Space variant filtering of optic flow for robust three dimensional motion estimation

Dirk Calow[1], Norbert Krüger[3], Florentin Wörgötter[2] and Markus Lappe[1]

1 Dept. of Psychology, Westf.- Wilhelms University, Fliednerstr. 21, 48149 Münster, Germany
2 Institute for Neuronal Computational Intelligence and Technology Stirling Scotland, UK
3 Dep. of Computer Science and Engineering Aalborg University Esbjerg, Denmark

Abstract: We test a biologically motivated filtering method [9] for noise decreasing in optical flow fields. We use the task of heading detection from optic flow as a way to estimate improvements of flow fields generated by a standard algorithm. The image sequences which we use for the testing are directly calculated from three dimensional real world data assuming a given self motion. Thus we retain the control about the exact heading and rotation and have ground truth. Not surprisingly, due to the noise and the aperture problem the results for the raw flows are often incorrect. In contrast the filtered flows allow correct heading detection.

1 Introduction

Optical flow fields generated by image analysis applied to image sequences are usually very noisy. Such noisy, low quality flow fields are what computer vision scientists and engineers have to deal with when developing vision based driver assistance systems or autonomous robots. Such systems should be capable to estimate self motion, i.e. the direction of motion (heading) and the rotation of a camera moving through a static scene. Biological systems are able to estimate self motion rather exact, but assuming that biological vision systems have no superior detectors they can not be better than the mathematics used in technical systems. Probably, on the input stage, biological systems face the same problematic optic flows as the computer vision scientist (or the robot prototype) does. Somehow, however, the brain has developed methods to remedy the shortcomings of the flow detectors. We searched for features which enable biological vision systems to handle noisy flow fields successfully. The visual system of primates contains an area that is specially devoted to processing of visual motion. This area, the middle-temporal or MT area, establishes a space variant map of the visual motion field. Based on the properties of the middle temporal (MT) area in [9] a method is proposed to decrease the noise of the optical flow by averaging flow vectors over image areas which increase in size d proportional to the eccentricity ϵ from the center of the field of view (see [1] and figure 1)

$$d = 0.018 + 0.61\epsilon.$$

While averaging over large areas is more favorable for noise reduction and smoothing, averaging over small areas save information. The spatial integration over peripherally increasing image areas is a compromise between both goals and well adjusted to the typical structure of the flow field elicited by self motion. The singular point, of the optic flow, i.e. the point with vanishing flow, is

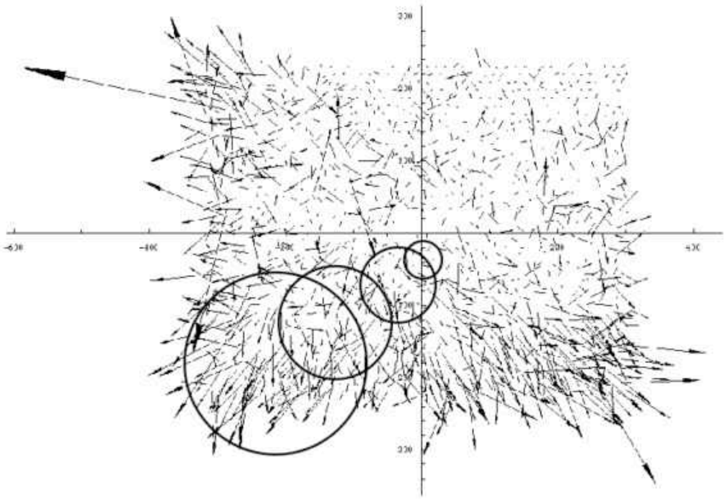

Figure 1: Increasing receptive fields sizes combined superimposed on a noisy optical flow stimulus

usually near the center of the visual field [8]. 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 [10]. In [9] one can find an application of this method and an implementation of the Heeger-Jepson [3] heading detection algorithm in terms of a network model. The network was tested with artificial motion fields of simulated self movements through three dimensional clouds of random dots or over a ground plane. The flow fields contained eye rotation components and noise was added. The results show that noise is reduced and heading detection is possible with errors up to 4 degrees, for the ground plane situation and a signal to noise ratio of 1.

The main goal of the present paper is to apply the *MT-like filtering* model to optical flow fields obtained from image sequences with an optical flow algorithm and to test its implication on the quality of heading detection. The optical flow algorithm we use is a combination of the Nagel-algorithm [11] with the concept of local multi-modal primitives [7]. These primitives are motivated by processing in the human visual system as well as by functional considerations and are regarded as the functional analogues of the hyper-columns in V1 [7]. However, rather then using sequences obtained by a camera we calculate the basic image sequences from three dimensional data sets of several natural environments (Brown Range Image Database, available on *http://www.dam.brown.edu/ptg/brid/range/*). 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 we are able to evaluate the results of the used method quantitatively against ground truth.

Part of the errors in estimation optical flow from image sequences stems from the aperture problem. This type of noise is dependent on the local 2D structure of the images which is described by the

Figure 2: 3D data of a range-image projected onto a sphere, the grey scale encode the intensity of the reflected laser beam, i.e white: high intensity, black: vanishing intensity

local intrinsic dimensionality. Image areas with rich two dimensional structure usually provide better flow measurements than image areas with predominantly one dimensional structure (edges) or largely homogeneous areas. Therefore in addition to the MT-like filtering we test an approach which takes into account a continuos formulation of intrinsic dimensionality (iD) (see [6]). The concept of the iD allows to define [6] with respect to noise and the influence of the aperture problem for each vector of the optical flow field generated by the motion algorithm. We impose several weighting functions which reflect the confidence expressed by the iD data on the method of MT-like filtering and compare the results of heading detection with the results after the original MT-like filtering.

2 Methods

We used for our investigations the Brown Range Image Database (brid) , a database of 197 range images collected by Ann Lee, Jinggang Huang and David Mumford at Brown University ([4]). The range images are recorded with a laser range-finder. Each image contains 44×1440 measurements with an angular separation of 0.18 degree. The field of view is 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 of the laser beam is $0.9\mu m$ in the near infrared region. 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 2 shows a typical range-image projected onto a sphere.

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. Figure 3 shows the projection of the 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 light sensitive sensors or matter. Figure 4 shows the transformation of the camera centered coordinates of a certain point during self motion.

The components of the camera motion, translation and rotation, simulate straight ahead moving

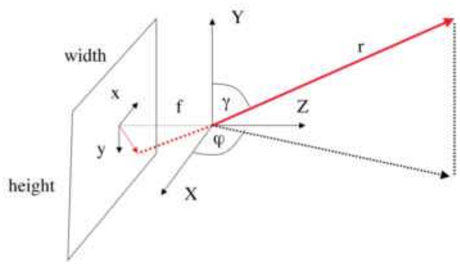

Figure 3: Projection of the 3D data onto a plane, f focal length, γ vertical angle, ϕ horizontal angle, range distance, r_x, r_y 2D coordinates of the plane

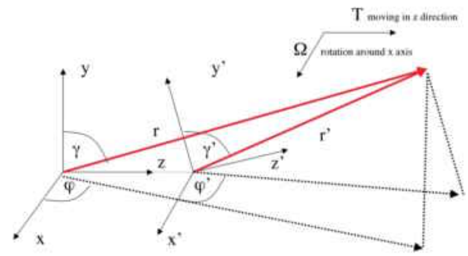

Figure 4: Simulation of a self movement through an environment given by the range image

(concern the world 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 common biological situations [10]. A human walks through an environment and fixates between objects or structures on surfaces deviating from the heading direction while retaining control about his self motion. The fixation direction that we used in our simulations were obtained from measuring fixations of observers that viewed the scenes on a computer monitor.

Image sequences and optical flow fields: Despite that the resulting images are not true grayscale images and the reflecting surfaces looks like floodlighted, it is possible to determine edges and to identify the objects which originally populated the measured environment and which are now encoded in the three dimensional data (Figure 5 shows an example). This simple fact is evidence for the correct positions of the identifiable basic image features like edges and corners used by visual systems to perform grouping in images. We obtain the optical flow fields from the 3D-generated image sequences by the Nagel-algorithm [11], modified with respect to the concept of local multi-modal primitives [7]. This technique calculates the local displacement by the Nagel-algorithm and performs a local averaging for each local primitive allowing to obtain either the correct flow for a near intrinsic two dimensional structure like corners or resulting in the normal flow for a near intrinsic one dimensional structure like an edge within the particular primitive. Since the Nagel-algorithm operates on local intensity changes over space and time, it is reasonable to assume that the application of the Nagel-algorithm to image sequences based on the brid-data gives results similar to results obtained with standard camera images. Figure 5 shows an image sequence based on a simulated self movement through an environment given by a brid-data set. The left picture of figure 6 pictures the correct optical flow from this self movement and the middle picture of figure 6 shows the result of the application of the Nagel-algorithm to the image sequence figure 5. The result is clearly noisy, but captures the overall structure of the true flow.
Heading detection is performed by the Heeger-Jepson algorithm [3]. During our calculations the algorithm operates on 150 randomly selected flow vectors for each run and the results are accumulated for 30 runs. As shown in the right picture of figure 6 the attempt to apply the Heeger-Jepson algorithm to the noisy flow in figure 5 leads to a distribution of estimated headings that is broad and hardly connected to the correct heading. Therefore a rectification of the flow by noise reduction

Figure 5: Three image frames selected from a sequence generated from the brid-data. The first image is the first of the sequence, the following images have a distance of 15 and 30 frames respectively from the first. The simulated self movement results in a translational shift of 0.01m and a rotation of 0.01 degree per frame. The ratios between width and hight and the focal length are chosen as 1 and 0.7 respectively.

would appear welcome.

3 Results

Reduction of noise by MT-like filtering: First we want to test the model of MT-like filtering to a fully controlled situation in which we have knowledge of the noise. For our example in figure 7 we applied random noise with a signal-to-noise ratio of 1 to the true optical flow field. Comparison of the two pictures in figure 7 gives the impression of a dramatically increased quality of the MT-like rectified flow field. This impression is confirmed by comparing the results of heading detection on both flow fields (figure 8). The heading directions obtained from the unfiltered flow are spread over a range of 25 degrees and the deviation of the mean value from the correct heading is 12 degree. The heading directions obtained from the MT-like filtered flow fall within a range of 6 degree and deviation of the mean value from the correct heading is about 4 degree. Thus the stability of the Heeger-Jepson heading detection algorithm is significantly improved compared with the unfiltered situation. Furthermore the results for the filtered flow match the observations in [9].

The next step is to apply the MT-like filter model to optical flows generated from the Nagel-algorithm. As the basis for the image sequences we chose three scenes (first row of figure 9) from the brid-database and simulated camera motion composed of translation and rotation through the scene. The second row of figure 9 shows the flow fields calculated by the Nagel-algorithm. Obviously these flow fields are very noisy and the results of heading detection algorithm applied to these flows are indeed unusable (figure 10). In contrast to the unfiltered flow the heading detection algorithm applied on the MT-like filtered flow (third row of figure 9)) leads to more stable results (figure 11). It is possible to recognize the correct heading much better than in the unfiltered case.

In the first picture in figure 11 the width of distribution is 7 degree, the deviation of the mean value (blue point) from correct heading (red point) is 2.5 degree. In the second picture the width of distribution is 4 degree and the deviation of the mean value is 1.5 degree. In the third picture

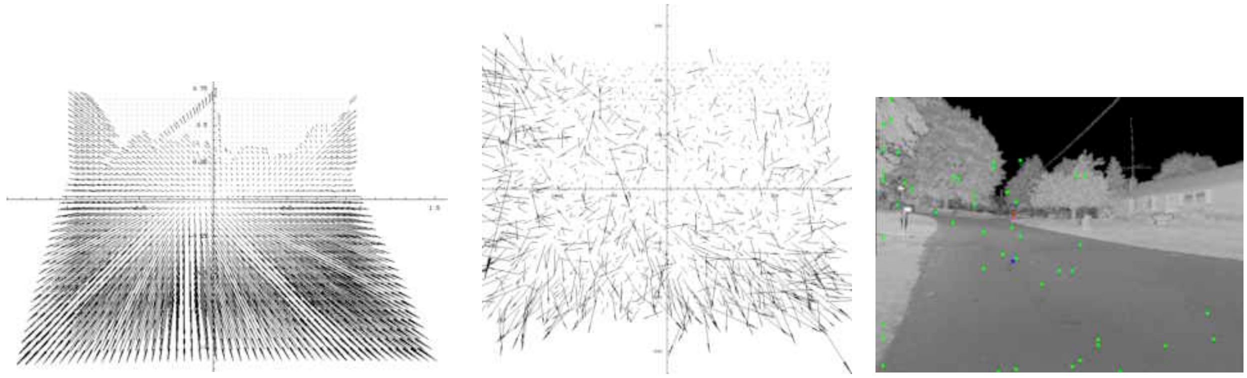

Figure 6: Left picture: Correct flow directly calculated from the 3D data and the simulated self movement in figure 5. Middle picture: Flow calculated with the Nagel-algorithm from the image sequence of figure 5. Right picture: Results of the heading detection applied to the flow in the middle. Green points represent the estimated heading directions of the individual runs. The blue point represent the mean value of the estimated headings.The red point represents the correct heading.

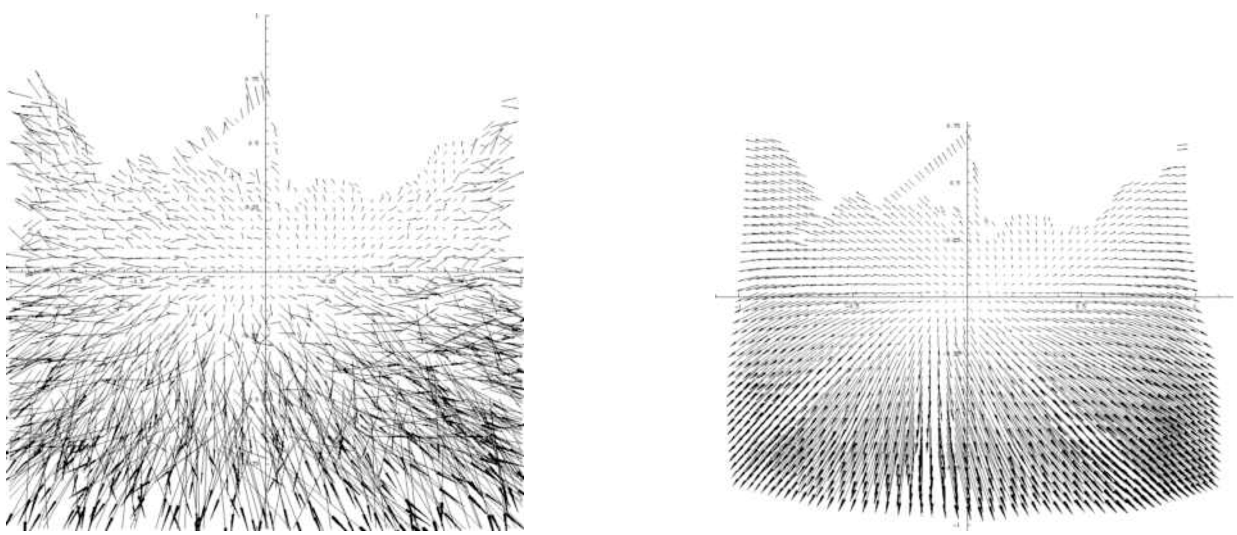

Figure 7: Left: the correct flow field in figure 6 with superimposed noise with a signal-to-noise ratio of 1, right: the same flow field rectified by MT-like filtering

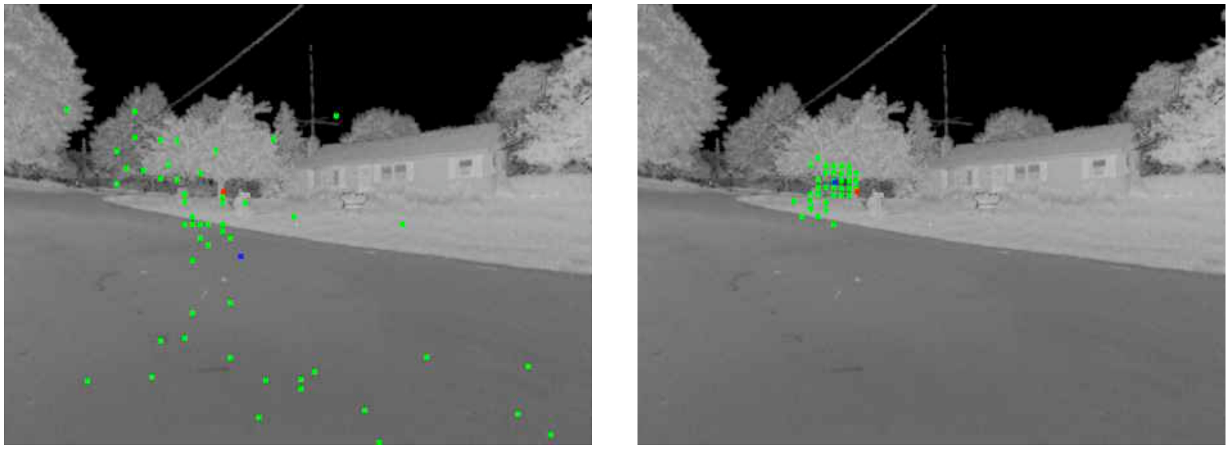

Figure 8: Results of heading detection, left: noisy flow, right: MT rectified flow. Green points represent the estimated heading directions of 30 individual runs each with sub-sampling of 150 flow vectors. The blue point represent the mean value of the estimated headings of the 30 runs. The red point represents the correct heading.

the width of distribution is 4.5 degree and the deviation of the mean value is 9.5 degree.

4 Confidence weighting by intrinsic dimensionality

Natural images are dominated by specific local sub–structures, such as edges, junctions, or texture. Sub–domains of computer vision have analyzed these sub–structures by making use of certain conepts (such as, e.g., orientation, position, or texture gradient). The intrinsic dimension (see, e.g., [13, 2]) has proven to be a suitable descriptor to distinguish between homogenous image patches, edge, or junction structures. We omit here the detailed mathematical formulation of the continuous concept of iD and refer for a detailed description to [6].
In brief, the iD-concept characterizes structures or features in images in terms of the distribution of the intensity in a local area around a single point.

- Local areas in which the intensity is constant in all directions have intrinsic dimensionality zero (i0D) and describes points of the image where there is no local structure.
- Local areas in which the intensity is constant in one direction have intrinsic dimensionality two (i1D). These are points of the image where the local structure resembles an edge.
- Local areas in which the intensity is varying in two certain direction have intrinsic dimensionality two (i2D). These are points of the image with a local 2D structure (corners, points etc.).

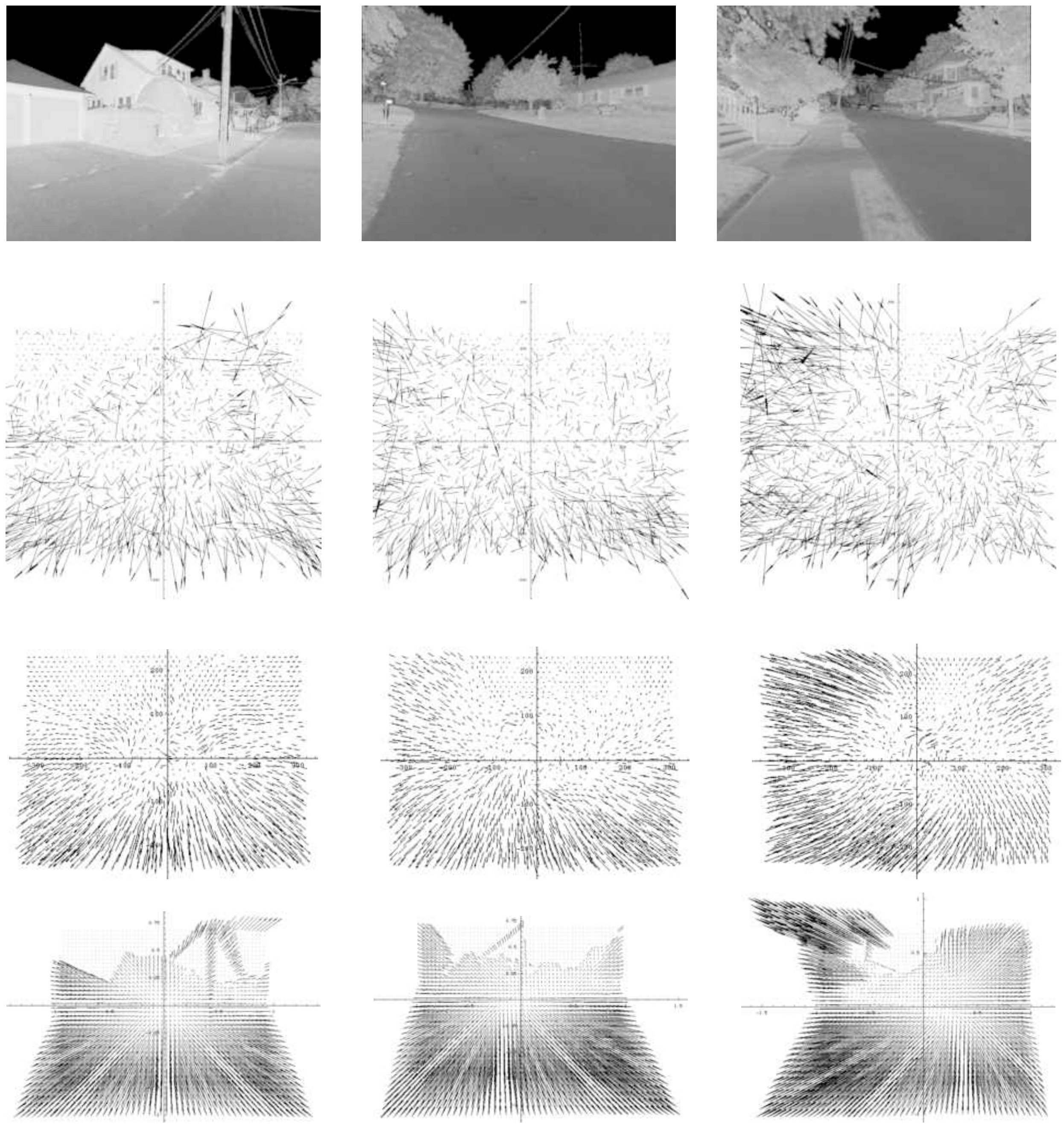

Figure 9: Top row: Basic scenes for the image sequences. Second row: Flow field obtained from the Nagel-algorithm. Third row: MT-like filtered flow. Fourth row: correct flow.

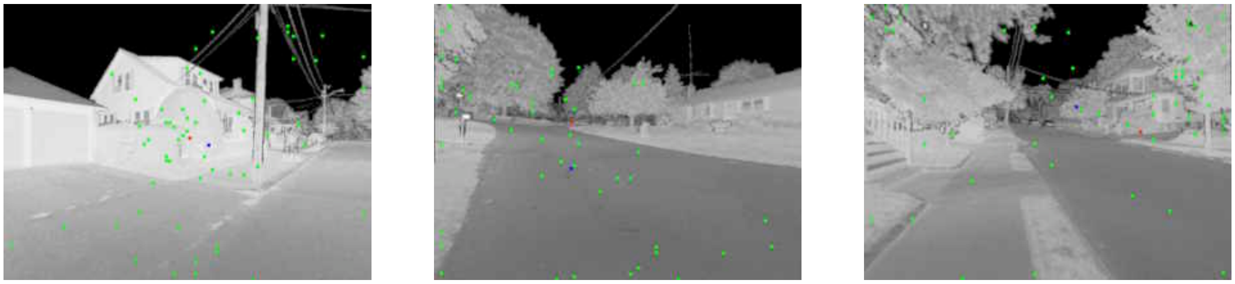

Figure 10: The heading results due to the unfiltered flow

Figure 11: The heading results due to the MT-like filtered flow

The association of intrinsic dimension to a local image structure has mostly be done by a discrete classification [13, 2, 5]. Homogeneous image patches have an intrinsic dimension of zero (i0D), edge–like structures are intrinsically 1–dimensional (i1D) while junctions and most textures have an intrinsic dimension of two (i2D). In [6] a continuous definition of intrinsic dimensionality has been introduced in which 3 confidences for each intrinsic dimension are associated to an image patch.

With respect to optic flow processing, i1D structures will be affected by the aperture problem while i2D structures are less affected and are more likely to provide the correct flow. Flow vectors obtained in areas of i0D are typically unreliable and introduce noise.
Within the continuous formulation in [6] the iD of a point or image patch is described by a point in a 2D space defined by three barycentric coordinates $(i0D, i1D, i2D)$, i.e. $\sum_{K=0}^{2} iKD = 1$. The barycentric coordinates can be interpreted as likelihoods or confidences. The continuous concept is not only a theoretical approach but is related to neuropsychological findings from so called end-stopping neurons in the primary visual cortex of macaques (see [12]). End-stopping neurons respond best to endpoints of long contours moving in a certain direction. These neurons solve the aperture problem for the long contour encircled by the endpoints. Results in [12] (figure 5) show that the distinction between end-stopped cells and ordinary oriented selective cells is not sharp, but rather continuous. This can be interpreted as a biological implementation of the continuos approach of iD. With respect to to our model of MT-like filtering of noisy flow, one can imagine that the end-stopping property of V1 cells modulates the contribution of cell input to the averaging procedure in MT in the sense that input from V1 neurons with strong end-stopping is weighted higher in order to obtain the more correct flow.
In the following we test this idea in the application of MT-like filtering to image sequences by

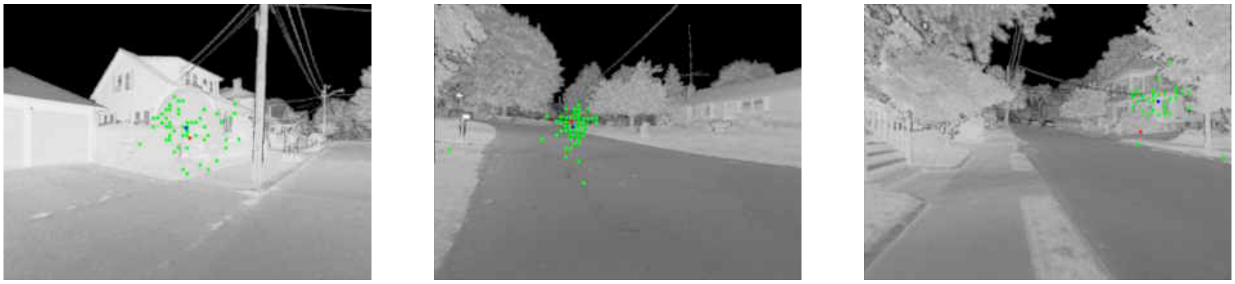

Figure 12: The heading results due to the iD-modulated MT-like filtered flow according to the weighting function $wf1_{12,6}$.

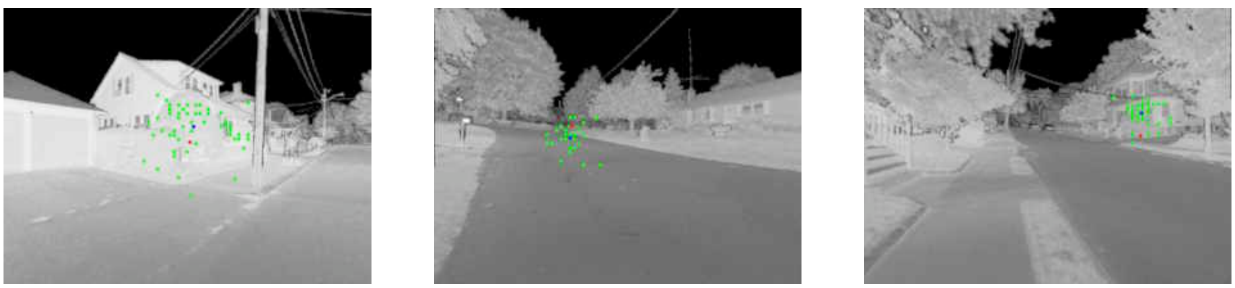

Figure 13: The heading results due to t the weighting function $wf2_{12,6}$.

imposing iD depended weighting functions on the MT-like filtering model. Recall that the iD of a point or image patch is described as a point in a 2D space defined by the barycentric coordinates $(i0D, i1D, i2D)$. Let e_1 and e_2 be two orthonormal vectors spanning the two dimensional space, the iD given by $(i0D, i1D, i2D)$ is the 2D-vector $(i1D + i2D)e_1 + i2De_2$. The weighting functions we test are

$$\begin{aligned} wf1_{g,s}(i0D, i1D, i2D) &= \frac{0.5}{(e^{(-(i1D+i2D)g+s)} + 1} + \frac{0.5}{e^{(-i2D)g+s)} + 1)}, \\ wf2_{g,s}(i0D, i1D, i2D) &= \frac{1}{e^{(-(i2D)g+s)} + 1)}. \end{aligned}$$

Here g and s are parameters governing the slope and the position of the jump of the function. Both functions approach one for $i2D = 1$ and go to zero for $i0D = 1$. The difference between the functions is that $wf2$ suppresses both the contribution of $i0D$ and $i1D$ whereas $wf1$ lets structures with $i1D = 1$ contribute with a weight of $0,5$. Now, let $V_K, (i0D_K, i1D_K, i2D_K)_{K=1,\ldots,N}$ be a set of N flow vectors calculated from features with certain iD. The averaging procedure is weighted by iD:

$$\bar{V} = \frac{\sum_{K=1}^{N} wf_{g,s}(i0D_K, i1D_K, i2D_K)V_k}{\sum_{K=1}^{N} wf_{g,s}(i0D_K, i1D_K, i2D_K)}.$$

We use two weighting functions $wf1_{12,6}, wf2_{12,6}$ and perform the modified MT-like averaging methods on the noisy flows (second row of figure 9). The results can be observed in figures 12, 13. In the first picture in 12 the width of distribution is 8 degree, the deviation of the mean value (blue point) from correct heading (red point) is 3 degree. In the second picture the width of distribution is 5,5 degree and the deviation of the mean value is 1.5 degree. In the third picture the width of distribution is 5 degree and the deviation of the mean value is 8,5 degree. In the first picture in figure 13 the width of distribution is 10 degree, the deviation of the mean value is 4.5 degree. In the second picture the width of distribution is 4,5 degree and the deviation of the mean value is 3.5 degree. In the third picture the width of distribution is 3.5 degree and the deviation of the mean value is 6 degree.

5 Discussion

Our results clearly demonstrate that MT-like 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 dramatically decreased and the mean is near to the correct heading.
Less clear is the potential role of the intrinsic dimensionality. In our tests the sets of results of heading detection on the ordinary MT-like filtering method and the iD-modified show no or only little differences in stability, correctness, and spread. This suggest that the use of iD in this particular manner is limited. We do not believe however that the intrinsic dimensionality has no relevance for the rectification of optic flow fields and the improving of heading detection. Likely, the importance of iD informations depends on the statistics of the particular scene. The original MT-like filtering method works well for rich structured scenes, with lots of corners, nearly equal distributed horizontal and vertical edges and complex but not uniformly structured textures, the situation we have for the most natural scenes. In these cases the averaging of the optical flow over large domains of the view field solves the aperture problem approximately and sufficiently exact for successful heading detection.
In contrast the iD information can give advantages in cases of sparse scenes with biases in the distribution of horizontal or vertical oriented edges and less i2D points. In this situation, the aperture problem is more damaging and can only be solved by using the i2D information. The question is strongly connected with the problem, whether biological visual systems faced with such a moving scene use the i2D information or, if not, fail while trying heading detection.
A further reason why increased weighting of i2D information shows no improvement over equal weighting of all flow vectors might be that the large size of the integration areas in the periphery of the visual field, relying on a sparse set of i2D vectors can cause location errors. If the only vector with high i2D within a particular integration area lies near its border, its high weighted contribution to the averaging procedure would lead to a locally incorrect vector in the center of the integration area. Equal weighting preserves locational correlations even if each individual vector is less reliable. We will investigate the meaning of signals with different intrinsic dimensionality more closely in future work.

Acknowledgment:

ML is supported by the German Science Foundation, the German Federal Ministry of Education and Research BioFuture Prize, the Human Frontier Science Program, and the EC Project ECoVision and Eurokinesis.

We thank Michael Felsberg for fruitful discussions.

References

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

[2] Felsberg, M., Low-level image irocessing with the structure multivector, Institute of Computer Science and Applied Mathematics, Christian-Albrechts-University of Kiel, (2002)

[3] Heeger, D. and Jepson, A., Subspace methods for recovering rigid motion 1: algorithm and implementation, *Int. Journal of Computer Vision*, **7:2** 95-117 (1992)

[4] J. Huang and A. B. Lee and D. Mumford, Statistics of range images, CVPR 2000

[5] Jähne, B., Digital image processing – concepts, algorithms, and scientific applications, Springer (1997)

[6] Krüger, N. and Felsberg, M., A continuos formulation of intrinsic dimension, Proc. of the British Machine Vision Conf. 2003

[7] Krüger, N., Lappe, M. and Wörgötter, Biologically motivated multi-modal processing of visual primitives, *AISB Journal* **1** (4) (2003)

[8] Lappe, M. and Rauschecker, J.P., Motion anisotropies and heading detection, *Biol. Cybern.* **72** 261-277 (1995)

[9] Lappe, M, Functional consequences of an integration of motion and stereopsis in area MT of monkey extrastriate visual cortex, *Neural Computation*, **8, 1449-1461** (1996)

[10] Lappe, M., Bremmer, F. and van den Berg, A.V., Perception of self-motion from visual flow, *Trends in Cognitive Since* **3** No. 9 329-336 (1999)

[11] Nagel, H. H., Displacement vectors derived from second-order intensity variations in image sequences, *Comput. Graph. Image Process.* **21** 299-324 (1983)

[12] Pack, C.C., Livingstone, M.S., Duffy, R. D. and Born, R. T., End-Stopping and the aperture problem: two-dimensional motion signals in macaque V1, *Neuron*, **39** 671-680 (2003)

[13] C. Zetzsche and E. Barth, Fundamental limits of linear filters in the visual processing of two dimensional signals, *Vision Research*, **30** (1990)