Journal of Neuroscience Methods 176 (2009) 157–165

Contents lists available at ScienceDirect

Journal of Neuroscience Methods

journal homepage: www.elsevier.com/locate/jneumeth

# Compensating for camera translation in video eye-movement recordings by tracking a representative landmark selected automatically by a genetic algorithm

Faisal Karmali[a,b,*], Mark Shelhamer[b,c]

[a] Department of Otolaryngology, Massachusetts Eye and Ear Infirmary, Boston, MA, United States
[b] Department of Biomedical Engineering, The Johns Hopkins University School of Medicine, Baltimore, MD, United States
[c] Department of Otolaryngology–Head & Neck Surgery, The Johns Hopkins University School of Medicine, Baltimore, MD, United States

ARTICLE INFO

*Article history:*
Received 21 February 2007
Received in revised form 15 August 2008
Accepted 5 September 2008

*Keywords:*
Eye
Video
VOG
Video-oculography
Cross-correlation
Genetic algorithm

ABSTRACT

It is common in oculomotor and vestibular research to use video or still cameras to acquire data on eye movements. Unfortunately, such data are often contaminated by unwanted motion of the face relative to the camera, especially during experiments in dynamic motion environments. We develop a method for estimating the motion of a camera relative to a highly deformable surface, specifically the movement of a camera relative to the face and eyes. A small rectangular region of interest (ROI) on the face is automatically selected and tracked throughout a set of video frames as a measure of vertical camera translation. The specific goal is to present a process based on a genetic algorithm that selects a suitable ROI for tracking: one whose translation within the camera image accurately matches the actual relative motion of the camera. We find that *co-correlation*, a statistic describing the time series of a large group of ROIs, predicts the accuracy of the ROIs, and can be used to select the best ROI from a group. After the genetic algorithm finds the best ROIs from a group, it uses *recombination* to form a new generation of ROIs that inherit properties of the ROIs from the previous generation. We show that the algorithm can select an ROI that will estimate camera translation and determine the direction that the eye is looking with an average accuracy of 0.75°, even with camera translations of 2.5 mm at a viewing distance of 120 mm, which would cause an error of 11° without correction.



## 1. Introduction

### 1.1. Motivation

Eye-movement studies are important for understanding specific neural control systems as well as general properties of sensorimotor processing in the brain. They provide an almost direct measure of vestibular (balance) function and provide insight into many forms of neural processing, such as sensorimotor integration, adaptation, and prediction. Video eye monitoring, using high-speed cameras and appropriate image processing, has the potential to provide non-invasive measurement of eye position with high spatial accuracy and temporal resolution (Clarke et al., 2002; Abadi and Gowen, 2004; MacDougall and Moore, 2005).

We recently conducted a study of human eye movements with different acceleration and gravity levels (*g* levels) aboard a NASA aircraft, which flies parabolic trajectories to provide alternating levels of reduced (∼0*g*) and enhanced (∼1.8*g*) *g* levels (Lackner and DiZio, 2000; Shelhamer et al., 2002; Karmali and Shelhamer, 2008). Subjects in these flight experiments had a vertical misalignment of the eyes that was dependent on *g* level (Karmali et al., 2006), which was confirmed with binocular eye-movement recordings using a video system (Fig. 1) consisting of two infrared cameras rigidly attached to a head-mounted frame (headset) (Clarke et al., 2002). Eye position as a function of time was determined by finding the pupil position in each video image. These recordings show that the eyes diverged vertically up to 3° as a function of *g* level.

One possible discrepancy in this result is that the cameras might have moved relative to the head, which could have created an artifactual difference in vertical eye position. Of particular concern in this study was roll rotation of the headset relative to the head, which would result in vertical translation of the two cameras by different amounts, causing an artifact of vertical eye misalignment, and which would be difficult to distinguish from the true ocular vertical misalignment that was being studied. Although we are confident that the result is not an artifact because the recorded eye misalignment was accompanied by double vision, we felt it necessary to

* Corresponding author at: 243 Charles St., JVPL 421, Boston, MA 02114, United States. Tel.: +1 410 218 7614; fax: +1 410 614 1746.
*E-mail address:* faisal_karmali@yahoo.com (F. Karmali).

0165-0270/$ – see front matter © 2008 Elsevier B.V. All rights reserved.
doi:10.1016/j.jneumeth.2008.09.010

**Fig. 1.** Binocular video eye tracker (Chronos Vision GmbH). Two side-mounted infrared cameras record eye movements via infrared-reflective mirrors mounted in front of the eyes. A bitebar is used the reduce movement of the headset relative to the head, although movement still occurs.

develop a quantitative approach to confirm our qualitative assessment. We note that a typical translation motion of 2.5 mm with a camera distance of 120 mm and a typical eye diameter would cause an artifact of change in eye gaze of 11°.

The solution to this problem has wide applicability. Many methods have been attempted in order to mechanically secure a camera relative to the head, such as goggles with tight elastic bands (MacDougall and Moore, 2005) and MRI thermoplastic masks (Clarke et al., 2002). However, short of drilling holes in the skull, none of these methods eliminate camera movement adequately, especially in environments with large accelerations where inertia causes larger relative displacements of the cameras and the head. A current interest in vestibulo-ocular research is the use of natural, rapid, impulsive head rotations and translations in which the most interesting component of eye movement occurs when head acceleration is highest (Walker and Zee, 2005), which is also the time when the largest movement of the cameras relative to the head occurs. Since there is no simple and feasible method to eliminate undesired camera movement, we aimed here to develop an algorithm to measure any such movement, so that it could be compensated for in the analysis of the related eye-movement recordings.

### 1.2. Previous approaches

Several previous techniques have been developed to address the problem of measuring camera motion relative to the eye using image processing of video eye recordings. These algorithms are applied to each video frame in the recording (the sequential set of images that make up a video recording), and many algorithms compare each video frame to a reference video frame. The general approach is to find and track a landmark in each video frame that moves when the camera moves, but not during other events, such as eye or eyelid movement.

In one implementation of this technique, the landmark that is tracked is a “glint,” the reflection on the cornea of infrared LEDs that are mechanically coupled to the camera (Young and Sheena, 1975; Hutchinson et al., 1989; Zhu and Ji, 2005; Kolakowski and Pelz, 2006; Guestrin and Eizenman, 2006). In theory, if one assumes a spherical eyeball, these reflections translate with the camera and do not move with eyeball rotation. However, in practice the cornea is not spherical, and precise calibrations are required for this technique to be successful. One system requiring subject-specific calibration is able to estimate gaze angles with an accuracy of 0.9° (Guestrin and Eizenman, 2006), while another system requiring less calibration is able to estimate vertical gaze angles with an accuracy of 1.7° (Zhu and Ji, 2005). We preferred to develop a technique that does not rely on critical calibration procedures.

A second approach is to apply markers to the skin around the eye and then assume that any translation of these landmarks is due only to movement of the cameras. There are two problems with this method. First, it requires marking of the skin before the experiment is performed, and in our specific case we have a library of data which was acquired without markers, and for which the experiments cannot be repeated. Second, there remains the question of how much the skin may move due to blinks or fatigue, which is discussed below.

A third technique is to detect and track movement of the medial canthus, where the upper and lower lids converge next to the nose. Unfortunately this potential landmark is not visible in many recordings due to a small camera field of view.

Another landmark that can be tracked is the upper eyelid. We previously developed an automatic image-processing technique to measure the positions of the eyelids, which were then used to make corrections to the pupil location so that the *difference* in eye positions could be determined with less than 0.8° of error (Karmali and Shelhamer, 2005). However, eyelid position is affected by blinks, and thus is valid for our purposes only when the *difference* in vertical positions between the two eyes is required, because the eyelids can be assumed to move by the same amount, and thus any movement of the lids will cancel when the difference in eye position is computed. This method also required manual modification of intensity thresholds because of variability in skin reflectance between subjects. In another technique we developed, the *en-bloc* movement of each video frame was estimated using two-dimensional cross-correlations (Russ, 2002), but this technique had a high error, which is expected because eye and eyelid movement interfere with the cross-correlation.

Optical flow techniques (Barron et al., 1994) were briefly considered as a means to estimate camera translation. These techniques estimate the motion of objects within a series of images by estimating the optical flow, or motion velocity, of each pixel in the image, using both spatial and temporal derivatives of intensity. The equations to estimate motion are underdetermined over small regions of the image because of the “aperture problem”: there are regions of constant intensity whose derivatives change only at the edges. To solve the aperture problem, additional constraints are added to the analysis. These constraints are based on assumptions that are violated in our application, which could reduce spatial accuracy. The Lucas–Kanade algorithm (Lucas and Kanade, 1981) assumes that the amount of movement between video frames is small. When comparing each video frame to a reference, some frames would have large movements, and thus the assumption of the Lucas–Kanade algorithm would be violated. This problem could be overcome by comparing the motion in each frame to the previous frame and integrating the velocity over time, but this would cause an accumulation of error. The Horn–Schunck algorithm (Horn and Schunck, 1981) assumes global velocity smoothness over the image, which is violated in our application at the boundary of the eye and skin, and thus these errors could be propagated through the images. Although each of these existing methods is adequate under certain circumstances, none of them provide an accurate, practical, and easy-to-use solution to estimate camera translation.

## 2. Methods

### 2.1. Proposed approach

Our goal is to develop an automatic algorithm to estimate the amount of camera translation relative to a deformable anatomical