High-frequency Optical Solution for Human Motion Tracking and Eye-tracking

V A Chertopolokhov\textsuperscript{1}, E V Khinenzon\textsuperscript{1}, A V Pravdivtsev\textsuperscript{2} and G L Grigoryan\textsuperscript{2}

\textsuperscript{1}Faculty of Mechanics and Mathematics, Lomonosov Moscow State University, 1 Leninskiye Gory, Moscow, Russia
\textsuperscript{2}LLC “Total Vision”, str. 1, d. 42, Bolshoi bul., Skolkovo Innovative center, Moscow, Russia. psvr@vrmsu.ru

Abstract. The high frequency optical tracker is discussed in the article. Test image processing algorithm and the math model of a marker image was proposed. It is based on the real images of typical objects for detection. The investigation of its statistical characteristics gives a good approach for image processing algorithms. The model shows good conformity to real data. A new gaze tracking system based on self-designed system on the crystal was developed and tested. The high frame rate allows us to detect and determine several neurologic diseases like Alzheimer's and Parkinson's disease. The device can be also used in wearable virtual reality headsets.

1. Introduction

Total Vision Ltd. in collaboration with the laboratory of MOIDS, Moscow State University is developing a virtual reality system and a solution for tracking the entire body of the user, as well as the direction of human gaze.

One well-known technology is a marker-based motion capture system. A marker system is a set of LED elements that are rigidly attached to a designated object. When an object moves with a set of cameras, the location of the LED markers is fixed, which allows you to determine the coordinates and orientation of the object in space.

Foveated rendering has become the next focus of our research as a way to optimize the performance of a virtual reality system. Total Vision has developed an eye tracker which can be used in monocular or binocular setup. This device allowed real-time gaze tracking. A high-frequency eye tracker (1000 Hz) allowed us to recognize fast saccadic eye movements. In order to achieve smooth quality for highly loaded virtual scenes, we significantly reduced the basic rendering quality and used an eye tracker to determine the field for high-resolution rendering.

In both problems, the question arises of determining the accuracy of the methods for finding the center of a light or dark spot in an image, taking into account various noise recording equipment.

2. Noise model
This section discusses the creation of a mathematical model of noise of a real marker, which is tested on artificially synthesized data and the hypothesis of a normal distribution of digital noise of real data is built. To confirm or refute the hypothesis, the Pearson test was used.

2.1. Mathematical model
Marker-based motion capture systems are divided into two main groups - with optically passive markers and optically active markers.

Optically passive markers are based on the technology of reflective elements, that is, each marker reflects the light sent to it. In such systems, there is a need to shine on markers from additional devices. The second large group - with optically active elements, is based on LED markers or simply on LEDs. The technology does not require the supply of light to the markers and ordinary cameras are sufficient to capture the movement of the diodes.

The chapter discusses the mathematical model of noise built for active markers, but since the features of active systems are not used, this mathematical model can also be applied to systems with passive reflective elements.

Synthesizing an artificial diode is the primary task of the algorithm for constructing a mathematical model. In the image, it was necessary to create a diode spot as similar as possible to a real diode. Under conditions of low shutter speed, the diode is displayed as a region of bright light, with a decrease in light intensity with distance from the center.

The solution to the problem was divided into two parts – the first is the drawing the center of LED with monochrome color and the second is drawing of the halo. For the center of the LED, was drawn a circle with a certain radius, halos were drawn around the center with a proportional decrease in brightness with distance from the center. The ability to set the size of the LED spot on the image made it possible to check the quality of the mathematical model at different distances from the camera. Below is an example of an artificial diode with a diameter of 15 pixels and a center at the point (100; 100).

![Synthesized diode with a diameter of 15 pixels.](image)

The center of the synthesized diode was considered known and selected in advance, which is one of the main positive features. This approach simultaneously solves the problem of determining the accuracy of the algorithm, that is, the coordinates of the center of the LED are set and it is drawn on the image. A sample of synthesized images is created with different sizes of diodes, different spot and brightness. After processing a large number of synthesized images, you can calculate the error model of a particular method.

But the data synthesized in this way do not correspond to reality. In real conditions, the concept of noise of photosensors arises. And after synthesizing the LED spot, you need to solve the problem – add noise so that it is as close as possible to the current conditions for recording video/photo data.

Color noise in the image is the interference caused by the sensors of digital video recording equipment. Noise can be caused by insufficient lighting and imperfect equipment. In the image, noise
appears as a chaotic grid of pixels of various colors on photo/video data. As digital noise, a Gaussian or normal noise was chosen [1].

Adding Gaussian noise is the masking of pixels made up of random color values in a given range on the original image.

\[ I' = I + \varepsilon \]  

where \( I \) - source image, \( I' \) - noised image, \( \varepsilon \) - Gaussian random variable.

Gaussian random variable is determined by the Gauss function:

\[ f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\bar{x})^2}{2\sigma^2}} \]

where \( \bar{x} \) is the average value, \( \sigma \) is the standard deviation, \( \sigma^2 \) is the distribution variance.

Below are images of a synthesized diode with normal noise (average - 0, variance = 10) and a real diode.

![Synthesized diode with normal noise and real diode](image)

**Figure 2.** Synthesized diode with the diameter (left), real diode (right).

2.2. **Image segmentation**

To assess the quality of the constructed model, it is necessary to separate the spot of the LED from the background, that is, to segment the image. There are several main groups of methods for separating an object from the background, such as: threshold filters, image clustering methods, and methods based on border detectors. For image segmentation, the k-means clustering method was used [2].

A significant drawback of the k-means method is that the result depends on the choice of the starting points, that is, the centers of the clusters, which does not guarantee the most optimal result, but despite this it is proved that the algorithm converges, it means that, with a large number of iterations, an acceptable result is achieved.

Below are example pictures of the operation of k-means clustering algorithm on an artificially synthesized diode and a real diode with the detection of cluster centers. For a real image, examples of separation into 2 clusters and 4 clusters are presented.
2.3. **Definition of characteristics**

After clustering, we know the centers of the clusters. The first cluster center for the background, the second for the LED spot in the case of two clusters. As a descriptive characteristic, the distance was calculated, as the difference between the value of the component of color value of all image points and the value of the points of the centers of the clusters, which are obtained by the k-means algorithm. The distance was calculated using the following formulas:

\[ r_{i,j} = x_i - M_j(r) \]

where \( r_{i,j} \) is the distance from the \( i \)-th pixel to the \( j \)-th center of the color; \( x_i \) are the values of the color component of the pixel; \( M_j(r) \) are the values of the center of the \( j \)-th cluster, \( i = 1, 2, \ldots, n \) are all pixels of the image, \( j = 1, 2, \ldots, m \) – all image clusters. This value may be negative. The value was calculated for color value and for each center of the cluster. In this case, 4 clusters were taken. And in the end, we got 4 parameters that determine the characteristics of the image. The purpose of the selection is to calculate the distribution. The data obtained for each of the above characteristics were recorded in an array for subsequent processing.

2.4. **Estimation of the math model**

To estimate the quality of the constructed mathematical model on real data, the following approach was implemented. From the data obtained, the mean and variance estimates were found, respectively, the density function and distribution function, a hypothesis of normal distribution of input data was built. To confirm the hypothesis, a chi-square agreement criterion or Pearson criterion was chosen [4]. According to the value of the obtained criterion, the hypothesis was accepted, which is shown below.

To apply the consent criterion, data processing was required. In general, the data is a sample - \( x_1 \ldots x_n \), where \( n \) is the number of image points. From the sample, the mean and variance values were found.

The data were normalized - it means that, the average value was taken away and divided by the variance and the number of points.
The set of samples $x_i'$ was divided into $N$ intervals - $\Delta_j = (a_j, b_i), j = 1 \ldots N$ and, accordingly, the number of hits $\xi_i$ in each segment was considered.

To apply the Pearson criterion, it is necessary to construct an empirical distribution function according to the normal law. The empirical distribution function was calculated using the following formula:

$$F(x) = \frac{1}{2} \left[ 1 + \text{erf} \left( \frac{x - \bar{x}}{\sqrt{2} \sigma} \right) \right]$$  \hspace{1cm} (3)

where $\text{erf}$ is the error function (Laplace function), $\bar{x}$ is the average value, $\sigma$ is the standard deviation, $\sigma^2$ is the distribution variance.

As a result of the above calculations, values were obtained - the density function of real data and the empirical distribution function. Pearson's criterion was considered as follows:

$$X_k^2 = k \sum_{i=1}^{N} \frac{1}{p_i^0} \left( \hat{p}_i - p_i^0 \right)^2$$  \hspace{1cm} (4)

where $X_k^2$ is the Pearson criterion with $k$ degrees of freedom, $N$ is the number of intervals, $\hat{p}_i$ is the density function of real values, $p_i^0$ is the probability function of falling into the $i$th interval, which is calculated by the formula

$$p_i^0 = F(b_i) - F(a_i)$$  \hspace{1cm} (5)

where $F(x_i)$ is the value of the empirical distribution function of the $i$th value.

By the obtained Pearson criterion - $X_k^2$, it was required to determine at what $\alpha$, or with what probability the chi-square value occurred in a known normal distributed population and accordingly confirm or refute the hypothesis of a normal distribution.

For each selected characteristic, for each center of the cluster — the values of the Chi-square agreement criterion — $X_k^2$ — were calculated and under what $\alpha$ the hypothesis of normal distribution was accepted. According to the results of the calculations, the average value of $\alpha$ was 88%. This means that if we get the samples from the generally known normally distributed population, then the value of $X_k^2$ with a large sample occurs with a probability near of 88%, which confirms the hypothesis of a normal distribution of digital noise.

These results allow us to determine the accuracy of the algorithms for finding the center of the marker on the image. We use it for motion and eye tracking solution.

3. Tracking solution

The common solution for motion and eye tracker system is a separate sensor and image processing system. The projected image on the sensor transfer to the external multi-purpose computer for tracking calculation. The negative side of the solution is a big amount of the data to transfer and process, low speed, high power consumption. The advantages are quick algorithm update, easy upgrade for new electronic components. To remove these disadvantages, the new system we created based on our design video system on crystal (SOC) (Figure 5). Characteristics of the system are in the Table 1.
Figure 5. The picture of the designed SOC.

Table 1. The characteristics of the SOC.

| Characteristic                           | Value                           |
|-----------------------------------------|---------------------------------|
| Resolution                             | 128 x 128 px                    |
| Accuracy for coordinate finding         | 0.01 px                         |
| Frame rate                              | 1000 Hz                         |
| Pixel size                              | 20.16 x 20.16 um²               |
| Active area size                        | 2.58 x 2.58 mm²                 |
| Fill factor                             | 90 %                            |
| Working spectral range                  | 0.36 - 1.05 um                   |

The SOC was designed to find black and white objects on complex background with frequency 1000 Hz. The crystal can show the coordinates, area and brightness up to point and longitudinal 50 objects in each frame with this frequency. All this information allows to determine the gaze direction. The system works in wide spectral range so different designs for haze tracking can be used. The designed eye tracker can be used not only for pupil detection but also to control the reflection of illumination LEDs. This method allows additionally increase the accuracy of the system. The assembled device has quite small dimensions, mass (less 36 grams for binocular system), high accuracy, low power consumption (<50 mW) so it is quite good for wearable systems.

3.1. The experimental setup

The prototype of the system was checked on experimental setup to determine the accuracy characteristics. The eye was imitated by a sphere with painted pupil on it. The object was fixed on a high precision rotating table (accuracy less than 0.5 arc minute). During the measurement the imitator was rotated to specific angle and position of the pupil’s image on the sensor was captured with additional image processing to find the energy sensor of the image. Internal algorithms were used to determine coordinates on the sensor. For each position 100 measurements were done for additional data processing and statistical calculation. The object was illuminated by a light source in visual spectral range. In this experiment the lens which have field of view ±20° was used. The linear field of view of the lenses was slightly less than the sensor size.
Figure 6. Setup to measure accuracy of eye tracker system.

The measured dependents of digital points (coordinate on the sensor) via angle are shown on the Figure 7.

Figure 7. Angular coordinates measurement curve.

As the position of the object can be determine with the subpixel accuracy, the maximum number of measurements over the sensor is 16000 digital points. Based on the experimental results the sensitivity was 3.5 arc minutes for a digital point. RMS error value was less than 2.5 arc minutes at 3σ level. So in the case of diagonal direction we can have field of view about 35° with calculated accuracy.

The accuracy of the measurements can be increase with data processing with corresponding decreasing the frame rate. The field of view of the system can be increase with another lens with specific focal length. The accuracy in this case should be scaled with field of view value.

Analyze of the result shows that there is some nonuniformity in the response curve. It can be explained by residual distortion in the lens that was used to project image on the sensor. The nonuniformity will be removed in calculation after the distortion calibration of the lens.

For the object detection the internal algorithm of the SOC was used. The video system can transfer signal with frequency 1 kHz so this video stream can be used for additional processing in external system. In this case the periodic calibration can be used to correct the results from internal processing.
The design eye tracker was implemented in Horizon system. The system is a mobile perimeter based on VR headset to detect glaucoma on an early stage.

3.2. Applications

The high frame rate eye tracker is used for gaze tracking. It allows us to determine the false measurement and increase the accuracy of the results, decrease the duration of the tests. In the device the NIR spectral range was used to avoid light adaptation of the eye and fit the standards for perimeters. As the eye tracker has quite high frame rate (1000 Hz) the system can be used to detect several neurologic diseases like Alzheimer's disease, Parkinson's disease and several others [4]. The saccade that specific for these diseases show quite high angular velocity (up to 500 °/s) with duration of saccade up to 100 ms.

Foveated rendering is another application of the eye tracker. High tracker frequency allows us to use saccade mathematical models [5][6] for signal filtering.

4. Conclusions

We’ve designed, manufactured and tested the performance of the tracker based on the video system on crystal (SOC). Due to the SOC it is possible to get very fast haze direction calculation with high accuracy. The system has 1kHz frame rate, small sized and power consumption less than 50mW. For the field of view 40° the accuracy 3.5 arc minutes was reached. So the designed that tracker can be used in wearable devices, such as AR and VR headsets.

For tracking purposes, we addressed the issue of constructing and estimating a mathematical model of the noise of a sensor of video recording equipment. A technology for synthesizing an artificial diode with a well-known center has been developed. The hypothesis about the normal distribution of digital noise was confirmed. These results allow us to determine the accuracy of the algorithms for finding the center of the marker on the image, as well as to build an estimation model of these algorithms.

The tracking device was implemented in medical VR helmet Horizon which allow to detect glaucoma on the early stages. Clinic tests confirmed the efficiency of the system. Additional improvements of the accuracy based on the combining results from internal algorithms and external video processing are under investigation now.

Acknowledgments

The first part of this research is supported by the Russian Foundation for Basic Research grant 18-00-01684 (K) (18-00-01590 and 18-00-01641).

5. References

[1] Gonzalez R and Woods R 2007 Digital Image Processing (New Jersey: Prentice Hall) 222
[2] Bradski G and Kaehler A 2008 Learning OpenCV (Sebastopol: O’Reilly Media) 479
[3] Lagutin M 2009 Visualising Mathematical Statistics (Moscow: BINOM. Laboratory of Knowledge) 273
[4] Belyakova-Bodina A I, Bril' E V, Zimnyakova O S et al. 2017 Annals of clinical and experimental neurology 11 (4) 52 DOI: 10.18454/ACEN.2017.4.6
[5] Rabichev I 2011 Functional model of eye movement in visual perception (Moscow: METHODS OF COMPUTER DIAGNOSTICS IN BIOLOGY AND MEDICINE, Materials of the annual All-Russian Scientific School-Seminar) 36
[6] Kruchinina A P and Yakushev A G 2018 Biophysics 63 (2) 241