electronics

MDPI

Article

An Improved Unscented Kalman Filtering Combined with Feature Triangle for Head Position Tracking

Xiaoyu Yu [1,2], Yan Zhang [2], Haibin Wu [2,*] and Aili Wang [2,*]

[1] College of Electron and Information, University of Electronic Science and Technology of China, Zhongshan Institute, Zhongshan 528402, China
[2] Heilongjiang Province Key Laboratory of Laser Spectroscopy Technology and Application, Harbin University of Science and Technology, Harbin 150080, China
* Correspondence: woo@hrbust.edu.cn (H.W.); aili925@hrbust.edu.cn (A.W.)

Abstract: Aiming at the problem of feature point tracking loss caused by large head rotation and facial occlusion in doctors, this paper designs a head-position-tracking system based on geometric triangles and unscented Kalman filtering. By interconnecting the three feature points of the left and right pupil centers and the tip of the nose, they form a coplanar triangle. When the posture of the doctor's head changes due to rotation, the shape of the corresponding geometric triangle will also deform. Using the inherent laws therein, the head posture can be estimated based on changes in the geometric model. Due to the inaccurate positioning of feature points caused by the deflection of the human head wearing a mask, traditional linear Kalman filtering algorithms are difficult to accurately track feature points. This paper combines geometric triangles with an unscented Kalman Filter (UKF) to obtain head posture, which has been fully tested in different environments, such as different faces, wearing/not wearing masks, and dark/bright light via public and measured datasets. The final experimental results show that compared to the linear Kalman filtering algorithm with a single feature point, the traceless Kalman filtering algorithm combined with geometric triangles in this paper not only improves the robustness of nonlinear angle of view tracking but also can provide more accurate estimates than traditional Kalman filters.

Keywords: head position estimation; viewpoint tracking; unscented Kalman filter; Kinect

Citation: Yu, X.; Zhang, Y.; Wu, H.; Wang, A. An Improved Unscented Kalman Filtering Combined with Feature Triangle for Head Position Tracking. *Electronics* **2023**, *12*, 2665. https://doi.org/10.3390/electronics12122665

Academic Editor: George A. Papakostas

Received: 25 March 2023
Revised: 5 June 2023
Accepted: 11 June 2023
Published: 14 June 2023

1. Introduction

With the rise of the era of artificial intelligence, there are many ways to realize human–computer interaction in the virtual world. The traditional mouse operation and keyboard input can no longer meet people's needs. Virtual Reality (VR) is developing multimedia and multi-channel direction, such as gesture interaction, body interaction, behavior capture, tactile anti-touch, eye movement tracking, voice response, etc. [1]. Among them, head-position-tracking technology is an important research topic in the VR field.

Visual angle tracking generally refers to the process of automatically detecting the relative position of the human pupil or estimating the direction of vision. However, the eye movement process is always accompanied by various kinds of noise from the external environment, such as external light interference, and mask occlusion, which makes the timing positioning of the eye tracking target error. There is also noise accompanied by the eye movement itself, such as the large deflection of the head, eye shaking, blinking, and other unconscious actions, which make the pupil-tracking algorithm need to be fast and robust. The purpose of visual-angle-tracking technology is to predict the target position and pose (position, rotation angle, motion speed, acceleration, and direction, etc.) of the following frames based on statistical or learning methods according to the initial state and characteristics of the eyes in the tracked human head posture and finally draw the motion trajectory of the tracked object in the entire video sequence.

Scholars at home and abroad have been engaged in the research of perspective tracking for decades, and a large number of excellent tracking algorithms and evaluation systems have emerged, aiming to design a tracker with good robustness and high accuracy. From the perspective of the development of perspective-tracking technology, the current perspective-tracking system is mainly divided into an invasive perspective-tracking system and a non-invasive perspective-tracking system [2]. In the intrusive system, users need to wear optoelectronic auxiliary equipment designed by the developer. Using the movement information of the head and eyeball collected by this equipment, the geometric model of the direction of the human eye line of sight is established to estimate the focus position of the human eye line of sight. Invasive visual-angle-tracking systems generally require relatively expensive hardware (similar to glasses and helmets) to be fixed on the human head.

Marquard [3] proposed the direct observation method to realize eye movement tracking, which aims to reduce patient recognition errors using eye movement tracking data and direct observation mode. MacNeil [4] described a method to estimate the corrected EOG signal by calibrating it with the industry-standard pupillary corneal reflex (PCR) eye tracker. In the same year, Katona [5] used an eye tracker to understand the code via the evaluation based on eye tracking, analyze the gaze route, and measure and evaluate the eye movement parameters to determine the readability and understandability of the code. With the eye movement instrument also playing a huge role in medicine, Fabio [6] and others used eye movement instruments for cognitive rehabilitation to analyze patients' attention, which studied the longitudinal effect of cognitive rehabilitation on patients. Recently, Chugh S. [7] developed a complete convolution neural network based on U-Net architecture, which solved the detection and matching problems when there are false and missing reflections. This is a significant improvement on the most advanced learning-based XR eye-tracking system, which reports 2–3° fixation errors.

Although eye trackers are widely used at present, there are many problems. For example, a complete set of eye trackers is expensive, and the construction environment is complex, which is not suitable for the perspective tracking of doctors' surgery. With the continuous emergence of emerging technologies, artificial intelligence, pattern recognition, and other new technologies are providing more possibilities for perspective-tracking technology [8], especially in the field of medical imaging, which can play an auxiliary role in doctors' surgery or assist patients with mental disorders via non-invasive technology.

In medicine, the commonly used method for tracking the position of the head is to use single or multiple sensors [9,10] in this positioning system and sample these sensors to measure the changes in the rotation of the doctor's head. The measurement quantities in this method are usually the size of the magnetic field, the reduction value of the light source projected by the camera after collection, and the inertia value of the sensor. These changes are calculated and demonstrated to obtain the translational and rotational information values of the target object [11,12].

These methods usually require the placement of equipment near the doctor's head or operating table, and the accuracy and robustness of tracking are subject to the resolution and accuracy of the equipment. Therefore, how to obtain the doctor's head position and posture with the help of a small amount of external equipment on the basis of existing medical equipment and without affecting the normal operation of the doctor and apply it to minimally invasive surgery has been a research hotspot for scholars at home and abroad in recent years.

In order to accurately obtain the position and posture of the human head, many scholars have conducted research on this. According to the different implementation methods in positioning, domestic and foreign scholars' research can be mainly divided into two parts. One is the head-positioning technology based on sensors, which obtains the position information of the doctor or patient by arranging sensor devices on their body, thereby achieving target tracking; The second is based on visual-head-positioning

technology, which mainly requires external devices such as Microsoft's Kinect [13–15]. The camera captures facial images and processes the images to achieve target tracking.

Non-invasive eye tracking based on video images has caused a lot of in-depth research in academia and industry [16]. New instruments and equipment, such as high-resolution image acquisition equipment and external light source, are increasingly applied to the information acquisition of the line-of-sight-tracking system. New technologies such as image processing, pattern recognition, and artificial neural network provide more feasible methods for line-of-sight tracking, which is moving towards a non-invasive, portable, and low-cost trend [17].

Some researchers rely on non-contact eye image acquisition equipment for line-of-sight tracking research. Zhang et al. [18] developed a simplified 3D line-of-sight tracking technology with stereo vision. Compared with other 3D systems, this paper uses the first image of two LEDs to estimate the 3D corneal center and pupil center. Secondly, the fixation point on the screen is calculated by intersecting with the estimated 3D gaze. Li's method based on depth sensor usually uses Kinect to obtain depth information and converts the POR from the center of the eye, and performs 3D gaze estimation under natural head motion, where the center of the eye is calculated using the inner eye angle as the anchor point for head posture tracking [19]. Although the model is simple, the proposed screen calibration method is too complex. Zanetti et al. [20] used Kinect equipment to reduce the line-of-sight tracking error based on pupil–cornea reflection in 3D scenes and improve the accuracy of the subject's eye feature tracking under infrared light irradiation. Elmadjian [21] revealed the limitations of the widely used 3D gaze estimation technology and proposed a new calibration procedure, which uses the uncalibrated head-worn binocular eye tracker and RGB-D camera to track the 3D gaze within the scene volume. The results show that although the interaction technology with a focus on gaze should be realized in 3D, it is still a challenge to accurately estimate in this setting.

In the past three years, Gonzal et al. [22] proposed two gaze area estimation modules integrated into the driving simulator. One uses the 3D Kinect device, and the other uses the virtual reality Oculus Rift device, which can detect the driver's gaze area in each route processing frame of the seven areas divided by the driving scene. For the stereo environment, Han et al. [23] used the 3D vision estimation method provided by Intel Realsense 3D camera for the 3D environment. By combining Move Eye Tracking (MET) with Visual Simultaneous Localization and Mapping (VSLAM) technology, we can capture individuals' gazes in open spaces and irregular viewing paths. Liu developed [24] a 3D model-based line-of-sight tracking method based on a single camera and single light source system, which uses iris features instead of pupil features used in most traditional model-based 3D gaze-tracking systems, which is simpler than the system configuration required by existing technical methods. Because the position of the fixation point in the perspective tracking requires the depth information of the face, Chen Qin [25] and others selected the RealsenseD455 depth camera to obtain RGB images and depth maps and used the three-dimensional TFGazeNet and SSRGazeNet neural networks to track the perspective, and the experiments verified the feasibility of the algorithm. However, the camera is expensive, and other depth cameras with the same function can be considered to replace it, and the perspective-n-points (PNP) method can be used to obtain the face position information.

The Kalman filter is widely used in the field of moving target detection [26]. The position and direction of the objects depend on the Kalman filter, extended Kalman filter, and unscented Kalman filter to reduce the error [27]. Sultan M.S. [28] combines robot kinematics, vision, and force sensors to independently perform hand–eye coordination tasks, and the proposed method accurately determines the camera pose of single-view objects. The corners are indirectly detected from the intersection of lines, and the lines are obtained using RANSAC (Random Sample Consumus) algorithm. The final pose correction is improved by the Kalman filter to gradually update. Munir F. [29] proposed an extension of spatiotemporal context learning based on Kinect and Kalman filtering. Spatiotemporal context learning provides the most advanced accuracy in general object tracking, but its

performance will be affected by object occlusion. By combining with the Kalman filter, the proposed method can simulate the dynamics of eye movement and provide reliable eye movement tracking under occlusion. Pan [30] designed a combination of a convolution neural network and Kalman filter to estimate and analyze the real-time changes in eye position. Kalman filter is used to track the human eye and eliminate jitter interference. In recent years, Garapati [31] has combined a modified gain extended Kalman filter (MGEKF) with a particle filter to improve the performance of the filter. Huang et al. [32] proposed a view-tracking algorithm based on a multi-model Kalman filter, which uses a multi-model Kalman filter to improve the efficiency of eye tracking in images and uses multiple models to estimate objects. The algorithm consists of two parts. The first step is to identify the initial position of the eyes using Support Vector Machine (SVM). In the second part, the multi-model Kalman filter is used to predict the eye position in the next frame, which is based on the constant speed and acceleration model of normal people.

The angle-of-view tracking method using Kalman filtering for the above algorithm is a linear system estimation algorithm. In a real visual-angle-tracking environment, head and eye movements have a high degree of nonlinearity of the likelihood model, and standard Kalman filters are no longer optimal in real environments. Feng [33] proposed a new improved UKF based on Double Unscented Transformation (DUT), starting with the inherent mechanism of UKF and the dynamic state model of the human hand, with the goal of improving the accuracy of human hand tracking. Zutao et al. proposed a new driver fatigue detection scheme based on nonlinear unscented Kalman filtering and eye tracking. Assuming that the probability distribution ratio approximates any nonlinear function or transformation, unscented transformation (UT) can be used to achieve nonlinear eye tracking, which uses a set of deterministic sigma points to match the posterior probability density function of eye movement. Hannuksela J. [34] studied human head posture estimation based on facial features. Special consideration is given to facial feature extraction methods suitable for real-time systems. The extended Kalman filtering (EKF) framework is used to track and estimate the three-dimensional posture of the rotating head in an image sequence. During the tracking process, the correspondence between the rigid head model and the three extracted facial features (eyes and mouth) is used to solve the pose. However, there are some limitations in the current implementation. Users who wear glasses, occlusion, and shadows can cause system problems. Bankar R. et al. [35] described performance improvements in head gesture recognition. Firstly, an improved Camshift-based face-tracking algorithm is used to improve head gesture recognition. UKF is used to suppress noise and correct the shortcomings of inaccurate positioning. To overcome the shortcomings of existing gesture-tracking methods in terms of accuracy and speed, Du X. proposed a new YOLOv4 model that combines Kalman filtering with real-time hand-tracking methods. The new algorithm can solve the problems of detection speed, accuracy, and stability in hand-tracking technology. The YOLOv4 is used to detect the current frame tracking target, and Kalman filtering is applied to predict the next position and boundary box size of the target based on the current position of the target. Tian L. et al. proposed [36] Consistent Extended Kalman Filtering (CEKF) for maneuvering target tracking (MTT) with nonlinear uncertain dynamics and applied it to manual position tracking.

In summary, the previous algorithm has two problems: (1) combining linear Kalman filtering with eye motion tracking by assuming that the human eye motion is with uniform motion mode; (2) Most visual-angle-tracking systems aim at the problem of errors in positioning and tracking a single feature point. Different visual-angle-tracking algorithms are proposed based on real-time images of the left and right eyes, respectively.

This paper introduces a head-position-tracking system, which uses a combination of geometric triangles and unscented Kalman filtering to analyze and predict head and eye movement information for the next frame. Firstly, aiming at the practical needs of doctors' surgical scenarios and limitations such as data collection, a simulation platform for operating room perspective tracking was built using Kinect v2 and combining different organ and soft tissue models. Secondly, a doctor's head posture model is established using

the geometric triangle method, using three relatively stable feature points: the nose and the center of the pupil, and the relative position changes with the change of the head posture, presenting certain geometric features to convert the change of the head posture into a geometric triangle deformation. Finally, the most suitable head position orientation is obtained by combining the unscented Kalman filter with geometric triangle tracking to track the doctor's visual angle based on the position and pose of the previous moment and the weight of the feature point position and pose at the current moment. The performance of the head-position-tracking system is evaluated qualitatively and quantitatively.

2. Materials and Methods

This section describes the overall architecture and methods of the proposed eye-movement-tracking system, using triangular models to obtain head posture, eye movement information, and unscented Kalman filtering for perspective prediction and tracking.

The geometric relationship of the head in the perspective-tracking system is shown in the Figure 1, R(camera–world) represents the matrix of the world coordinate system under the camera coordinate system, R(world–head) represents the posture matrix of the doctor's head relative to the world coordinate system, and R(camera–head) represents the posture matrix of the doctor's head under the camera coordinate system.

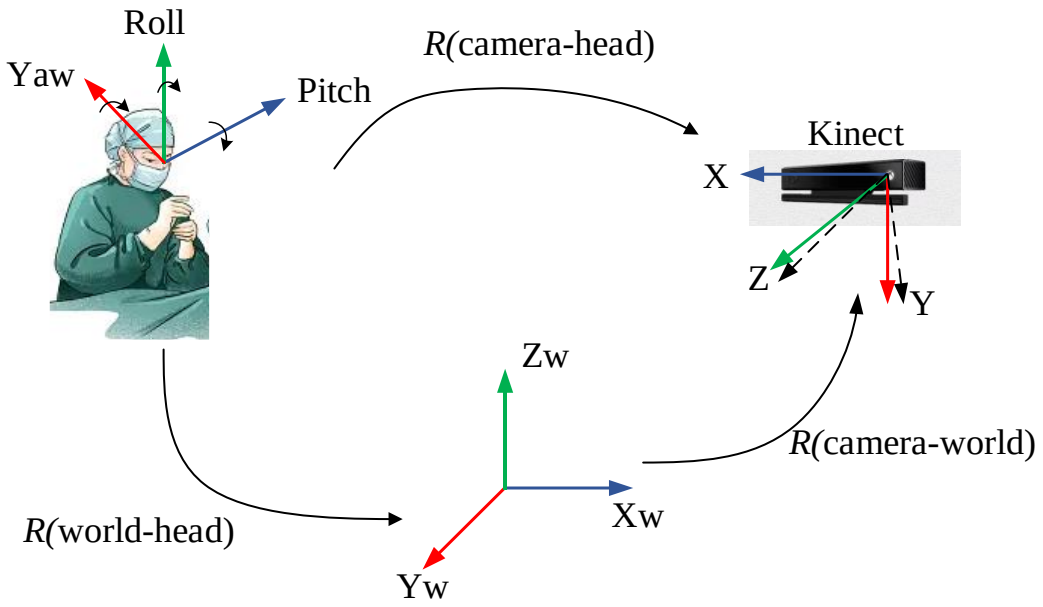


Figure 1. Geometric relations in head pose estimation.

In order to more intuitively describe the doctor's perspective-tracking system, this article has built an operating room perspective-tracking simulation platform, as shown in Figure 2. The entire platform includes Kinect v2 sensors to obtain doctor's head information, patient's body surface and internal cavity, and surgical forceps. Due to the non-invasive head position tracking used in this article, doctors do not need to wear complex and heavy equipment on the head, which is convenient for doctors to operate. This method can achieve both single feature point tracking and multiple feature point tracking. Faced with the problem of losing tracking of a certain frame in the video, it is not necessary to recalibrate and is more suitable for long-term perspective-tracking research [37].

The software module is responsible for performing different tasks, mainly including the following sub modules: face detection and feature point localization, triangular model acquisition of head posture, and unscented Kalman filter prediction and tracking.

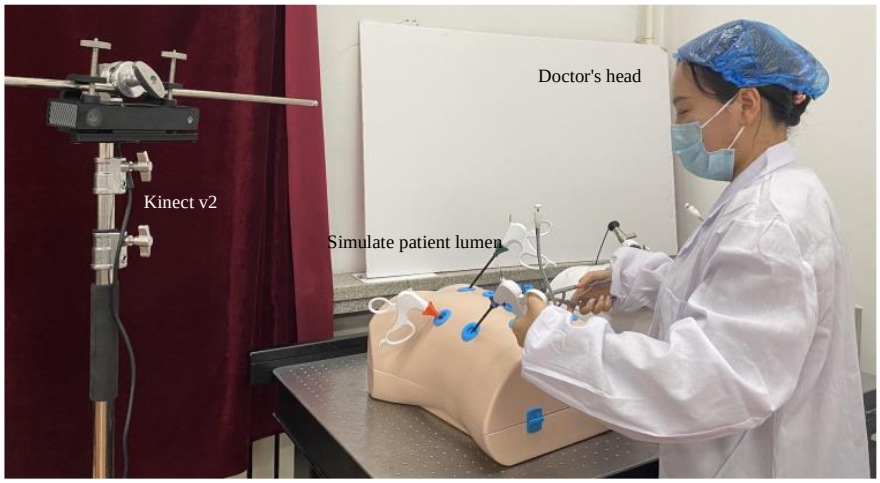


Figure 2. Construction of head-position-tracking simulation platform in operating room.

In this paper, the geometric model method is used to estimate the head posture. A doctor's head posture model is established using a geometric triangle method, using three relatively stable feature points of the nose and pupil center, and their relative positions change with the head posture, presenting certain geometric features. Therefore, after locating the eyes and mouth, the estimation of the head posture is achieved. The flow chart of the proposed head position tracking is shown in Figure 3. Firstly, the current measurement value of Kinect was used as input, and the facial information of doctors with different gestures was obtained via Adaboost face detection algorithm combined with MB-LBP features. The angle-positioning model of feature triangle was analyzed and established via ASM feature-point-positioning method and pixel coordinates combined with depth information to solve the problem that the head position and posture of doctors wearing masks change frequently, resulting in the loss of feature point tracking. Finally, combining the head-attitude-positioning algorithm of feature triangle and the tracking algorithm of untraced Kalman filter, the real-time prediction and tracking of the doctor's perspective during surgery are carried out, and the predicted value of the next frame is output.

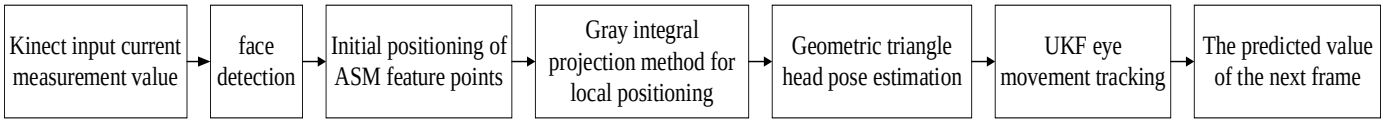


Figure 3. The flow chart of the proposed head position tracking.

2.1. Location of Left and Right Pupil and Nasal Tip Feature Points

In view of the complex surgical environment of doctors, ASM and AAM algorithms are used to obtain 68 feature points of the face. Due to the excessive computation of feature points, which can increase the speed of program operation, it is necessary to reduce the number of feature points in order to analyze and experiment with the real-time stability of feature points. Therefore, based on the global ASM model [38] localization, local secondary localization is performed for this region to obtain three feature points, namely, the center of the pupil and the tip of the nose.

For the positioning of the tip of the nose, the gray value of the lower nostril generally differs greatly from the neighbors. Therefore, the approximate region of the tip of the nose is determined using the gray integration projection method of the nostril region [39]. The grayscale integral projection is a projection distribution feature of the grayscale value of an image in certain directions. It is set $F(x, y)$ as the grayscale value of the pixel (x, y) in the

image, and the size is M × N. The horizontal and vertical grayscale integration functions of an image are shown in Equations (1) and (2), respectively, as follows:

$$H(y) = \sum_{x=0}^{N-1} F(x,y) \tag{1}$$

$$V(x) = \sum_{x=0}^{M-1} F(x,y) \tag{2}$$

Typically, the nostrils have the lowest brightness values in the approximate area of the nose. Therefore, first, a simple graying process can be performed on the color image of the nose region to convert it into a grayscale image that reflects the brightness level; Then, convert the grayscale image into a binary image to find the location of the nostrils.

If the set of all pixels in the left nostril region is X which contains N (X) pixel points. The centroid of the left nostril region can be expressed as (x_i, y_i), where $x_i = (\sum_{x \in X} x)/N(X)$, $y_i = (\sum_{y \in X} y)/N(X)$. Similarly, the centroid of the right nostril region is (x_r, y_r).

If the lateral distance between the left and right nostril points is d, it is generally believed that the tip of the nose is located at $d/2$ above the middle perpendicular of the two nostrils. Let the midpoint of the left and right nostril points be (x_o, y_o), where $x_o = (x_i + x_r)/2$, $y_o = (y_i + y_r)/2$ and then $d = x_r - x_i$. The coordinates of the center of the nostril are (x_m, y_m), where $x_m = x_o$, $y_m = d/2$.

The integral projection method is also used for locating the center of the pupil. According to the gray distribution characteristics of the left and right eye images, the boundary of the binocular region is extracted, and the binocular region is demarcated separately to achieve independent pupil localization. Figure 4a,b are grayscale integrated projections in the vertical and horizontal directions, respectively. As the gray values of both the eyebrow and eye regions are lower than the gray values of their surrounding regions, it can be seen from Figure 4a that the first valley bottom corresponds to the eyebrow region, and the second valley bottom corresponds to the pupil region. As the gray value of the pupil is lower than the gray value of its surrounding area, it can be seen from Figure 4b that the first valley corresponds to the left pupil area, and the second valley corresponds to the right pupil area.

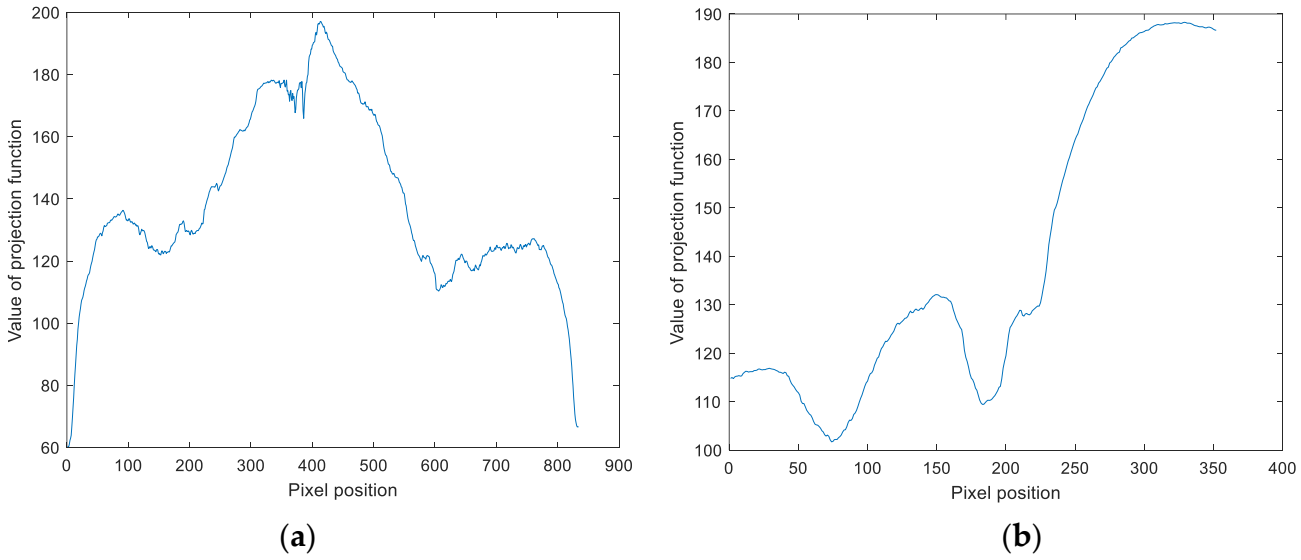
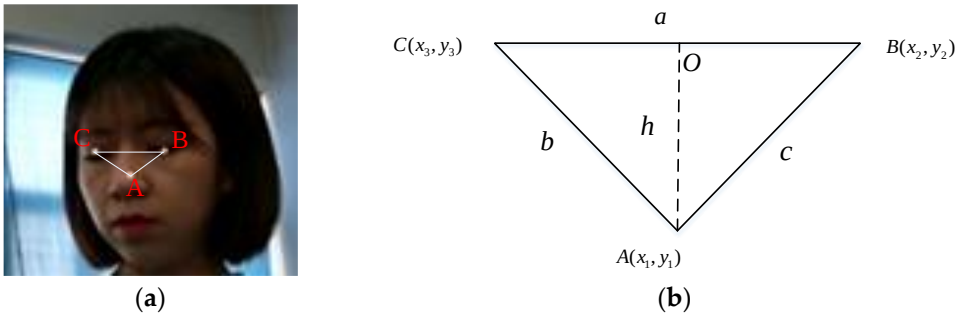


Figure 4. Gray integral projection in different directions: (**a**) vertical direction and (**b**) horizontal direction.

2.2. Geometric Triangular Head Pose Estimation Model

Using Kinect combined with geometric triangles to track the three degrees of freedom parameters of the head direction around the X, Y, and Z axes, the mathematical models are established, respectively. As shown in Figure 5, changes in the rotation angle in the state of head posture deflection such as pitch, yaw, and roll are converted into triangular deformations, which are represented by Pitch, Yaw, and Roll, respectively.



(**a**) (**b**)

Figure 5. Characteristic triangle model. (**a**) Face triangle model; (**b**) feature triangle.

During head position tracking, the geometric triangle is formed by the center of the pupil, and the tip of the nose changes with the doctor's head and eye movements and presents certain geometric characteristics. According to the geometric features of the face, when the head is facing the camera, the Euler angle is defined as 0, and the left and right eyes and nose form an isosceles triangle ΔABC. When the head is deflected, the triangle also undergoes a corresponding position shift as the geometric position of the feature points changes, resulting in deformation. By analyzing the geometric features of ΔABC, you can determine the current head posture.

According to the coordinates of three characteristic points, namely nose tip coordinates $A(x_1, y_1)$, left eye coordinates $B(x_2, y_2)$, and right eye coordinates $C(x_3, y_3)$, the three side length of ΔABC can be obtained as follows:

$$a = \sqrt{(x_3 - x_2)^2 + (y_3 - y_2)^2} \tag{3}$$

$$b = \sqrt{(x_3 - x_1)^2 + (y_3 - y_1)^2} \tag{4}$$

$$c = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2} \tag{5}$$

The three side length a, b, c correspond to $\angle A$, $\angle B$, and $\angle C$, and their angular sizes are expressed as follows:

$$\angle A = \pi - \angle B - \angle C \tag{6}$$

$$\angle B = \arccos(\frac{1}{2c}(\frac{1}{a}(c^2 - b^2) + a)) \tag{7}$$

$$\angle C = \arccos(\frac{1}{2b}(a - \frac{1}{a}(c^2 - b^2))) \tag{8}$$

The height corresponding to the side length a is expressed as follows:

$$h = \frac{2\sqrt{p(p-a)(p-b)(p-c)}}{a} \tag{9}$$

where, $p = \frac{a+b+c}{2}$. Based on the obtained pixel coordinates of the eyes and nose tip, a facial feature triangle is constructed, and a mathematical model of head posture is established

and derived. The doctor's head posture parameters are estimated based on the changes in the position of the triangle.

In order to accurately study the technology of doctors' head angle of view and orientation tracking, we establish triangular models for the rotation of the head around different coordinate axes to determine the head and angle of view posture.

(1) Head posture positioning relative to the X-axis (pitch)

The pitch angle is the solution corresponding to the rotation of the doctor's head about the X-axis. When the doctor raises his head or nods, the pitch angle changes to α, and for the frontal image in the triangle model, the side length a of the feature triangle changes, but the corresponding height h does not change. Based on prior knowledge of the distribution of facial organs, the positional relationship between the nose and the eyes can be obtained as follows:

$$h' = 0.8\mathrm{a} \tag{10}$$

where h' is the change in the height of the triangle when the doctor's head changes, and a is the distance between the eyes after turning the head.

$$a = \arccos\frac{2\sqrt{p(p-a)(p-b)(p-c)}}{0.8a^2} \tag{11}$$

When the pitch angle of the doctor's head changes, the angle between the two triangles corresponding to the change will also change. When the doctor raises his head, the corresponding angle α is a negative number; When the doctor lowers his head, the corresponding angle α is positive.

(2) Head posture positioning relative to the Y-axis (yaw)

The solution of the yaw angle corresponds to the rotation of the doctor's head around the Y-axis. When the doctor's head turns to the left or right, the pitch angle changes marked as Ψ, and the shape of the feature triangle changes.

If $y_2 > y_3$, the head is deflected to the left, and the deflection angle Ψ is expressed as follows:

$$\Psi = \arccos(\frac{y_2 - y_1}{c}) - \frac{1}{2}\angle C \tag{12}$$

If $y_2 < y_3$, the head is deflected to the right, and the deflection angle Ψ is expressed as follows:

$$\Psi = \arccos(\frac{y_3 - y_1}{c}) - \frac{1}{2}\angle C \tag{13}$$

When the doctor's head is tilted to the left, the corresponding angle Ψ is positive; When the doctor's head is tilted to the right, the corresponding angle Ψ is negative.

(3) Head posture positioning relative to the Z-axis (roll)

When the doctor turns his head left or right, the change of the roll angle is set to φ as follows:

$$\varphi = \arccos\frac{a}{a'} \tag{14}$$

where a is the distance between the pupils after turning the head, and a' is the distance between the pupils in the forward direction. During head turning analysis, the side length a of the feature triangle changes, but the corresponding height does not change.

$$a' = \frac{h}{0.8} \tag{15}$$

The solution for obtaining the final roll angle is obtained as follows:

$$\varphi = \arccos\frac{0.8a}{h} \tag{16}$$

When the head's roll angle changes, the corresponding change φ between the two triangles will also change. When the doctor turns left, the corresponding angle φ is negative; When the doctor turns his head to the right, the corresponding angle φ is a positive number. By modeling the feature triangles formed by the left and right pupil and nose tip, a head posture estimation model is established based on the geometric deformation of the head shape and the relative position of the facial feature points. The geometric model composed of multiple feature points overcomes the disadvantage of losing the location of a single feature point and converts the angle of head posture transformation into geometric surface deformation.

2.3. Traceless Kalman Filtering Tracking Combined with Triangles

Using the traditional classic Kalman filtering algorithm, we assume that the head and eye movements are uniform, thereby introducing linear Kaman filtering to complete head position tracking. However, under the interference of different factors, it is difficult to ensure a strictly linear system. Especially in the doctor's head-position-tracking system applied in this paper, the doctor's eye movements are frequent and complex and usually exhibit nonlinear motion. The assumption of linear motion is unreasonable, which results in the head deflection. The deviation of the eyes can lead to the loss of tracking, resulting in significant errors.

In order to improve the accuracy and robustness of head position tracking, we need to introduce nonlinear Kaman filtering for motion optimization. Currently, the commonly used nonlinear Kalman filtering algorithms include Extended Kalman Filter (EKF) and unscented Kalman filter. EKF assumes that the expected value of the result distribution is equal to the expected value of the map and considers only its local environment via linearization. The difference is that UKF [40] has a strong modeling uncertainty capability using unscented transformation (UT), only requires the first few moments of random parameters, and is easy to implement, which can overcome the shortcomings of local linearization of EKF to estimate expected values and distributions more accurately. For linear motion systems, the tracking and prediction effects using EKF and UKF are not significantly different from those using linear KF; However, for nonlinear systems, UKF has better performance than EKF, and as the degree of nonlinearity increases, the gap between the two algorithms will become increasingly large [41]. This paper considers using UKF algorithm to study a head-position-tracking system when doctors' heads undergo significant head deflection when transferring surgical forceps to nurses and adjusting the position of projectors.

Theoretically, UKF uses traceless transformation to process nonlinear system models, uses system state expansion to obtain sampling points and perform nonlinear transformation, thereby obtaining the probability distribution and corresponding statistical characteristics of random variables, and then combines them with the standard Kalman filtering framework for calculation. The first frame image in the video is collected via Kinect to obtain the iterative initial value of the observation state, and then UKF is used to track the trajectory of the doctor's head and eyes. The proposed algorithm completes the modeling process of three-dimensional coordinates and three rotation angles using the UKF algorithm for the centroid of the feature triangle. The specific process of the UKF algorithm is as follows.

(1) Initialization.

$$\overline{x}_0 = E[x_0] \tag{17}$$

$$P_0 = E[(x_0 - \overline{x}_0)(x_0 - \overline{x}_0)^T] \tag{18}$$

The above formula is to substitute a non-linear function with a sigma point to obtain the corresponding non-linear function values, and then calculate the transformed mean $\overline{x}_0$ and covariance P_0 from these point sets.

(2) Select $2n + 1$ sigma points x_k in set form.

$$\varepsilon_{k-1} = [\overline{x}_{k-1|k-1}\overline{x}_{k-1|k-1} \pm (\sqrt{(n+1)P_{k-1|k-1}})] \tag{19}$$

Since the perspective-tracking model has $n = 6$ states in this paper, the number of sampling points is set to $2 \times n + 1 = 13$.

(3) Time update.

$$\varepsilon_{k|k-1} = f(\varepsilon_{k-1}) \tag{20}$$

$$\hat{x}_{k|k-1} = \sum_{i=0}^{2n} W_{i,e}\varepsilon_{i,k|k-1} \tag{21}$$

$$P_{k|k-1} = \sum_{i=0}^{2n} W_{i,c}(\varepsilon_{i,k|k-1} - \hat{x}_{k|k-1})(\varepsilon_{i,k|k-1} - \hat{x}_{k|k-1})^T \tag{22}$$

$W_{i,e}$ and $W_{i,c}$ represent the mean and covariance of sigma points.

(4) Measurement update.

$$\widehat{Z}_{k|k-1} = h(\varepsilon_{k|k-1}) \tag{23}$$

$$\widehat{Z}_{k|k-1} = \sum_{i=0}^{2n} W_{i,e}Z_{i,k|k-1} \tag{24}$$

(5) Estimate the state and covariance.

$$P_{Z_KZ_K} = \sum_{i=0}^{2n} W_{i,c}(Z_{i,k|k-1} - \widehat{Z}_k)(Z_{i,k|k-1} - \widehat{Z}_k)^T \tag{25}$$

The above formula represents the point set $Z_{k|k-1}$ obtained by substituting sigma points into the nonlinear function in the UT transformation.

$$P_{\varepsilon_KZ_K} = \sum_{i=0}^{2n} W_{i,c}(\varepsilon_{i,k|k-1} - \hat{x}_k)(Z_{k|k-1} - \widehat{Z}_k)^T \tag{26}$$

$$K_k = P_{\varepsilon_{k|k-1}Z_{k|k-1}}P^{-1}_{Z_{k|k-1}Z_{k|k-1}} \tag{27}$$

Formulas (26) and (27) represent the mean and covariance of the weighted nonlinear function $Z_{k|k-1}$ in the UT transform, respectively. $W_{i,e}$ and $W_{i,c}$ are specifically defined as follows:

$$\begin{cases} W_{0,\mathrm{e}} = \frac{l}{n+l} \\ W_{0,c} = \frac{l}{n+l} + (1 - \alpha^2 + \beta) \\ W_{i,e} = W_{i,c} = \frac{l}{2(n+l)}, i = 1, 2, \ldots, 2n \end{cases} \tag{28}$$

where $l = \alpha^2(n + m) - n$. m is the scale factor, and l indicates the scaling parameter. If x is a multidimensional variable, the scale factor is taken as $m = 3 - n$; α is a positive number, and β is a non-negative weight coefficient.

$$\hat{x}_k = \hat{x}_{k|k-1} + K_k[Z_k - \widehat{Z}_{k|k-1}] \tag{29}$$

$$P_k = P_{k|k-1} - K_kP_{Z_KZ_K}K_k^T \tag{30}$$

Repeat the above steps to achieve the calculation of the unscented Kalman filtering algorithm.

The proposed algorithm in this paper aims to obtain the optimal estimation of three-dimensional coordinates and angles in the perspective-tracking environment. Therefore, an unscented Kalman filtering perspective-tracking model combining feature triangles is proposed. An overall model is established for the characteristic triangle centroid composed of three characteristic numbers, namely, the center of the left and right pupil and the tip of the nose. Similar to linear Kalman filtering, UKF also includes three stages: state prediction, prediction of measured values, and state update. The main difference is that for the prediction stage, UT transform is used to approximate the probability distribution of head and eye movement states, making traceless Kalman filtering suitable for nonlinear doctor's head-position-tracking systems.

In this paper, the following nonlinear equations are used to simulate visual angle motion, namely, the following:

$$x = x_0 + vt + \frac{1}{2}at^2 \tag{31}$$

$$x'_{k+1} = v_0 + A_k \sin(w_k t) \tag{32}$$

$$a_{k+1} = x''_{k+1} = A_k w_k \cos(w_k t) \tag{33}$$

where the initial values of x_0 and v_0 are both 0, the acceleration a conforms to a sinusoidal distribution function, and process noise v_k is taken into account. $A_k = 0.08$ m/s and $w_k = \pi rad/s$. In the eye movement state, Newton's motion formula is used, and the results are shown in the following state matrix:

$$F = \begin{bmatrix} 1 & 0 & \Delta t & 0 \\ 0 & 1 & 0 & \Delta t \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \tag{34}$$

$$H = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \tag{35}$$

where Δt is the interval from k to $k + 1$, and F and H are the state transition matrices.

Compared to the previously assumed perspective-tracking system with uniform motion, this section uses a nonlinear model that changes over time and uses a nine-dimensional vector space $[p_x, p_y, p_z, v_x, v_y, v_z, \omega_x, \omega_y, \omega_z]^T$ as the state variable of the process equation, where $[p_x, p_y, p_z]^T$ is the three-dimensional coordinate of the facial feature point or feature triangle, $[v_x, v_y, v_z]^T$ is the component of the linear velocity of the facial feature point on the X, Y, and Z coordinate axes, and $[\omega_x, \omega_y, \omega_z]^T$ is the component of the rotational angular velocity on the X, Y, and Z coordinate axes. This article focuses on the study of coordinates and rotation angles. Therefore, the state variable of the process equation of a UKF-based visual tracking nonlinear motion system is $k[p_x, p_y, p_z, \omega_x, \omega_y, \omega_z]^T$, that is, $[X_W, Y_W, Z_W, \alpha, \Psi, \varphi]^T$. The form of the state equation and measurement equation is expressed as follows:

$$x_{k+1} = Fx_k + v_k \tag{36}$$

$$y_{k+1} = Hx_k + n_k \tag{37}$$

where x_k represents the state vector of the system time k, which is the input vector of a known time k, that is, $x_k = [X_W, Y_W, Z_W, \alpha, \Psi, \varphi]^T$.$y_k$ is the measurement vector, v_k is the system process noise, and the measurement noise matrix is represented by n_k, which are all zero mean noise. F represents the transfer matrix, and H represents the observation matrix.

As shown in Equations (36) and (37), Q is the noise covariance matrix of the transfer process, and R is the noise covariance matrix of the measurement process.

$$Q = E\{w_k \, w_k^T\} \tag{38}$$

$$R = E\{v_k \, v_k^T\} \tag{39}$$

The advantages of the proposed algorithm are summarized as follows:

(1) Real-time performance: In the process of perspective tracking, the video stream collected by Kinect camera is 30 fps, and the detection time of the whole system starts from the incoming of each frame image until the output interface displays the perspective pose information. According to the experiment, the average processing time of a single frame is 27.66 ms, which meets the real-time head pose estimation of doctors at a speed of 30 fps.

(2) Robustness: whether under dark light, strong light, or face-blocking environment, the feature triangle can be positioned more accurately, its position and pose information can be obtained, and the corresponding rotation angle changes are consistent with the pre-results.

(3) Accuracy: When forecasting and tracking the same dataset, the root-mean-square error of the improved UKF algorithm combined with feature triangles in this paper is less than that of the KF algorithm of single feature points, which is because when the head is greatly deflection and the face is blocked, the linear KF algorithm will lead to the loss of feature point tracking and generate a large error. The improved UKF algorithm combined with feature triangle overcomes the shortcomings of linear KF, has strong anti-noise ability, and is suitable for nonlinear perspective-tracking models.

Among the evaluation criteria for target-tracking algorithms, a high-performance target tracker mainly includes Multiple Object Tracking Accuracy (MOTA), Multiple Object Tracking Precision (MOTP), and Root Mean Square Error (RMSE) [42].

MOTA is used as the evaluation standard for target-tracking accuracy, which is defined as follows:

$$\text{MOTA} = 1 - \frac{FN + FP + IDSW}{GT} \in (-\infty, 1] \tag{40}$$

FN is the total number of missed reports in the whole video; FP is the total number of false positives in the whole video; IDSW is the number of times that are not tracked correctly in frame t, and GT is the sum of the number of the entire video. The closer MOTA is to 1, the better the performance of the filter tracker is. Moreover, MOTA can provide an intuitive measurement of the performance of the head-tracking algorithm independent of the accuracy of target detection.

MOTP is another measure of tracking error, which describes the ability to accurately locate the target. Its calculation formula is as follows:

$$\text{MOTP} = \frac{\sum_{t,i} d_{t,i}}{\sum_t c_t} \tag{41}$$

$d_{t,i}$ represents the distance between the predicted pose and the real pose of the feature points of the detected target i in frame t, and the larger the value, the stronger the accurate positioning ability of the target. c_t indicates the number of successful traces in the current frame. The closer the MOTP is to 1, the higher the accuracy of the nonlinear Kalman filter is.

In the field of target tracking, RMSE is used to evaluate tracking performance as follows:

$$\text{RMSE} = \sqrt{\frac{1}{N}\sum_{\text{i}=1}^{N}(x_i - \hat{x}_i)^2} \tag{42}$$

The closer the RMSE is to 0, the better the prediction effect of the tracking point is, and the better the running effect of the algorithm is.

3. Results

3.1. Dataset Description

In order to verify the applicability of the proposed methods from multiple perspectives, this section conducts experiments and analyses on the public dataset Biwi Kinect Head Pose Data [43] dataset, which uses Kinect to collect nearly 15,000 images, including depth images and RGB images of 20 experimental objects with heads rotating in different directions, as shown in Figure 6. The real dataset collects facial images of eight people (four boys and four girls) using Kinect v2 sensor, each with 100 images of different poses.

Figure 6. Samples of public and real datasets for tracking. (**a**) Samples of public dataset; (**b**) samples of real dataset.

Since the dataset in this paper is a video of different faces, the current state of all frames needs to be listed to show all changes in each dataset, which cannot be shown in this current paper. This paper only chooses some frames as representatives for description.

3.2. Experiments Comparison and Analysis

In order to test the effectiveness and performance of the proposed algorithm in tracking human eyes, this paper tested the tracking performance of the geometric triangle combined with an unscented Kalman filtering algorithm composed of multiple feature points and the single feature point combined with linear Kalman filtering algorithm in video sequences captured by Kinect cameras.

As shown in Figure 7, the visualization effect of tracking arbitrary frame images using single feature points combined with linear Kalman filtering is shown. Real dataset1 and real dataset2 in Figures 7 and 8 represent the true values obtained from different faces selected in the laboratory, and each true value in this paper was obtained by measuring the

state of each frame via Kinect. In Figure 7, the blue color represents the actual measurement value obtained by Kinect, and the red color represents the Kalman filter prediction value. The measured value (represented in blue font) and Kalman prediction value (represented in black font) are output in real time. The serial number represents the number of human faces in the selected dataset, respectively.

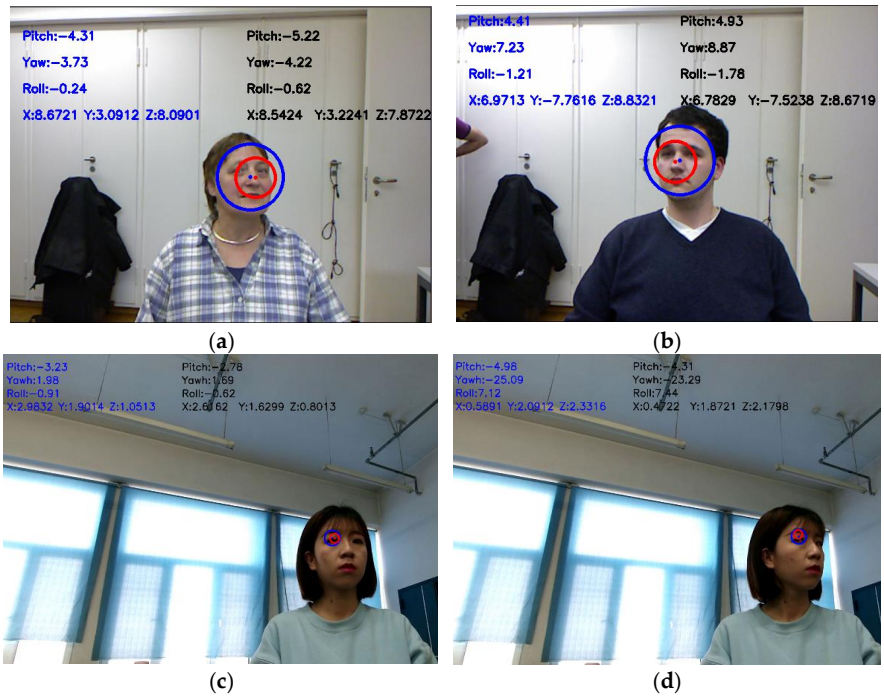


Figure 7. Single feature point tracking using Kalman filter. (**a**) Turn right (public dataset 11), (**b**) forward direction (public dataset 13), (**c**) forward direction (real dataset 1), and (**d**) turn right (real dataset 1).

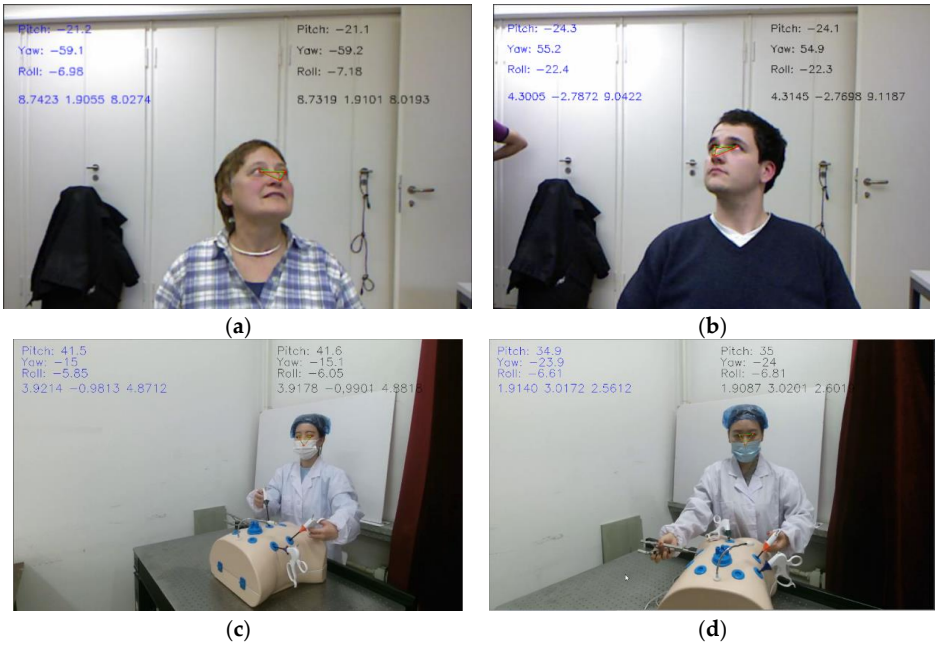


Figure 8. Feature triangle + UKF attitude estimation. (**a**) Right deflection (public dataset 1), (**b**) left deflection (public dataset 13), (**c**) left deflection (real dataset 2), and (**d**) forward direction (real dataset 2).

As seen from Figure 7, using linear Kalman filter single point tracking can better obtain posture for facial and head postures with slight deflection without wearing a mask. The predicted value and the tracking point of the measured value have a good fitting effect, and the measured value and predicted value of the current frame are output in real-time.

The above experiments assume that the head motion is a linear model, with no significant deflection of head and eye movements, and no significant occlusion of the face. However, this model has the following drawbacks: (1) The method based on a single feature point is susceptible to noise interference and has poor robustness; (2) Linear Kalman filtering cannot meet the nonlinear motion of large head deflection. Therefore, this paper proposes an improved nonlinear Kalman filter combined with geometric triangles.

In order to verify the feasibility and implementation effect of a geometric triangle combined with an unscented Kalman filter in a doctor's head-position-tracking system, this study determines the initial position of the head posture and sets the parameters of the unscented Kalman filter, and the simulation results are shown in Figure 8. The green triangle represents the measured value of the head posture, and the red triangle represents the tracking prediction value of UKF. The blue text on the left represents real-time measurement values, and the black text on the right represents real-time Kalman tracking prediction values. From the numerical results, it can be seen that the UKF algorithm can accurately estimate the posture of the next frame regardless of head deflection or face occlusion.

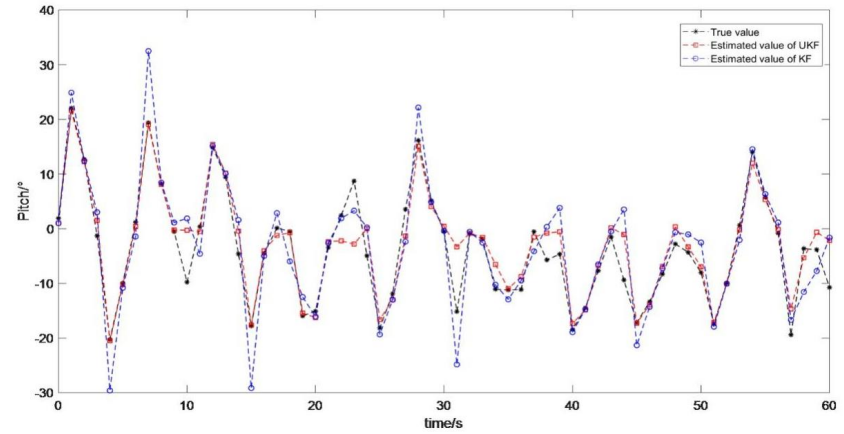
From Figure 8, it can be seen that the improved feature triangle-based unscented Kalman perspective-tracking environment in this article can overcome the limitations of traditional Kalman filtering and is more suitable for head tracking from a doctor's perspective for estimating the position and posture of a face with a large head deflection, as shown in Figure 8a,b. Both the Biwi public dataset and the real dataset collected by the surgical simulation platform can accurately track posture information.

Figure 9 shows the comparison between the rotation angles Pitch, Yaw, and Roll around the X, Y, and Z axes, respectively, and the true values after the traditional KF algorithm and the improved UKF tracking correction combining feature triangles in this article. As seen from the figure, the deviation between the angle information obtained by UKF is represented by the red line, and the actual value is small, while the deviation of the KF algorithm is represented by the blue line is large. The results show that the UKF algorithm combined with feature triangles can effectively reduce noise in non-linear head movements and has better tracking performance than traditional KF algorithms with single feature points in environments with large head deflections or occluding faces. The results in Figure 9 show that the UKF algorithm combined with feature triangles can effectively reduce noise in non-linear head movements and has better tracking performance than traditional KF algorithms with single feature points in environments with large head deflection or occlusion of the face.

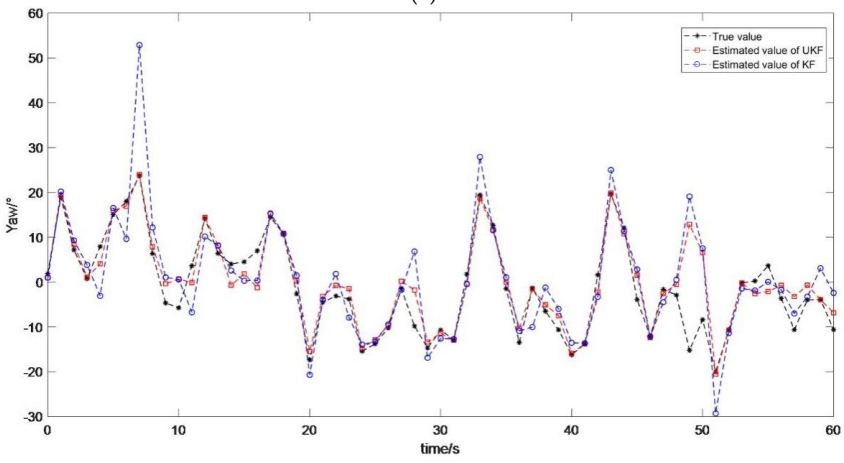
In order to more clearly compare the error between the predicted value and the measured value after using different algorithms, the center of gravity of the linear Kalman filter tracking feature triangle, the center of gravity of the UKF tracking feature triangle, and the actual value obtained by Kinect are extracted, thereby outputting three 3D trajectory data in the world coordinate system. Figure 10 gives three tracking trajectories of Kinect true values and estimated KF and UKF values under conditions of small unobstructed head deflection, large unobstructed head deflection, and occluded head deflection.

As seen from Figure 10, when the head posture movement is slow and approximately linear, both filtering methods can achieve tracking and filtering of measured data. It is shown that the classical KF algorithm and the improved feature triangle combined with the UKF algorithm in this paper have similar tracking effects in linear systems with approximate uniform motion, and both algorithms have good fitting effects with real trajectories. Moreover, linear Kalman filtering is more suitable for linear systems due to its simpler structure and lower algorithm complexity. However, when there is a significant deflection or occlusion of the head posture in nonlinear motion, the predicted value based

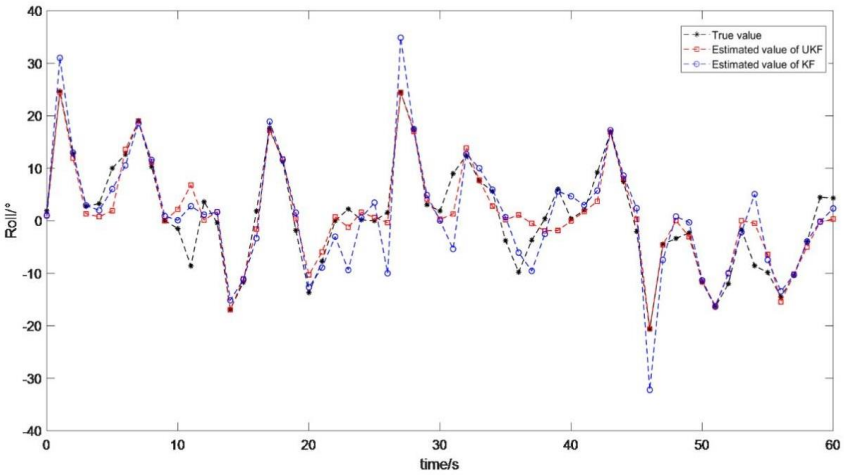
on the KF algorithm will have a significant error from the actual value, while the UKF algorithm still has a good tracking effect. This is because the UKF algorithm approximates the posterior mean and variance of the nonlinear system via a traceless transformation and still maintains good tracking performance in nonlinear motion systems with head deflection and occlusion.



(**a**)



(**b**)



(**c**)

Figure 9. Comparison chart of rotation angle tracking of different algorithms: (**a**) Pitch, (**b**) Yaw, and (**c**) Roll.

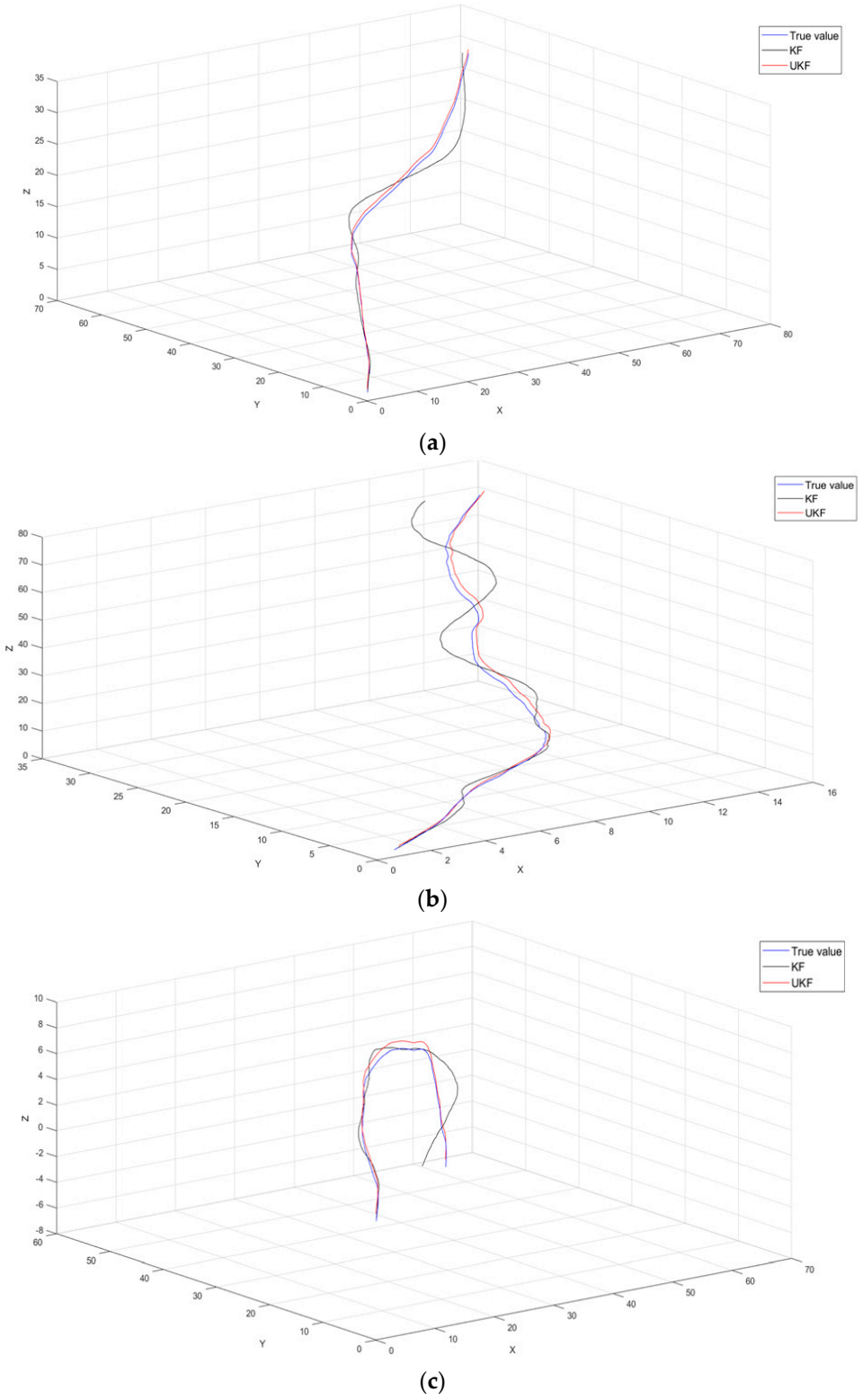


Figure 10. Comparison chart of 3D trajectory prediction in different attitudes: (**a**) Head's small rotation without occlusion, (**b**) Head's large rotation without occlusion, and (**c**) Head's rotation with occlusion.

4. Discussion

Although eye tracking using linear Kalman filtering in the past has made interference judgments and predictions, the tracking results still cannot meet expectations when the

target is severely occluded or occluded for a long time, and there may still be tracking failures. Moreover, in the actual tracking process, there are often situations where multiple interference factors occur simultaneously. Compared to single-factor interference, multiple-factor interference (occlusion, lighting, deflection, etc.) has a more severe impact on target tracking and can even directly lead to tracking failure. In contrast to the nonlinear unscented Kalman filter used in this paper, the probability density distribution function similar to the nonlinear function is used, a large number of determined sample data is used to approach the posterior probability density function of the state variable, and the derivative is not calculated by the Jacobian matrix, which reduces the amount of calculation. Additionally, the UKF algorithm combined with feature triangles can effectively reduce noise and has better tracking performance than traditional KF algorithms with single feature points in environments with large head deflection or occluded faces. The UKF algorithm improved in this article, which combines the feature triangle method compared to the traditional single point linear Kalman filter for target tracking and significantly improves the perspective tracking ability.

In order to test the performance of the algorithm in this paper, the following two sets of experiments were carried out using different algorithms on the same dataset, respectively: comparing the tracking performance of the feature triangle center of gravity combined with the linear KF algorithm and the nonlinear UKF algorithm; The single feature point UKF algorithm for the left pupil, right pupil, and nose tip is averaged based on the results and compared experimentally with the center of gravity of the feature triangle formed by multiple feature points. The comparison results are shown in Tables 1 and 2.

Table 1. Comparison of tracking accuracy between KF and UKF algorithms.

Methods	MOTA (%)	MOTP (%)
KF + The center of gravity of characteristic triangle	76.2	80.9
UKF + The center of gravity of characteristic triangle	81.7	84.9

Table 2. Comparison of UKF combined with single and multiple feature points tracking.

Methods	MOTA (%)	MOTP (%)
Pupil of the left eye	78.9	83.8
Pupil of the right eye	78.4	83.1
Nose	77.9	82.5
Average of three feature points	78.4	83.1
The center of gravity of characteristic triangle	81.7	84.9

By analyzing Table 1, it can be seen that the feature triangle combined with the UKF tracking algorithm used in this article approximates the posterior mean and variance of nonlinear systems via traceless transformation. Compared to the KF algorithm, it improves the MOTA index by 5.5% and MOTP by 4%. Therefore, the UKF algorithm combined with feature triangles performs better in head posture estimation.

According to Table 2, using the UKF algorithm for single-point tracking of the left pupil, right pupil, and nasal tip, and averaging the three tracking results, the MOTA and MOTP for target tracking are 78.4% and 83.1%, respectively. Using the UKF algorithm to track the center of gravity of the feature triangle improves the MOTA to 81.7% and MOTP to 84.9%. Therefore, the improved UKF algorithm combined with the feature triangle method in this paper has a higher tracking effect and accuracy for complex occlusion environments compared to the traditional linear Kalman filter tracking.

Table 3 gives the root mean square error of the KF algorithm combined with triangle tracking and the UKF algorithm combined with triangle tracking. As seen from Table 3, when predicting and tracking the same dataset, the root mean square error of the improved

UKF algorithm combined with feature triangles in this article is smaller than that of the KF algorithm with a single feature point. This is because when the head is deflected significantly and the face is occluded, the linear KF algorithm will cause feature point tracking loss and generate large errors. The improved UKF algorithm combined with feature triangles overcomes the shortcomings of linear KF and has strong noise resistance, which is suitable for nonlinear head-position-tracking models.

Table 3. Error comparison of different tracking algorithms.

Results	KF + Triangle	UKF + Triangle
X/cm	1.53	1.39
Y/cm	1.76	1.44
Z/cm	1.39	1.17
Pitch/°	1.67	1.24
Yaw/°	1.38	1.12
Roll/°	1.57	1.16

The main comparison in this article is between the true values provided in the public dataset and the predicted values obtained by the algorithm in this article, as well as the measured and predicted values obtained from the experimental dataset via camera calibration, model establishment, parameter conversion, and other work. KF and UKF tracking comparisons are conducted on the above data, and the issue of inaccurate feature point localization caused by the head deflection of doctors wearing masks is addressed, which makes traditional linear Kalman filtering algorithms unable to accurately track feature points. The improved feature triangle and UKF tracking algorithm in this paper approximate the posterior mean and variance of the nonlinear system via traceless transformation to update the time and measurement of the doctor's head state.

However, the perspective tracking mentioned in previous literature mostly focuses on the driver's head posture, which is different from the application environment in this article. The improved feature triangle-based unscented Kalman perspective-tracking environment in this article can overcome the limitations of traditional Kalman filtering for large head deflection and pose estimation of masked faces, making it more suitable for head tracking from a doctor's perspective. Accurate tracking of pose information can be achieved on both the Biwi public dataset, and the dataset collected on the surgical simulation platform. Thus, the effectiveness of the algorithm proposed in this paper was verified from two aspects: the actual test dataset and the public test dataset.

5. Conclusions

Aiming at the problem that the head deflection of doctors wearing masks can easily lead to the loss of feature point positioning, and the linear Kalman filtering algorithm cannot accurately track feature points. This paper implements an improved perspective-tracking model that combines feature triangles and unscented Kalman filtering algorithm. From the real-time output data, it can be seen that the maximum angular error between the predicted value and the measured value of the improved unscented Kalman filter is 0.3°, and the maximum coordinate error is 1.74 cm, which meets the experimental requirements. Compared to the linear Kalman filtering algorithm, the tracking accuracy of the geometric triangle-based unscented Kalman filtering algorithm in this paper has been improved by 3.3%, and the accuracy has been improved by 4%. Compared to tracking the average of three feature points, the algorithm in this paper improves the accuracy of target tracking by 5.5% and accuracy by 1.8%, which can meet the needs of doctors for surgery.

Author Contributions: Conceptualization, X.Y., Y.Z., H.W. and A.W.; methodology, X.Y. and Y.Z.; software, Y.Z.; validation X.Y. and Y.Z.; writing—review and editing X.Y., Y.Z., H.W. and A.W. All authors have read and agreed to the published version of the manuscript.

Funding: This work was funded by the high-end foreign experts introduction program (G2022012010L) and Major Science and Technology Projects of Zhongshan City in 2022 (2022A1020).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: https://data.vision.ee.ethz.ch/cvl/gfanelli/head_pose/head_forest.html#db (accessed on 24 March 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Klaib, A.F.; Alsrehin, N.O.; Melhem, W.Y.; Bashtawi, H.O.; Magableh, A.A. Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and Internet of Things technologies. *Expert Syst. Appl.* **2021**, *166*, 114037. [CrossRef]
2. Liu, M.; Li, H.; Dai, H. Appearance-based Gaze Estimation Using Multi-task Neural Network. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *806*, 012054. [CrossRef]
3. Marquard, J.L.; Henneman, P.L.; He, Z.; Jo, J.; Fisher, D.L.; Henneman, E.A. Nurses' behaviors and visual scanning patterns may reduce patient identification errors. *J. Exp. Psychol. Appl.* **2011**, *17*, 247. [CrossRef] [PubMed]
4. MacNeil, R.R.; Gunawardane, P.D.S.H.; Dunkle, J.; Zhao, L.; Chiao, M.; de Silva, C.W.; Enns, J.T. Using electrooculography to track closed-eye movements. *J. Vis.* **2021**, *21*, 1898. [CrossRef]
5. Katona, J. Clean and dirty code comprehension by eye-tracking based evaluation using GP3 eye tracker. *Acta Polytech. Hung.* **2021**, *18*, 79–99. [CrossRef]
6. Fabio, A.; Giannatiempo, S.; Semino, M. Longitudinal cognitive rehabilitation applied with eye-tracker for patients with Rett Syndrome. *Res. Dev. Disabil.* **2021**, *111*, 103891. [CrossRef]
7. Chugh, S.; Brousseau, B.; Rose, J.; Eizenman, M. Detection and Correspondence Matching of Corneal Reflections for Eye Tracking Using Deep Learning. In Proceedings of the 2020 25th International Conference on Pattern Recognition (ICPR), Milan, Italy, 10–15 January 2021.
8. Pauszek Joseph, R. An introduction to eye tracking in human factors healthcare research and medical device testing. *Hum. Factors Healthc.* **2023**, *3*, 100031. [CrossRef]
9. Yu, B.; Zhang, X.; Zhang, Y.; Wang, L.; Chen, X.; Lin, X. The Application of Sensors in Medical Fields. *China Med. Device Inf.* **2017**, *23*, 17–18.
10. Li, Z. Indoor human body localization method based on IMU and infrared sensor network. *Transducer Microsyst. Technol.* **2018**, *37*, 24–27.
11. Atallah, L.; Lo, B.; King, R.; Yang, G.Z. Sensor positioning for activity recognition using wearable accelerometers. *IEEE Trans. Biomed. Circuits Syst.* **2011**, *5*, 23–33. [CrossRef]
12. Gabela, J.; Kealy, A.; Li, S.; Hedley, M.; Moran, W.; Ni, W.; Williams, S. The Effect of Linear Approximation and Gaussian Noise Assumption in Multi-Sensor Positioning Through Experimental Evaluation. *IEEE Sens. J.* **2019**, *19*, 10719–10727. [CrossRef]
13. Zhang, Z. Microsoft Kinect Sensor and Its Effect. *IEEE MultiMedia* **2012**, *19*, 4–10. [CrossRef]
14. Smisek, J.; Jancosek, M.; Pajdla, T. 3D with Kinect. In Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops) Computer Vision Workshops (ICCV Workshops), Barcelona, Spain, 6–13 November 2011; pp. 1154–1160.
15. Khoshelh, K. Accuracy Analysis of Kinect Depth Data. *Int. Arch. Photogramm. Remote Sens. Spat. Inf. Sci.* **2012**, *XXXVIII–5/W12*, 133–138. [CrossRef]
16. Zhang, R.H.; Walshe, C.; Liu, Z.; Guan, L.; Muller, K.; Whritner, J.; Zhang, L.; Hayhoe, M.; Ballard, D. Atari-head: Atari human eye-tracking and demonstration dataset. In Proceedings of the 34th AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 6811–6820.
17. Li, Z.; Zheng, X.; Liu, K. Research Progress in Mobile Device Sight Tracking Technology. *Comput. Eng. Appl.* **2018**, *54*, 6–11.
18. Zhang, K.; Zhao, X.; Ma, Z.; Man, Y. A simplified 3d gaze tracking technology with stereo vision. In Proceedings of the 2010 International Conference on Optoelectronics and Image Processing, Hainan China, 11–12 November 2011; pp. 131–134.
19. Li, S.; Liu, Z.; Sun, M.T. Real time gaze estimation with a consumer depth camera. *Inf. Sci.* **2015**, *320*, 346–360.
20. Zanetti, M.; De Cecco, M.; Fornaser, A.; Leuci, M.; Conci, N. The Use of INTER-EYE for 3D Eye-Tracking Systematic Error Compensation. In Proceedings of the International Symposium ELMAR-2016, Zadar, Croatia, 12–14 September 2016; pp. 173–176.
21. Elmadjian, C.; Shukla, P.; Tula, A.D.; Morimoto, C.H. 3D gaze estimation in the scene volume with a head-mounted eye tracker. In Proceedings of the Workshop on Communication by Gaze Interaction, Warsaw, Poland, 15 June 2018; pp. 1–9.
22. González-Ortega, D.; Díaz-Pernas, F.J.; Martínez-Zarzuela, M.; Antón-Rodríguez, M. Comparative analysis of Kinect-based and oculus-based gaze region estimation methods in a driving simulator. *Sensors* **2020**, *21*, 26. [CrossRef]
23. Han, E. Integrating mobile eye-tracking and VSLAM for recording spatial gaze in works of art and architecture. *Technol. Archit. Des.* **2021**, *5*, 177–187. [CrossRef]

24. Liu, J.; Chi, J.; Hu, W.; Wang, Z. 3D Model-Based Gaze Tracking via Iris Features with a Single Camera and a Single Light Source. *IEEE Trans. Hum.-Mach. Syst.* **2021**, *51*, 75–86. [CrossRef]
25. Chen, Q. Research of Attention Analysis System Based on Gaze Estimation Neural Networks. Ph.D. Thesis, University of Electronic Science and Technology of China, Chengdu, China, 2022; pp. 29–35.
26. Chi, J.N.; Xie, L.H.; Zhang, P.Y.; Lu, Y.F.; Zhang, G.S. Hybrid Particle and Kalman Filtering for Pupil Tracking in Active IR Illumination Gaze Tracking System. *Math. Probl. Eng.* **2014**, *2014*, 426234. [CrossRef]
27. Janabi-Sharifi, F.; Marey, M. A kalman-filter-based method for pose estimation in visual servoing. *IEEE Trans. Robot.* **2010**, *26*, 939–947. [CrossRef]
28. Sultan, M.S.; Chen, X.; Ma, G.; Xue, J.; Ni, W.; Zhang, T.; Zhang, W. Hand-eye 3D pose estimation for a drawing robot. In Proceedings of the 2013 IEEE International Conference on Mechatronics & Automation, Takamatsu, Japan, 4–7 August 2013.
29. Munir, F.; Jalil, A.; Jeon, M. Real time eye tracking using Kalman extended spatiotemporal context learning. In Proceedings of the Second International Workshop on Pattern Recognition, SPIE, Singapore, 1–3 May 2017; Volume 10443, pp. 254–258.
30. Pan, Z.; Liu, R.; Zhang, M. *Human Eye Tracking Based on CNN and Kalman Filtering*; Transactions on Edutainment XV; Springer: Berlin/Heidelberg, Germany, 2019; pp. 265–273.
31. Vaishnavi, B.; Rao, S.K.; Jahan, K. Underwater bearings-only tracking using particle filter. *Int. J. Innov. Technol. Explor. Eng.* **2019**, *8*, 451–455.
32. Bagherzadeh, S.; Toosizadeh, S. Eye tracking algorithm based on multi model Kalman filter. *HighTech Innov. J.* **2022**, *3*, 15–27. [CrossRef]
33. Feng, Z.; Yang, B.; Li, Y.; Wang, Z.; Zheng, Y. Research on Human Hand Tracking Aiming at Improving Its Accurateness. *J. Comput. Res. Dev.* **2008**, *45*, 1239–1248.
34. Zhang, Z.; Zhang, J. A new real-time eye tracking based on nonlinear unscented Kalman filter for monitoring driver fatigue. *J. Control. Theory Appl.* **2010**, *8*, 181–188. [CrossRef]
35. Hannuksela, J. *Facial Feature Based Head Tracking and Pose Estimation*; Department of Electrical & Information Engineering, University of Oulu: Oulu, Finland, 2003.
36. Bankar, R.; Salankar, S. Improvement of Head Gesture Recognition Using Camshift Based Face Tracking with UKF. In Proceedings of the 2019 9th International Conference on Emerging Trends in Engineering and Technology—Signal and Information Processing (ICETET-SIP-19), Nagpur, India, 1–2 November 2019.
37. Du, X.; Chen, D.; Liu, H.; Ma, Z.; Yang, Q. Real-time hand tracking based on YOLOv4 model and Kalman filter. *J. China Univ. Posts Telecommun.* **2021**, *28*, 86.
38. Tian, L.; Xu, Y.; Xue, W.; Cheng, L. Consistent Extended Kalman Filter Design for Maneuvering Target Tracking and Its Application on Hand Position Tracking. *Guid. Navig. Control.* **2022**, *2*, 26. [CrossRef]
39. Li, P.; Liang, L.; Hui, D.; Wang, G. Head Pose Estimation of Patients with Monocular Vision for Surgery Robot Based on Deep Learning. *Chin. J. Biomed. Eng.* **2022**, *41*, 537–546.
40. Li, H.; Chutatape, O. Boundary detection of optic disk by a modified ASM method. *Pattern Recognit.* **2003**, *36*, 2093–2104. [CrossRef]
41. Putriany, D.; Rachmawati, E.; Sthevanie, F. Indonesian ethnicity recognition based on face image using gray level co-occurrence matrix and color histogram. *IOP Conf. Ser. Mater. Sci. Eng.* **2021**, *1077*, 012040. [CrossRef]
42. Singh, U.K.; Singh, A.K.; Bhatia, V.; Mishra, A.K. EKF-and UKF-based estimators for radar system. *Front. Signal Process.* **2021**, *1*, 704382. [CrossRef]
43. Papakon, G.; Amir, M.; Warn, G. A scaled spherical simplex filter (S3F) with a decreased n + 2 sigma points set size and equivalent 2n + 1 Unscented Kalman Filter (UKF) accuracy. *Mech. Syst. Signal Process.* **2022**, *163*, 107433.