Rotation-invariant Optical Flow by Gaze-depended Retino-cortical Mapping

Markus A. Dahlem and Florentin Wörgötter

Computational Neuroscience
Department of Psychology
University of Stirling
Stirling FK9 4LA
Scotland / UK
{mad1, faw1}@cn.stir.ac.uk
http://www.cn.stir.ac.uk/

Abstract. Computational vision models that attempt to account for perception of depth from motion usually compute the optical flow field first. From the optical flow the ego-motion parameter are then estimated, if they are not already known from a motor reference. Finally the depth can be determined. The better the ego-motion parameters are known by extra-retinal information to be restricted to certain values before the optical flow is estimated, the more reliable is a depth-from-motion algorithm. We show here, that optical flow induced by translational motion mixed with specific rotational components can be dynamically mapped onto a head-centric frame such that it is invariant under these rotations. As a result, the spatial optical flow dimension are reduced from two to one, like purely translational flow. An earlier introduced optical flow algorithm that operates in close approximation of existing brain functionality gains with this preprocessing a much wider range of applications in which the motion of the observer is not restricted to pure translations.

1 Introduction

One of the chief problems in computational vision is the three-dimensional reconstruction of a static scene from two-dimensional images [1]. Motion parallax is one of the depth cues that can be used to recover the three-dimensional structure of a viewed scene [2], [3]. Motion induces a velocity field on the retina called the optical flow [4]. In the most general motion case, i. e., object plus ego motion, the resulting curved optical flow field pattern cannot be resolved for depth analysis without additional assumptions [5] and even if simplifying assumptions are made, the problem of depth-from-motion remains rather complex.

The goal of this study is to generalize a neuronal algorithm one of us (FW) developed earlier [6]. So far this algorithm analysis radial optical flow fields, obtained by translational motion towards an object. The central advantage of this algorithm is that all computations remain local, which permits parallelization (see below). We now introduce a preprocessing step, that maps the visual data such that the algorithm can still operate in parallel even when the flow fields are more complex. The preprocessing is actually a global operation in the sense that it effects the whole optical flow field, but it is strictly independent of any external information, such as the viewed scene, and can therefore be implemented separately. The main focus for our depth-from-motion algorithm is on a close approximation of existing brain functionality, i. e. (cortical) sensor maps, and local (neuronal) operations.

Radial flow fields are one of the simplest optical flow fields and are obtained when the observer is moving straight ahead with fixed gaze towards the direction of the translation. The optical flow has then a fixed point, called the focus of expansion (FOE). All optical flow trajectories move outwards from the FOE. It is readily seen that the motion in such a radial flow field (RFF) is one-dimensional in polar coordinates, along the radial coordinate. The possible reduction of the spatial dimensions of the optical flow field can be seen as the main reason why there is a simple relation between the flow velocity and the distance between objects and observer. However, RFF exits only for rather restricted motion, i. e., purely translational motion.

The flow velocity v_p at a certain point p of the RFF is reciprocally proportional to the Cartesian coordinate Z (depth).

$$Z \sim \frac{1}{v_p} \tag{1}$$

One of the Cartesian coordinates (X,Y,Z) of objects can therefore be recovered from the optical flow, except for a scaling constant. Two other coordinates are actually already implicitly known as the eccentricity θ and the azimuth ϕ of the retinal frame. These can be interpreted as the polar and azimuthal angles in three-dimensional spherical coordinates (ρ,θ,ϕ). The unknown radius ρ, that is, the only spherical coordinate that was lost by the central perspective projection, is given by $\rho = Z\theta/f$, where f is the focal length of the visual system. At last, the scaling constant in the Z coordinate (Eq. 1) can be eliminated, if the velocity of the ego-motion is known to the observer, for example by a motor reference. With these relations the three-dimensional world can be reconstructed.

The optical flow field and its relation to the distance between objects and observer is far more complicated, when there are rotational components in the ego-motion. When, for example, the direction of gaze changes while the body is moving straight, the directions of the optical flow field are not independent of the distance of objects. Therefore the direction of the optical flow field can not be known *a priori*. In this case, a two-dimensional correlation problem must be solved to obtain the optical flow direction and magnitude. Additionally, all ego-motion parameters must be known to re-construct a three-dimensional scene from the optical flow [7], [8]. Instead of analyzing the two-dimensional optical flow field, we suggest a dynamical mapping of the visual data such that the spatial dimensions of the resulting flow field are reduced to one and the same neuronal algorithm can be applied after this preprocessing as already used for purely translational motion.

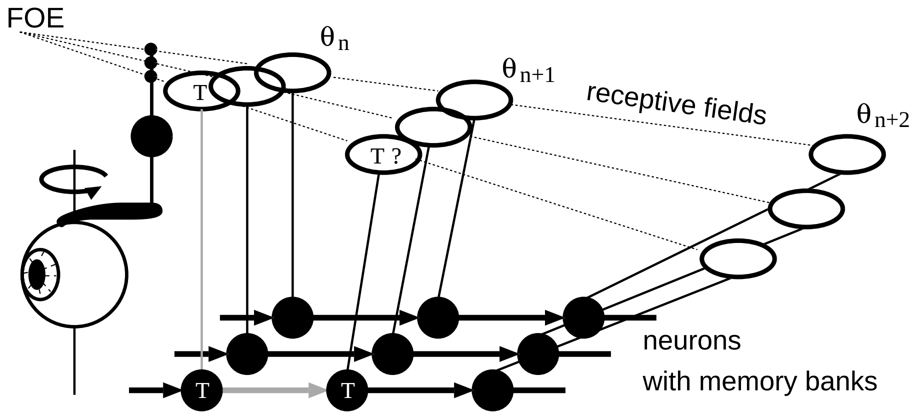

Fig. 1. Architecture of the two layer neuronal network. The input layer consists of receptive fields sampling the optical flow. Each receptive field projects to a neuron with a memory bank in the processing layer. A separate neuron represents a structure mapping eye-positions. A visual tokens (T) is passed from the receptive field along the exemplarily shown grey connections towards the memory bank of a consecutive neuron. A head-centric representation of visual input is achieved by dynamically mapping the receptive field positions according to the direction of gaze. To re-construct three-dimensional position of viewed objects, the processing layer needs only locally exchanged information in one spatial direction (from left to right).

2 The RFF-Algorithm

One of us (FW) introduced earlier an algorithm that efficiently analysis an RFF and reconstructs the viewed three-dimensional scene [6]. Details of the algorithm should be taken from that reference, but we will shortly describe its basics. Since the optical flow directions are fixed in an RFF, only the flow velocity is unknown. The velocity is measured by the time a specific visual token takes to pass successive points ("receptive fields") located on the retina at eccentricity θ_n and θ_{n+1} on a single radial line with azimuth ϕ. As the correspondence token changes in gray-level were chosen. When a significant change in gray-value is registered at a receptive field, this gray-level value is passed from θ_n, via a "neuron" n_n to a memory bank of a neuron n_{n+1} with the adjacent receptive field at θ_{n+1} (Fig. 1). The time taken to "see" this expected gray-level value on the receptive field θ_{n+1} is proportional to the depth of the object generating the gray-level token (Eq. 1). The RFF-algorithm was successfully tested on real images in real time, in other words it is sufficiently fast and noise robust. Head-centric maps [9], [10] are used now for the RFF-algorithm, and therefore straight head motion can be combined with eye-gaze movements. Any other algorithm, that is developed for one-dimensional RFFs, can as well profit from the dynamical mapping strategy introduced here. However, we would like to emphasize, that the main motivation of preprocessing the optical flow is that the remaining computations are strictly performed locally and thus can be done in parallel.

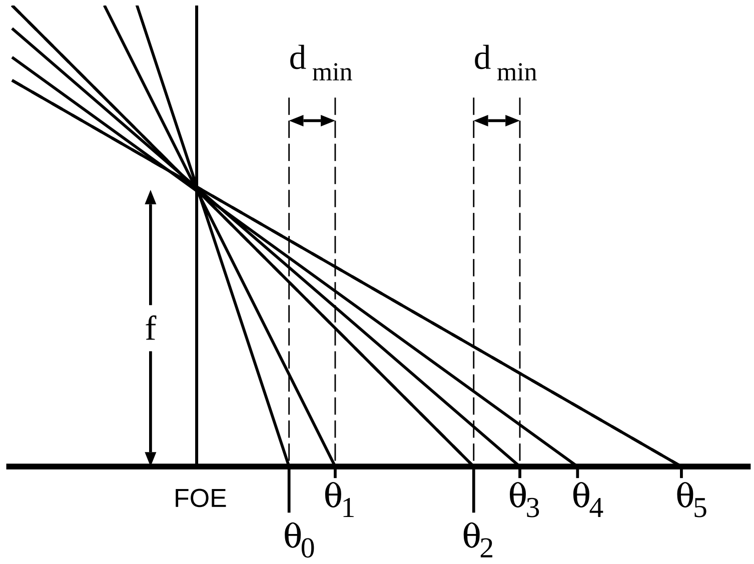

Fig. 2. Geometry of the layout of receptive fields along one radial component of the optical flow. The first receptive field θ_0 is set close to the focus of expansion (FOE) and defines the starting field of the first hyperbolic section. The second receptive field θ_1 is placed at the distance d_{min} from the first in the direction away from the FOE. In the example shown here, only one more receptive field θ_2 fits on this hyperbolic section before the distance of successive receptive field becomes too large. θ_2 is therefore the starting receptive field for the next hyperbolic section.

3 Dynamical Mapping

To map the retinal flow field to a head centric frame, the retina is initially sampled by point-like receptive fields (top layer in Fig. 1), such that the layout of the receptive fields matches the RFF. The receptive fields are placed on a polar grid defined by m radial axes expanding from the FOE. If the distance of successive receptive fields increases hyperbolically, the optical flow is sampled uniformly along one radius (Eq. 1). However, only few receptive fields would fit on a radial axis, when their positions increase hyperbolically. Therefore, the overall layout is composed of pice-wise hyperbolic sections. The design is arranged in the following way. The first receptive field of a hyperbolic section (θ_0) is set close to the FOE at position d. The second receptive field is placed at $\theta_1 = \theta_0 + d_{min}$, where d_{min} is the minimal allowed distance of receptive fields. All subsequent positions increase hyperbolically, until a maximal receptive field distance d_{max} between neighboring receptive field positions is reached. The next receptive field position is then set again the minimal distance d_{min} away from the former and a new hyperbolic section starts. This leads to:

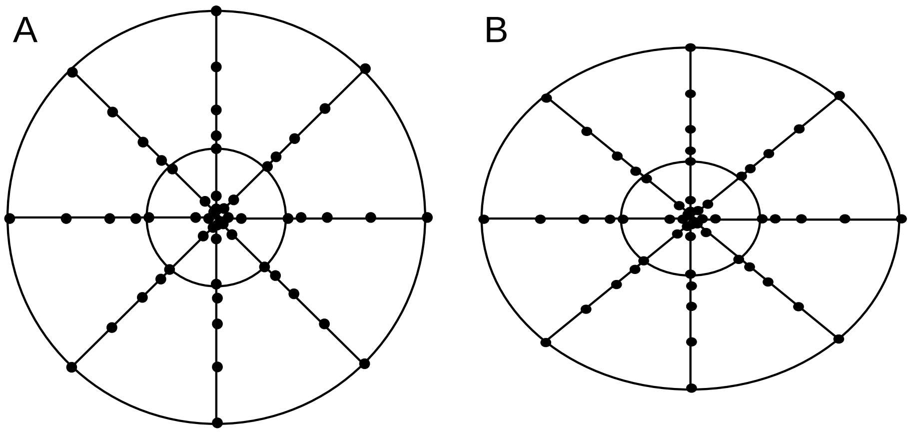

Fig. 3. Layout of receptive field grid with $m = 8$ radial axes. (A) If heading direction and direction of gaze coincide all radial axes of the receptive field grid are identical. (B) In the oblique case (α = constant $\neq 0$) the FOE is shifted and the receptive field grid has only a two-fold instead of a 8-fold cyclic symmetry for rotations about the FOE.

$$\theta_{n+1} = d \frac{d + \quad d_{min}}{d - n\, d_{min}}\,, \tag{2}$$

where d denotes at each hyperbolic section the first receptive field position.

Receptive field positions, that are placed according to Eq. 2, optimally sample optical flow, only if the motion direction and direction of gaze coincide. When these directions differ by a constant angle α, these receptive field positions must be re-mapped to sample the flow equally well. Since in artificial visual systems the projection plane is usually flat, the transformation due to eye rotation about α must be lifted to a transformation of a flat plane. (Note that all equations concerning the projection plane, starting from Eq. 1, were derived for a flat retina, but these equations can be adapted to curved projection planes.) Two-dimensional Cartesian coordinates on the flat plane will be denoted with small letters (x,y). After a gaze shift α about the Y axis (angle of yaw) the optical flow is transformed according to lifted rotation:

$$x^{hc}(\alpha) = f \frac{x \cos\alpha - f \sin\alpha}{f \cos\alpha + x \sin\alpha} \qquad \text{and} \qquad y^{hc}(\alpha) = f \frac{y}{f \cos\alpha + x \sin\alpha} \tag{3}$$

To derive this equations see any book on projective geometry or computer graphics (e. g. [11]). It is a handy but not necessary feature of these mapping functions that straight lines are conserved: the optical flow is still along *straight* radial lines, thus justifying the term RFF also for α = constant $\neq 0$ (oblique case). The FOE is shifted by $f \tan(\alpha)$, but an oblique RFF has only a two-fold

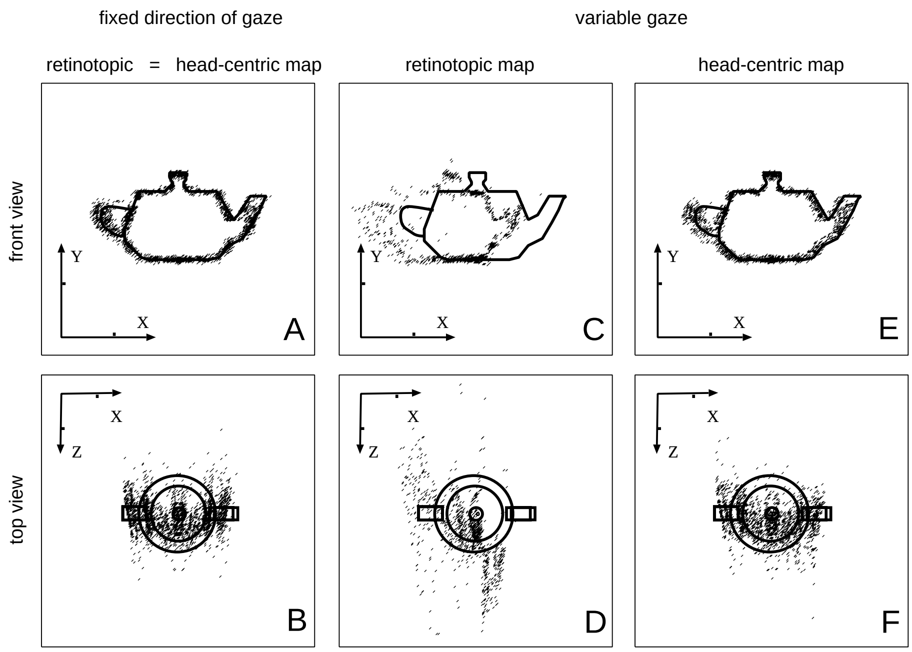

Fig. 4. A teapot viewed with stable and variable gaze. The position of the teapot in Cartesian coordinates (X,Y,Z) can be detected on a retinotopic map by the RFF-algorithm only when the gaze is pointing toward a fixed direction (A and B). The solid line is a reference to the actual position of the teapot, while the grey data points are the computed position of the visual tokens. If the direction of gaze varies while the observer is moving, the algorithm makes systematic errors (C and D). If the detected position of the teapot is to remain stable, this algorithm must operate on a head-centric map (E and F). See also text.

instead of a m-fold cyclic symmetry for rotations about the FOE, because the mapping is not conformal (Fig. 3).

The index hc in Equation 3) indicates that these coordinates are head-centric while without index they are retinotopic. On a head-centric map, the optical flow induced by translational motion combined with gaze shifts is congruent with an RFF. In other words, the RFF-algorithm is invariant under eye-gaze movements.

4 Results

An observer moving straight without changing the direction of gaze can adequately detect the three-dimensional position of the edges of objects in view by the RFF-algorithm [6]. For example, viewing a teapot and determining the optical flow field velocities in the corresponding RFF, produces the three-dimensional coordinates, shown in front view (Fig. 4 A) and top view (Fig. 4 B). These coordinate points outline the contour of a simulated three-dimensional teapot as

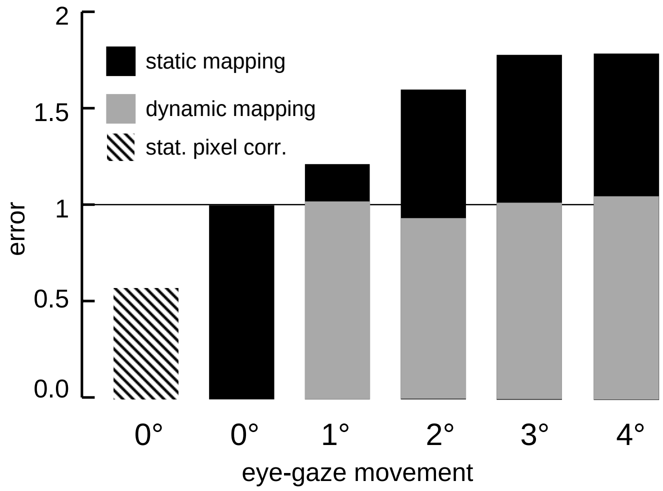

Fig. 5. Performance of the RFF-algorithm operating on a retinotopic map compared to a head-centric map obtained by dynamically mapping both with increasing eye-gaze movements. While on a head-centric map the performance is stable, on the retinotopic map is fastly deteriorates. On a retinotopic map one can take realizable receptive field positions positions into account improving the accuracy almost by a factor of two (left bar).

seen and detected by the RFF-algorithm. Note that the depth coordinate Z, shown in the top view (Fig. 4 B), is the actual output of the RFF-algorithm. The position of the edges in the other two spatial coordinates, X and Y (the contour of the object in front view of Fig. 4 A) are projected onto the retina. Therefore they are already implicitly known, except for a scaling constant.

The detection of the teapot deteriorates when the straight body motion is combined with eye-gaze movements (front view Fig. 4 C, and top view D). There is even a shift of the projection of the teapot in the X-direction, that is, in the direction of one implicitly known coordinate (Fig. 4 C). This shift is inherently in the retinotopic map. Such a map can not statically store spatial locations, because the spatial registry between retina and external space changes every time the eyes move. To be precise, edges of the teapot that are located on the retina right (left) from the FOE are accelerated (slowed down) by the additional rotational flow component, when the gaze rotates clock-wise about the Y-axis. This change in the flow velocity is falsely interpreted by the RFF-algorithm as an edge too near (far), as shown by the tilt in Fig. 4 D. If the retinal coordinates are mapped on a head-centric map by continuously checking the eye position, thus taking the rotational shift into account, the RFF-algorithm can operate on the resulting head-centric optical flow field. On a head-centric map, the performance of the RFF-algorithm is invariant under gaze sifts. (Fig. 4 E and F).

To quantify the performance of the RFF-algorithm on both the retinal flow field and the head-centric flow field, we defined a standard detection task:

the three-dimensional reconstruction of a centric viewed square plane. For fixed direction of gaze this corresponds to a situation where edges move with hyperbolically increasing velocity along the receptive fields of an individual radial line. The angles between the edge and the radius vary between $0°$ and $45°$. The average error in the detected three-dimensional position of the square plane was normalized to 1 for fixed direction of gaze (Fig. 5). If the gaze direction rotates step wise by a total angle between $1°$ to $4°$ about the Y-axis, the error increases when the RFF-algorithm works on retinal optical flow fields, as expected (see Fig 5). On head-centric optical flow fields the performance of the standard detection task is stable. Note that for fixed direction of, gaze one can introduce correction terms taking into account the actual static location of the receptive fields, which must lie on a square pixel grid. Therfore a receptive field location can not exactly obey Eq. 2. The error decreases by almost a factor of 2, with these correction terms, but these terms are not straightforwardly obtained for a head-centric map.

5 Discussion

Purely translational ego-motion induces an RFF in the retina, that contains reliable and rather easily accessible information about the structure of the viewed three-dimensional scene. In polar coordinates the RFF is only along the radial dimension and therfore this flow field is essentially one-dimensional in any retinotopic map for a specific curve-linear coordinate system. For example, in the primary visual cortex radial flow is mapped along parallel alined neurons [12]. And indeed some animals, e.g. the housefly [13] or birds [14], seem to reduce the optical flow to a single, translational component. However there will often be additional rotational components in the optical flow, foremost in form of small saccades or smooth pursuit eye movements. As soon as a rotational component is mixed with translation motion, the optical flow is two-dimensional in any coordinate system of a retinotopic map. In this case, deducing depth from optical flow is far more complicated. We showed that with a dynamical mapping strategy of visual space, the effect of eye-gaze movements on the optical flow can be eliminated. The resulting flow field on a head-centric map is congruent to the one induced by pure translational motion. In other words, dynamical mapping induces an optical flow invariant under eye-gaze shifts.

Can other rotational components than eye-movements, e.g. head rotation that could be mapped into a stable body-centric frame, be accounted for in a similar way? In all cases in which rotational components come together with translational in the ego-motion the resulting optical flow is two-dimensional. The condition to reduce the spatial dimension of the flow field by dynamically mapping is that the axis of rotation must include the view point of the perspective projection. This is in good approximation true for eye-gaze movements and with less accuracy also for head rotations, or body rotation about the central body axis. Generally, the larger the distance between rotation axis and view point is, the farer away must objects be, to be accurately detected by the RFF-algorithm.

If the heading direction changes slowly the center of rotation is far away from the view point but then the resulting trajectory can approximately split into linear parts in which the translation is again translational. Furthermore, if the rotation angle is too large, viewed object disappear on one side of the visual field and new objects come into existence on the other. Therefore the rotation must be small enough to have sufficient time to determine the distance of the object. Taken these facts together, eye-gaze movements are the most likely rotation that could be filtered from the optical flow by dynamical mapping.

References

1. Marr, D.: Vision. W.H. Freeman and Company, New York, 2000.
2. Nakayama, K., Loomis J.M.: Optical velocity patterns, velocity-sensitive neurons, and space perception: a hypothesis. Perception **3** (1974) 63–80.
3. . Longuet-Higgins H.C., Prazdny K.: The interpretation of a moving retinal image. Proc. R. Soc. Lond B Biol. Sci. **208** (1980) 385–397.
4. Gibson, J.J. The perception of the visual world. Houghton Mifflin, Boston, 1950.
5. Poggio G.F., Torre V., Koch C.: Computational vision and regularization theory. Nature, **317** (1985) 314–319.
6. Wörgötter, F., Cozzi, A., Gerdes V.: A parallel noise-robust algorithm to recover depth information from radial flow fields. Neural Comput. **11** (1999) 381–416.
7. Horn, B.K.P.: Robot Vision. The MIT Press, Boston, 1986.
8. Barron, J.L., Fleet, D.J., Beauchemin, S.S.: Performance of optical flow techniques. International Journal of Computer Vision, **12** (1994) 43–77.
9. Andersen R.A., Essick G.K., Siegel R.M.: Encoding of spatial location by posterior parietal neurons. Science. **230**(1985) 456–458.
10. Zipser D., Andersen R.A.: Related Articles A back-propagation programmed network that simulates response properties of a subset of posterior parietal neurons. Nature. **331** (1988) 679–684.
11. Marsh D.: Applied Geometry for Computer Graphics and CAD. Springer-Verlag, Berlin Heidelberg New York, 1999.
12. Schwartz, E.: Spatial mapping in the primate sensory projection: analytic structure and relevance to perception.: Biol. Cybern. **25** (1977) 181–194.
13. Wagner H.: Flight performance and visual control of flight of the free-flying housefly (*Musca domestica l.*). I. Organization of the flight motor. Phil. Trans. R. Soc. Lond., B**312** (186) 527–551.
14. Wallman J., Letelier J.-C.: Eye movments, head movments and gaze stabilization in birds. In Zeigler H.P., Bishop H.J. (Eds) Vision brain and behavior in birds. MIT Press. Cambridge, 1993.