Research on the Real-time 3D Image Processing System using Facial Feature Tracking

Jae-gu Song, Yohwan So, Eunseok Lee, Seoksoo Kim

Abstract—This research is on the real-time 3D image processing system using facial feature tracking and how the system works. When transferring an input 2D image to a 3D stereoscopic image, this system provides real-time 3D synthetic images. It also provides measures to trace a face in the input image in order to distinguish a person from the background and includes measures to digitize positional values within the face by tracking colors and facial feature points. The real-time 3D image processing system in this study that uses facial feature tracking is the preprocessing system for special effects. Firstly, it allows users to utilize basic positional values obtained from a person and a background in the collected images when applying special effects. Secondly, by checking a 3D stereoscopic image in real-time, users can verify composition and image effects prior to application of special effects. Lastly, data that successfully generated facial areas can be constantly improved and used as foundation to standardize facial area detecting data as well as to create the plug-in.

Keywords— Imaging Processing, 3D, Face Recognition, Tracking, Real-time processing

I. Introduction

Among computer image studies, those on methods for tracking and animating human face recognition, facial expression change, and movement detection have been constantly conducted along with growth of the computer vision field. Up to now, there have been various researches on facial movement, change, and detecting eyes, nose and mouth based on 2D images and outcomes of such researches have been used to identify the same person through pattern analysis or for the photograph security guarding systems. Most of processing technologies have been introduced to the special purpose study field and commercialized. These image processing technologies is lately supported to process various and complicated effects as a result of recent growth of the computer hardware technology. Particularly, studies on movies, virtual reality technology, animation and VFX are noticeably being conducted. These areas need solutions for general users to access easily without the limit that occurs because special equipment is required for production.

In order to process stereoscopic images, it is required to conduct researches on each modularization. In this process, automation of the CG production environment reduces production costs and period, which increases competitiveness of the contents. However, it is difficult to establish an efficient process by post-processing basic image shot sources after checking every single cut. By applying the real-time monitoring system from shooting, improvement in time and quality can be achieved as outcomes of shooting can be monitored immediately, which enables revise and reshoot immediately if the image is beyond the error range.

However, there is a drawback when producing 3D contents as it requires constant monitoring of the whole screen and unnecessary data processing time that may result in system overload and inefficient service. Thus, it is necessary to find a method for applying a system that allows of tracking and expressing only specific areas, characters or objects suitable for 3D contents and monitoring in real-time from the shooting stage.

In this study, 2 units of web cameras were used to create 3D images and the fundamental data generating system for synthesizing images by tracking facial areas to apply VFX effects was established. Facial area generating method using color-based face recognition and facial feature points was used in this study.

So far, most of studies have been conducted to improve the identification rate by eliminating secondary information such as lighting values that is not required for face recognition or by tracking specific values after generating facial feature that can only be seen in the face or detecting colors of facial skins[1][2]. These studies show improved identification rate by generating facial feature points of the user's. In this study, a fundamental research was conducted to recognize facial expressions using the bottom-up feature-based characteristic as well as color value generating algorithm. Through this, more detailed facial expressions can be analyzed by acquiring positional values based on a specific point in the face and analyzing the characteristic of the area using color values[3].

Considering such characteristics, the objective of this study is to detect facial areas automatically by using color values and feature points obtained from two input images and to digitize positional values by creating control lines responding to each image.

II. Fundamental research

A. Stereoscopic Image Display

The finest stereoscopic image is created by coordinating two units of cameras with binocular disparity information for a 2D realistic image to generate a 3D stereoscopic image and by applying special effects and computer graphics technology to express the depth and the impression of space[4].

The basic of such stereoscopic images is binocular disparity. It occurs because location of the left and right eyes are separated, image location of an object seen differently by each eye and a retinal image of the left and right eyes are different.

In order to realize characteristics well when creating such stereoscopic images, depth perception, sharpness, resolution, naturalness and visual comfort should be taken into account. Such stereoscopic images can be classified into various categories according to display methods and viewpoint shooting conditions.

The anaglyph method is one of the oldest methods for creating stereoscopic images. It separates images that are on the left and right to reproduce images using the difference in colors of the left and right. In this study, the anaglyph method was used to express stereoscopic images. Figure 1 displays the conceptual diagram of anaglyph method[5].

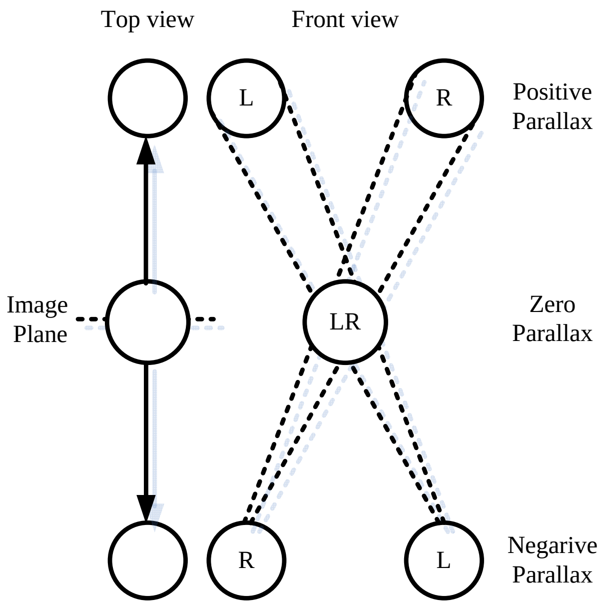

Fig. 1 Conceptual Diagram of Parallax

In this study, the anaglyph method was used to express stereoscopic images from images that were shot with two web cameras.

B. Face Recognition

Face recognition is the most accurate when analyzing a frontal 2 dimensional image[6, 7]. However, in actual circumstances, there is diverse information that interrupts face recognition such as various image information, side image analysis and impact of lights. Particularly, additional information that occurred in the shooting setting such as background images, lighting, and shadow interrupts accurate face recognition.

To solve this problem, a number of studies have been conducted on how to extract face and those include regularized face symmetry, color information, a flux of light, Gabor Wavelet, neural network and template matching[8][9]. Those methods attempt to match extracted features with the most similar shapes and use the area with the most symmetry as a feature region. Through substitution, distinct outcomes can be generated. However, it is necessary to carry out studies to track 3D-based face regions based on features of stereoscopic images as per recent demand[10][11]. To meet such demand a method for tracking change in facial expressions using an infrared display device or a probe, a method for tracking 3D position change on the facial surface by marking in the face, estimating method for linear tracking feature points by estimating parameters of a 3D face model. However, these methods require special marking in the face.

B-1. Color-based Face Recognition

Color analysis based methods have been used and proven to be an effective feature in the application of face detection and tracking. Although different people have different skin color, several studies have that the major difference lays largely between their intensity rather than their chrominance[12].

B-2. Color centroids segmentaton model

Different methods are introduced that apply widely-known methods to recognize color values in image processing. The approach of this study is based on "Face detectionand tracking in clolor images using color centroids segmentation" (Zhang, Q.) technique[12].

The procedure of color centroids segmentaton model is as followings;

Color triangle is created by following steps:

Step 1 : create a standard 2-D polar coordinate system;

Step 2 : create three color vectors to reflect R, G and B, the range of them is [0, 255] and alternation reciprocally as follows:

$$\begin{cases} R\ Component\ \varphi_R = 90°, & r_R \in [0, 255] \\ G\ Component\ \varphi_G = 90°, & r_G \in [0, 255] \\ B\ Component\ \varphi_B = 90°, & r_B \in [0, 255] \end{cases}$$

Step 3 : connect the three apexes.

B-3. Face Detection Technique using RGB Color

In general, an algorithm that can recognize the most simple skin color can be created by analyzing RGB values by pixel, which is excellent in sensing various colors by the light. The pixels for skin region can be detected using a normalized color histogram, and can be further normalized for changes in intensity on dividing by luminance[13].

B-4. Face Detection Technique using YCbCr Color

YCbCr Color space algorithm detects pixel values of a limited region by extracting Cb and Cr values and analyzing their similarity [14]. YcbCr Color space can detect any skin color of a person.

The thresholds be chose as $[\mathrm{C}_{r1}, C_{r2}]$ and s $[\mathrm{C}_{b1}, C_{b2}]$,a pixel is classified to have skin tone if the values $[\mathrm{C}_{r}, C_b]$ fall within the thresholds. The skin color distribution gives the face portion in the color image[15].

B-5. Face Detection Technique using HSI Color

HIS color technique is very similar to YCbCr Color technique, but it uses hue(H) value and saturation (S) value. The threshold be chose as $[H_1, S_1]$ and $[H_2, S_2]$, and a pixel is classified to have skin tone if the values $[H, S]$ fall with the threshold and this distribution gives the localised face image.

C. Feature-based Face Recogntion

To extract specific regions in the image processing, corner detection method is most commonly used. A disadvantage of corner detection is that it responds to background image values so sensitively that it cannot extract accurate corner values. In other words, it is very sensitive to noise. But with limiting image regions, accurate positional values of feature points can be identified. In this study, a method using local feature was applied. Local feature method consists of 4 steps as following. This method adopts the division concept instead of coordination when measuring similarity.

1. feature detection
2. feature description
3. feature clustering
4. frequcency histogram construction
   (for image representation)

Bag of features can be the most typical research.

C-1. Bag of Features[16]

Given N training images belonging to M classes, an interest point detector is employed to detect representative interest points. As basic elements of object representation, image patches around detected points are extracted and normalized to uniform scale. Then SIFT descriptor is used to represent features of these local regions. However, sliding window, regular grid and random sampling can also be applied to obtain patches, as well as other feature descriptors. Given a collection of image patches, codebook is formed by performing K-means clustering algorithm. K is the size of codebook. Codewords are then defined as the centers of clusters. Thus descriptors in each training image can be coded by hard assignment to the nearest codeword, yielding a histogram $n(\omega_i, d_j)$ ounting the frequency of occurrence of each codeword, where ω_i denotes j^{th} visual word in the K-size codebook and d_j is j^{th} class. The histogram is treated as a “bag”. After representing a test image as a histogram, the most similar training histogram can be found by some similarity metric and corresponding object class label is also returned.

III. SYSTEM DESIGN

This study was conducted on the real-time 3D image processing system using facial feature tracking which provides the real-time 3D synthetic images while converting input 2D images into 3D stereoscopic images. Characteristically this study provides methods for tracking face in order to separate a person from background in the input image. This includes methods for saving after digitizing positional values of face regions by tracking color and face feature, methods for revising images to allow of color-based face region detection when it is hard to detect face region in the input image as well as methods for extracting face region by utilizing successfully detected data as side information and synthesizing when required. The real-time 3D image processing system that uses facial feature tracking specified in this study is the preprocessing system for application of special effects in 3D image processing. It allows users 1) to utilize initial positional values acquired from separation of a person and background in the collected image for application of special effects, 2) to immediately monitor composition and image results prior to application of special effects by monitoring 3D stereoscopic images in real-time, and 3) to improve constantly successful face region extraction data so it can be used as the fundamental data for standardized face region extraction as well as for creation of plug-in.

A. System Configuration

The fundamental configuration diagram of this study is as following:

Figure 2 shows steps as following:

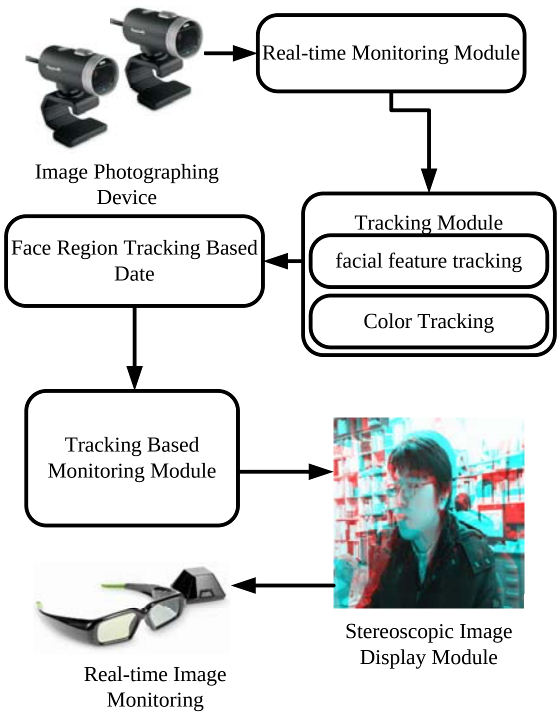

Fig. 2 System Diagram of the Real-time 3D Image Processing System using Facial Feature Tracking

STEP 1: Image Photographing Device
STEP 2: Real-time Monitoring Module
STEP 3: Tracking Module
STEP 4: Face Region Tracking Based Date
STEP 5: Tracking Based Monitoring Module
STEP 6: Stereoscopic Image Display Module

STEP 7: Real-time Image Monitoring

System configuration is the real-time monitoring module where image photographic device information is collected and real-time image is displayed and consists of the tracking module which extracts face from transmitted images and collected data. The tracking module includes facial feature tracking and color tracking in this process.

The positional values extracted in the tracking module shall be saved as face region tracking based data. Digitalized images shall be transmitted to the tracking based monitoring module in which face regions are traced based on face region tracking based data in real-time and be displayed as the stereoscopic display module in which the end user can check in real-time. Lastly, the system is comprised of the real-time image verification step to verify images generated from the stereoscopic image display module.

B. *Image Input Procedure*

The procedure of the image input step in this study is as following. This flowchart displays how the images collected from two units of web cameras are displayed on two videos.

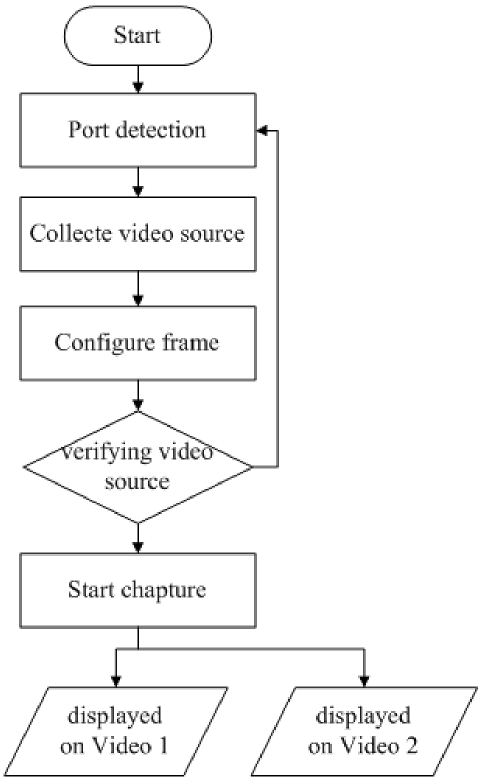

Fig. 3 Monitoring Processing Flowchart

The Real-time Monitoring Module is comprised of the step for creating video source in which port detection and data are collected in order for image photographing devices to receive data. After the frame is configured to set up video display, it proceeds to the step for verifying video source display status to check if images are displayed properly on the video. If displayed properly, images shall be displayed on Video 1 and Video 2.

C. *Tracking Module Procedure*

In this study, the tracking module was used for tracking face region in collected images.

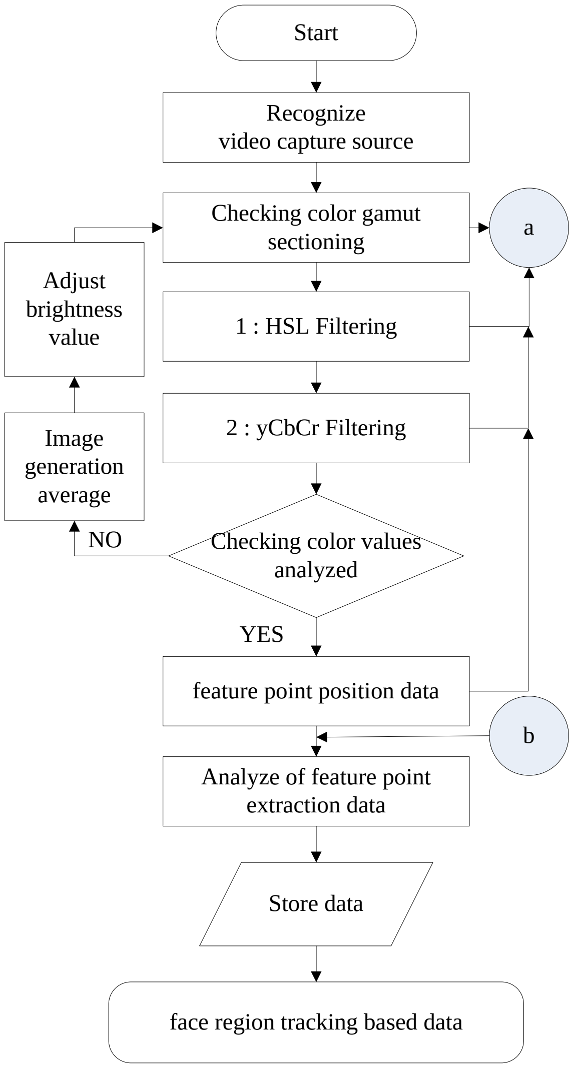

Fig. 4 Tracking Module Flowchart

In the image tracking module, clarity of color gamut sectioning is ensured by checking intensity values to determine whether video capture source is recognized and color values are separable. If color gamut sectioning is clear: 1) HSL Filtering is applied for the primary sectioning of face region that later substitutes as an auxiliary step for HSL Filtering step where the preliminary face region sectioning was attempted; 2) yCbCr Filtering is applied to extract tracking values for the secondary detection of color gamut. If color values cannot be analyzed in the previous steps including the color value analytical judgment step, brightness value shall be forcibly adjusted using data received from the image generation average verification step and the image generation average step. This is the process where an attempt to re-extract is made in case it is difficult to separate face region through color values.

If color value can be analyzed in the color value analytical judgment step, there will be the operation process for feature point position data in which the feature point extraction data is analyzed from the extraction step. The last step is the data storing step in which operated data is stored in the table. Stored data shall be saved as face region tracking based data.

D. Face Region Tracking Based Data Structure

The face region tracking based data structure contains information that can be used as the future pattern based data as it continuously stores fundamental data required to configure this system as well as new data generated when face tracking is conducted.

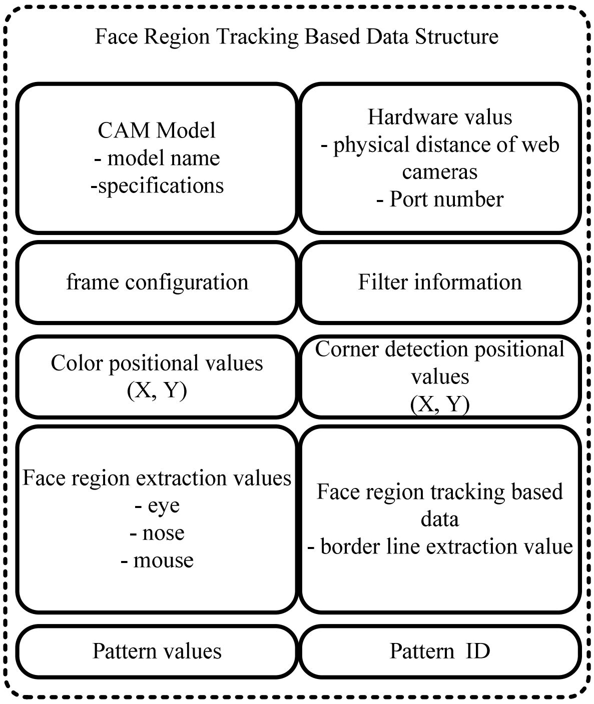

Fig. 5 Face Region Tracking Based Data Structure

This structure stores information that is collected from the real-time monitoring module including web camera model name, support resolution and port number required for connection to the system and physical distance of web cameras, frame configuration that is applied in the tracking module, corner detection positional values (X, Y), and face region extraction values and border line extraction value that were acquired by face region tracking based data. This stored information is utilized as the previous pattern data for similar pattern values so it can be used as correction information in case when extracted information is insufficient.

E. Pattern Application Real-time Tracking Method

This method is to identify uncollectable data during the process where information is being stored based on the data stored in the face region tracking based data table, and to verify data collected for detecting face region from the image that is difficult to recognize by using existing information as additional information. The procedure is as following:

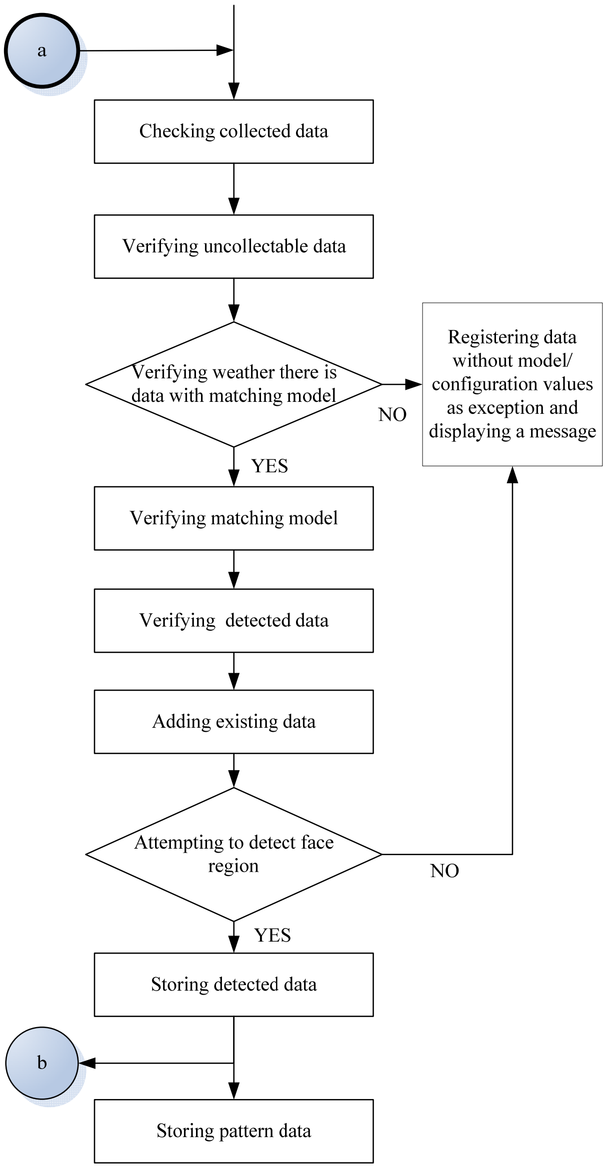

Fig. 6 Real-time Pattern Application Tracking Flowchart

Step 1: Verifying uncollectable data

Step 2: Verifying weather there is data with matching model/configuration values among images with uncollectable data

Step 3: Registering data without model/configuration values as exception and displaying a message

Step 4: Verifying matching model/configuration value data in case there is data with matching model/configuration values

Step 5: Verifying data in which matching data was detected in verified data

Step 6: Adding existing data to the detected uncollectable data

Step 7: Attempting to detect face region using added data

Step 8: Extracting detected data from the detectability verification step

Step 9: Storing detected data as pattern information in to face region tracking based data table.

IV. REALIZATIO

In this stage, actual Proto type program is configured based on configuration in the design stage to describe test results. Configuration of hardware and software used in this test is as following:

Hardware
- pc (intel core 2 cpu 2.13ghz, 2gb Memory)
- 2 units of WEB CAM (Microsoft Live Carcinoma)

Software
- os(windows 7)
- development tool (Microsoft visual studio)
- language (c#)
- module (emgu, aforge.net, openCV)

This program receives images respectively from two separate units of cameras as designed in the monitoring module. Figure 7 demonstrates how the program that receives an image in real-time and displays.

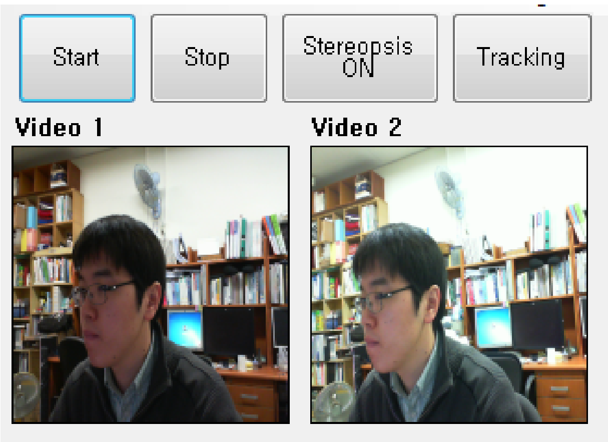

Fig. 7 Monitoring

Once it was ensured with monitoring that images from Video 1 and 2 are displayed properly on the screen, the tracking process to detect face region commences. Face region detection by color information is the most fundamental step for extracting information of skin color. By this, a facial skin color filter is created based on HSL and YCBCR information and the optimal color distribution is optimized for each image. This program determines if color values are separable and detects face region through HSL Filtering. Figure 8 shows images to which HSL Filtering is applied.

Fig. 8 HSL Filtering Application

And then, YCBCR Filter is applied. Figure 9 shows YCBCR Filter applied images.

Fig. 9 YcBCR Filter Application

Using 2 types of filter, background and face region are separated for color-based face region extraction. Then, face region tracking commences using feature points. Figure 10 shows images with extracted feature points.

Fig. 10 Feature Point Extraction in the Full Area

Only positional values of those extracted feature points that are determined to be within the range of color-based face region shall be digitalized. All feature points extracted from background shall be excluded. Figure 11 shows extracted feature points only that are within face region.

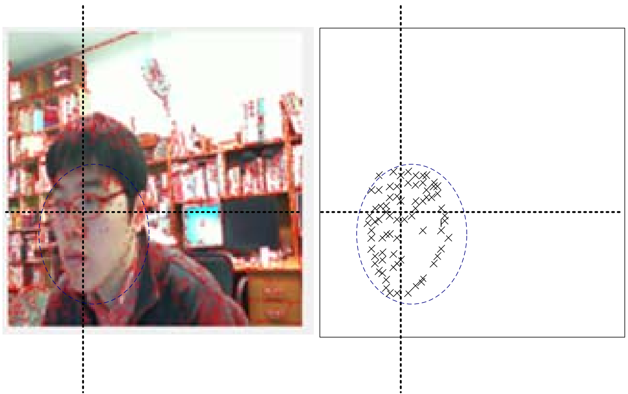

Fig. 11 Face Regional Feature Point Extraction

Lastly, stereoscopic images are displayed. Figure 12 shows images in the final process.

Fig. 12 Stereoscopic Image Display

V. Conclusion

This study was conducted on the real-time 3D image processing system and methods for extracting the fundamental data that can be used to shoot images and apply 3D effects. The results of this study can be practically utilized. Firstly, this study can be used when producing diverse types of contents including UCC in inexpensive ways without limitation of expensive equipment for displaying 3D images. Secondly, this study provides the fundamental image processing data for interweaving an actual person’s image with the background by extracting face region and storing positional values of objects in the image. Lastly, in case it is difficult to extract face region from an image, images can be shot with existing supplementary data of extraction information to extract more elaborate face region. As such data is digitalized into pattern data, it can be expanded into various data and utilized in the plug-in type.

ACKNOWLEDGMENT

This research is supported by Ministry of Culture, Sports and Tourism(MCST) and Korea Creative Content Agency(KOCCA) in the Culture Technology(CT) Research & Development Program 2010.

REFERENCES

[1] HSU, R.-L, ABDEL-MOTTALEB, M., AND JAIN, A. K., “Face detection in color images,” *IEEE Trans. Pattern Analysis and Machine Intelligence 24*, pp. 696–706, 2002.
[2] W. Chen, M. J. Er, and S. Wu, “Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain,” *IEEE Trans. on System, Man and Cybernetics-Part B: Cybernetics*, vol. 36, no. 2, pp. 458–466, Apr. 2006.
[3] Jae-gu Song, Sungmo Jung, Yohwan So, Seoksoo Kim, “A Study on Face Recognition by Extracting CT-based Color Values,” *Recent Researches in Circuits, Systems, Electronics, Control & Signal Processing*, pp. 132-136, 2010.
[4] Chang Ok Yun, Tae Soo Yun, Dong Hoon Lee, “Integrated editing system for 3D stereoscopic contents production,” Journal of the Korea Industrial Information System Society, vol.13, no.1, pp 11-21 , 2008
[5] EunYoung Choi, “3D Stereoscopic CGI Production Pipeline-Focus on Making Process of Avatar,” *Journal of Korea Contents Association*, vol. 10, no. 8, pp. 159–167, Aug. 2010.
[6] A Samal and P. A. Iyengar, “Automatic recognition and analysis of human faces and facial expressions: a survey,” *Pattern Recog. 25*, pp. 65–77, 1992.
[7] R. Chellappa, C. L.Wilson, and S. Sirohey, “Human and machine recognition of faces: A survey,” *IEEE vol. 83, Issue 5*, Aug. 2002.
[8] Kelsey Ramirez-Gutierrez, Daniel Cruz-Perez, Jesus Olivares-Mercado, Mariko Nakano-Miyatake, Hector Perez-Meana, “A Face Recognition Algorithm using Eigenphases and Histogram Equalization,” NAUN:INTERNATIONAL JOURNAL of COMPUTERS, vol.5, no.1, pp 34-41 , 2011.
[9] Hamido Hourani, Mohammed Al Rawi, Abd El-Latif Abu Dalhoum and Sabina Jeschke, “The Human Face Semantic Web,” NAUN:INTERNATIONAL JOURNAL of COMPUTERS, vol.1, no.3, pp 218-230 , 2008.
[10] Mario Malcangi, “Text-Driven Avatars based on Artificial Neural Networks and Fuzzy Logic,” NAUN:INTERNATIONAL JOURNAL of COMPUTERS, vol.4, no.2, pp 61-69 , 2010.
[11] Gualberto Aguilar-Torres, Karina Toscano-Medina, Gabriel Sanchez-Perez, Mariko Nakano-Miyatake, Hector Perez-Meana, “Eigenface-Gabor Algorithm for Features Extraction in Face Recognition,” NAUN:INTERNATIONAL JOURNAL of COMPUTERS, vol.3, no.1, pp 20-30 , 2009.
[12] Qieshi Zhang and Jun Zhang, “RGB Color Analysis for Face Detection,”, Advances in Computer Science and IT, pp.119-126, 2009.
[13] Crowley, J. L. and Coutaz, J., “Vision for Man Machine Interaction,” Robotics and Autonomous Systems, vol. 19, pp. 347-358, 1997.
[14] Cahi, D. and Ngan, K. N., “Face Segmentation Using Skin-Color Map in Videophone Applications,” *IEEE Transaction on Circuit and Systems for Video Technology*, vol.9, pp. 551-564 1999.
[15] Sanjay Kr. Singh, D. S. Chauhan, Mayank Vatsa, Richa Singh, “A Robust Skin Color Based Face Detection Algorithm,” Tamkang Journal of Science and Engineering, vol.6, no.4, pp.227-234, 2003.
[16] Fan Yang, Huchuan Lu, Yen-Wei Chen, “Bag of Features Tracking,” *2010 International Conference on Pattern Recognition*, pp. 153-156, 2010

Jae-gu Song received a B.S. degree in Multimedia from Hannam University, Korea, 2006 and M.S. degree in Multimedia from Hannam University, Korea, 2008. And currently, on the Ph.D. course in Multimedia Engineering from Hannam University. In 2011 He joined Korea Atomic Energy Research Institute, where he is now researcher. His research interests include multimedia network system, ubiquitous system, medical information system, network security.

Yohwan So received a B.S. degree in Fine Art from Hongik University, Korea, and M.S. degree in Fine Art from Hongik University, Korea, and M.S degree in Engineering from New York Institute of Technology. And currently, he joined the faculty of Hannam University, Korea. His research interests include multimedia contents, 3D modeling, VFX.

Eunseok Lee received a B.S. degree in Fine Art from Daegu University, Korea, and M.S. degree in Fine Art from Sejong University, Korea, and ABD Doctor of Philosophy in Design form Sejong University, Korea. And currently, he joined the faculty of Hannam University, Korea. His research interests include multimedia contents, Visual communication design.

Seoksoo Kim received a B.S. degree in Computer Engineering from Kyungnam University, Korea, 1989 and M.S. degree in Information Engineering from Sungkyun-kwan University, Korea, 1991 and Ph.D. degree in Information Engineering from Sungkyun-kwan University, Korea, 2002. In 2003, he joined the faculty of Hannam University, Korea, where he is currently a professor in the Department of Multimedia Engineering. His research interests include multimedia communication systems, distance learning, multimedia authoring, telemedicine, multimedia programming, computer networking, and information security. He is a member of KCA, KICS, KIMICS, KIPS, KMS, and DCS. He is editor-in-chief of IJMUE.