Head movement estimation for wearable eye tracker

Constantin A. Rothkopf*
Center for Imaging Science
Rochester Institute of Tecnology

Jeff B. Pelz†
Center for Imaging Science
Rochester Institute of Tecnology

Abstract

In the study of eye movements in natural tasks, where subjects are able to freely move in their environment, it is desirable to capture a video of the surroundings of the subject not limited to a small field of view as obtained by the scene camera of an eye tracker. Moreover, recovering the head movements could give additional information about the type of eye movement that was carried out, the overall gaze change in world coordinates, and insight into high-order perceptual strategies. Algorithms for the classification of eye movements in such natural tasks could also benefit form the additional head movement data.

We propose to use an omnidirectional vision sensor consisting of a small CCD video camera and a hyperbolic mirror. The camera is mounted on an ASL eye tracker and records an image sequence at 60 Hz. Several algorithms for the extraction of rotational motion from this image sequence were implemented and compared in their performance against the measurements of a Fasttrack magnetic tracking system. Using data from the eye tracker together with the data obtained by the omnidirectional image sensor, a new algorithm for the classification of different types of eye movements based on a Hidden-Markov-Model was developed.

Keywords: head movement, eye movement classification, natural task

1 Introduction

Research in visual perception and attention using eye movements has moved from signal detection paradigms and the assessment of the mechanics and metrics of eye movements to the study of complex behavior in natural tasks. In such tasks, subjects are able to move their head and body and interact in a purposeful way with a changing environment. Also, their vision is not restricted to a small field of view so that peripheral vision can influence subsequent actions.

This type of research, has helped to understand how humans gather information from their environment, how they store and recover visual information, and how they use that information in planning and guiding actions [M. Land 2001] [M. Hayhoe 2002]. Apart from the impact of this research on the understanding of human visual perception and attention, the study of eye movements has contributed to developments in other fields that range from biologically inspired vision such as robot and animat vision systems to gaze dependent media systems and application in human computer interaction.

Accordingly, there has been a shift in interest from experiments being carried out in a laboratory situation with constrains on the ability of the subjects to move their head and body towards experimental settings in which natural tasks are studied. Such environments can range from every-day life locations to virtual reality environments. This development has profited from the development of small and lightweight portable video based eye trackers. Using such system under several new problems arise.

First, the subject will be able to use its entire field of view to acquire information from the scene. Usually, the scene camera in an eye tracker captures only a field of view of about 45°. This means, that the video recording of the eye tracker may not capture important features in the visual scene that contribute to the following actions. An offline evaluation of the recording showing only the restricted field of view of the scene camera may lead to misinterpretations about the involved attentional processes.

Secondly, depending on the task, it may be desirable to capture a video from the entire surroundings of a subject. In a task where the subject is allowed to freely move around in an environment that is constantly changing, a conventional eye tracker will not capture a representation of the surrounding space. It would be advantageous to monitor how the environment changed in time in order to relate actions of the subject to these changes. A question that could be addressed is under which circumstances a subject decided to carry out a movement and focused attention on a region of the scene that had previously been out of sight.

Thirdly, it is also of interest to measure the head and body movements that are carried out by a subject during an extended task. It may be of interest to analyze the coordination of head and eye movements or to infer in which frame of reference actions may be linked. Typically such movements can be measured with high accuracy in a laboratory using a tracking system as e.g. a Fastrack magnetic tracking system. For natural tasks that are set outside the laboratory, such a device can not be used.

Fourthly, given that the eye and head movements could be measured, the classification of eye movement types could be improved. In general, the classification is an important task in the study of human visual perception and attention, because the different types of eye movements can be related to different cognitive processes. Usually, the analysis of eye movements has focused on the extraction of fixations and fixation regions. For a fixed head position that can be achieved using a bite bar, the observed eye movements will indeed be either fixations or saccades. But in a natural task with no restrictions on the subject's head and body movements also smooth pursuits and Vestibular Ocular Reflexes (VOR)will occur. Both are eye movements that are necessary in order to stabilize the image on the fovea. While smooth pursuits follow a smoothly moving target while the head is relatively stable, VORs keep a target centered during rotational head movement. Also, with large field motions, subjects will carry out optokinetic nystagmus (OKN). These eye movements can be described as repeated VORs that occur when a subjects makes large, rotational movements. OKN is characterized by a slow phase, which is essentially a smooth pursuit of the tar-

*e-mail: cxr6017@cis.rit.edu
†e-mail:pelz@cis.rit.edu

get, interposed with the quick phase, which is a saccade made to re-center gaze on the target[Carpenter 1991]. The typical example is watching a passing train at a railroad crossing.

Several different algorithmic approaches for the identification of fixations and saccades that were developed during the last 30 years are described in the literature. For a review see [Salvucci 1999] and [Duchowski 2002]. Most of these algorithms, especially those based on spatial clustering, assume that the head is at rest. Other algorithms that operate with a velocity or an acceleration criterion are aimed at detecting saccades. In cases where the subject carries out a smooth pursuit, a velocity based algorithm will most likely not detect a saccade and therefore assume that a fixation was continued. But in this cases the eyes moved in space in order to follow a target and may overlap with the position of other parts of the scene. Therefore, it would be desirable for an algorithm to be able to handle and identify these four types of eye movements that occur under natural tasks.

We propose to use an omnidirectional image sensor that is mounted on top of a regular eye tracker. The video camera is pointing upwards into the mirror that captures a circular image in which the subject is at the center. The parameters of the mirror and the lens influence the field of view i. e. the angle of elevation up to which the surroundings of the subject are captured. This device is able to monitor the surroundings of the subject, and the captured video can be analyzed offline in order to study the involved attentional processes. Moreover, head and body movements can be estimated from the image sequence of the omnidirectional image sensor, using known ego motion estimation algorithms.

Our solution is similar to the one chosen by [Rungsarityotin and Starner 2000], who used an omnidirectional image sensor on a wearable computing platform for the purpose of localization. In contrast, their system was first trained on a specific location and a Bayesian estimation algorithm was used to find the most likely position within the space given the acquired prior knowledge and an image captured insitu.

2 Image formation with omnidirectional image sensor

In the following paragraphs we provide a brief review of the image formation for an omnidirectional vision sensor. We are mainly following the treatment of [T. Svoboda 1998b], [S. Baker 1999], and [A. Makadia 2003].

[S. Baker 1999] have shown, that an omnidirectional image sensor with a single viewpoint can be constructed using a hyperbolic mirror and a perspective camera and with a parabolic mirror and an orthographic camera. The single viewpoint implies, that the irradiance measured in the image plane corresponds to a unique direction of a light ray passing through the viewpoint. Here, the solution using a hyperbolic mirror together with a perspective camera has been used.

The fact that such an image sensor possesses a single viewpoint has been used to show [C. Geyer 2000], that the resulting projection is equivalent with a projection on the sphere followed by a projection to a plane. Accordingly, an omnidirectional image can be remapped onto a sphere that can be thought of as an environment map.

The origin of the coordinate system is chosen to coincide with the focal point F of the hyperboloid. The focal point of the camera therefore coincides with second focal point of the hyperboloid. The hyperboloid can be expressed as:

$$\frac{(z+e^2)}{a^2} - \frac{x^2+y^2}{b^2} = 1 \tag{1}$$

where a and b are the parameters of the hyperboloid and $e = \sqrt{a^2+b^2}$ is the eccentricity. The camera is modeled as a pinhole camera. A world point with coordinates $\mathbf{X} = [x,y,z]^T$ is imaged by the camera according to:

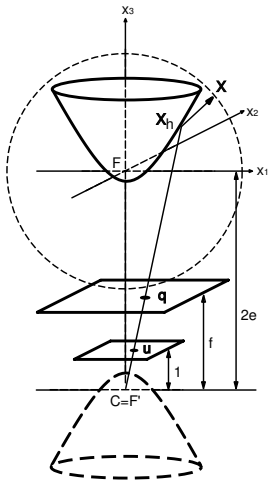

Figure 1: Coordinate system of omnidirectional sensor.

$$\mathbf{q} = \mathbf{K}\mathbf{u} = \begin{pmatrix} f\cdot k_u & 0 & q_{u0} \\ 0 & f\cdot k_v & q_{v0} \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x/z \\ y/z \\ 1 \end{pmatrix} \tag{2}$$

where $\mathbf{q}$ is the vector representing pixel coordinates, $\mathbf{u}$ represents normalized image coordinates, f is the focal length of the lens of the camera, k_u and k_v are the scale factors in x and y direction, and q_{u0} and q_{v0} are the center coordinates in pixels. In order to relate points in space to the respective pixels in the captured image, the intersection of a vector $\mathbf{X} = [x,y,z]^T$ from a world point $\mathbf{X}$ through the origin with the hyperboloidal surface $\mathbf{X}_h$ can be calculated as:

$$\mathbf{X}_h = \left(\frac{b^2(-ez + a \parallel \mathbf{X} \parallel)}{-(ax)^2 - (ay)^2 + (bz)^2} \right) \mathbf{X} \tag{3}$$

The point $\mathbf{X}_h$, which lays on the surface of the hyperboloid, is then projected according to equation (2) so that the entire transformation can be expressed as:

$$\mathbf{q} = \frac{1}{(0,0,1)^T(\mathbf{X}_h - \mathbf{t}_c)} \mathbf{K}(\mathbf{X}_h - \mathbf{t}_c) \tag{4}$$

where $\mathbf{X}_h$ is given by equation (3), $\mathbf{K}$ is the camera calibration matrix from equation(2), and the translation vector $\mathbf{t}_c$ is determined by the system setup as $\mathbf{t}_c = [0,0,-2e]^T$.

In reconstructing a spherical environment map, the surface of the unit sphere can be sampled in θ, ϕ space and the corresponding pixel in the captured omnidirectional image can be calculated according to equation (4). The advantage in using the mapping from the sphere onto the captured image is that the calculations that have to be carried out are determined by the required angular resolution of the environment map. Additionally, linear interpolation is used in order to obtain a smooth image.

The system consisting of the camera, the lens, and the mirror has to be calibrated so that the effective focal point of the camera is located in the focal point of the F' of the mirror. This can be done by calculating the size in pixels of the known diameter of the mirror's top rim as imaged by the camera. Using equations (1) and (2) one obtains:

$$r_{pix} = \frac{r_{toprim}}{z_{top} + 2e} \tag{5}$$

3 Estimation of Rotational Movement

Originally we had implemented an algorithm to estimate the rotational and translational egomotion of the subject from the image sequence captured by the scene camera using methods described e.g. in [Irani et al. 1994]. Due to the well known problem that the flow field for rotations and translations with a point of expansion that is located outside the field of view can be difficult to distinguish, the resulting accuracy was not satisfying. In order to test the usefulness of the device we aimed at the recovery of the rotational motion that could be used for the classification of the types of eye movements. The used motion model assumes, that transformation between two captured images is due to a rotation of the sensor and that the translation can be neglected.

The angular inter frame head movements have to be estimated from the image sequence of the omnidirectional camera. In the past, several algorithms have been developed to obtain such estimates for catadioptric camera systems using different approaches. While [T. Svoboda 1998b] first identified point correspondences within two images, [J. Gluckman 1998] estimated the optical flow between image pairs. The former method requires identifying and tracking at least eight corresponding points within subsequent images in order to calculate an estimate of the fundamental matrix, which relates the image points in two separate views of a scene captured by the same camera from different viewpoints. The second method estimates the optical flow between subsequent images and projects the image velocity vectors onto a sphere. Several methods for estimating ego-motion from spherical flow fields can then be used as described by [J. Gluckman 1998].

While [T. Svoboda 1998b] used images of size 768 x 512 and [J. Gluckman 1998] used images of size 512 x 480 the images captured by the miniature camera are limited by the size of the NTSC image with 360 lines of resolution. Due to this limitation, initial attempts to use the above mentioned optical flow method did not yield satisfying results.

Independently from [A. Makadia 2003] the use of spherical harmonics decomposition of the omnidirectional images was pursued. The transformation of spherical harmonics coefficients under rotations are well described e.g. in the fields of astronomy and geodesy. Here we follow the exposition of [Sneeuw 1992], while the notation was adapted from [G.S. Chirikjian 2001]

The set of spherical harmonics function constitutes an orthonormal set of basis functions on the sphere. They are obtained as the eigenfunctions of the spherical Laplacian:

$$Y_l^m(\theta,\phi) = \sqrt{\frac{(2l+1)(l-m)!}{4\pi(l+m)!}} P_l^m(cos\theta) e^{im\phi} \tag{6}$$

where the P_l^m are associated Legendre functions, θ is the polar, and ϕ is the azimutal angle. A function on the unit sphere can therefore be expanded using this set of basis functions. Due to the fact that the obtained images are discrete representation, the discrete spherical harmonic decomposition has to be carried out. For a bandlimited function which is sampled in θ,ϕ space with N samples in each dimension, this can be expressed as:

$$f(l,m) = \frac{2\sqrt{\pi}}{N} \sum_{j=0}^{N-1} \sum_{i=0}^{N-1} a_j \cdot f(\theta_j,\phi_i) \cdot \overline{Y_l^m(\theta_j,\phi_i)} \tag{7}$$

The a_j reflect the infinitesimal element of integration on the sphere in θ,ϕ space, which depend on the sampling density and are given by:

$$a_j = \frac{2}{N} sin(\frac{\pi}{N} j) \sum_{k=0}^{\frac{N}{2}} \frac{1}{2k+1} sin\big((2k+1)\frac{\pi}{N} j\big) \tag{8}$$

A rotation of a function on the sphere can be described using Euler Angles. [Sneeuw 1992] parameterized the rotation using ZYZ angles, i.e. the rotation is expressed as a concatenation of three separate rotations: first a rotation of magnitude α about the original z-axis, then a rotation of magnitude β around the new y-axis, and finally a rotation about the final z-axis. Under such a rotation the spherical harmonic function of order l transforms according to:

$$\overline{Y_l^m(\theta',\phi')} = \sum_{k=-l}^{l} D_{mk}^l(\alpha,\beta,\gamma) \overline{Y_l^k(\theta,\phi)} \tag{9}$$

where prime denotes the transformed coordinates. While [Makadia & Daniilidis 2003] expressed the D_{mk}^l using generalized associated Legendre polynomials, [Sneeuw 1992] used the following expressions:

$$D_{mk}^l(\alpha,\beta,\gamma) = e^{im\alpha} \overline{d}_{mk}^l(\beta) e^{ik\gamma} \tag{10}$$

with

$$d_{mk}^l(\beta) = \left[\frac{(l+k)!(l-k)!}{(l+m)!(l-m)!}\right]^{\frac{1}{2}} \sum_{t=t_1}^{t_2} \binom{l+m}{t} \cdot \binom{l-m}{l-k-t} (-1)^t cos\Big(\frac{\beta}{2}\Big)^{2l-a} sin\Big(\frac{\beta}{2}\Big)^a \tag{11}$$

where $a = k-m+2t$, $t_1 = max(0, m-k)$, $t_2 = min(l-k, l+m)$. Details on how these representations are connected can be found in [Chirikjian & Kyatkin 2001]. The coefficients of a function decomposed into spherical harmonics after a rotation can then be expressed as:

$$f'(l,m) = \sum_{k=-l}^{l} f(l,k) D_{km}^l \tag{12}$$

Assuming that two subsequent images from the recorded sequence are related by a transformation that can be described as a rotation, the task is then to minimize an error function that measures the difference between the application of a rotation to the spherical harmonics coefficients from an image in the sequence and the calculated coefficients from the next captured image in the sequence. In this case, the squared Euclidean distance of the error over all coefficients up to order l was used so that the minimization problem can be expressed as:

$$\sum_{n=1}^{l} \sum_{m=0}^{l} \Big[\Big(\sum_{k=-l}^{l} D_{km}^n(\alpha,\beta,\gamma) \cdot f_{t_1}(n,k)\Big) - f_{t_2}(n,k)\Big]^2 = 0 \tag{13}$$

This function was minimized using Levenberg-Marquardt nonlinear iterative minimization. For completeness it should me mentioned, that [Makadia & Daniilidis 2003] reparameterized the rotation as a concatenation of two rotations in order to simplify the calculation of the D_{mk}^l coefficients in equation (9).

The restriction to the method described above is that the image captured from the hyperbolic mirror is not a true omnidirectional image. Accordingly, [A. Makadia 2003] reported, that for small rotational angles, the accuracy of the algorithm decreases. For rotations about the original z-axis the algorithm will perform best. It was expected, that this was not a significant disadvantage for the classification of the eye movements, because the fastest head movements are those around the perpendicular axis of the body. Given a sampling rate of 30 frames per second and the fact that head movements are reported to reach peak velocities of approximately 300

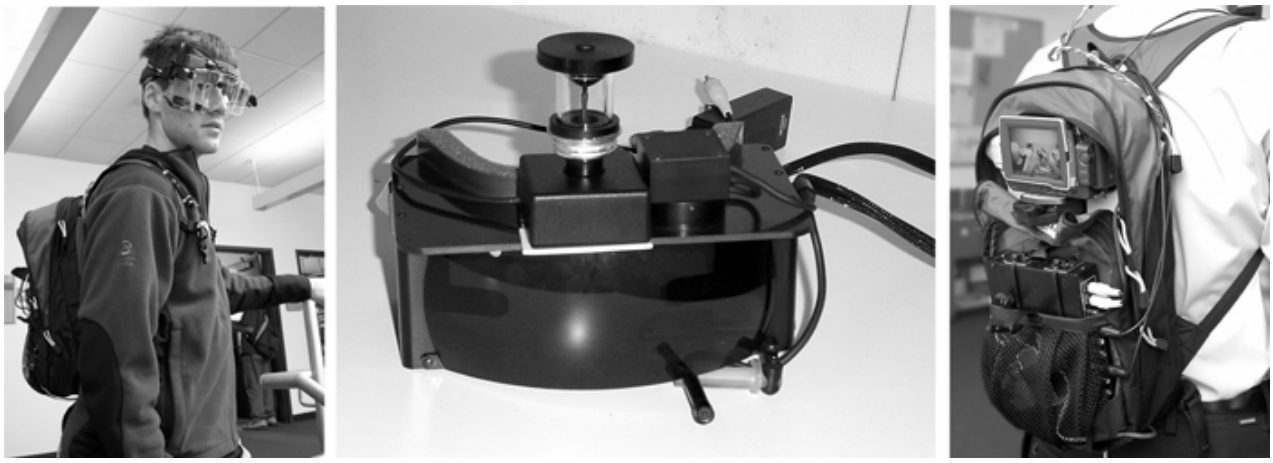

Figure 2: The wearable eye tracker: headgear and backpack, headgear with omnidirectional vision sensor, and detail of backpack.

degrees per second, the corresponding inter frame rotational angles are expected to be smaller than 10 degrees. In addition, it is possible to apply the algorithm on image pairs that are more then one time sample apart.

4 Eye Movement Classification with Hidden Markov Model

A Markov Model is a stochastic model, which assumes that a system may occupy one of a finite number of states, that the transitions between states are probabilistic, and that the transition probabilities, which are constant over time, depend only on the current state of the system. A Hidden Markov Model further assumes that the system emits an observable variable, which is also probabilistic, depending on the state the system is in. In the continuous case, the state of the system determines the parameters of this continuous random variable. Hidden Markov Models have been used for fixation identification by [Salvucci 1999]. A two state Hidden Markov Model representing a 'saccade' state and a 'fixation' state was used. The velocity distributions corresponding to the two states were modeled with a single Gaussian with a higher mean velocity for the saccade state and a lower velocity for the fixation state in accordance with physiological results (e.g.[Carpenter 1991]). The parameters for the transition probabilities and the parameters for the Gaussian distributions were estimated from the collected data using the Baum-Welsh algorithm [Rabiner 1989].

We have extended this model in order to also represent smooth pursuits and VORs. This is accomplished by integrating the head movement velocities as a second observation variable and introducing two additional states. Accordingly, the probability distributions for the observation variables are bivariate. The four states S_j represent states for fixations, saccades, smooth pursuits, and VORs. A single observation $\mathbf{O}_t$ at time t consists of a two component vector. The first component represents the angular eye velocity and the second component represents the angular head velocity. The observation sequence is denoted as $O = \{\mathbf{O}_t, \mathbf{O}_{t+1}, \mathbf{O}_{t+2}, ...\}$. The assumption of the HMM is, that each observation $\mathbf{O}_t$ at time t corresponds to one of the states S_i at time t from the state sequence denoted by $Q = \{q_t, q_{t+1}, q_{t+2}, ...\}$ and that the next state that the system will be in depends only on the current state of the system. Accordingly, the transition probabilities a_{ij} between the states S_i and S_j are expressed as:

$$a_{ij} = P(q_{t+1} = S_j \mid q_t = S_i), 1 \leq i, j \leq 4 \tag{14}$$

The distributions of eye and head velocities for the eye movements in question have been studied empirically and can be found in [Carpenter 1991]. These distributions are approximated with single bivariate Gaussian distributions for the use in the HMM, so that the probability density function can be expressed as:

$$p_j(\mathbf{O}_t) = \frac{1}{\sqrt{2\pi}|\Sigma|} e^{-\frac{1}{2}(\mathbf{O}_t - \mu_j)\Sigma^{-1}(\mathbf{O}_t - \mu_j)} \tag{15}$$

where j denotes the state that the system is in and Ot denotes the observation vector at time t.

The goal of using a HMM in this specific application is first to use a sequence of observations O in order to reestimate the transition probabilities a_{ij} and the parameters μ, Σ of the distributions of the observed variables. The second step is then to use these parameters to decode the observation sequence O, such that to each observation $\mathbf{O}_t$, a corresponding eye movement type is assigned. For details regarding the reestimation and decoding calculations see [Rabiner 1989].

5 Construction and Calibration of the Device

The RIT-Wearable-Eye-Tracker is based on a pair of racquetball goggles and a backpack. The custom headgear uses an infrared illuminator, a miniature, IR-sensitive CMOS camera, and beam splitter to capture an image of the eye. The eye is illuminated by the IR source along an optical path directed by a first surface mirror and hot mirror. The bright-pupil image is reflected back to the CMOS camera along the same path. A second miniature camera is located just above the right eye and is used to capture the scene from the subject's perspective. A battery is mounted on the left side of the goggles to power the cameras, and also to balance the weight of the optical components. The headgear is connected to a customized Applied Science Laboratory (ASL) Model 501 controller unit contained in the backpack, which carries out the data processing for the gaze position calculation. The camera used for the omnidirectional vision sensor is a PCX-169 NTSC video camera with a resolution of 360 lines. The lens used with the camera has a focal length of 6mm. The hyperbolic mirror is a 'Miniature' mirror from Accowle, Ltd. with parameters a=8.37mm and b=12.25mm and a top rim radius of rtop=26mm. The mirror is connected to the camera and both are mounted on top of the eye tracker.

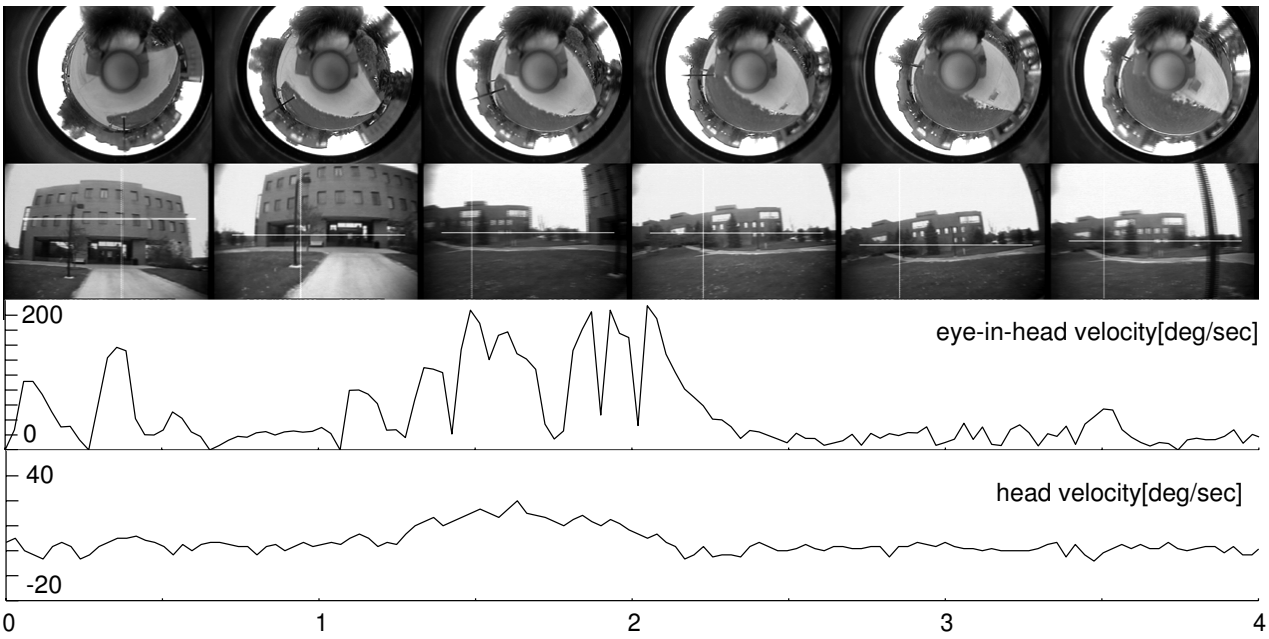

Figure 3: From top to bottom: Image sequences from omnidirectional camera, scene camera, eye in head velocity, and estimated head velocity. The timescale represents 4 seconds.

6 Results

The system was first used in a laboratory setting where ground truth was obtained using head movement measurements from a Polhemus Fastrak magnetic tracking system, in order to assess the performance of the algorithm for the rotation estimation. A subject first carried out sequences of head movements around all three axes while standing still. In a second sequence, the subject, still not moving the body, carried out sequences of controlled eye movements. These comprised fixations, saccades with varying amplitudes, smooth pursuits of targets moving at different velocities, and sequences of VORs. In addition, similar sequences were repeated while the subject walked around the room.

The algorithm performs well for azimutal rotations. This is because such motions can be regarded as true omnidirectional rotations although the field of view is not totally omnidirectional. On the other hand, the performance of the algorithm declined rapidly for large motions in roll. Also, very fast head and body motions can blur the recorded video image so that the algorithm fails completely. Under such cases an inertial sensor would be the more adequate choice for a pure ego motion estimation.

The resulting recordings from the eye tracker suffer from a significant noise level compared to experimental conditions in which the subject does not move neither the body nor the head. Also, visual inspection of the recordings shows that much more complex patterns of eye and head movements have to be considered. One example is given in figure 4. where a series of OKNs can be observed.

At first, several of the algorithms described by [Salvucci 1999] were implemented and compared in their performance. Algorithms based on a velocity threshold resulted in a significant number of false alarms and tended to classify VORs as saccades. Algorithms based on an acceleration threshold were superior to those using a velocity criterion but also suffered from a large number of false positives. In addition, these algorithms were sensitive to a number of internal parameters, that had to be adjusted for each individual and each trial in order to perform satisfactorily.

We obtained better results by using a velocity based algorithm that was altered in such a way that the threshold was adjusted depending on the average velocity in a running window. The average angular eye in head velocity over a time window of 0.5 seconds was calculated for each data point in a single trail. The minimum and maximum values for this average were then used to fit a line between a minimal and maximal threshold value, which would then be used for detecting the saccade. This algorithm empirically performed better than other velocity based criteria used. Two parameters have to be set and empirically these were determined only by the setup in use. Our choice was 20 deg/sec for the lower and 100 deg/sec for the upper bound. It should be mentioned that these values were not adjusted on a trial by trail basis or for different subjects.

The four state HMM was trained on the recording of the subject carrying out sequences of controlled eye movements. Before training, the parameters for the bivaraite Gaussians were adjusted according to prior knowledge and exploratory data analysis where subjects carried out sequences of different eye movements. The initialization of the parameters for the reestimation process is crucial for the 4 state HMM. After training, the estimated parameters were used to classify the recordings. The preliminary results show that the algorithms is well able to disambiguate VORs and Saccades in many of the considered cases. Smooth pursuits were sometimes missed especially when they had small velocities. The HMM classified these slow smooth pursuits consistently as fixations. These results are preliminary in that extended scoring by trained human experts should be carried out in order to assess the performance of the algorithm.

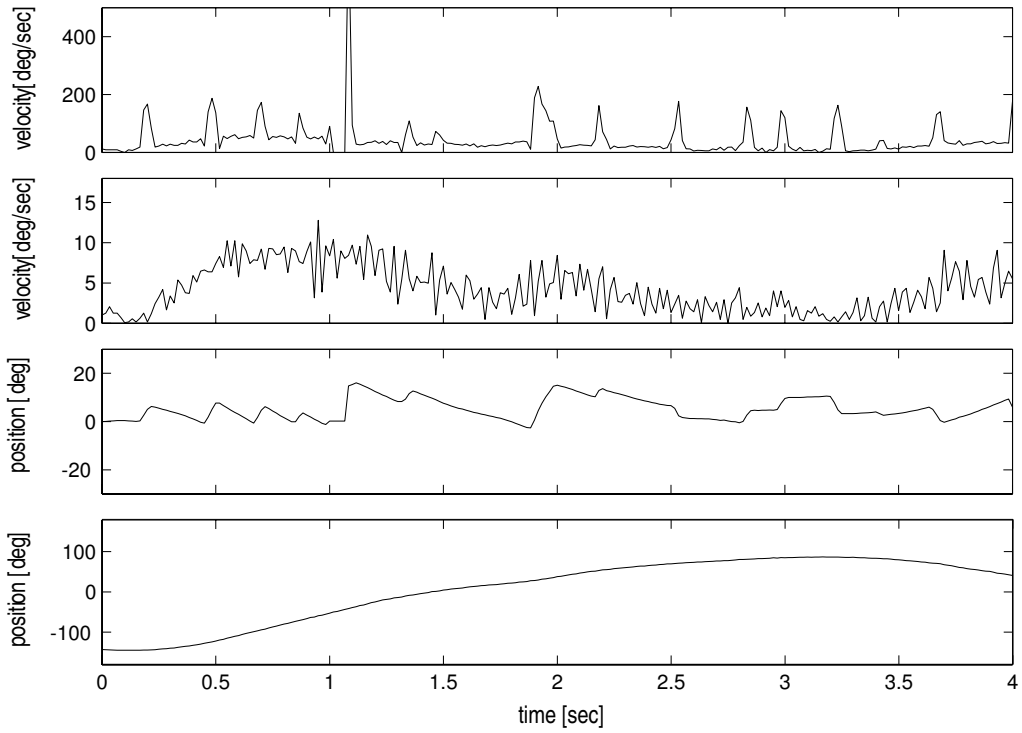

Figure 4: Example of Optokinetic Reflex: angular displacement and velocities for head and eyes.

7 Discussion

7.1 Omnidirectional Image Sensor and the Wearable-Eye-Tracker

The developed device is able to capture a scene representation that includes a large field of view which is always centered at the location of the subject. The parameters for the mirror and the lens can be adjusted in order to obtain the desired field of view depending on the application. The captured omnidirectional image can be unwarped to a spherical or a cylindrical image to give a representation of the subject's surroundings. It is therefore possible to monitor the environment of the subject during the experiment. In addition, it is possible to analyze the captured image sequence offline after the experiment has been ended in order to analyze and extract additional information about the dynamics of the surroundings and the actions of the subject within its changing environment.

7.2 Estimation of Rotational Head and Body Movement

Currently, the developed algorithm can not distinguish between body and head movements. For the application that was pursued in this project, namely the classification of eye movements using eye and head movements, the obtained precision and accuracy would be sufficient. Further testing will be required in order to assess the accuracy of the classification algorithm. A further evaluation of the estimation algorithm should analyze the performance of the estimation of rotational movement when more complex motions as translation are present.

7.3 Eye Movement Classification Including Head Motion

Head movements can be used as an additional source of information in order to disambiguate the different types of eye movements that can occuor when the subjects are able to move their head and body. The developed Hidden Markov Model that models fixations, saccades, smooth pursuits, and VORs with bivariate Gaussian distributions of the head and eye velocities was assessed using only recordings from sequences of controlled saccades, fixations, smooth pursuits, and VORs. The preliminary results suggest that the additional information about the rotational head movement can indeed be used to classify these eye movements. These results are preliminary in the sense that extended comparisons between the classification by the algorithm and manual scoring of video recordings from the eye and scene camera by an expert will have to be carried out.

7.4 Further Work

It is desirable to increase the resolution of the used camera and use a system that has a maximal field of view. It would therefore be advantageous to try an omnidirectional camera using a mirror giving a larger field of view than the one used in this project. A further development of the algorithm should incorporate the estimation of translational motion. This could be accomplished either using the image sequence obtained by the omnidirectional camera or by integrating the image sequence from the scene camera of the eye tracker. Finally, the classification algorithm should undergo extensive testing, which can be accomplished through comparison of classification results by expert coders.

8 Conclusion

We have shown that new techniques need to be developed when performing eye tracking in extended natural tasks outside the laboratory. The proposed omnidirectional vision sensor is a device that aids in monitoring the surroundings of the subject during such tasks. Offline evaluation of the attentional influences in the surroundings of the subject and also real time monitoring are possible possible application. Also, we have shown that the additional knowledge of head movements can fruitfully be used in order to classify the occurring eye movements. Further validation of the developed HMM algorithm should assess the degree to which smooth pursuits and VORs can be distinguished from fixations and saccades.

This work was funded in part by grants from the Naval Research Laboratories, the New York State Office of Science, Technology, and Academic Research, and the Eastman Kodak Company.

References

A. MAKADIA, K. D. 2003. Direct 3d-rotation estimation from spherical images via a generalized shift theorem. In *2003 Conference on Computer Vision and Pattern Recognition (CVPR '03)*, 15–22.

C. GEYER, K. D. 2000. Equivalence of catadioptric projections and mappings of the sphere. In *IEEE Workshop on Omnidirectional Vision*.

CARPENTER, R. H. S. 1991. *Eye Movements*. CRC Press.

DUCHOWSKI, A. T. 2002. *Eye Tracking Methodology*. Springer Verlag.

G.S. CHIRIKJIAN, A. K. 2001. *Engineering applications of noncommutative harmonic analysis*. CRC Press.

IRANI, M., ROUSSO, B., AND PELEG, S. 1994. Recovery of Ego-Motion Using Image Stabilization. In *ICVPR*, 454–460.

J. GLUCKMAN, S. K. N. 1998. Ego-motion and omnidirectional cameras. In *IEEE International Conference on Computer Vision (ICCV'98)*, 999–1005.

M. HAYHOE, D. H. BALLARD, E. A. 2002. Vision in natural and virtual environments. In *Eye Tracking Research and Application(ETRA 2002)*.

M. LAND, M. H. 2001. In what ways do eye movements contribute to everyday activities? *Vision Research, special issue on Eye Movements and vision in the Natural World 41*.

RABINER, L. 1989. A tutorial on hidden markov models and selected applications in speech recognition. *Proceedings of the IEEE 77*, 257–286.

RUNGSARITYOTIN, W., AND STARNER, T. 2000. Finding location using omnidirectional video on a wearable computing platform. In *ISWC*, 61–68.

S. BAKER, S. K. N. 1999. A theory of single-viewpoint catadioptric image formation. In *International Journal of Computer Vision(IJCV 1999)*.

SALVUCCI, D. D. 1999. *Mapping Eye Movements to Cognitive Processes*. PhD thesis, Carnegie Mellon University.

SNEEUW, N. J. 1992. Representation coefficients and their use in satellite geodesy. *Manuscripta Geodaetica*.

T. SVOBODA, T. PAJDLA, V. H. 1998. Epipolar geometry for panoramic cameras. *Lecture Notes in Computer Science 1406*, 218–??

T. SVOBODA, T. PAJDLA, V. H., 1998. Motion estimation using central panoramic cameras. In IEEE International Conference on Intelligent Vehicles, October 1998. to appear.