the parameters of ego-motion [12]. Area MT establishes a space variant map of the visual motion field, i.e the size d of the receptive fields of the neurons in MT increases proportionally to the eccentricity ϵ from the center of the field of view [1].

$$d = 0.018 + 0.61\epsilon. \tag{1}$$

Based on the properties of this map Lappe [10] proposed a method to decrease the noise of the optical flow by averaging flow vectors over image areas which increase in size with eccentricity ϵ from the center of the field of view. In this model, optic flow is represented in a population code over direction selective neurons at any position in the visual field. The motion signal in the center of the receptive field of a single neuron is derived from the average of all V1 motion signals within its receptive field. In the center of the visual field, the integration is restricted to a small area. In the periphery the integration area becomes very large (Figure 6).

The spatial integration over peripherally increasing image areas is well adjusted to the typical structure of the flow field. For predominantly forward motion and restricted camera rotation the singular point of the optic flow, i.e., the point with vanishing flow, is usually near the center of the visual field [14]. The singular point is the center of a basically radial structure of the flow field, which may however be distorted by superimposed rotation of the camera. Therefore, small areas surrounding the center of the flow field contain sets of vectors with large deviations in the local flow direction. The periphery of the flow field is more homogeneous allowing spatial averaging over a large scale without loosing to much information. In human vision this is true even when the direction of heading deviates from the direction of gaze since eye rotation reflexes in this case

introduce rotational flow that nulls the motion in the direction of gaze [13, 14]. Since averaging over large areas is more favorable for noise reduction and smoothing while averaging over small areas retain information if neighbored signals are different, the space variant mapping is a compromise that satisfies both goals to the degree necessary for the structure of the flow field.

In [10] one can find an application of this method, and an implementation of a standard heading detection algorithm [6] in terms of a neural network model. This network was tested with artificial motion fields of simulated self movements through three dimensional random scenes. The flow fields contained translation and rotation components, and uniform noise was added. The results show that noise is reduced, and heading detection is possible with errors up to 4 degrees for a signal to noise ratio of 1, similar to human performance with identical stimuli in psychophysical experiments [21, 22]. These results demonstrate that the space-variant filtering stage is a sensible model for the flow representation in the primate visual system and that it can capture performance characteristics of the human visual system. But, it remains unclear whether this method is beneficial for the processing of flow fields derived from camera images. The noise that is inherent in motion fields derived from camera images is often very different from the uniform noise that was artificially added in model simulations and in the psychophysical and physiological studies that the model was based upon. Furthermore, for the types of motion performed by a rigidly installed camera in a moving car, for example, the method might be suboptimal when the singular points is shifted far away from the center of view.

The present paper is concerned with the examination of the improvement of noisy flow fields derived from image sequences by space-variant filtering. Since we concentrate on the processing of flow fields for egomotion estimation, we use the task of heading estimation from the filtered flow fields as a criterion for improvement. The space variant filtering method is compared with results obtained from the raw, unfiltered motion field. We start our investigation with image sequences recorded by a camera rigid installed behind the wind shield of a moving car. The car is moving forward and encounters rotational motions caused by steering in curves or bumps in the road. Without ground truth about the correct heading, our main criterion to evaluate the efficiency of the space variant filtering method is the consistency of the heading estimation for subsampled flow fields and the constancy of the expectation value for heading over time. Since we can roughly see the original motion of the car from the image sequence gross errors in the heading estimation would be observable. To better quantify improvements with respect to true heading, for the second part of our analysis we use image sequences constructed from three dimensional data sets of several natural scenes. Image sequences are directly calculated from the scene data assuming a given self motion. Flow fields are calculated from the image sequence. The advantage of this procedure is that we have the exact knowledge about the correct heading direction of the simulated self movements – and of the true flow field –and are thus, able to evaluate the results of the filtering method against ground truth.

3 Space-variant filtering

3.1 Estimation of optic flow from image sequences

Since many algorithms and models of motion estimation have been proposed (cf. [4]), the first step in our analysis involves the choice of a flow algorithm to begin with. For several reasons, we have chosen a version of the Lucas-Kanade flow estimation algorithm [15] as the input stage to the filtering procedure. First, many flow algorithms that are based on differential techniques are modifications of the Lucas-Kanade algorithm [2]. Second, the Lucas-Kanade algorithm shows the typical range of errors caused by the aperture problem, numerical problems, weak contrasts, etc., which occur in optic flow estimation from image sequences. Third, the algorithm is simple, requires only a few samples in space and time, and generates dense flow fields which are well suited for the filtering procedure. The algorithm is not intended as a model for early stages of the visual system but rather as a generic representative of early motion estimation.

Our implementation of the Lucas-Kanade algorithm follows the description in [4]. We use a spatial neighborhood of 3×3 pixel to calculate the spatial luminance gradient, and a temporal neighborhood of 3 frames for the temporal derivative. The spatial neighborhood for the least-squares minimization is 4×4 pixels. In certain degenerate cases the algorithm will not find a solution. In these cases, the respective pixels are omitted from the subsequent analysis. This also applies to image areas with very low contrast. As this happens only infrequently, the resulting flow field is dense, but noisy (second row Fig. 2).

3.2 Mathematical description of the filtering method

The space variant filtering is achieved by averaging flow vectors over domains of position depending sizes. The filtering method adopts the domain sizes from eq. (1). This has to be converted into a pixel based formulation. Let f be the focal length and (i, j) be a position in a pixel composed image and (px, py) the principal point of the image. For eccentricities smaller than 30 deg the radius r in pixel can be approximated as

$$r(i, j) = 0.009f + 0.4\sqrt{(i - px)^2 + (j - py)^2}. \tag{2}$$

To generate the filtered flow from the raw flow a subset of positions in the pixel grid is selected, mainly to reduce calculation time and because the heading estimation procedure is based on a subsampling of the flow field anyways. At these positions, the filter procedure is performed over all signals of the original pixel grid falling into the averaging domain. Mathematically, this can be described as follows: Let $W \times H$ be the dimension of the image and $P = \{(i, j)\}_{i=1,...,W;\ j=1,...,H}$ be the set of pixel positions. Further let $S = \{(\sigma, \rho)\} \subset P$ be a subset of P. Let $P \supset \tilde{P}(\sigma, \rho) = \{(i, j) \in P | \sqrt{((i - \sigma)^2 + (j - \rho)^2)} < r(\sigma, \rho)\}$. The filtered vector $v^f(\sigma, \rho);\ (\sigma, \rho) \in S$ is determined by

$$v^f(\sigma, \rho) = \frac{\sum_{(i,j)\in\tilde{P}(\sigma,\rho);\|v^r(i,j)\|>0} v^r(i, j)}{\sum_{(i,j)\in\tilde{P}(\sigma,\rho);\|v^r(i,j)\|>0} 1}, \tag{3}$$

where $v^r(i, j)$ means the raw flow at position (i, j). The condition $\|v^r(i, j)\| > 0$ in the sum ensures that pixel positions with vanishing motion signals are not included in the averaging procedure.

Positions $(\sigma, \rho) \in S$ with vanishing motion signals or with radiuses $r(\sigma, \rho)$, for which averaging domains would cross the borders of the image, are also omitted from S. Therefore, regions which present sky in the images and regions close to the image boundaries do not contribute to the filtered flow.

3.3 Estimation of heading

For the estimation of heading from the optic flow we use a version of the Heeger-Jepson subspace algorithm [6] as implemented in [10]. Briefly, the subspace algorithm combines a small set of flow vectors into a residual function $R(t)$ of the heading t. Then for a given flow field the direction of translation (heading) can be estimated by minimizing the residual function

$$R(t) = \|(\Phi^t C^\perp(t))\|^2$$

by variation over t. Since the residual function involves only a matrix product and rectification it can be easily implemented in a neural network to yield a map of heading direction likelihoods [12, 14]. The minimization is usually performed over several small sets of flow vectors ($m > 5$) and a compound residual function is constructed by adding the individual matrices. This strategy increases the robustness of the heading estimation against outliers and reduces the complexity of the orthogonalization. We use $m = 10$ and draw 5×10 random samples from the flow field. The 5 individual residual functions for each collection of 10 flow vectors are summed into a compound residual function. Thus, a single heading estimate of the algorithm is based on a subsampling of the flow field of 50 vectors.

4 Results for real camera sequences during car driving

We tested the filtering method for 3 image sequences of different car driving situations (Fig. 2). The first image sequence was taken during straight open road driving. The second sequence was recorded on a motorway. The third sequence was taken during driving in a town. During the recording of this sequence, the camera car slowed down and turned left to bypass another car that was braking and turning right in front of the camera car. This introduced two common difficulties of driving which the algorithm has to deal with. First, the motion of the car involves a rotation of the path and a slightly sideward component of translation. Second, the other car to some degree disturbs the global flow pattern obtained from the otherwise static scene. The image sequences were recorded by a camera rigid installed closely behind the front shield of a moving car. The view direction of the camera was approximately parallel to the longitudinal axis of the vehicle. The camera had a focal length of 2388 pixels. The image resolution was 1276×1016 pixels. The position of the principal point was $(627, 551)$.

From each sequence, 10 successive frames were used in order to allow the tracking of the estimated heading over a length of time of car motion. Raw flow fields for the image sequences were estimated with the Lucas-Kanade algorithm. Filtered flow fields were derived from the raw flow. Figure 2 shows the raw and the filtered flow for the first frame of each driving scenario. In all three cases, the singular point of the flow pattern is close to the center of the field of view. Note that the singular

point coincides with the heading only without rotational motion of the camera. If the camera translates and rotates the singular point is different from the heading direction. A comparison of the raw and filtered flow fields in Figure 2 shows that the filtered flow is much less noisy and shows the typical pattern of a mainly forward motion. This pattern is less recognizable in the unfiltered flow fields.

The panels in the forth and fifth row of Figure 2 show the results of heading estimation from the unfiltered and the filtered flow fields seen in the second and third row of Fig. 2. For each flow, 60 random subsamples of 5×10 flow vectors were drawn and heading was estimated as described in section 3.3 based on the particular subsampling. For the raw flow, heading estimates from different subsamplings of the flow field show large variability. The standard deviations of over 60 runs are 17 degree for the first sequence, 16 degree for the second and 17 degree for the third. This suggest that the heading algorithm applied directly to the raw flow is rather unstable. In contrast, for the filtered flow the same heading estimation procedure yields estimates that are tightly clustered. Standard deviations are 5 degree, 3 degree and 1 degree for the filtered flow. The large standard deviation for the first scene is caused by one outlier which is not visible in the image because it is not within the field of view. Without this outlier the standard deviation is reduced to only 1 degree. Thus, the consistency of the heading estimate is much improved by the space-variant filtering procedure.

The first row of figure 3 shows the consistency of the heading estimate over frames of the sequence. The plots give the standard deviation over 60 subsamplings for each frame of the space variant filtered flow and the raw flow. In almost all cases, the standard deviation is lower for the filtered flow than for the raw flow field. Excluding frames 8 and 9 of scene 2, the differences of the standard deviations between the raw and filtered flow is between 7 and 17 degree. The large standard deviations for the filtered flow in frames 8 and 9 of scene 2 are probably due to the car being in an "extreme" motion situation where the driver is performing a fast navigation correction. Therefore, the camera encounters a sudden large rotation. This rotation shifts the singular point far away from the center of view, and the space variant filtering method fails to improve the flow field in this case. Indeed, the flow field extracted from these frames shows no singular point within the field of view.

Since it is unlikely that the motion of the car changes significantly over the ten successive frames of each sequence we can use the coherence over time of the expectation value of heading over the 60 subsamplings for each frame as a measure of the quality of the heading estimate and, indirectly, of the flow representation. Figure 3 shows the expectation values of heading estimation over the 10 successive frames. The heading estimate from the raw motion field is quite variable and does not capture the properties of continuous motion of the car. In contrast, the curves presenting the temporal evolution of the heading estimate from the space variant filtered flow are much smoother and consistent with continuously changing motion. For instance, the motion in scene 1 is going through a slight left curve and bumping down. The decreasing absolute value of the slope of the curve of the vertical heading component between frames 1 to 5 and between frames 8 to 10 respectively indicate that the bumping is damped by the anti-shocks of the car.

These results suggest that the space-variant filtering method can improve the consistency of heading estimation for camera sequences taken during car driving. However, the evidence is indirect since the true heading of the car is not known. Moreover, the applicability of the method to more

biological motion situations is not fully explored. We have therefore devised a second test which involves different image sequences and camera motions.

5 Results for range image sequences

Although the results obtained with real camera sequences are promising, we have no ground truth about the correct motion. Thus, the errors of heading estimation can only be evaluated indirectly and qualitatively. Therefore, we decided to test the space variant filter method also with flow fields estimated from image sequences that are calculated from the three dimensional data of natural scenes. For comparison, we also include true motion fields from these scenes with superimposed artificial noise. The direct calculation of the image sequences and the true motion fields from the 3D data provides the ground truth to quantitatively evaluate the efficiency of the noise reduction by space-variant filtering.

5.1 Construction of image sequences and flow fields from range image data

Our image sequences were derived from the Brown Range Image Database, a database of 197 range images available from Brown University ([7]). The range images were recorded with a laser range-finder. Each image contains 444×1440 measurements with an angular separation of 0.18 degree. The field of view covers 80 degree vertically and 259 degree horizontally. The distance of each point is calculated from the time of flight of the laser beam, where the operational range of the sensor is $2-200$m. The laser wavelength is in the near infrared region ($0.9\mu m$). Thus, the data of each point consist of 4 values, the distance, the horizontal angle and the vertical angle in spherical coordinates and a value for the reflected intensity of the laser beam. Figure 4 pictures panoramically a typical range-image. The knowledge of the 3 dimensional data of a given environment makes it possible to simulate the view of a moving camera in this scene and calculate both the image on the camera as well as the true motion field. The panel A of Figure 5 shows an examples of the projection of range image data onto a plane identical to the situation in cameras, where the intensity of light coming from the reflecting surfaces of the environment and bundled in the lens is projected onto planes of a light sensitive sensor. All images used in the investigation are composed of 350000 pixels and are generated with a projection of focal length of 341 pixel. The projection plane is 683 pixels wide and 512 pixels high. The resulting images are not true grayscale images but rather near infrared intensity images. The optical properties and the contrast values are sufficient, however, to estimate motion between successive frames in a matter identical to camera images.

To simulate the motion of the camera within the range image scene the camera centered coordinates are transformed by a shift and a rotation (see panel B of Figure 5). Let $T = (T_x, T_y, T_z)$ be the translational component and $\Omega = (\Omega_x, \Omega_y, \Omega_z)$ the rotational component of the camera motion. Further let I_t be the number of images taken by the camera over the course of motion. Thus between two frames the camera is translated over the distance $d = \frac{\|T\|}{I_t}$ and rotated over the angle $\phi = \frac{\|\Omega\|}{I_t}$. Let $s = (\frac{T_x}{I_t}, \frac{T_y}{I_t}, \frac{T_z}{I_t})$ be the translation vector. The original position vector r of any point

in the environment is transformed to the new camera position by

$$r' = \frac{\Omega \cdot (r-s)}{\|\Omega\|^2}\Omega + \cos(\phi)\left((r-s) - \frac{\Omega \cdot (r-s)}{\|\Omega\|^2}\Omega\right) + \sin(\phi)\frac{(r-s)\times\Omega}{\|\Omega\|},$$

where $\cdot$ and $\times$ denote the scalar product and the vector product respectively. The image sequences we investigate are generated by translations of $0,01$m/frame and rotations between 0deg/frame and 0.2deg/frame. The components of the camera motion, translation and rotation, simulate straight ahead movement (with respect to scene coordinates) and rotation such that the point in the center of the image is stabilized. The view direction towards this stabilized point is different from the heading direction. This form of motion is similar to the common biological situation in which an observer is moving but keeps gaze onto a particular object of the scene [13]. In this case the singular point is in the center of view.

The natural scenes from the range image data base can be subdivided in urban scenes and forest scenes.The panels A of Figure 5 shows a typical example for the urban case. The distribution of objects in the urban scene is more structured than in forest scenes. Also, the depth statistics of the scenes differ. Thus, as a by-product of our analysis, we investigate whether scene structures influences the efficiency of the space variant filtering method.

The data set we use comprises 70 different urban scenes and 34 different forest scenes. For each scene 16 different motion situations are tested. The motion situations are distinguished by their respective magnitude of rotation of the camera. Psychophysical findings [13] and computational considerations [9] suggest that rotation rate has a critical influence on heading estimation. A given value of rotation can be achived by a particular motion direction relative to the camera (or view) axis in conjunction with the distance of the fixated object from the camera. For each such set of motion parameters both rotation in the negative and the positive direction are evaluated. For each scene and and motion situation the optic flow estimated by the Lucas-Kanade algorithm from the image sequences and the correct optic flow directly calculated from the range image data and superimposed by uniformly distributed noise with a signal to noise ratio of 1 are considered. Then the raw and the noisy motion fields are filtered according to the space variant filter procedure. Figure 5 show examples for the flow fields obtained from the range image data.

5.2 Quality of heading estimates

Figures 6 and 7 depict examples of the heading estimation results obtained from raw and filtered flows for urban and forest scenes, respectively. For each scene 20 single runs are shown. The results are similar to those of the car driving scenes in that the estimates are quite variable for the raw flow and more tightly clustered for the filtered flow. Similar calculations were performed for all available scenes. Ten heading estimates with different sub-samplings of the flow field were used for each motion situation and scene. For rotation values higher than 8 deg/s, some scenes had to be withdrawn because the depth of the nearest object in the scene was too large to achieve the required rotation rate. The total number of heading estimates in each different motion situation ranged from 400 to 1400 for the urban scenes and from 380 to 680 for the forest scenes.

The left panels of Figure 8 show the mean errors for raw and filtered flow of the heading over all scenes and motion situations. In all cases the filtered flow provides superior performance, although

the error rises with rising rotation rate. Thus, space variant filtering leads to more reliable heading estimation also with respect to ground truth heading. Moreover, there seem to be no principal differences in the efficiency of the performance of the space variant filtering method between urban and forest environments.

The right panels of Figure 8 show the mean errors for raw and filtered versions of the true motion field with added artificial noise. These true motion fields were directly calculated from the 3D data of the scene and superimposed with uniformly distributed noise with a signal to noise ration of 1. Evidently, for the artificial noise flow fields the space variant filtering is more efficient than for the real flow fields. This is to be expected because the spatial filtering is statistically optimal for this type of noise. This results confirm the findings in [10] with respect to natural scenes.

6 Conclusion and discussion

Our results clearly demonstrate that space-variant filtering is a reasonable strategy to decrease noise in optical flow fields and to improve heading detection. The method works well on optical flow fields based on natural scenes affected by strong noise and the aperture problem. The stability of the heading detection algorithm is increased, the spread of the resulting heading directions is decreased and the mean is more reliable than in the unfiltered case.

A single flow vector in the filtered case combines information from a large set of raw flow vectors. In the raw flow field each single vector is only one measurement. Thus, it may be that similar improvements in heading estimation occur if more flow vectors are used in the heading estimation algorithm from the raw flow field. Likewise, combining a larger number of heading estimates from different subsamplings into a compound estimate may improve the heading estimation from the raw flow field. However, the complexity of the heading procedure described in section 3.3 increases considerably with the number of flow vectors. The space variant filtering stage is thus, a simple and effective mid-level method to recode and condense information in a sensible way for subsequent heading estimation.

The stability of the heading estimate is clearly improved after space variant filtering. However, this does not mean that the filtered flow field matches the true motion field in all aspects. The smoothing properties of the method, particularly in the periphery, are too strong to accomplish this. Direct comparisons of the true motion field and the filtered flow in Fig 5 shows the differences. We rather believe that the space-variant filtering method reproduces the global structure of the correct flow and therefore allows to better estimate the correct components of self motion. Hence, our method is a task specific optimization that enhances the applicability of the flow field to the task of heading detection, but not for other task. For instance, the segmentation of objects in depth would rather become more difficult after the space-variant filtering. Nevertheless, our results encourage the view that space variant mapping of the visual field is advantageous for certain tasks of visual processing [3, 16, 19].

The filtering approach is based on the expectation that averaging over many independent measurements of essentially the same signal reduces noise. Because of the radial structure of the optic flow averaging can be performed over large image areas in the periphery but only small image areas in

the center of the visual field. Thus, noise reduction is particularly effective in the periphery. However, in the visual pathway of primates a space variant mapping of the visual field already exists on the early stage of visual processing, i.e., in the retina and primary visual cortex V1. Thus, for primate area MT the input is already organized such that averaging over small image areas in the center of the field of view includes many individual measurements of the motion signal. In this case, the noise reduction properties of the filtering procedure are probably equally effective across the visual field. Therefore, our filtering method would become even more efficient if the input consists of a space variant representation of the image rather than a standard camera representation. Such space variant sensors have been developed and may be used in our approach as well [5, 18].

The filtering method easily lends itself to further improvements by adding for instance disparity information from a second camera [6]. Therefore, the successful performance of heading estimation on the filtered flow should only be the first step of a larger agenda. Certainly, the next step is to estimate the rotational component of the self motion from the filtered flow. Further, although the filtered flow is not applicable to extract grouping information, to perform background-foreground segregation, or to identify moving external objects, the knowledge of the global parameters of self motion from the filtered flow gives the possibility, supposing one has the disparity information of the scene, to reconstruct the correct flow. Thus, apart from the relevance for the reliable estimation of the global motion parameters from noisy flow, the space-variant filtering method could be used as a part of an improved optic flow algorithm.

References

[1] T. D. Albright and R. Desimone. Local precision of visuotopic organization in the middle temporal area (MT) of the macaque. *Exp.Brain Res.*, 65:582–592, 1987.

[2] S. Baker and I. Matthews. Lucas-Kanade 20 years on: A unifying framework. *I.J.Computer Vision*, 56:221–255, 2004.

[3] G. Baratoff, C. Töpfer, and H. Neumann. Combined space-variant maps for optical-flow-based navigation. *Biol.Cybern.*, 83:199–209, 2000.

[4] J. E. Barron, D. J. Fleet, and S. S. Beauchemin. Performance of optical flow techniques. *I.J.Computer Vision*, 12:43–77, 1994.

[5] N. Franceschini, J. M. Pichon, and C. Blanes. From insect vision to robot vision. *Philos.Trans.R.Soc.Lond. B*, 337:283–294, 1992.

[6] D. J. Heeger and A. Jepson. Subspace methods for recovering rigid motion I: Algorithm and implementation. *I.J.Computer Vision*, 7:95–117, 1992.

[7] J. Huang, A. B. Lee, and D. Mumford. Statistics of range images. In *Proc. CVPR*, 2000.

[8] B. Jähne, editor. *Digital image processing – concepts, algorithms, and scientific applications*. Springer, 1997.

[9] J. J. Koenderink and A. J. van Doorn. Facts on optic flow. *Biol.Cybern.*, 56:247–254, 1987.

[10] M. Lappe. Functional consequences of an integration of motion and stereopsis in area MT of monkey extrastriate visual cortex. *Neural Comp.*, 8:1449–1461, 1996.

[11] M. Lappe, editor. *Neuronal Processing of Optic Flow*, Int.Rev.Neurobiol.44. Academic Press, 2000.

[12] M. Lappe, F. Bremmer, M. Pekel, A. Thiele, and K.-P. Hoffmann. Optic flow processing in monkey STS: A theoretical and experimental approach. *J.Neurosci.*, 16:6265–6285, 1996.

[13] M. Lappe, F. Bremmer, and A. V. van den Berg. Perception of self-motion from visual flow. *Trends.Cogn.Sci.*, 3:329–336, 1999.

[14] M. Lappe and J. P. Rauschecker. Motion anisotropies and heading detection. *Biol.Cybern.*, 72:261–277, 1995.

[15] B. D. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. *Proc.DAPRA Image Understanding Workshop*, pages 121–130, 1981.

[16] H. A. Mallot, W. von Seelen, and F. Giannakopoulos. Neural mapping and space-variant image processing. *Neural Networks*, 3:245–263, 1990.

[17] K. Pauwels and M. van Hulle. Segmenting independently moving objects from egomotion flow fields. In *Early Cognitive Vision Workshop, Isle of Skye*, 2004.

[18] G. Sandini and G. Metta. Retina- like sensors: motivations, technology and applications. In T. Secomb, F. Barth, and P. Humphrey, editors, *In Sensors and Sensing in Biology and Engineering*. Springer-Verlag, 2002.

[19] E. L. Schwartz. Spatial mapping in the primary sensory projection: Analytic structure and relevance to perception. *Biol.Cybern.*, 25:181–194, 1977.

[20] L. M. Vaina, S. A. Beardsley, and S. Rushton, editors. *Optic Flow And Beyond*. Kluwer Academic Press, 2004.

[21] A. V. van den Berg and E. Brenner. Humans combine the optic flow with static depth cues for robust perception of heading. *Vision Res.*, 34:2153–2167, 1994.

[22] A. V. van den Berg and E. Brenner. Why two eyes are better than one for judgements of heading. *Nature*, 371:700–702, 1994.

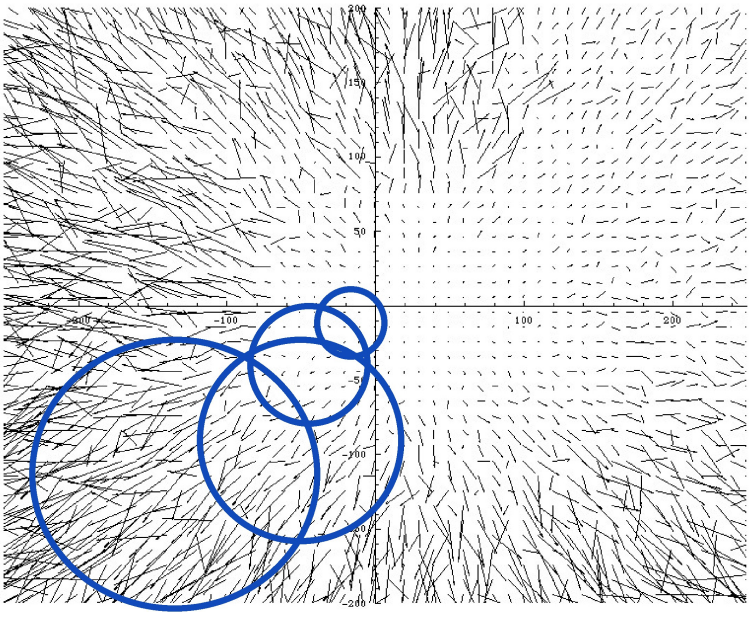

Figure 1: Increasing filter fields sizes with respect to a noisy optic flow stimulus

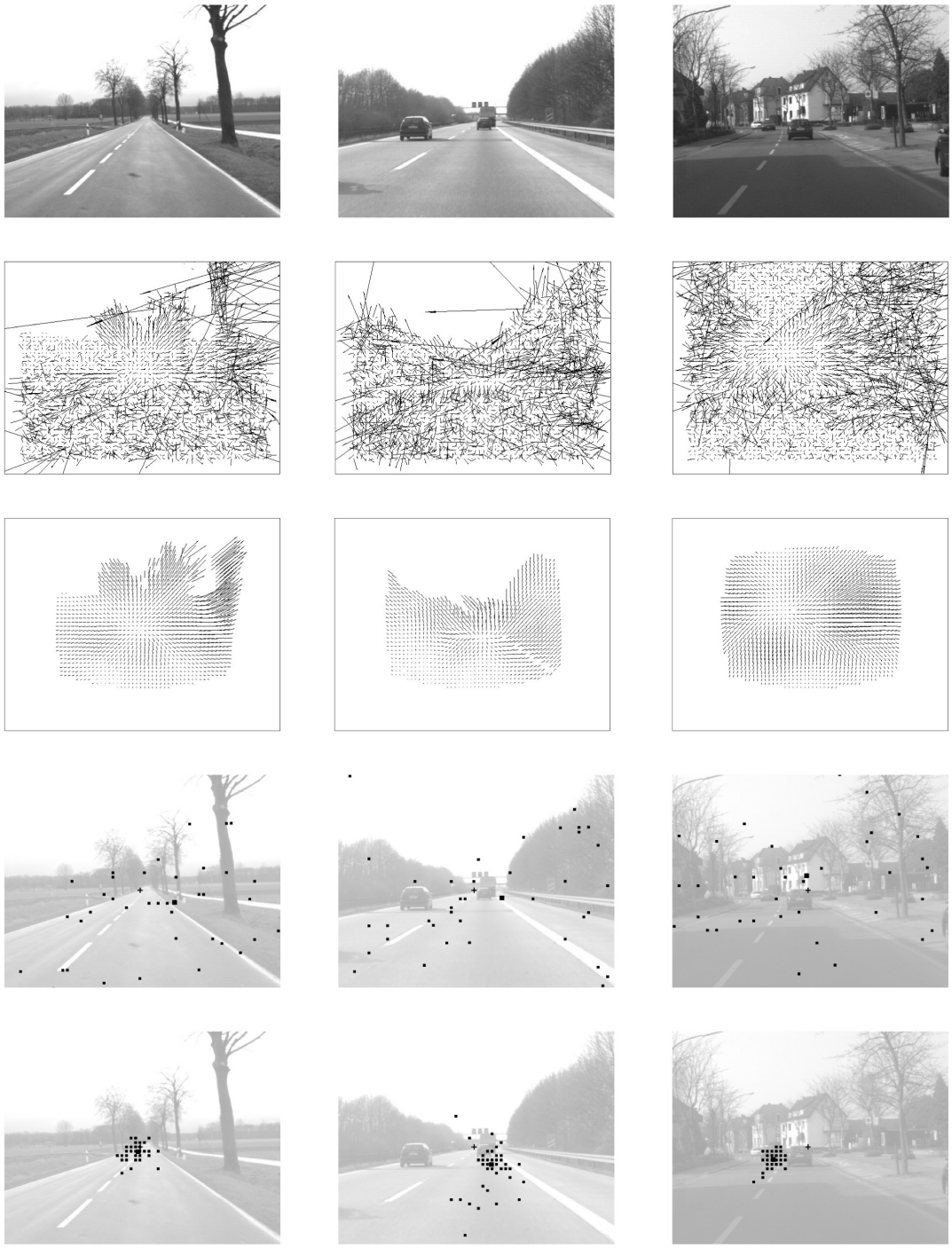

Figure 2: Top row: First frames of the camera sequences for different driving scenarios, Second row: Optic flow estimated from the first three frames by the Lucas-Kanade Algorithm, Third row: Space variant filtered flow, Forth row: Single heading estimations from the raw flow, The large black square denotes the mean. Fifth row: Single heading estimations results from the filtered flow

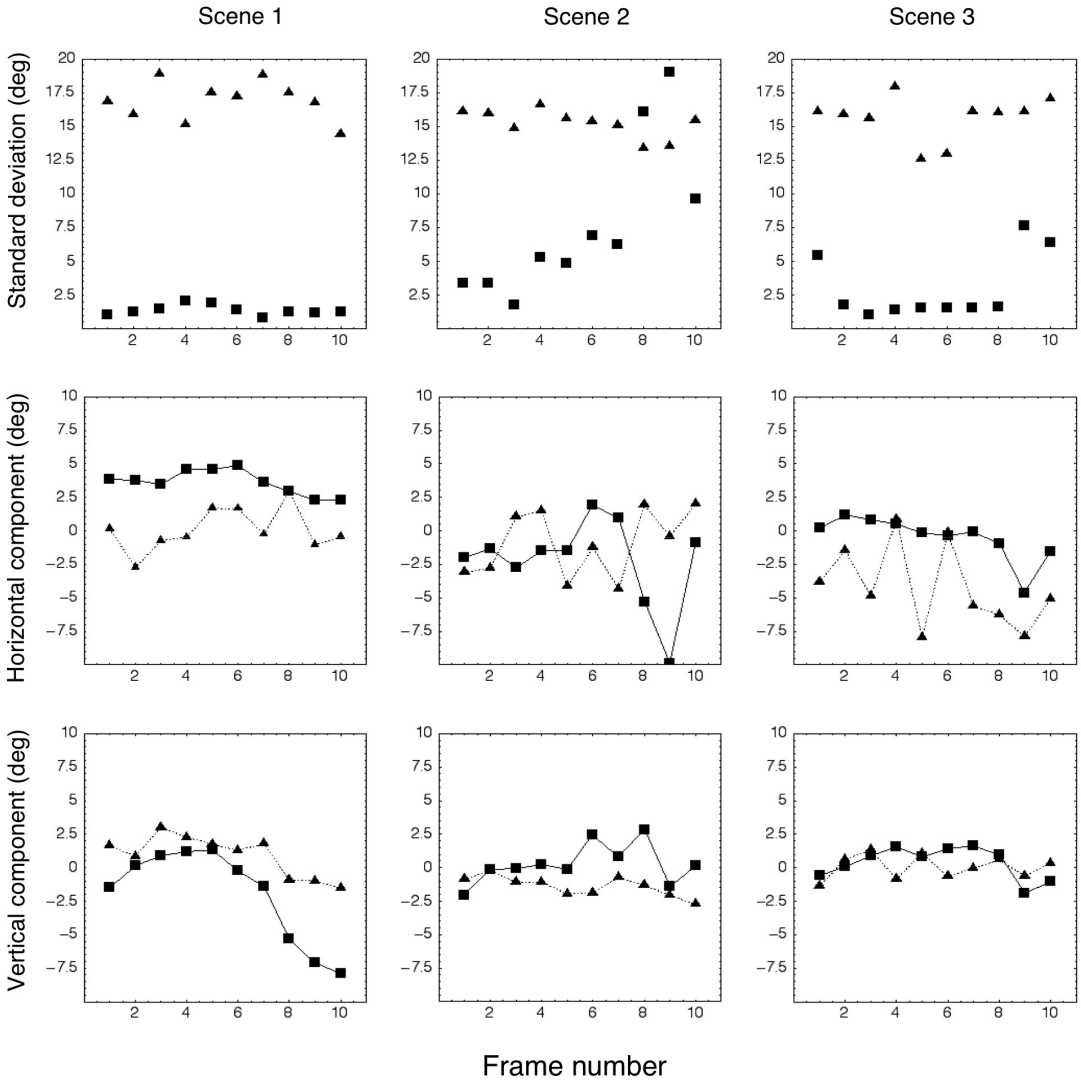

Figure 3: Top row: Standard deviation over 60 heading estimation runs for the different driving scenarios, Middle and bottom rows: Horizontal and vertical components of the mean for 10 subsequent frames, Positive and negative angles mean leftward and rightward and upward and downward translation respectively. Squares: Heading estimations from the filtered flow, Triangles: Heading estimations from the raw flow

Figure 4: Panoramic projection of the reflectance data of a range-image.

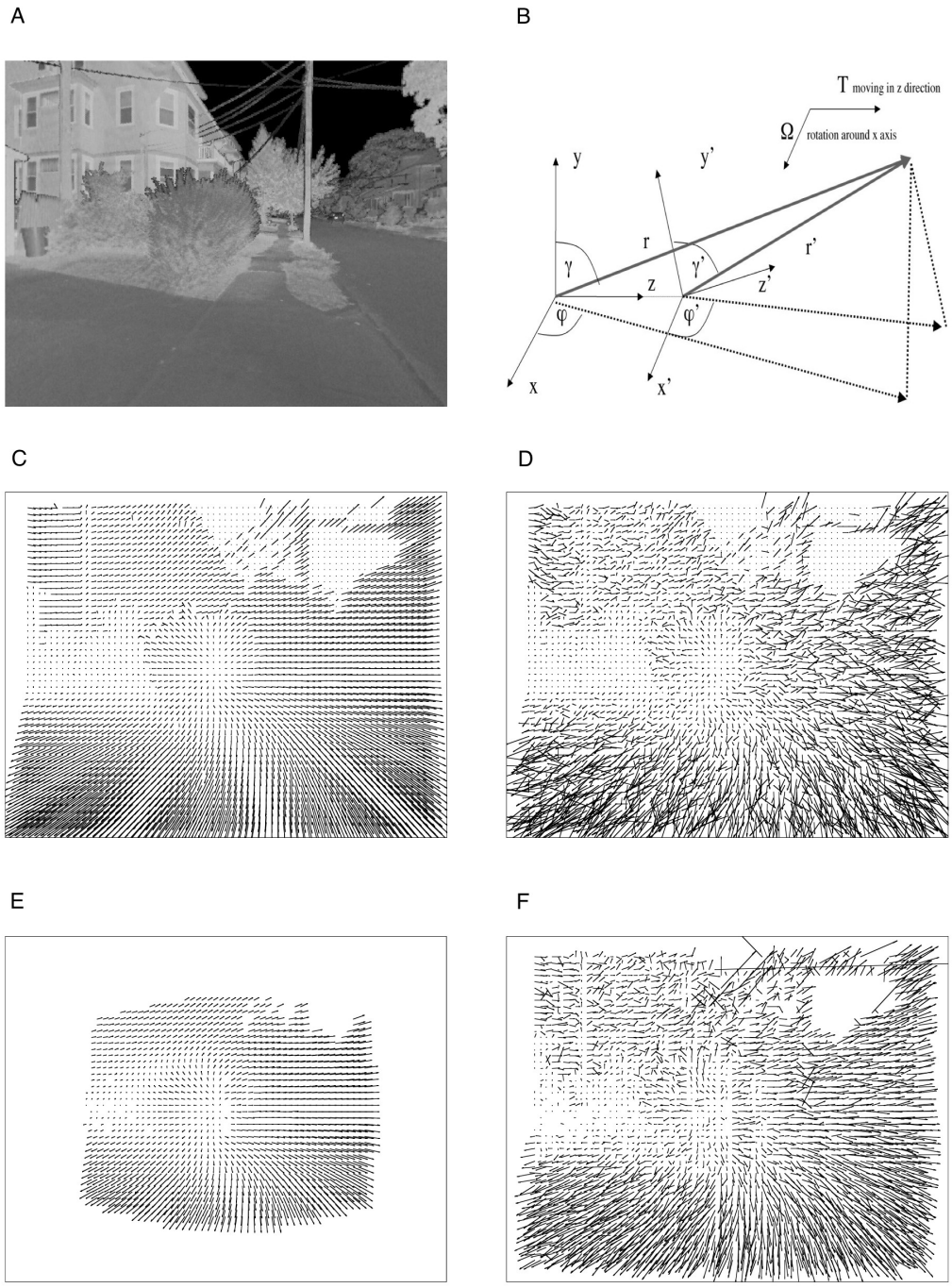

Figure 5: A: Reflectance data of a range-image projected onto a plane simulating a camera shot. B: Transformation of camera centered coordinates during ego-motion, C,D,E and F: Optic flow fields generated from range image data for a rightward translation of 27 degree and a rotation around the vertical axis of 7deg/s, C: True flow, D: True flow with superimposed uniformly distributed noise (SNR=1), E: Space variant filtered flow, F: Raw flow estimated from the image sequence,

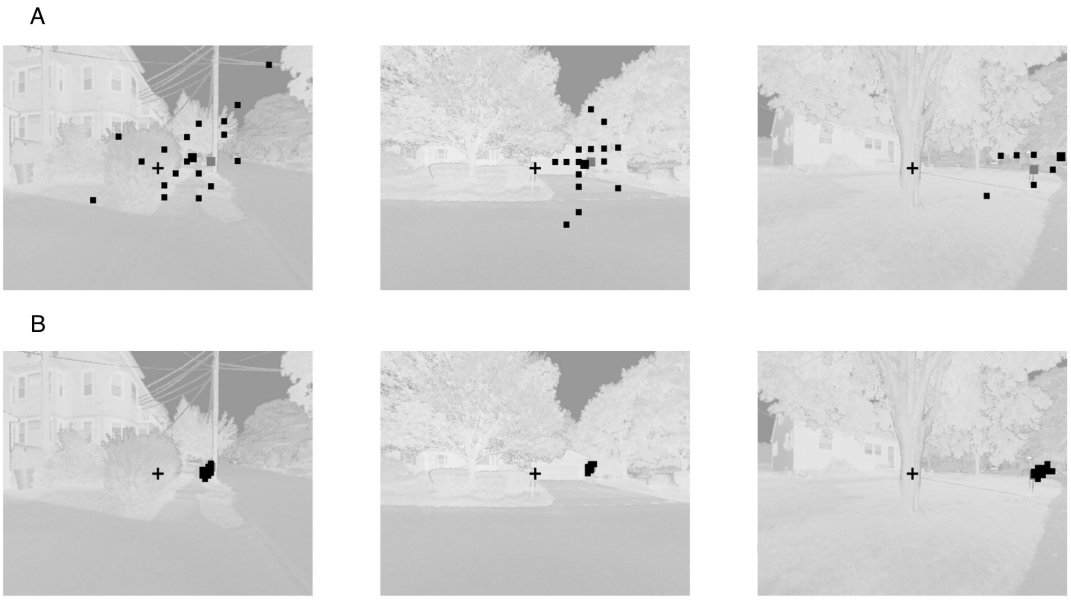

Figure 6: Results of heading estimation for urban scenes, Translation of the camera motion is towards the grey point (horizontal component between 10 degree and 35 degree rightwards), Rotations are 3 deg/s, 5 deg/s and 9 deg/s. The black cross denotes the principal point of the camera. The small black points show the results of single runs. The large black point marks the mean over 20 runs. The grey point denotes the correct heading. A: Raw flow, Standard deviations are 10 degree, 19 degree and 11 degree, B: Space variant filtered flow, Standard deviations are around 1.6 degree.

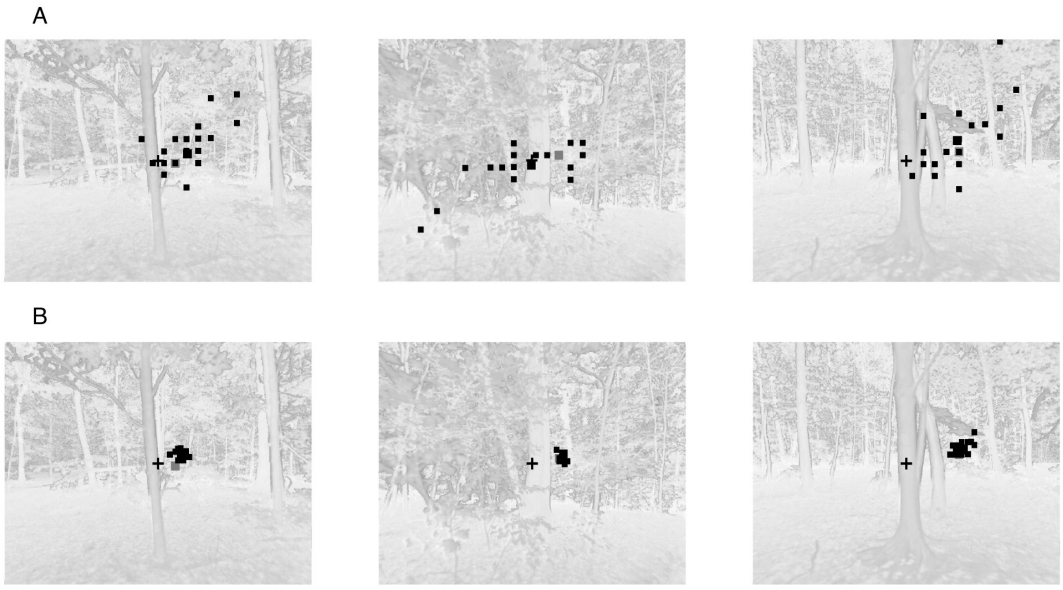

Figure 7: Results of heading estimation for forest scenes, Translation of the camera motion is towards the grey point (horizontal component between 4 degree and 20 degree rightwards), Rotations are 3 deg/s, 5 deg/s and 9 deg/s. The black cross denotes the principal point of the camera. The small black points show the results of single runs. The large black point marks the mean of 20 runs. The grey point denotes the correct heading. A: Raw flow, Standard deviations are 16 degree, 17 degree and 13 degree, B: Space variant filtered flow, Standard deviations are 1.9 degree, 1.6 degree and 2.8 degree.

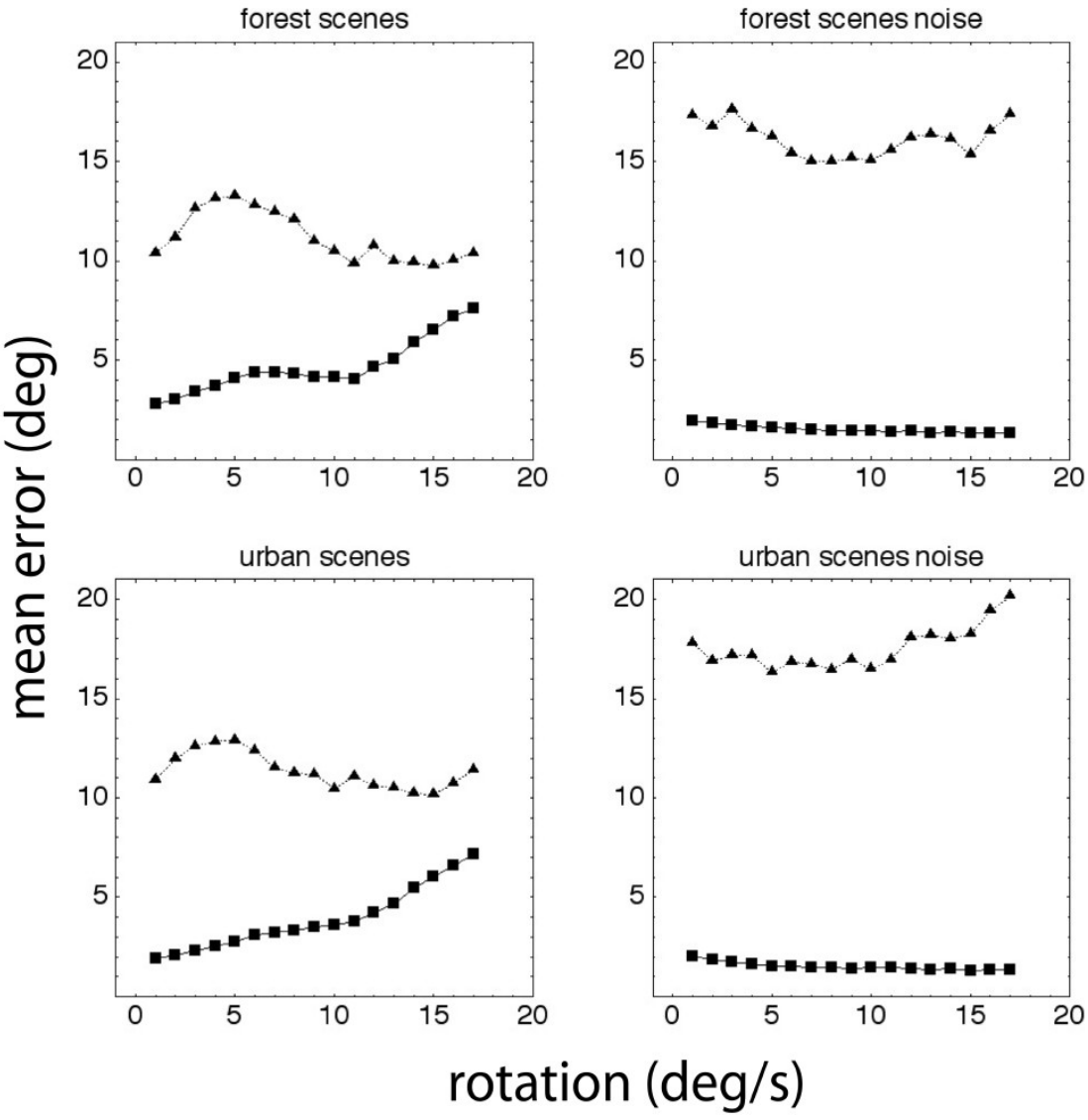

Figure 8: Mean error for heading estimation, Boxes: space variant filtered flow, Triangles: Raw flow, Left: Flow obtained with the Lucas-Kanade algorithm, Right: True motion field derived from the scene data with uniformly distributed noise added (SNR=1)