Network: Computation in Neural Systems
December 2005; 16(4): 323–340

Taylor & Francis
Taylor & Francis Group

Biologically motivated space-variant filtering for robust optic flow processing

D. CALOW,[1] N. KRÜGER,[3] F. WÖRGÖTTER,[2] & M. LAPPE[1]

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

(Received 26 November 2004; revised 2 November 2005; accepted 29 November 2005)

Abstract
We describe and test a biologically motivated space-variant filtering method for decreasing the noise in optic flow fields. Our filter model adopts certain properties of a particular motion-sensitive area of the brain (area MT), which averages the incoming motion signals over receptive fields, the sizes of which increase with the distance from the center of the projection. We use heading estimation from optic flow as a criterion to evaluate the improvement of the filtered flow field. The tests are conducted on flow fields calculated with a standard flow algorithm from image sequences. We use two different sets of image sequences. The first set is recorded by a camera which is installed in a moving car. The second set is derived from a database containing three dimensional data and reflectance information from natural scenes. The latter set guarantees full control of the camera motion and ground truth about the flow field and the heading. We test the space-variant filtering method by comparing heading estimation results between space-variant filtered flow, flow filtered by averaging over domains of the visual field with constant size (constant filtering) and raw unfiltered flow. Because of noise and the aperture problem the heading estimates obtained from the raw flows are often unreliable. Estimated heading differs widely for different sub-sampled calculations. In contrast, the results obtained from the filtered flows are much less variable and therefore more consistent. Furthermore, we find a significant improvement of the results obtained from the space-variant filtered flow compared to the constant filtered flow. We suggest extensions to the space-variant filtering procedure that take other properties of motion representation in area MT into account.

Keywords: *Visual motion, optic flow, visual ecology, heading estimation, scene statistics*

Introduction

The patterns of optic flow fields received by the optical detectors of visual systems during self-motion encode a substantial amount of information about the physical parameters of self-motion and about the three-dimensional structure of the environment. It is assumed that biological systems exploit such information for path planning, obstacle avoidance, ego-motion control and foreground – background segregation (Lappe 2000; Vaina et al. 2004). The utilization of optic flow information also makes sense for vision-based technical applications like driver assistance systems or autonomous robots. For instance, when driving a car a correct heading detection is necessary to assess the current driving situation. A driver assistance system capable of estimating heading could warn the driver of any unintended direction changes of the car. Optic flow fields obtained from flow algorithms applied to camera image sequences often contain errors (Barron et al. 1994; Jähne 1997; Kalkan et al. 2005). Optic flow estimation is plagued by ambiguities due to the aperture problem, the correspondence

ISSN: 0954-898X print / ISSN 1361-6536 online
DOI: 10.1080/09548980600563962

problem, noise emerging from resolution and quantization effects, lack of signals in homogeneous image areas, and ambiguities caused by depth discontinuities. Therefore, estimates of self-motion obtained from raw optic flow are error-prone. Since motion estimation in biological systems also commences with spatial-temporal variations of image intensity, it is likely that it faces the same problematic optic flow at the input stage. However, biological systems are clearly able to estimate self-motion very precisely (overview in Lappe et al. 1999). Somehow, therefore, the brain must have developed methods to overcome the shortcomings of the early motion detectors. We searched for features which enable biological vision systems to handle noisy flow fields successfully.

In the visual system of primates, the output of the set of motion-sensitive cells in the primary visual cortex (V1) can be regarded as the raw optic flow field. The output of these cells is further processed by higher order motion sensitive cells in the medial temporal area (MT). Thereafter, the signals are transferred from area MT to the medial superior temporal area (MST), which is thought to analyze the entire optic flow pattern and to extract the parameters of ego-motion (Lappe et al. 1996). Area MT establishes a space-variant map of the visual motion field as the diameters d of the receptive fields of the neurons in MT increase proportionally with the eccentricity from the center of the field of view (Albright & Desimone 1987).

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

Based on the properties of this map Lappe (1996) proposed a method for decreasing the noise in optic flow fields by averaging flow vectors over image areas, the sizes of which increase with the 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. The receptive field sizes depend on the position of the center of the receptive fields according to Equation 1 (see also Figure 1).

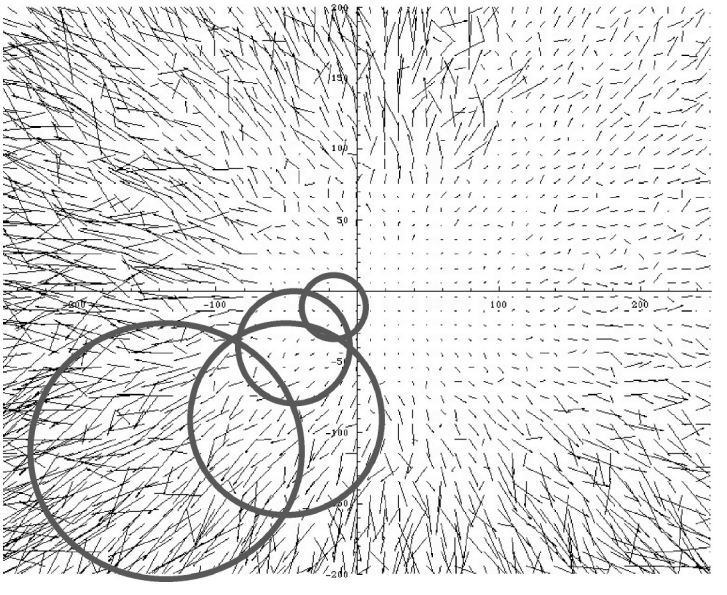

Figure 1. Increasing filter fields sizes.

The spatial integration over peripherally increasing image areas appears 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 (Lappe & Rauschecker 1995). The singular point is the center of an essentially 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 too much information. In human vision this is true even when the direction of heading deviates from the direction of gaze since eye rotation reflexes introduce rotational flow that nulls the motion in the direction of gaze (Lappe & Rauschecker 1995; Lappe et al. 1999). Averaging over large areas is more favorable for noise reduction and smoothing. Averaging over small areas retains information if neighbored signals are different. Thus, the space-variant mapping is a compromise that satisfies both goals to the degree necessary for the structure of the flow field.

In Lappe (1996), one can find an application of this method and an implementation of a standard heading detection algorithm (Heeger & Jepson 1992) 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 that heading detection is possible with errors up to 4 degrees for a signal to noise ratio of 1. These data resemble the human performance with identical stimuli in psychophysical experiments (van den Berg & Brenner 1994a, 1994b). The results demonstrate that the space-variant filtering stage is an applicable model for the flow representation in the primate visual system and that it can emulate the performance characteristics of the human visual system. It remains unclear, however, 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 on. Furthermore, the method might be suboptimal for the types of motion performed by a rigidly installed camera in a moving car. For example, when driving through curves, the singular points of the flow fields can be 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 ego-motion estimation, we use the task of heading estimation from the filtered flow fields as a criterion for improvement. To prove the assertion that the space-variant filtering method is well adapted to the typical structure of the flow field, and allows more precise heading estimation than other averaging methods, we compare the space-variant filtering method with a filtering procedure using averaging domains with constant radiuses in the visual field (constant filtering method). We chose the radius of these averaging domains as the mean radius of the space-variant averaging domains. For completeness, results of heading estimation obtained from filtered flows are 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. Since we have no ground truth information about the correct heading, our main criteria to evaluate the efficiency of the filtering methods are the scatter of individual heading estimates from different sets of flow vectors and the variability 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 from visual inspection of the results.

The second part of our analysis is based on image sequences constructed from three dimensional data sets of several natural scenes. These image sequences are directly calculated from the scene data assuming a given self-motion. Flow fields are calculated from the image sequence. Since the correct heading is given, this procedure allows one to evaluate the estimated heading results against ground truth information. Thus, we can precisely quantify the performance of the different filtering methods with respect to heading detection.

Methods

Estimation of optic flow from image sequences

Since many algorithms and models of motion estimation have been proposed (cf. Barron et al. 1994), 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 (Lucas & Kanade 1981) 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 (Baker & Matthews 2004). 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 estimations 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. A dense flow field in this context means that almost each pixel carries a motion signal, even in areas with low contrast where the motion signal would be noisy. Obviously, in image areas devoid of any contrast – such as sky – the algorithm will not compute a flow measurement. These areas are withdrawn from further calculations. Note that 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 Barron et al. (1994). Briefly, let $\nabla I(x, y, t)$ and $I_t(x, y, t)$ be the spatial gradient and the temporal derivative of the luminance at image position (x, y). The estimated flow vector $v(x, y)$ at position (x, y) is the solution of the weighted least-square fit

$$v(x, y) = \min_{v} \left(\sum_{(x', y') \in \Omega(x, y)} (\nabla I(x', y', t)v + I_t(x', y', t))^2 \right), \tag{2}$$

where $\Omega(x, y)$ denotes a small neighborhood of (x, y). In our implementation the window function is set to unity.

The calculation of the spatial luminance gradients and the temporal luminance derivations are based on super-pixels, which are formed by a sub-sampling procedure. The luminance of a super-pixel at a certain position is the averaged luminance of all pixels within a quadratic neighborhood around the position. To calculate spatial luminance gradients we use a spatial neighborhood of 3×3 super-pixels. The edge lengths l of the super-pixels take the values $l = 2n + 1$. Because l is an odd number, each super-pixel has a definite central position. The central positions of super-pixels belonging to the same neighborhood take positions shifted $(l - 1)$ pixel from the central position of the central super-pixel and have the same edge length. Let $(\bar{I}_{i,j})_{i=1,2,3;j=1,2,3}$ be the set of averaged luminance values of the super-pixels of a the neighborhood, where $(1, 1)$ denotes the super-pixel in the top-left corner and $(3, 3)$ denotes the super-pixel in the down-right corner. The spatial derivatives of the luminance at

the respective position are obtained from

$$I_x = \frac{1}{2}\left(\frac{\bar{I}_{2,3} - \bar{I}_{2,1}}{l-1} + \frac{\bar{I}_{3,3} - \bar{I}_{1,1}}{2(l-1)} + \frac{\bar{I}_{1,3} - \bar{I}_{3,1}}{2(l-1)}\right)$$
$$I_y = \frac{1}{2}\left(\frac{\bar{I}_{1,2} - \bar{I}_{3,2}}{l-1} + \frac{\bar{I}_{1,1} - \bar{I}_{3,3}}{2(l-1)} + \frac{\bar{I}_{1,3} - \bar{I}_{3,1}}{2(l-1)}\right), \tag{3}$$

where the distance between super-pixels lying on the diagonal are considered as $\sqrt{2}(l-1)$. This procedure takes all of the four possible numerical directional derivatives of a 3×3 neighborhood into account. The spatial luminance gradients are calculated for each position/pixel of the image by pixel-wisely shifting the neighborhoods. The calculation of the temporal derivatives is based on the super-pixels described above and a temporal neighborhood of three frames. The sizes of the super-pixels belonging to a certain neighborhood increase linearly with the distance of the position to the principal point. This approach takes into account the smaller spatiotemporal shift of the luminance close to the center of view and the larger spatiotemporal shift of the luminance in the periphery. The specific implementation of this approach is adapted to the expected maximum flow speeds in the periphery.

The spatial neighborhood for the least-square minimization at a certain position depends on the size of the super-pixel at this position and comprises the pixels within a quadratic neighborhood with edge length $3l$. In certain degenerate cases, the algorithm will not find a solution and 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 Figure 2).

Space-variant and constant filtering methods

The filtering is based on averaging flow vectors over domains of visual space. The domains are defined as follows. Let f be the focal length of the camera and the pair (x, y) denotes a position on the image. Without loss of generality, let the principal or foveal point being at $(0, 0)$. The position vector $r(x, y)$ in visual space (a unit sphere) corresponding to the position (x, y) on the image is

$$r(x, y) = \frac{1}{\sqrt{x^2 + y^2 + f^2}}(x, y, f). \tag{4}$$

Then $d_{\phi,(x,y)}$ is a spatial domain with radius ϕ at the position (x, y):

$$d_{\phi,(x,y)} := \{(x', y') | \arccos(r(x, y)r(x', y')) \leq \phi\}. \tag{5}$$

Formula 5 simply means that (x', y') lies within $d_{\phi,(x,y)}$ if the angle between the vectors $r(x, y)$ and $r(x', y')$ is smaller than ϕ. $d_{\phi,(x,y)}$ covers a circular section of the visual space, which corresponds to a distorted disc in the image plan.

To reduce calculation time the filtered flow is generated only for a subset of positions of the image. This is sufficient because the heading estimation procedure is based on a subsampling of the flow field. At the selected positions the filter procedure is performed over all signals of the original pixel grid lying within the averaging domain. Pixel positions with vanishing motion signals ($|v| < 0.0001$pixel/frame) are not incorporated in the filtered flow and are not regarded for the averaging procedure for other positions. Positions (x, y) with radiuses ϕ, for which averaging domains $d_{\phi,(x,y)}$ would cross the borders of the image, are also omitted from the set of filtered flow vectors. Therefore, regions which present the sky in the images and regions close to the image boundaries do not contribute to the filtered flow.

Figure 2. A: First frames of the camera sequences for different driving scenarios, left: town sequence, right: road sequence. B: Raw optic flow estimated from the first three frames by the Lucas – Kanade algorithm and heading estimates computed from 100 random sub-samples of the flow field. The black cross marks the principal point. The small boxes marks the singular heading estimates. The large boxes marks the mean heading estimate of the 100 runs. C: Flow fields and heading estimates after constant filtering. D: Flow fields and heading estimates after space-variant filtering.

According to Formula 1, for space-variant filtering the radius ϕ_{sv} at position (x, y) is

$$\phi_{sv}(x, y) = 0.006 + 0.305 \arccos\left(\frac{f}{\sqrt{x^2 + y^2 + f^2}}\right). \tag{6}$$

For constant filtering we use a radius ϕ_{const}, which is the average radius of the subset of positions P on which the space-variant filtering is performed on:

$$\phi_{const} = \frac{1}{N} \sum_{(x,y)\in P} \phi_{sv}(x, y). \tag{7}$$

N is the number of positions in P.

Estimation of heading

For the estimation of heading from the optic flow we use a version of the Heeger – Jepson subspace algorithm (Heeger & Jepson 1992) as implemented in Lappe (1996). 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

$$R(t) = \|(\Phi^t C^{\perp}(t))\|^2 \tag{8}$$

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 (Lappe & Rauschecker 1995; Lappe et al. 1996). 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 five individual residual functions for each collection of ten flow vectors are summed into a compound residual function. Thus, a single heading estimate of the algorithm is based on a sub-sampling of the flow field of 50 vectors.

Results for real camera sequences during car driving

We tested the filtering method for two image sequences of different car driving situations. The first sequence was taken whilst driving in a town. The second image sequence was taken during straight whilst open road driving. The first images of the sequences can be seen in Figure 2A. While recording the first image sequence the car is going through a slight curve to the left. Thus, the motion of the car involves rotation, and the singular point of the motion pattern is shifted to the left (see Figure 2B). Furthermore, approaching cars disturb to some extent the global flow pattern obtained from the static scene. Please note that even during curved motion as in the town sequence, the heading is the tangential vector of the curve. Thus, the true heading vector is parallel to the camera provided that the camera is aligned with the longitudinal axis of car. The motion underlying the second image sequence is mainly straight ahead. The scene is only sparsely populated with objects and dominated by ground and sky. The long edges of the trees lead to errors in flow estimations from the aperture problem. Thus, the selected image sequences represent a large amount of problems occurring in real image sequences. The image sequences were recorded by a camera rigidly 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 (p_x, p_y) was $(627, 551)$.

Ten successive frames are used from each sequence to allow the tracking of the estimated heading over a length of time of car motion. Raw flow fields for the image sequences are estimated with the Lucas – Kanade algorithm as described previously. The edge length l of the quadratic super-pixels at pixel-position (x, y) are calculated with

$$l = 5 + 2 f_{int}(0.03\sqrt{(x - p_x)^2 + (y - p_y)^2}), \tag{9}$$

where the function f_{int} gives the integer fraction from the argument. The specific implementation of Equation 9 relies on the assumption that the speed of the car is around 50 km/h, which

leads to displacements up to 50 pixel/frame in the periphery. Note that small displacements up to 5 pixel/frame can also be detected for positions close to the principal point.

Space-variant filtered and constant filtered flow fields are derived from the raw flow. According to Equation 7 the radius ϕ_{const} of the averaging domains of the constant filtering method is $\phi_{const} = 3.7$ degree. The positions of the flow vectors which contribute to the heading estimation are randomly selected from the pixel-grid of the image, where each position in the field of view has the same likelihood to be picked. Figure 2B shows the raw flows estimated from the first three frames of the image sequences. The flow fields show the typical pattern of a mainly forward motion. The influence of rotation due to curved motion can be seen by the left-ward shifted point of vanishing motion and the longer flow vectors in the right part of the image. The flow algorithm does not detect motion signals within regions that present sky. Such regions are withdrawn from further analysis. The contrast in regions that present the ground (road, meadows) is low but sufficient to detect motion. These motion signals might be erroneous, however.

The third row and the fourth row show the constant filtered flow and the space-variant filtered flow, respectively. In contrast to the raw flows, the filtered flow fields are very smooth and the noise occurring in the raw flow fields is strongly decreased. Differences between the optic flows filtered with different methods are hard to recognize. The panels next to the pictured flow fields in the second, third and fourth row of Figure 2 show the results of heading estimation based on the respective flow fields. One hundred random sub-samples of 5×10 flow vectors were selected for each flow field and heading was estimated as described in the previous section based on the particular sub-sampling. For the raw flow, heading estimates from different sub-samplings of the flow field show large variability. The standard deviations over 100 runs are 6.4 degrees horizontally and 2.3 degrees vertically for the town scene and 7.1 degrees horizontally and 3 degrees vertically for the road scene. These suggest that the heading algorithm applied directly to the raw flow is rather unstable. In contrast, for the filtered flows the same heading estimation procedure yields estimates that appear tightly clustered. Standard deviations for the constant filtered flow are 0.7 degrees horizontally/0.5 degrees vertically and 0.8 degrees horizontally/0.7 degrees vertically for the town scene and the road scene respectively. Standard deviations for the space-variant filtered flow are 0.6 degrees horizontally/0.5 degrees vertically and 0.7 degrees horizontally/0.7 degree vertically for the town scene and the road scene respectively. Thus, the spatial consistency of the heading estimate is much improved by both filtering procedures.

Figure 3 shows the temporal consistency of the heading estimates of the town image sequence (A) and the road image sequence (B). The panels show the means of 100 singular heading estimates over ten subsequent frames of the respective sequence based on the space-variant filtered flows, the constant filtered flows, and the raw flows. The panels of the upper rows show the horizontal angle and of the lower rows the vertical angle of the mean heading estimates. The bars denote the standard deviation. If both the horizontal and vertical angles are equal to zero the heading is parallel to the camera ($(0, 0)$-direction). Figure 3 shows that for all frames the standard deviations are lower for the filtered flow fields than for the raw flow fields. Thus, the singular heading estimates based on the filtered flows are less scattered and more clustered. The mean heading estimate based on the filtered flows appears consistent and smooth over the frames and coincides with our expectation that the true heading is approximately aligned with the camera axis. In contrast, the course of the mean heading estimate from the raw flow jumps between the frames and strongly deviates from the $(0, 0)$-direction particular for the horizontal angle.

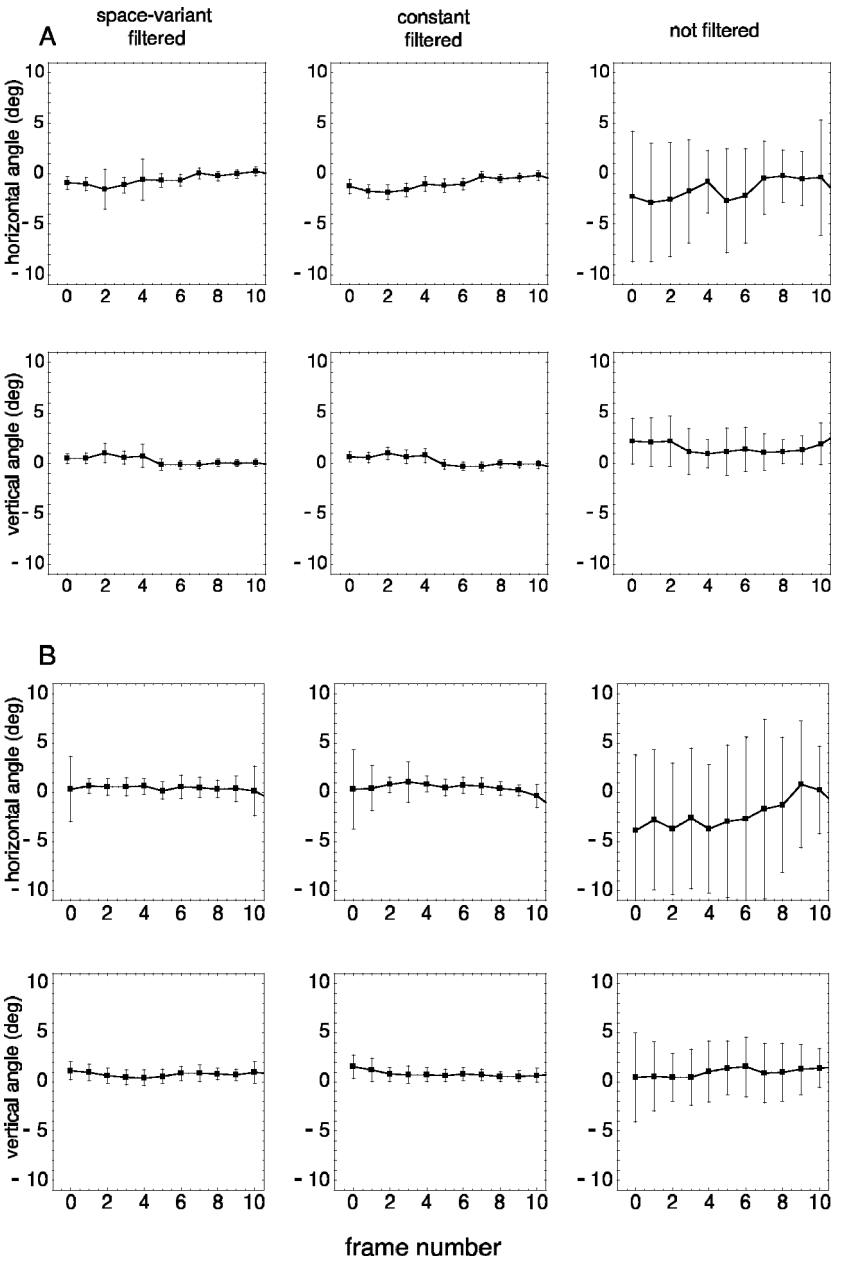

Figure 3. A: Mean heading estimates from the town image sequence for 10 sequential frames B: Mean heading estimates from the road image sequence for 10 sequential frames, left: space-variant filtered flow, middle: constant filtered flow, right: raw flow, upper rows: horizontal angle of the heading vector, lower rows: vertical angle of the heading vector. The bars denote the standard deviations of 100 singular heading estimates.

Although the results suggest that filtering methods based on averaging are feasible, there are no obvious differences between the qualities of heading estimates obtained from the space-variant and the constant filtered flow. The courses of the mean heading estimates resemble each other, and the standard deviations of the sets of singular heading estimates

are reduced to the same level for both filtering methods. There are several possible reasons why both filtering methods are equally effective for these image sequences. First, the space-variant filtering method is only sub-optimal for car motions, where the singular point of the generated flow fields is shifted away from the center of view (see Figure 2, particular the left row). Second, the scenes are sparsely structured and the planar ground (road and meadow) is predominant. We will see in the next section that flow fields generated from scenes with predominant ground show fewer but significant differences in the performance of heading estimation between the different filtering methods than flow fields generated by motion within more structured scenes. Third, the set of vectors on which the heading estimation is performed contains many vectors from the periphery. The image sequences we used cover only a very small field of view (approximately 30×26 deg^2). Differences may occur only for larger visual fields.

Results for range image sequences

Although the results obtained with real camera sequences indicate that filter approaches relying on averaging make sense, we have so far found no evidence that the space-variant filtering method is more advantageous than the constant filtering method. Possible reasons are discussed in the foregoing section. Very likely the space variant filtering in primate area MT is adapted to specifics of the flow fields occurring in ecological conditions of primates. A larger field of view is more adapted to the vision of primates and since the space-variant filtering method adopts properties of the visual systems of primates, we assert that the space-variant filtering method performs better for larger fields of view. Second, the flow field is derived from a combination of body-motion and eye-movements, unlike the motion of a car. Third, the structure of a typical scene may be different. Furthermore, the errors of heading estimates obtained from the real camera sequences can only be evaluated indirectly and qualitatively, because we have no ground truth information about the correct motion. Therefore, we extend our investigation to image sequences, which possess a larger field of view and different scene structure, and which guarantee ground truth information about the correct motion.

These image sequences are calculated from the three dimensional data of natural scenes. The direct calculation of the image sequences and the true motion fields from the three dimensional data allows to quantitatively evaluate the efficiency of the noise reduction by the filtering methods. Thus, we can investigate motion fields obtained from self-motion parameters imitating the typical biological situation, in which an observer is moving but keeps gaze on a particular object of the scene (Lappe et al. 1999). In this case, the singular point is in the center of view.

We investigate heading performance with two different procedures of selecting the sub-samples of filtered flow vectors. The first procedures (uniform sub-sampling) randomly selects the positions of the flow vectors. Each position of the visual field has the same probability to be picked. The second strategy picks with a higher likelihood positions close to the center of view (non-uniform sub-sampling). The non-uniform sub-sampling is governed by the following probability density $p(x, y)$

$$p(x, y) = \frac{1}{4\sigma} \exp\left(- \frac{|x - p_x|}{f\sigma} \right) \exp\left(- \frac{|y - p_y|}{f\sigma} \right), \tag{10}$$

where f is the focal length and $\sigma = 0.2$. The non-uniform sub-sampling qualitatively copies certain properties found in visual areas such MT, where the number of cells coding for equally sized domains in the visual field decreases with the eccentricity according to the cortical

magnification factor. We compare the performance of both strategies of heading estimation with the results of heading estimation based on constant filtered flows. The radiuses of the averaging domains of the constant filtering procedure are the mean radius of the respective space-variant filtering method.

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 (Huang et al. 2000). 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 degrees vertically and 259 degrees 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 μm). Thus, the data of each point consist of four values, the distance, the horizontal and the vertical angles 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 three dimensional data of a given environment makes it possible to simulate the view of a moving camera in this scene and to calculate both the image on the camera as well as the true motion field. Panel A of Figure 5 shows an example 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 light sensitive planes. All images used in the investigation are composed of 350000 pixels and are generated with a projection of focal length of 341 pixels. The projection plane is 683 pixels wide and 512 pixels high and covers a field of view of approximately 90×74 deg^2. The position of the principal point (p_x, p_y) is $(341, 512)$.

The resulting images are not true gray scale 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.

The natural scenes from the range image data base can be subdivided into urban scenes and forest scenes. Figure 4 shows a typical example of the urban case. Panel A of Figure 5 shows a picture from a forest scene. Overall, there are fewer objects in the urban scene than in forest scenes. Also, the depth statistics of the scenes differ. Urban scenes are dominated by the ground. Forest scenes are densely populated with brushwood in the lower field of view and trees in the upper field of view. Thus, as a by-product of our analysis, we investigate whether scene structure influences the efficiency of the filtering methods.

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

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

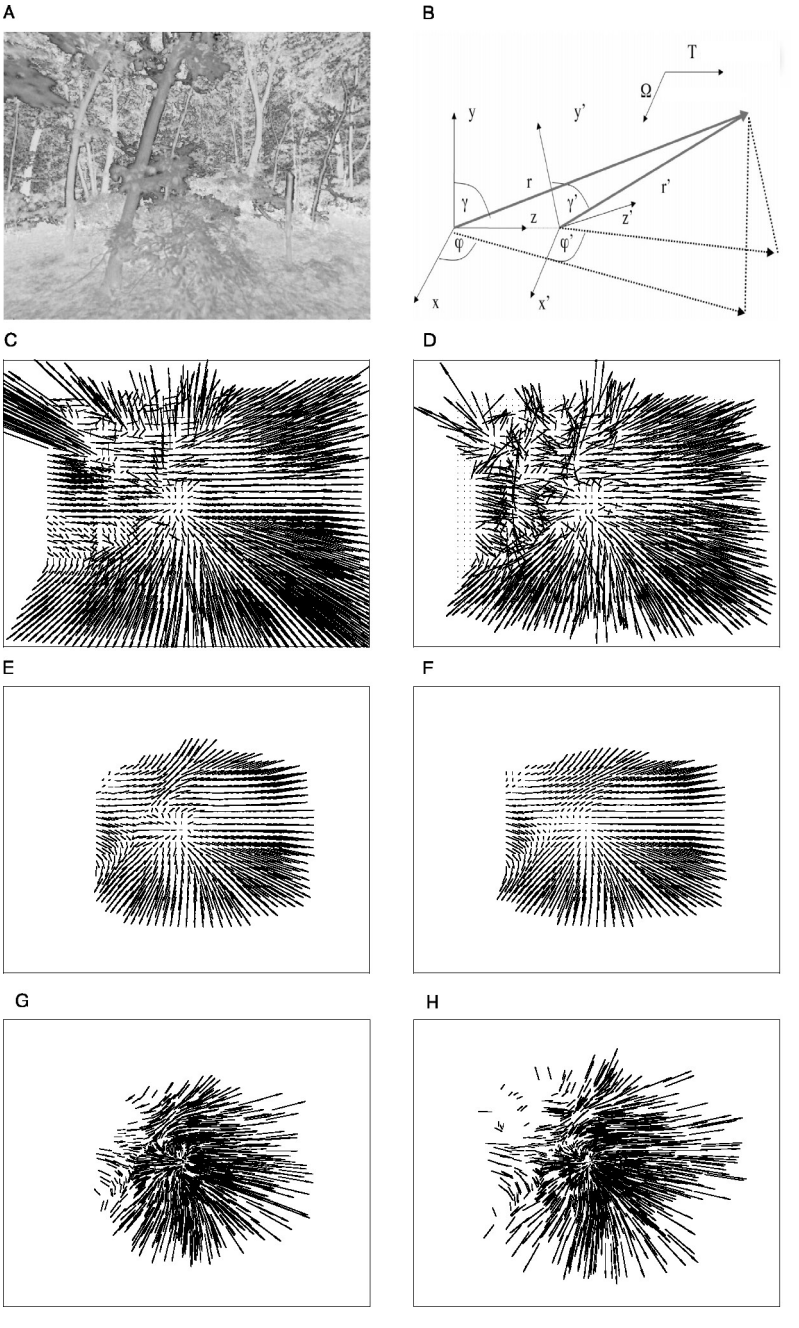

Figure 5. A: Reflectance data of a range-image projected onto a plane simulating a camera shot. B: Transformation of camera centered coordinates during ego-motion. C, D, E and F: Optic flow fields generated from range image data for a leftward translation and a rotation of 8 deg/second. C: Actual flow. D: Raw flow estimated from an image sequence. E: Space-variant filtered flow, positions uniformly selected. F: Constant filtered flow, positions uniformly selected. G: Space-variant filtered flow, positions non-uniformly selected. H: Constant filtered flow, positions non-uniformly selected.

$T = (T_x, T_y, T_z)$ 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

$$r' = \frac{\Omega \cdot (r - s)}{\|\Omega\|^2}\Omega + \cos(\phi)\left((r - s) - \frac{\Omega \cdot (r - s)}{\|\Omega\|^2}\Omega\right) + \sin(\phi)\frac{(r - s) \times \Omega}{\|\Omega\|}, \qquad (11)$$

where $\cdot$ and $\times$ denote the scalar product and the vector product respectively. The image sequences we investigate are generated by translations of 0.06 m/frame and rotations between 0.04 deg/frame (1 deg/second) and 0.6 deg/frame (15 deg/second). 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 the stabilized point is different from the heading direction. The data set we use comprises 70 different urban scenes and 32 different forest scenes. For each scene 30 different motion situations are tested. The motion situations are distinguished by the magnitudes and direction of rotation of the camera. Psychophysical findings (Lappe et al. 1999) and computational considerations (Koenderink & van Doorn 1987) suggest that rotation rate has a critical influence on heading estimation. If the distance between the fixated object and the camera is known, a certain value of rotation can be achieved by a particular motion direction relative to the axis of the camera (or view). For each scene and motion situation the optic flow is estimated by the Lucas – Kanade algorithm from the image sequences as described earlier. The edge length l of the quadratic super-pixel at pixel-position (x, y) is calculated with the equation

$$l = 3 + 2 f_{int}(0.02\sqrt{(x - p_x)^2 + (y - p_y)^2}). \qquad (12)$$

The specific implementation of Equation (12) relies on the assumption, that the chosen translation and rotations lead to displacements up to 15 pixel/frame in the periphery. Figure 5D shows an example of a raw flow field obtained from the range image data for a leftward motion and a rotation of 8 deg/second. The actual flow field can be seen in Figure 5C. Figure 5E shows an example of the space-variant filtered flow, where positions are selected by uniform sub-sampling. The mean radius of the averaging domains is 10.5 degrees, which is also the radius of averaging domains of the respective constant filtering method. Figure 5F shows an example of the respective constant filtered flow. Figure 5G shows an example of the space-variant filtered flow, where the positions are selected by non-uniform sub-sampling. Figure 5H shows an example of the respective constant filtered flow. The radius for the averaging domains is 5.3 degrees.

Quality of heading estimates

Figure 6 depicts examples of the heading estimation results obtained from raw and filtered flows for an urban and a forest scene with uniform sub-sampling. For each scene, 25 single runs are shown. The results are similar to those of the car driving scenes in that the estimates are quite erratic for the raw flow and more tightly clustered for the filtered flow. Particularly for the forest scene the single heading estimates obtained from the constant filtered flow appear more variable and deviate to a greater extent from the correct heading than the single

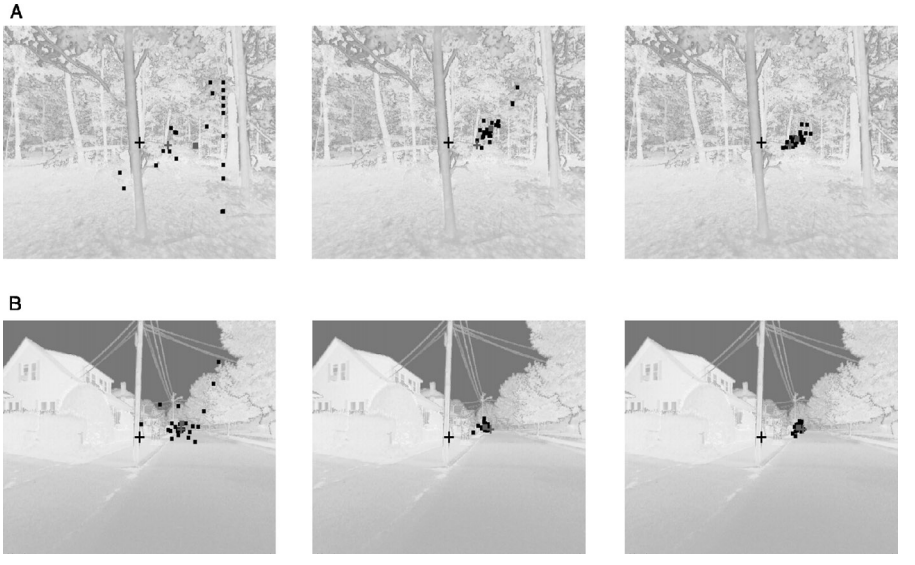

Figure 6. Results of heading estimation for two example scenes. A: Forest scene. B: Urban scene. Rotations for both scenes are 8 deg/s. The black cross denotes the principal point of the camera. The small black points show the results of single runs. The large grey point marks the mean over 25 runs. The grey cross denotes the correct heading. Left: Unfiltered flow. Middle: Constant filtered flow. Right: Space-variant filtered flow.

heading estimates obtained from the space-variant filtered flow. Differences are less clear for the urban scene.

We calculated heading performance for both sub-sampling procedures and for all available scenes. 25 heading estimates with different sub-samplings of the flow field were used for each motion situation. 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. Thus, the total number of heading estimates for each rotation value ranged from 2000 to 3500 for the urban scenes and from 1150 to 1600 for the forest scenes. The panels of Figure 7 show the mean errors of the single heading estimations over all scenes and motion situations for the raw and the filtered flows. Panels A and B present the results based on the uniform and the non-uniform sub-sampling for the forest scenes. Panels C and D present the results based on the uniform and the non-uniform sub-sampling for the urban scenes. The bars denote the 95 percent confidence intervals for the mean errors of each rotation value. In all cases, the filtered flow provides superior performance over the unfiltered flow, although the error increases with a rising rotation rate. Thus, filtering methods based on averaging procedures lead to more reliable heading estimations also with respect to the correct heading. This result confirms the findings in Lappe (1996) with respect to natural scenes. The diversity in performance of the space-variant and constant filtering method can also be seen in Figure 7. It is clearly discernible that space-variant filtering performs significantly better than constant filtering. The advantage of the space-variant filtering method is clear for rotation values less than 10 degrees/second. It is more pronounced for the non-uniform sub-sampling. The decrease of the mean errors produced by the space-variant filtering method compared to the mean errors produced by the constant filtering method range from 0.5 degrees to 1.5 degrees. Panels C and D present the results based on the uniform and the non-uniform sub-sampling for the urban scenes. Generally, the errors for both filtering methods are lower

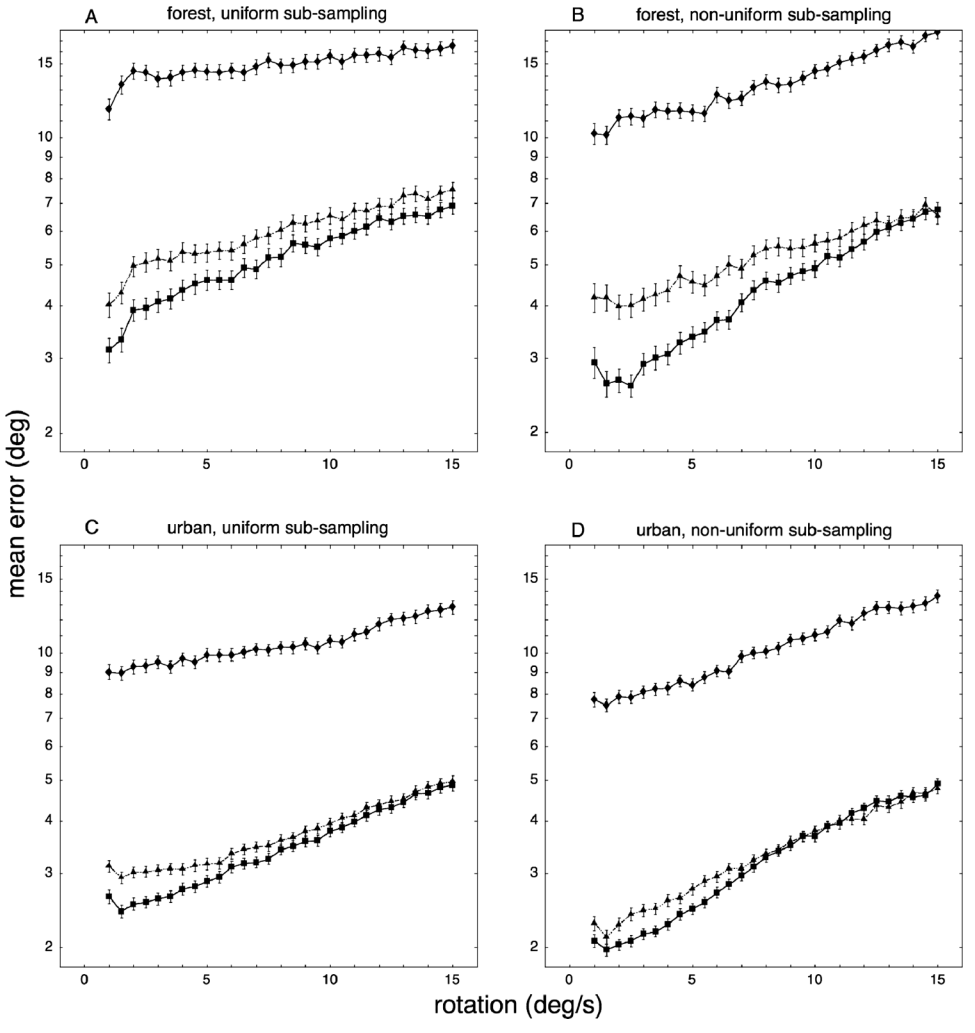

Figure 7. Logarithmic plot of mean errors for heading estimation. Boxes: Space-variant filtered flow. Triangles: Constant filtered flow. Diamonds: Raw flow. Error bars denote the 95 percent confidence intervals. A: Forest scenes, uniform sub-sampling method, mean domain radius 10.5 degree. B: Forest scenes, non-uniform sub-sampling method, mean domain radius 5.3 degree. C: Urban scenes, uniform sub-sampling. D: Urban scenes, non-uniform sub-sampling.

than the errors for the forest scenes. Space-variant filtering performs significantly better than constant filtering for low rotation values up to 6 degrees/second. However, the advantage of the space-variant filtering method is not as strong as for the forest scenes. The decrease of the mean errors produced by the space-variant filtering method compared to the mean errors produced by the constant filtering method is less than 0.5 degrees. Finally, we find that the non-uniform sub-sampling procedure generally provides better results than the uniform sub-sampling procedure. This is true for the raw and for both filtered flows.

Conclusion and discussion

Our results demonstrate that filter methods based on averaging procedures are reasonable strategies to decrease noise in optic flow fields and to improve heading detection. The

methods work well on optic flow fields based on natural scenes affected by 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 heading is more reliable than in the unfiltered case.

For typically biological motion situations in cluttered scenes, the space-variant arrangement of the sizes of averaging domains leads to an additional advantage over constant filtering. We adopted the spatial arrangement of the sizes of averaging domains from the space-variant mapping that area MT of primates imposes on the raw motion field conveyed from the area V1. Our results suggest that the space-variant map, which provides a filtering procedure, is a favorable adaptation to biological motion situations and scenes which resemble the natural habitat of primates (forest scenes). The space-variant filtering method is advantageous for urban scenes only to a lesser degree. Since the space-variant filter method does not require additional calculation time compared with the constant filtering method, for artificial visual systems which are intended to perform well in both kinds of scenes, the realization of the space-variant filter method is advisable.

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 over small image areas in the center of the visual field. Thus, noise reduction should in principle be particularly effective in the periphery. However, we observed that heading estimates were more reliable when the sub-sampling of the flow was non-uniform such that a higher proportion of flow vectors with lower eccentricity contributed. 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 our simulations, this is similar to the increasing size of the super-pixels in the periphery. In this case, the selection of a higher proportion of low eccentricity flow vectors for the heading estimation balances the larger convergence of motion measurements in the periphery, such that the total contribution of measurements is uniform across the visual field. Effectively, each early flow measurement is equally likely to contribute to the heading estimate but peripheral flow measurements are more strongly filtered. Therefore, our filtering method becomes more efficient when 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 as a front end to our approach (Franceschini et al. 1992; Sandini & Metta 2002).

Filtering based on averaging procedures may not be the only way to obtain a rather exact heading estimate from optic flow. A single flow vector in the filtered case comprises information from a large set of raw flow vectors. In contrast, a single vector picked up from the raw motion field is only one measurement. Therefore, equivalent improvements of heading estimation may be reached if more flow vectors from the raw flow field are used in the heading estimation algorithm. Also, combining a larger number of heading estimates from different sub-samplings into a compound estimate may improve the heading estimation from the raw flow field. However, the requirements of calculation time of the heading procedure increases considerably with the number of flow vectors. Therefore, the space-variant filtering stage is a simple and effective mid-level method to restore 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 reconstruct the actual optic flow. Direct comparisons of the

true motion field and the filtered flow in Figure 5 shows the differences. We rather believe that the space-variant filtering method saves the structure of the flow field only to a sufficient extent to decode the 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 (Schwartz 1977; Mallot et al. 1990; Baratoff et al. 2000).

The successful performance of heading estimation on the filtered flow should only be the first step of a larger scheme. Adding disparity information from a second camera is proposed to further improve heading estimation (Lappe 1996) and should be tested also for natural scenes. Also, reliability measures such as the intrinsic dimensionality (Kalkan et al. 2005) may be used to weigh the contribution of inputs. One may also be interested in estimating 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 external moving 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, the space-variant filtering method could be used as a part of an improved optic flow algorithm.

Acknowledgements

M.L. is supported by the German Science Foundation La952/2 and La952/3, the German Federal Ministry of Education and Research BioFuture Prize, and the EC Projects ECoVision and Eurokinesis. We thank the Hella KG Lippstadt for preparing the camera sequences.

References

Albright TD, Desimone R. 1987. Local precision of visuotopic organization in the middle temporal area (MT) of the macaque. Exp Brain Res 65:582–592.

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

Baratoff G, Toepfer C, Neumann H. 2000. Combined space-variant maps for optical-flow-based navigation. Biol Cybern 83:199–209.

Barron JE, Fleet DJ, Beauchemin SS. 1994. Performance of optical flow techniques. I J Computer Vision 12:43–77.

Franceschini N, Pichon JM, Blanes C. 1992. From insect vision to robot vision. Philos Trans R Soc Lond B 337:283–294.

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

Huang J, Lee AB, Mumford D. 2000. Statistics of range images. CVPR, 1:1324–1331.

Jähne B. 1997. Digital image processing – concepts, algorithms, and scientific applications. New York: Springer.

Kalkan S, Calow D, Felsberg M, Wörgötter F, Lappe M, Krüger N. 2005. Local image structures and optic flow estimation. Network: Comp Neural Systems 16:341–356.

Koenderink JJ, van Doorn AJ. 1987. Facts on optic flow. Biol. Cybern 56:247–254.

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

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

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

Lappe M, Bremmer F, van den Berg AV. 1999. Perception of self-motion from visual flow. Trends Cogn Sci 3:329–336.

Lappe M, Rauschecker JP. 1995. Motion anisotropies and heading detection. Biol Cybern 72:261–277.

Lucas BD, Kanade T. 1981. An iterative image registration technique with an application to stereo vision. Proc DAPRA Image Understanding Workshop, pp 121–130.

Mallot HA, von Seelen W, Giannakopoulos F. 1990. Neural mapping and space-variant image processing. Neural Networks 3:245–263.

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

Schwartz EL. 1977. Spatial mapping in the primary sensory projection: Analytic structure and relevance to perception. Biol Cybern 25:181–194.

Vaina LM, Beardsley SA, Rushton S, editors. 2004. Optic Flow And Beyond. Amsterdam Kluwer Academic Press.

van den Berg AV, Brenner E. 1994a. Humans combine the optic flow with static depth cues for robust perception of heading. Vision Res 34:2153–2167.

van den Berg AV, Brenner E. 1994b. Why two eyes are better than one for judgements of heading. Nature 371:700–702.