Head Tracking System Implementation Using a Depth Camera

YANG-KEUN AHN, KWANG-MO JUNG
Realistic Media Platform Research Center
Korea Electronics Technology Institute
121-835, 8th Floor, #1599, Sangam-Dong, Mapo-Gu, Seoul
REPUBLIC OF KOREA
ykahn@keti.re.kr

Abstract: - In this paper, a method to track the head of a user regardless of the number of users using a depth camera is proposed. The proposed method tracks the user head using only the depth information without considering color information. The tracking is performed through experimental data where the depth image differs for each user. Also, the proposed method has the advantage of being able to track the user head regardless of the camera type. Kinect for Windows by Microsoft and DS311 of SoftKinetic were used for the experiments in this study.

Key-Words: - Head Tracking, Head Tracking System, Head Determination, Highest Point Search, Template matching, Square Head Area Operation, Depth Camera

1 Introduction

Recently the demand for user contents has transformed from various content provided by companies to content tailored to the users. Accordingly, user recognition within the space where the content is provided has become a basic problem, and image data are being utilized in order to solve this problem.

The conventional methods to track the user’s head include recognizing the face of the user using color information and determining the area of skin by using color information. Such methods, however, are limited by the number of users that can be recognized, the lighting, and the distance. Among the conventional methods, the method of tracking the user head by using the depth information has a limitation regarding the number of users that can be recognized.

In this paper, a method to track the user head by using the depth information and data acquired through experiments is proposed. The proposed method only uses depth information and hence is not affected by the lighting and is not limited regarding the number of users that can be recognized.

In this study, the following constraints are defined as the area characteristic is used to track the user head.

Assumption 1. The recognition range for the head and body to appear is set according to the camera.

Assumption 2. When the users overlap, the upper body below the head has to appear in the image for that user to be recognized.

The head tracking method proposed in this study determines the head based on the highest point of the depth image and minimizes computation operations by deleting the image below. Fig.1 shows a flowchart of the head tracking method proposed in this paper.

Fig.1 Head tracking flowchart
2 Preprocessing
For the depth image, noise removal is performed because salt and pepper noise occurs significantly depending on the camera.

![Fig.2 Original image](image)

Second, the pepper noise is removed through the dilation operation and the dilated image is restored through the erosion operation. In this paper, the erosion and dilation functions are provided by the open source library OpenCV for the operations needed in the noise removal.

![Fig.3 Noise removal](image)

3 Highest Point Search and Square Head Area Operation
The highest point in the image where the noise was removed through the method shown in Fig.6 was found next. After that, using the depth information of the highest point, the head tracking square area was determined. Since the head is always located at the highest part of the human body, the area within the corresponding square area is set as the candidate for head determination.

![Fig.4 Erosion operation](image)

Through the depth information of the highest point, the distance between the camera and the object was calculated and the face length data according to the
distance given in Table 1, which shows the experimentally obtained head length according to distance data, was used to calculate the size of the square area. When an object existed between the experimental distance, linear interpolation was used to calculate the length. Table 1 shows the data acquired using DS311 of SoftKinetic with the setting for recognition distance range of 1.2m ~ 2.8m, Long Range Mode, and Confidence 500. The head length according to distance data can differ depending on the camera and the setting configuration.

Table 1 Head length according to distance

<table>
<thead>
<tr>
<th>Distance (m)</th>
<th>Length (pixel)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.2m</td>
<td>130</td>
</tr>
<tr>
<td>1.6m</td>
<td>110</td>
</tr>
<tr>
<td>1.9m</td>
<td>90</td>
</tr>
<tr>
<td>2.3m</td>
<td>70</td>
</tr>
<tr>
<td>2.8m</td>
<td>53</td>
</tr>
</tbody>
</table>

### 4 Head Determination

The following 6 conditions are used to determine whether the image within the head candidate square area is generated about the highest point.

- **Condition 1.** The height and width of the head tracking square area has to be the same.
- **Condition 2.** The face length has to be greater than the defined minimum length.
- **Condition 3.** Pixels have to be exist at the center of the head tracking square area.
- **Condition 4.** Pixel values have to exist at the center of the head tracking square area.
- **Condition 5.** The image area within the head tracking square area has to be 60% ~ 80% of the entire head tracking candidate area.
- **Condition 6.** The image within the head tracking square area has to be over 80% symmetric.

For condition 1, a case is disregarded when the subject to be tracked is located at the edge of the camera view, as shown in Fig.7. For objects that exist at the edge of the view it is difficult to determine the head through the bilateral symmetry determination and operations explained below, and thus such objects are disregarded as candidates.

For condition 2, a case is disregarded when problems with the head tracking square size due to errors of the highest point depth information from noise or exceptions that cannot be processed occur.

For condition 3, a case is disregarded when a head tracking square area is generated even though the object is not a difficult tracking subject such as the case of salt noise. For cases like Fig.8 where an object exists in the background, it is not considered a head candidate according to condition 2.

For condition 4, images not corresponding to the fact that the neck that connects the head and the rest of the human body is located in the lower center of the head are disregarded. The cases, as shown in Fig.9, where conditions 1 ~ condition 3 are satisfied but the image within the head tracking square area is not connected to the body through the lower center area are disregarded.

For condition 5, the cases with images not of the head when considering the area characteristic of the head are disregarded. In the case of condition 5, the assumption that the head area of people comprise a
normal distribution with small deviation is contained. Condition 6 determines the head based on the assumption that the head has bilateral symmetry. For cases like Figs.10 and 11, objects such as the hand are disregarded because although they satisfy conditions 1 ~ 5 they do not have bilateral symmetry.

5 Image Removal
In order to reduce the number of computation operations, the image was deleted for both cases when the head was determined and when it was not. For the case where the head was identified, the body area below the head was deleted, as shown in Fig.14. The body image was deleted based on the assumption that the length of a person’s shoulders is 3 times the length of the face. The area for image deletion includes the body area and the area within the head tracking square area, and the pixels with depth values similar to that of the user based on the highest point depth information were removed. In this paper, the pixels corresponding to distances of ±30cm from the highest point were removed.

For the cases when the subject was determined to not be a head, only the area within the head tracking square area was set for image deletion. Pixels within the range of the user thickness based on the highest point pixel were deleted. In this case, image information valid for the effective tracking subject within the head tracking square area may exist and therefore only the images in proximity are deleted.

6 User Information Supplementation
Noise occurs in the depth information inputted from the camera where error in the coordinate values of
specific frames can occur. Therefore, the reliability of the height, coordinates, and width of the head tracking square area calculated using the depth information decreases and compensation becomes necessary.

In every frame, the coordinates and radius of the head of the current users are recorded to perform compensation in the next frame. In order to compensate the coordinates of the tracking subject head in the current frame, brute force one to one matching of the coordinates the previous frame with the closest Euclidean distance was performed. Compensation was performed by reflecting the coordinates and radius of the previous frame to the coordinates and the radius of the user head tracking in the current frame. In this paper, compensation was performed with a weighting ratio of 8:2 for the previous frame information to the current frame information, and the compensated data were recorded for the next frame compensation.

7 Experimental Results
The development environment used the development tool Visual Studio 2010 in a Windows 7 OS while the hardware composition was a desktop PC Intel i7-2600k CPU, 3.40GB, and 8G RAM. As shown in Table 2, the camera was a SoftKinetic DS311 that has RGB VGA resolution, RGB Frame-Rate of 30fps, Depth Sensor QQVGA resolution, and a Depth Frame-Rate of 24 ~ 33fps. Also, development was done using OpenCV 2.4.1 and DepthSense SDK.

### Table 2 DS311 specifications

<table>
<thead>
<tr>
<th>SoftKinetic DS311</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Default resolution, Color stream</td>
<td>VGA (640x480)</td>
</tr>
<tr>
<td>Default resolution, Depth stream</td>
<td>QQVGA (160x120)</td>
</tr>
<tr>
<td>Color stream frame rate</td>
<td>30 Frames Per Second</td>
</tr>
<tr>
<td>Depth stream frame rate</td>
<td>24-33 Frame Per Second</td>
</tr>
<tr>
<td>Illumination Type</td>
<td>LED</td>
</tr>
</tbody>
</table>

For the camera selected for experimentation, the confidence was set to 500 with a minimum recognition distance of 1.0m, maximum recognition distance of 5.5m, and depth images with depth information normalized from 0 to 255 were used. For the head tracking square area, the length of one side of the head tracking square area was set to 30 pixels according to condition 2, the area characteristic was set to be 60%~80% of the head tracking square area according to condition 5, and the bilateral symmetry was set to 80% according to condition 6. For the image deletion, the thickness of people was set to 30cm. In the experimental environment with the above configuration, a minimum 30fps performance was obtained.

8 Conclusion
In this study, a method to track the heads of all users located in the view of a camera without using information other than the depth, such as color information, was proposed. Also, by proposing a unique method compared to the conventional user tracking method, recognition problems from lighting differences and the limitation in the number of people recognizable were overcome. For further study, hard tracking research guaranteeing a low frame rate in regular environments rather than a fixed experimental environment by supplementing exemption handling will be performed.

References: