Hand Detection at a Close Distance Using Depth Information of a TOF Camera

Yang-keun Ahn*, Jun-ho Park
Realistic Media Platform Research Center
Korea Electronics Technology Institute (KETI)
Seoul, Republic of Korea
ykahn@keti.re.kr*, bluedawn@msl.yonsei.ac.kr

Abstract: - A TOF camera is a useful device for obtaining 3 dimensional information of an object. At a very close distance, however, a TOF camera cannot obtain depth information. This study proposes a method to address this disadvantage of the TOF camera by skin color extraction, differences in images, and other methods to correct depth information at very close distances for the TOF to obtain depth information. The combination of corrected depth image and the conventional TOF image can detect a hand, regardless of the distance, with histogram matching.

Key-Words; - Hand Detection; Depth Image; Regions Isolation;

1 Introduction

Applications of virtual reality, a three-dimensional space created in computers designed to allow users to experience virtual environments in a realistic manner, are actively being researched in many fields of studies. Virtual environments are one of the most notable fields. The applications of VR technology include a wide range of fields such as entertainment, healthcare, education, architectural design, science, computing, communications, and almost every field of technology.

Recent increased interest in HCI (Human Computer Interaction) is triggering active research on more comfortable and instinctive user interfaces than conventional user interfaces, such as keyboards and the mouse. In order to avoid the conventional means of interactions, which rely on particular input and output devices, between users and computers, while fulfilling the need for more comfortable and instinctive interactions between users and computers, studies on handwriting recognition, speech recognition, image recognition, and others have been conducted. Technology for a ubiquitous computing environment, which allows users and computers to not rely on hardware for interactions but rather on natural and instinctive actions, is especially in demand.

Hand tracking technology is receiving much attention. Studies in the past attempted hand tracking with various technologies, and the methods of tracking have been improved. With recent improvement in hardware, research on hand tracking with 3D cameras is becoming more prominent.

A TOF (Time Of Flight) camera measures the distance between an object and a camera by shooting laser light or LED light and measuring the time the particle takes to hit the object and return to a sensor in the camera.

TOF cameras are cheaper than other 3D cameras and quickly provide relatively accurate depth information. Fig. 1 compares a stereo camera, a representative 3D camera, and a TOF camera. As illustrated in Fig. 1, the stereo camera is heavily influenced by light and other environmental factors whereas the TOF camera functions more independently of environmental factors and provides more credible results.

Figure 1. Compare two Cameras in the test (a) Stereo Camera, (b) TOF Camera

However, the TOF camera cannot obtain the depth image of an object that is located too far away from the laser or where LED light cannot reach or when it is too close and the light consequently returns too fast.
This paper proposes a technology that will correct the depth information that TOF cannot obtain independently. The technology utilizes color information of a TOF camera and corrects the depth information of an object at a very close distance, allowing hand tracking without the limit of distance by applying simple image processing to the depth information. This technology is not easily affected by environmental factors and is operated more lightly than other tracking technologies, because it only utilizes depth information at most distances. In addition, the technology is easy to implement, because it only applies a simple imaging process at very close distances.

The content of this paper is as follows. Chapter 2 briefly describes hand tracking technology and chapter 3 explains how to separate the object only with the use of depth information from a TOF camera. Chapter 4 explains the method to correct a depth image of a TOF camera at a very close distance, where accurate depth information is not provided. Chapter 5 presents experiment results of corrected depth images and of deriving the region of the hand by using only histograms and threshold values. Chapter 6 concludes the paper.

2 Hand Tracking Research

Hand tracking has been researched by many researchers. Most studies have started by limiting the environment. For example, skin colors are limited in terms of the colors for which color based hand detection is possible. [1-3] However, skin color detection is not a credible method. A hand has to be separately recognized from other objects of the same skin color and it is dependent on the lighting environment. Action flow is an alternative method, but the method induces problems when the camera is not fixed.[4] Statistical information of the locations of the hand, based on the previous probability, can provide useful information but this approach requires prior learning.[5] Model model is effective with sufficient contrast between the environment and the object, but it has a problem with complex environments and detecting concave objects.[6] There is also a method of detecting the hand with grey-level based on the shapes and textures of the hand. Wu and Huang have conducted extensive research on methods of classification for recognition of poses of the hand independent from the view. However, the results of hand detection performance were not satisfactory without the help of skin color.[7]

Cameras that offer a depth value provide additional information to assist hand detection; however, this method requires special and costly hardware. Nevertheless, relatively numerous studies have been conducted with the recent decrease in the cost of 3D cameras that offer depth values. A depth value is useful in object detection and other background isolation based on locations, because it yields an interest region with a relatively simple algorithm.[8]

3 Hand Tracking Using Depth Information

3.1 Background Isolation via Difference Image

The established depth region of an object within a certain distance can be limited with a setting in the TOF camera. This limitation allows easy background isolation; however, when an object is located outside the set region, the object is not detected. Without the limitation, the camera collects information up to the depth the camera can support. In a complex environment, however, background information is included in the information the camera collects and it is hard to isolate the object only with depth information. Fig. 2 is a histogram of depth information in a complex environment. The figure shows that the depth information histogram values from every region are evenly distributed. Therefore, it is difficult to isolate an object by setting a specific threshold value.

![Figure 2. Histogram of Depth information in a Complex Environment](image)

(a) Color Image, (b) Depth Image, (c) Depth Histogram

Using the difference image of a previously extracted background image and current image frames is a possible method to solve this problem [12]. Fig. 3 illustrates the process of isolating the background using a difference image. First, the background image should be extracted, as shown in Fig. 3(a), and when the difference image is extracted from current image, an image with an isolated background is obtained, Fig. 3(c). Fig. 3(d) illustrates the histogram value of the extracted
image and it confirms that the object close to the camera has been extracted.

Figure 3. Background extract from current image using histogram value (a),(b) Background images, (c),(d) Histogram values of the extracted image

3.2 Object Isolation via Depth Histogram
Information of a region isolated from an object within a certain distance from a camera can be extracted by setting an arbitrary threshold value in the foreground extracted from the difference image and eliminating the pixel regions under the threshold value. However, the depth image will vary as the object moves, and therefore, dynamic setting of the threshold value is required. The histogram of the depth image can be a good method of setting a dynamic threshold. A histogram of a depth image indirectly shows an object at a certain distance by graphing the pixel values of input grey image values. A hand, an arm, and a body can be separated, as in Fig. 4(c), when a depth histogram, Fig 4.(b), obtained from a depth image, Fig. 4(a), is utilized. The hand, the arm, and the body are separated by differentiating the histogram and setting the local maxima and minima as the separation values. A high performance TOF camera can easily separate the objects through histogram differentiation, but a Kinest camera may not be able to separate the objects in this manner. In such cases, the method described in chapter 3.1, that is, extracting the objects' region from a background image and a difference image, can extract the foreground, or the objects' region.

4 Depth Image Correction at a Close Distance
As explained in the previous chapter, the TOF camera enables relatively easy hand tracking. However, the TOF camera does not extract accurate depth information at close distances, as seen in Fig. 6, and returns error images over various regions. Utilizing color information, provided with depth information by the TOF camera, can correct the depth information at a close distance. This chapter describes a method of correcting the depth information of a TOF camera with skin color at a close distance. The proposed algorithm is executed in the following order.

Figure 4. Object isolation via Depth Histogram in the test (a) Depth Image, (b) Histogram of depth image, (c) Object isolation

Figure 5. Our Proposed Algorithm Flow
4.1 Hand Region Correction Using Color Information

The depth image of a TOF camera offers more accurate depth information of an object than other 3D cameras. However, the depth image of a 3D camera is unable to offer depth information at a close distance. Furthermore, when an object is very close or far away, it does not recognize the object at all, and it is hard to recognize objects at very close or far distances.

4.1.1 Isolation of Interest Region Using Difference Image and Skin Color Extraction

Hand and face regions can be found using skin color in a color image [11][12]. However, an object with color that is similar to skin color presents a challenge in finding a hand or face region. In a complex environment, as in Fig. 7, with objects colored similar to skin color, it is difficult to extract the hand region only with skin color information.

In such cases, the method described in chapter 3.1 can extract the foreground object without the background by background isolation using difference information. Fig. 8 illustrates the processes of extracting the foreground image and difference image from the previously extracted background image, the current image frame, the YCbCr color model, which is less influenced by light than other models, and the foreground object from the difference image. Figs. 8(a) and (b), respectively, are a color image of the previously extracted background image and an image with the Cr value of the extracted YCbCr color model. Figs. (d) and (e), respectively, are a color image of the background and foreground combined and Cr image. (c) is the difference image of the background and foreground combined image and background image, and (f) is the foreground obtained from a simple imaging process of the difference image.

The skin color of Fig. 8 is extracted by converting the obtained foreground image and RGB color image of a TOF camera into a YCbCr color model and applying Cb and Cr threshold values.[8] Fig. 9 illustrates the interest region extracted by the process previously described.

4.1.2 Hand and Face Regions Isolation via Labeling

We extract the hand region by extracting the skin color from the interest region obtained from chapter 4.1.1. Relatively credible results are expected in extracting the hand region using skin color at a close distance, because accurate information of the hand is obtained at a close distance. The results, however,
require a process to separate the hand region from the face region, because the face region is extracted along with the hand region.

The hand and face regions are separated by extracting the face and selecting it as the face region using the depth information. It is likely that the face will be within the depth information extraction range; unless the depth information within the interest region is 0, the region with a high value must be the face region. When the face region is excluded from the skin-colored regions, hand regions are extracted. Fig. 10 illustrates an image of a face and a hand region separated in the extracted skin-colored regions.

![Figure 10. Separated Image of Hand region and Face region](image)

**4.2 Depth Information Correction Using Color Information**

Chapter 4.1 showed that the extracted hand region can be corrected at a very close distance, i.e., close enough that the TOF camera is unable to calculate, with a color image. The depth of the hand is assumed as the maximum value in the hand’s depth image because the hand is at a very close distance.

However, the correction of depth information using color information should only be conducted when an object is at a very close distance, because relatively fast and credible information is provided when the object is not at a very close distance.

The point at which depth information is corrected was decided using the property of TOF cameras. A TOF camera is unable to measure the depth of an object when it is at a very close distance; therefore the camera treats the values of the corresponding pixels as 0. The point of application of color information is when the ratio between the nearest points and points with 0 depth values in the depth image from the TOF camera increases or decreases radically from a certain ratio. The ratio radically decreases when an object very close to the camera moves even closer to the camera, where the camera is unable to recognize the object, and the corresponding region changes to 0. This point is the point at which the TOF camera tries to measure the depth of the object but fails. The ratio radically increases when the camera was previously unable to recognize the depth of an object but starts to recognize the object in the frame.

\[
f(t) = \frac{\sum_{n=0}^{255} \text{hist}(n)}{\text{hist}(0)}
\]

(1)

\[
\text{Ratio} = f'(t) = f(t) - f(t-1)
\]

(2)

Fig. 11 illustrates the change rate, i.e., the difference between the ratio values of the current frame and the previous frame. The x-axis of the graph below is the image frame captured by a real time camera and it is the unit time. The y-axis represents the ratio value per unit time. The ratio value decreases radically at frame 196, illustrating the sudden disappearance of the hand, which results from a rapid increase in the number of pixels with 0 values compared to the previous frame. The 234 frame point in the blue box is the point where the pixels with 0 values in the image decreased rapidly. Therefore, corrected depth information should be used instead of the previous depth information in the range of frames 196 to 234 in hand tracking.

![Figure 11. The correct Point of depth information using Hand Area information](image)

**4.3 Results**

Hand region extraction experiments were conducted at a close distance with the results of correction and corrected depth information, and with depth histograms regardless of the distance in order to verify the proposed method.
The left figure of Fig. 12 is a conventional depth image of a TOF camera. The absence of depth information of the hand region is obvious. The figure on the right is the figure with the corrected depth information of the hand region using the proposed algorithm. It is evident that the hand region, closest to the camera and excluded in the depth information of the TOF camera, is successfully corrected.

Fig. 13 illustrates the result of hand region extraction at a very close range, which does not yield a normal depth image, corrected with the proposed image correction method. The hand region is easily extracted by eliminating the face region with simple labeling on the corrected image.

5 Conclusion
This paper describes a method of addressing the drawback of a TOF camera; a TOF camera cannot obtain depth information at close distances. Objects beyond a certain point are detected using depth information, because depth information from beyond a certain point offers relatively accurate 3D information and objects closer than a certain point are detected with color information. Skin color extraction technology using a YCbCr color model was used in background isolation and extraction of skin color information of the hand for accurate object isolation.

The proposed method allows the detection and tracking of an object regardless of its distance from the camera simply by analyzing the depth information, because the method offers corrected depth information at very close distances. However, the proposed method has a shortcoming that it is unable to extract cascade depth information at a very close distance, because the method limits the location of the object by the maximum value of the depth information. Future applications of object estimation and filter technology in order to obtain cascade depth information will increase the applicability of the TOF camera as a HCI device in even more areas.

References:
