Implementation of a Smart TV System with Context Awareness

YANG-KEUN AHN, KWANG-SOON CHOI, YOUNG-CHOONG PARK
Korea Electronics Technology Institute
121-835, 8th Floor, #1599, Sangam-Dong, Mapo-Gu, Seoul
REPUBLIC OF KOREA
ykahn@keti.re.kr

Abstract: - This thesis uses various types of recognition devices to research and suggest different methods of recognizing a person watching a TV. It suggests how the obtained data will be provided to the user and states how recognition devices were used to infer the situation of the data collected.

Key-Words: - Face Detection, Face Direction, Skeleton Model, Posture Recognition

1 Introduction

As technology has advanced, people have increasingly come to want more convenient products. People are more interested in products that are easier to use rather than products that require tools and come with complex instructions.

Existing TV's can be inconvenient to use without a remote control. Although you could manually push the buttons on the TV, this has two disadvantages: you would have to physically move in front of the TV, and not as many functions would be provided as can be found with the remote. For these reasons, the remote has become an essential tool when watching TV.

Unlike with earlier TV's, these and other complications can be avoided when watching a Smart TV, even without a remote. In the case of a Smart TV that supports gesture recognition, a camera mounted on the TV is used to recognize specific gestures that prompt the TV to perform correlating actions. Additionally, voice recognition can also be used. When a verbal command for the related action is given, the appropriate action is carried out by the TV. In this manner, the Smart TV dynamically adjusts to people's expectations.

This thesis documents the research and implementation conducted to develop technology to provide convenience to the Smart TV user. The area of focus is related to the use of controlling movements; in this case, using only the user's physical body. The approach does not simply recognize gestures as are currently done in existing Smart TV's, but takes this method a step further in an attempt to automatically address the user's current needs.

This thesis is divided into two parts: one part relates to the collection of information about the user and the user's surroundings using recognition devices; the other part relates to providing the collected information to the user. To collect data a 2D camera, 3D camera, thermo-hygrometer, and etc. were used as recognition devices. The set-up also included a function that used the collected data to draft information to be provided to the user. AGUI, android application, and wireless bulb can be used to provide this information to the user, or to allow the user to view the data. The recognition devices and functions that provide information utilize UDP communication to send and receive information.

2 Recognition using a 2D camera

The 2D camera is used to recognize information related to the user's face. Once an image inputs from the camera, it then searches to find a face in the image. Once the face is found, it extracts, special characteristic points from the face to identify the user. This information is also used to determine the direction the face is pointing.

2.1 Face Navigation

To find a more exact location of the face than simply searching for the face in the image, the following conditions are applied.

First, the higher the ratio of the face in the image, the better. Second, a crisp image showing facial characteristics is best.

In this thesis, it was decided to increase the accuracy of facial recognition, thereby satisfying the first condition. As a method to increase the ratio of the face in the image, it was decided to perform the face recognition process twice.

The initial facial recognition process will only search for the location of the face. Using the image input from the camera and with the facial location serving as the center, 2.5 times the width of the face is cut into an area and considered to be one side. The function provided by Luxand's Face SDK was used for the face recognition and for the image cutting process in the area around the face. Image 1 shows the location of the face found from the image after the first facial recognition process.

Fig. 1 Location of the face found from the image input from the camera

In the second facial recognition process, not only the location of the face is searched, but points of special characteristics are also searched. The face is searched with in the area cut from the image during the initial facial recognition process, and that image section is then used to find points of special characteristics. The function provided by Luxand's Face SDK was used to extract points of special characteristics. The threshold value of face recognition was set at 5.

Image 2 shows the location of the face after the second facial recognition process, from the area that was cut from the image

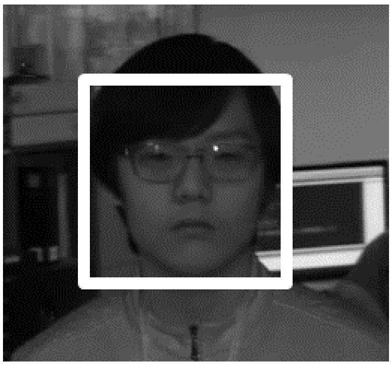

Fig. 2 shows the location of the face from the area of the image that has been cut

2.2 User Identification

After registering information from the searched image, a function to determine if the image is of the same person was established in a later search. The Tracker function, provided by Luxand's Face SDK, was used to distinguish the user and determine the level of similarity with the points of special facial characteristics from the stored information. The threshold value of the similarity ratio, to determine if it is the same user, was set to 0.995 for this study.

2.3 Facial Direction Determination

The special characteristic points taken from the previously searched face were used to determine the facial direction of the user. Since the special characteristic points change along with a change in facial direction, the relative location of the special characteristic points can be used to determine the direction of the face.

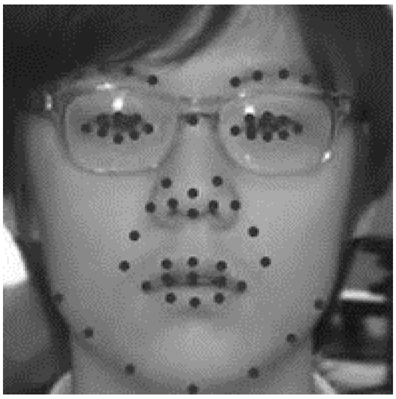

Fig. 3 Special characteristic points found from the searched face

There are four points used to determine the left or rightward direction of the face: the middle of the forehead, tip of the nose, center of the left eye, and center of the right eye. First, two triangles are created using the center point of each eye, with each connecting to the middle of the forehead and the tip of the nose, and then the area of each triangle is calculated and the ratio of the areas are used to determine the left or rightward direction of the face. If the area of the triangle on the left is smaller, then the face is facing left. If the area of the triangle on the right is smaller, the face is facing right. Image 4 shows the connected special characteristic points of the nose and eye. It can be seen that, depending on the direction the face is facing, the area of A and B are different.

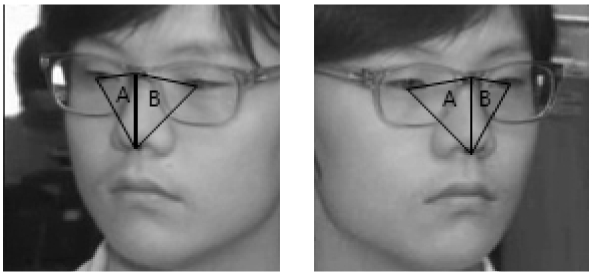

Fig. 4 The change in area when facing right (A) and left (B)

Special characteristic points in the middle of the forehead, and the left, right, and tip of the nose, are used to determine the direction of the upper portion of the face. This is based on the fact that, when the face is facing upward, the height of the nose decreases when looking directly at the nose, but the length remains the same. This relation is used to determine the direction of the upper portion of the face. With the length of the nose as the standard, if the height of the nose decreases, then this shows the face is facing upward, but an absolute cannot be used because the shape of the nose differs from person to person. During the search to see if it is the same person, the maximum and minimum value resulting from the input of the ratio of the length and height of the nose, is recorded. With the minimum and maximum values as the standard, the upward direction of the face is determined by using the current ratio of the length and height of the nose. Image 5 shows the length of the connected special characteristic points on the nose when looking upward and forward, respectively.

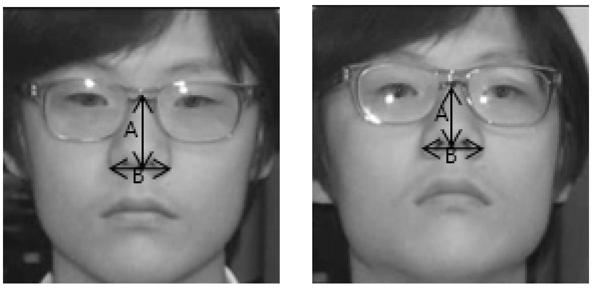

Fig. 5 The height (A) and length (B) of the nose are shown when looking forward and upward.

3 Recognition using a 3D Camera

The 3D camera is used to collect information relevant to the user's actions and location. Microsoft's Kinect SDK 1.7 was used to extract the skeleton model to gain information regarding the user's location and actions.

3.1 User Location Search

The skeleton model, provided by Microsoft's SDK 1.7, was used to search for the location of the user. The skeleton model extracted from the image of the 3D camera consists of information from a total of 20 joints, and each joint has spacial information x, y, and z.

Fig. 1 6 Skeleton model extracted from the image input by the 3D camera

Information about the joints from the skeleton model is used to determine the location of the user. Of the joint values of the skeleton model, the x-axis value can be used to determine the horizontal location, and the z-axis value can be used to determine the distance from the camera.

If there are more than two results for the tracking information, the result closer to the camera is used as the location of the user, after checking the location information.

3.2 User Posture Recognition

The current posture of the user is determined using the skeleton model of the user. The posture of the user can be determined by considering the location relation of the points in the vicinity of the joints taken from the skeleton model.

The knee joint location value is used when differentiating between standing and sitting. When the z-value of the knee joint and waist joint are compared and the knee joint is situated in front of the waist joint, it can be determined that the user is currently sitting. On the contrary, if the knee joint and waist joint are in similar locations, it can be determined that the user is currently standing.

To determine that the user is currently using the phone, the posture of the user's hand near the ear, while holding the phone, is used. The joints of the shoulder to the wrists are checked to examine the angle created. Also the distance between the hand and the head is calculated, and if it is close, it concludes that the user is on the phone.

Fig. 7 Skeleton model when sitting (Left) and skeleton model when on the phone (Right)

4 Thermo-Hygrometer

Information regarding the temperature and humidity can be obtained through the sensors on the thermo-hygrometer. Through the sensor, the temperature and humidity values can be read, but the temperature value must be read first and then applied to calculate the humidity value, as it changes depending on the temperature.

5 Android Application

Information can be received or input by linking the mobile phone of the user. The android application is able to display, on the screen of the mobile device, information received from the recognition devices. If the user's face is searched by the camera on the mobile phone, it will conclude that the user is using their mobile phone, and this information will be sent.

6 Communication

In this thesis, the portion regarding the drafting of information and the portion providing this information to the user were drafted separately. The method used to transmit the data involves requesting data through UDP communication, and once the prepared data has been received by the socket, the requested data would then be transmitted to the requester. If the character string, "RD"(Request Data), is made into a packet and transmitted, the received packet will be searched. Once "RD" is confirmed, the address of the sender of the information will be confirmed and the information will be relayed to that address.

7 GUI

This function provides contents to the user. Its role is to show the user information received from the recognition devices, and drafted information inferred from the situation.

Fig. 8 GUI Screen

After periodically receiving data by requesting communication the input data will appear on the screen, or activate contents according to control commands.

8 Wireless Bulb

Another device exists that displays information to the user other than the GUI. Philip's Wireless Bulb - HUE, which informs you of the user's color preference and changes the color of the lighting, was used. By using the CURL functions, an open source, a message is sent to the wireless bulb to make changes in color.

9 Situation Inference

By using information obtained from the recognition devices, processed information is provided for the convenience of the user. Its function drafts commands to control GUI contents by importing registered information about the user or using the user's action or posture.

After distinguishing the user, user information is used to draft information to construct the screen for the GUI. It drafts information regarding the type, color, and size of the layout, founded on the preference of each user, to provide different GUIs.

It drafts information to control the menu of the GUI by using information related to the facial direction of the user. By using the left and rightward facial directions it drafts the movement control of the menu, and by using the up and downward facial direction it selects the command.

If the user appears to be using the phone, or if it recognizes information showing that the user is on the phone, it will command the VOD that is playing to pause. If it recognizes an action other than the user being on the phone, or determines that the user has stopped using the phone, a command is sent to start the paused VOD.

10 Conclusion

This thesis introduced and examined a system that uses various recognition devices to collect information about the user and their surroundings, and processes that information to provide convenience to the user. It uses a 2D camera to collect facial information about the user and calculates facial direction in order to operate the menu. It uses a 3D camera to collect information about the user's location and actions, and when a specific action is conducted it automatically carries out functions related to that action.

The following research will be conducted in the future: a more precise method to utilize and expand information relating to the facial location and actions of the user, and methods to provide information to the user from other recognition devices.

References:

[1] Eun-ju Kim, Seong-lyeol Song, Myung-won Kim "Convergence Technique for Personalized Recommendation in Smart TV Environment"

[2] Young-sul Lee, Gyeon-mo Yang, Seong-bae Cho, "Ontology-based Smart TV Environment Modeling for Situational Smart TV Service"

[3] Microsoft Corporation, "Human Interface Guidelines v1.8.0"

[4] Abhishek Kar, "Skeletal Tracking using Microsoft Kinect"

[5] Sait Celebi, Ali S.Aydin, Talha T.Temiz, Tarik Arici, "Gesture Recognition Using Skeleton Data with Weighted Dynamic Time Warping"

[6] Sun-young Cho, Hye-lan Byeon, Hee-kyung Lee, Ji-hun Cha, "Arm Gesture Recognition for Kinect Sensor Based Shooting Game"

[7] Luxand, "Luxand FaceSDK Documentation v5.0"

[8] Abu Sayeed Md. Sohail and Prabir Bhattacharya, "Detection of Facial Feature Points Using Anthropometric Face Model"

[9] Matthias Dantone, Juergen Gall, Gabriele Fanelli, Luc Van Gool, "Real-time Facial Feature Detection using Conditional Regresstion Forests"

[10] Woo-gi Lee, Jong-tae Baek, Hwa-gi Lee, Young-mo Lee, "Angular Feature Vector Based Face Recognition Model"