An advanced Kalman filter for gaze tracking signal

Miika Toivanen *

Brain Work Research Centre, Finnish Institute of Occupational Health, Topeliuksenkatu 41 a A, 00250 Helsinki, Finland

A B S T R A C T

This paper considers the problem of removing unwanted noise from a gaze tracking signal real-time. The proposed remedy is a linear dynamic model for the gaze and a Kalman filter for estimating its optimal solution in closed form. The location and velocity of gaze are treated as independent parameters of the model. Two alternative methods for estimating the velocity are presented; the first is based on the difference in the subsequent eye images and the second on the PCA model and an affine mapping from the principal component space to the gaze space. The covariance matrix of the measurement noise distribution is modified real-time based on the estimated velocity. The presented filtering algorithm can be utilized with any eye camera based gaze tracker. Here, its ability to decrease noise of two published gaze tracking methods is demonstrated.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

As human's gaze reveals a person's focus of (visual) attention and interest, gaze tracking has a number of potential usages. Example fields include market research (e.g., how a person observes certain package), education (e.g., an expert can demonstrate where she looks while performing a professional operation), safety (e.g., how well a bus driver observes traffic while driving), and human–computer interaction (e.g., replacing a computer mouse with gaze). Gaze tracking systems usually use eye camera(s) for assessing the gaze. A reliable and well-performing gaze tracking system should have high accuracy and precision. Accuracy is defined as the average distance between the gaze point, estimated by the gaze tracking system, and the actual gaze point whereas precision is defined as the amount of fluctuation around the mean value, usually in terms of root-mean-square (RMS) value of subsequent distances [1].[1] For instance, if gaze is used as an input for activating symbols on a display, such as entering passwords [2], a too noisy gaze signal will make the usage impractical.

Various methods for improving the precision, i.e., decreasing the fluctuation have been presented. For instance, Kumar et al. [3] try to improve precision real-time by detecting saccades, using a threshold for the derivative of the signal, and smoothing fixations separately – this kind of approach is prone to errors with small saccades and smooth pursuits which are defined as the slow motion that eye makes when following a moving object [1]. Other similar

methods, with differing filtering shapes, were reviewed in [4]. The Kalman filter has also been used; Ji and Yang [5] use a Kalman filter for tracking the pupil and Komogortsev and Khan [6] use a Kalman filtering directly on the gaze signal. These works, however, had no observation model for the velocity of the pupil or gaze point which leads to problems if the gaze signal contains severe noise. The method of Komogortsev and Khan [6] was outperformed by other methods in the comparison of Špakov [4] in terms of accuracy and precision.

This paper presents a well-founded Kalman filter based solution for smoothing the gaze signal, recorded with any (video-based) gaze tracking device. The model has observation models for both location and velocity of the gaze point. For best performance, the location and velocity measurements should be independent. Here, two methods for estimating the gaze velocity are presented. The simpler one uses the pixel-wise difference between subsequent eye images as the amount of change. The other method constructs a principal component based model between the eye image and gaze point; this method can also be used as such for a coarse but lightweight robust gaze estimator. The covariance matrix of the measurement noise is modified real-time so that there would be less filtering during saccades than during fixations. The results show that the presented solution gives a manyfold increase in the precision when applied to the signal of two published gaze tracking methods.

2. Method

This section presents the dynamic model of gaze location and velocity and shows how they can be filtered with a Kalman filter.

---

* Tel.: +358438259572.
  E-mail address: miika.toivanen@ttl.fi

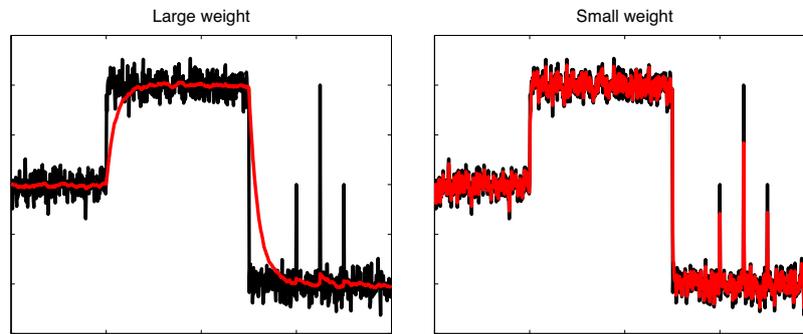[1] In this paper, increased accuracy and precision means lower error and RMS.

Fig. 1. An artificial signal (black line) which is filtered (red line) with a simple IIR filter: $\tilde{x}^t = w\tilde{x}^{t-1} + (1-w)x^t$, having a large value for w on the left panel and small on the right panel. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

The observed gaze point can be reported by any gaze tracking algorithm. For measuring the gaze velocity, two alternative schemas are given, based on the eye difference between subsequent frames and eigen eyes which are the principal components of eye images. The purpose of the gaze velocity measurement is to give information about the movement of the eye; at simplest, immobility of the eye can inform the system that the gaze point is not moving and the possible observed fluctuation of the gaze point should be treated as measurement noise.

First, the problem statement is presented and then the dynamical model and its Kalman filter solution. The following two subsections present the two methods for estimating the velocity and the last subsection presents the method for modifying the measurement noise covariance.

2.1. Problem statement

Any signal can be filtered real-time (that is, without causing a delay) as a weighted sum of the latest and past observations or estimates. A single pole IIR filter is one of the simplest such filter:

$$\tilde{x}^t = w\tilde{x}^{t-1} + (1-w)x^t \qquad (1)$$

where x^t is the input sample of the signal at time t and $\tilde{x}$ is the filtered signal. Giving an infinitely long signal which fluctuates around a constant value, the filter (1) converges to that value. However, with a changing signal, such as the gaze signal, this simple approach leads into problems, as is illustrated in Fig. 1; the weight must be large enough to achieve the desired level for filtering which causes the estimate to be inaccurate for some period after the signals has changed. If the weight is decreased, the difference between the signals also decreases with a cost of the filtered signal following the noise too much. Having a more complex filter or a FIR filter fails to solve the problem. One could build a heuristic algorithm which adapts w according to the derivative of the signal. However, if the signal has a varying noise distribution with occasionally large

fluctuation, as a gaze tracking signal often has, such an approach would also fail as the noisy samples (like the three "spikes" in Fig. 1) would probably be considered as signal changes.

Hence, the problem is to (automatically) infer the correct amount of filtering. This paper presents an algorithm which solves the problem. The algorithm is based on the Kalman filter method which recursively produces a statistically optimal closed form point estimate for the unknown state of a linear dynamic model which has Gaussian distributions for process and observation noises (see, e.g., [7]). Kalman filter can predict the state also in case of missing observations which occasionally occur in gaze tracking. Additionally, as opposed to heuristically detecting saccades and filtering the signal in between them (i.e., fixations), Kalman filter is a theoretically sound method which has been used in probably thousands of signal processing applications.

The presented method bases also on the idea that the velocity observations should be independent of the actual signal to be filtered. If the velocity would be estimated as the derivative of the location, the velocity estimate would follow the noise of the gaze tracking signal as in the right panel of Fig. 1. Instead, the gaze velocity is computed by using the whole eye image as input. If the gaze point is estimated by detecting the pupil and possibly glints of one or more LED sources, as it typically is [8–12], the velocity estimate becomes independent on the gaze location estimate. This leads to much more robust performance as opposed to estimating the velocity from the gaze points. For instance, a LED glint that is suddenly falsely detected gives an erroneous gaze location; by independently assessing from the eye image the amount of change in the gaze location, such false location estimate can be filtered out.

Lowering the fluctuation (by filtering) will improve the precision by definition. It will also improve the accuracy if the distance between the mean value of the signal and the real value is smaller than the amount of fluctuation. This simple phenomenon is illustrated in Fig. 2 where the mean signal represents a "perfectly" filtered signal.
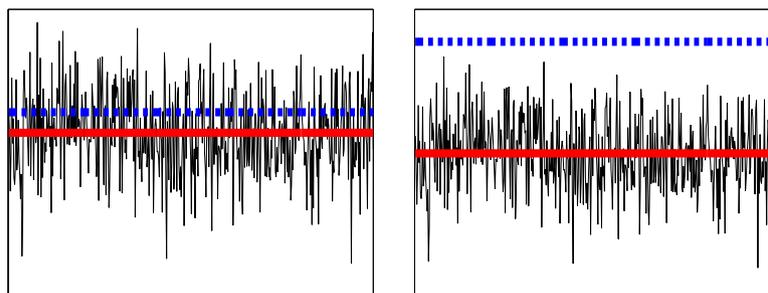

Fig. 2. An artificial noisy signal (thin black line), its mean value (thick red line), and the "true" value (dashed blue line). Replacing the signal with its mean value ("perfect filtering") naturally always improves the precision. In the left panel, the accuracy is better for the mean value than for the signal because the true value is within the noise level so the average distance between the mean and the true value is smaller than the average distance between the signal samples and the true value. In the right panel, the accuracies are the same. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)