Vision-Based 3D Fingertip Interface for Spatial Interaction in 3D Integral Imaging System

Nam-Woo Kim, Dong-Hak Shin, Dong-Jin Kim, Byung-Gook Lee
Dept. of Visual Contents, Dongseo Univ.,
Busan, Korea
{d5302010,shindh2,einekies,lbg}@dongseo.ac.kr

Eun-Soo Kim
3DRC, Kwangwoon Univ.,
Seoul, Korea
eskim@kw.ac.kr

Abstract

In this paper, we propose a novel interactive integral imaging system using vision-based 3D fingertip interface. This system consists of the real 3D image generation system based on integral imaging technique and the interaction device using a real-time finger detection interface. The proposed system can be used in effective human computer interaction method for real 3D image. To show the usefulness of the proposed system, we carry out the preliminary experiment and the results are presented.

Keywords: Integral 3D imaging, spatial interaction, computer vision, fingertip detection

1. Introduction

Recently, computers have increasingly influenced the way we communicate and the way we interact with environments. Although computer techniques have been advanced largely, the development of interfaces for human has lagged behind. The most commonly available interfaces between humans and computers are keyboards, mice and so on. These have limited the effectiveness and naturalness of the human-computer interaction (HCI) [1]. To overcome these limitations, many recent HCI research activities have devoted considerable effort to the development of new input interfaces based on natural ways of human communication. On the other hand, various new ideas for computer interface are also emerging. Examples of these new generation interfaces include augmented desks that offer a large and open screen for immersive and collaborative experiences, volumetric displays that illuminate voxels in 3D space to produce images akin to real world objects, and haptic devices that incorporate the sense of touch to enhance the understanding of visual data [2-4].

Recently, real 3D image displays present interesting challenges with demand special attention such as virtual reality environment. Many research works are actively being done on the development of real 3D display devices for virtual reality [4-6]. However, most of the currently developed real 3D reality systems require the viewers to wear special glasses to feel the depth, which is called stereoscopic 3D display. Even though its hardware system is very simple, it cannot gain the wide acceptance until now because of the problem of human factors. Therefore, many researchers have studied auto-stereoscopic 3D display system, because auto-stereoscopic display system can present real 3D image to the viewers without a need for any special glasses. Among them, an integral imaging technique has been actively researched for the next-generation 3D imaging and display technology since it was proposed by G. Lippmann for the first time in 1908. This is because it can provide real 3D images with full parallax, full color and continuous viewing points [7-12]. Basically, an integral imaging system consists of two parts; pickup and display. In the pickup part, the rays coming from a 3D object through a lenslet array is optically recorded as elemental images representing different perspectives of a 3D object. On the other hand, in the display part, the recorded elemental images are displayed on a display panel and then the 3D image can be optically reconstructed and observed through a display lenslet array.

In this paper, we propose a novel interactive integral imaging system using vision-based 3D fingertip detection interface. This system consists of the real 3D image generation system based on integral imaging and interaction system using real-time 3D fingertip detection device. The proposed system can be used in effective human computer interaction method that can be applied in realizing virtual system.

978-0-7695-3575-3/09 $25.00 © 2009 IEEE
DOI 10.1109/CISIS.2009.123

2. Review of integral imaging technique

In general, a conventional integral imaging system consists of two processes; pickup and display as shown in Fig. 1. In the pickup process, intensity and direction information of the rays coming from a 3D object through a lenslet array is optically recorded by use of a CCD camera as elemental images representing different perspectives of a 3D object. On the other hand, in the display process, the recorded elemental images are displayed on a display panel such as the liquid crystal display (LCD) and then the 3D image can be optically reconstructed and observed through a display lenslet array as shown in Fig. 1(b).

In the integral imaging system, the viewing angle is the maximum angle where a real 3D image can be viewed freely without any distortion [20, 21]. Figure 2 shows the geometry of calculating the viewing angle of integral imaging. In Fig. 2, Ψ means the diverging angle of the integrated point. Moreover d, p and g represent the viewing region, the pitch of the elemental lenslets, and the distance between the lenslet array and the image sensor, respectively. Then, the diverging angle of Ψ can be obtained from the relationship of (1).

$$\psi = 2\arctan\left(\frac{p}{2g}\right) \tag{1}$$

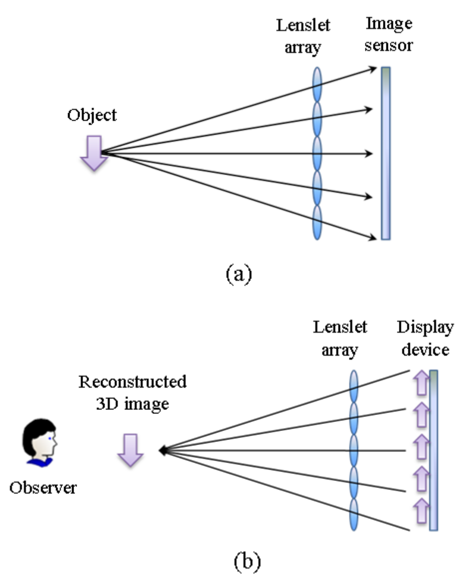


Figure 1. Operational principle of integral imaging

As you can see in (1), as the pitch of the elemental lenslet p is getting larger and the distance between the lenslet array and the image sensor g is getting smaller, the viewing angle Ψ will get wider. When the same lenslet arrays are used in the pickup and display processes, the pickup angle is just equal to the viewing angle.

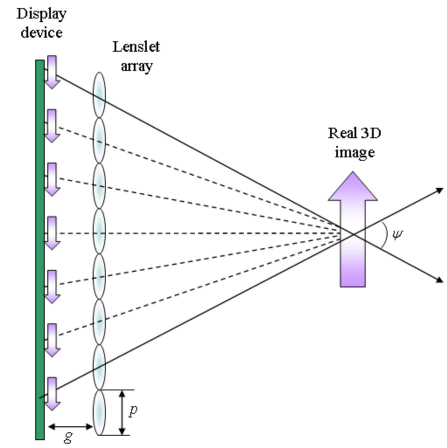


Figure 2. Viewing angle of the integral imaging system

3. Interactive integral imaging system

Figure 3 shows proposed interactive integral imaging system using vision-based 3D fingertip detection interface. The proposed system is divided into two parts. First part is the 3D image generation system based on integral imaging technique. Here the real 3D image is generated in real image field by using a lenslet array. Second part is interaction system using real-time fingertip detection device which is a vision-based hardware for detecting the user's finger motion.

3.1. Generation of real 3D images

To implement an interactive integral imaging system, real 3D images should be display in a space. In general, the integral imaging can provide a real 3D image with a real depth in the space. That is, the real 3D image can be viewed directly in a free space without a need of any special glasses.

So far, there are two kinds of integral imaging systems according to the gap distance between the lenslet array and the display panel depth-priority integral imaging (DPII) and resolution-priority integral imaging (RPII) systems [13]. Here the DPII system can be implemented by setting the gap equal to the focal length of the lenslets. The DPII system can provide 3D images with low-resolution and large-depth through both real and virtual image fields. On the other hand, the RPII system can be obtained when the gap distance is not equal to the focal length of the lenslets. This RPII system can give us 3D images with high-resolution and small-depth contrary to the DPII system.

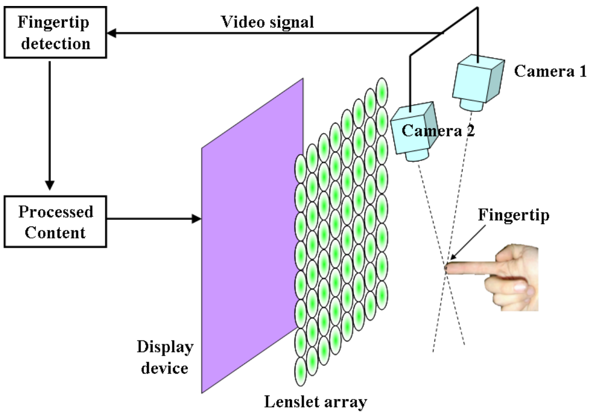


Figure 3. Proposed interactive integral imaging system

For implementing our interactive integral imaging system, the real 3D image may be displayed within a large depth. Thus, the DPII scheme is much proper in our system. In DPII, elemental images for an object to be displayed can be also digitally calculated. That is, computational modeling of integral imaging can be performed based on the ray-optics by using a virtual pinhole array instead of a lenslet array. Here, elemental images can be synthesized by recording the rays coming from 3D object through a virtual pinhole array. Figure 4 shows the schematics of a pinhole array-based computational pickup method. Rays are mapped inversely through each pinhole on the pickup plane. Suppose the kth pinhole, then the light intensity of the kth elemental image from the 3D object, $O(x_k, y_k)$ becomes

$$E(x_k, y_k, -g) = O\left(\frac{x_k g}{-z}, \frac{y_k g}{-z}, z\right) \qquad (2)$$

where x_k and y_k are the local axes for the kth pinhole. Entire elemental images of the 3D object can be obtained by repeating this for all the pinholes in the same way.

After obtaining the elemental images, a real 3D image is obtained by use of lenslet array and elemental images as shown Fig. 1(b). Each elemental image is projected through the corresponding lenslet and integrated as real 3D image with real depth in space. Figure 5 shows the elemental images for presenting real 3D image. By displaying it in the display panel, the real 3D image is reconstructed in front of lenslet array.

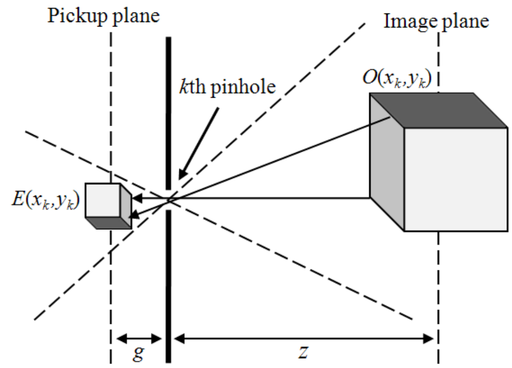


Figure 4. Schematic of the pinhole array-based computational pickup method

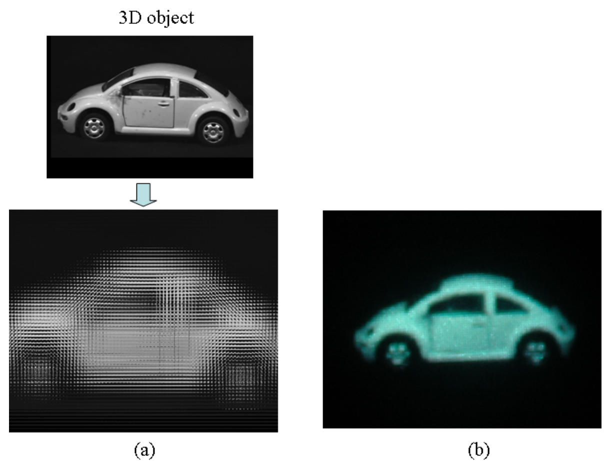


Figure 5. (a) Synthesized elemental images of 3D object (b) Real 3D image reconstructed from lenslet array

3.2. Fingertip detection interface

Our interaction system, which is combined with 3D image generation system, is based on the stereo vision technique. To allow our vision-based interface to have real time performance, we used simple processing of the initial images. Our strategy was first to extract the binary images of the finger by thresholding the luminance of the captured images. Then by working on the simple searching algorithm, we were able to compute the fingertip positions for both images. Based on the extracted positions we calculate the 3D position of finger. Finally, we can display real 3D image again by synthesizing new elemental images using the 3D position information. By repeating this process, we implement our fingertip tracking system. The detailed diagram is shown in Fig. 6.

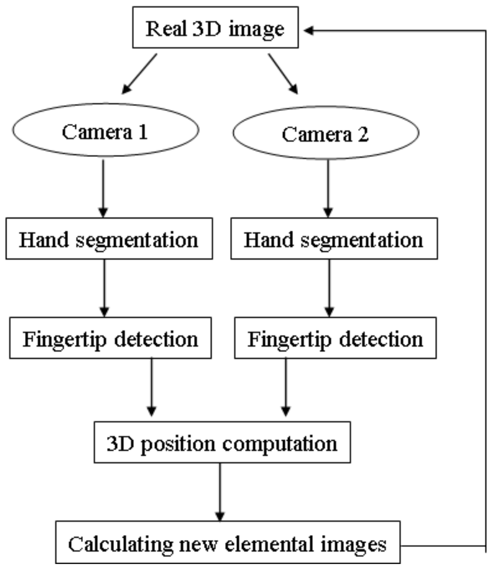


Figure 6. Diagram of the fingertip tracking interface

Hand segmentation is the first step in our fingertip detection interface. For real-time process, fast segmentation is needed without assumptions on the background color. Various skin color detection methods have been now commonly used in hand segmentation. We have chosen to detect skin color pixels in the YCbCr color space, with the fast and simple approach presented in [14]. Figure 7 shows an example of our hand segmentation. In order to reduce the noise effect, a median filter is applied.

Next step is fingertip detection. When the finger enters in the cameras field of view, it is necessary to detect accurately the fingertip position in order to compute 3D position for our fingertip tracking. Let us consider the case of pointing gesture in this paper. And we assume that only one finger is pointing and so the fingertip is the point located at the extremity of the hand region. With the hand's binary images obtained from the previous step, we search the boundary of the hand region. The fingertip is the point of the boundary that maximizes the distance from the center of gravity of the hand region. The center of gravity is obtained with the computation of the geometrical moments. Then we can detect the positions of fingertip in both images, which is denoted by (x_1, z_1) and (x_2, z_2), respectively.

With the position of the fingertip in each of both images, we can compute its 3D position [15]. Figure 7 shows the convergent stereo camera model used in our experiments. The x axis is defined by the baseline denoted by b. And f is focal length of the camera. It is assumed that rotation of either camera occurs about an axis parallel to the z axis. The convergent angles are denoted by θ_1 and θ_2. Then, we can calculate the 3D coordinates of a point (X, Y, Z) using two positions (x_1, z_1) and (x_2, z_2). This is given by

$$X = \frac{b(x_1 \cos\theta_1 + f \sin\theta_1)(x_2 \cos\theta_2 + f \sin\theta_2)}{T} \quad (3)$$

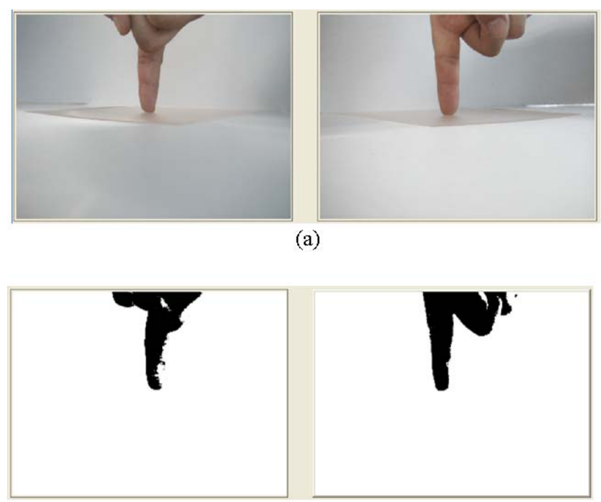
$$Y = \frac{b(-x_1 \cos\theta_1 + f \sin\theta_1)(x_2 \cos\theta_2 + f \sin\theta_2)}{T} \quad (4)$$

$$Z = \frac{bz_1(x_2 \sin\theta_2 + f \cos\theta_2)}{T} \quad (5)$$

where

$$T = f(x_1 - x_2)\cos(\theta_1 + \theta_2) + x_1 x_2 \sin(\theta_1 + \theta_2) + f^2 \sin(\theta_1 + \theta_2).$$

By using (3)-(5), we can get the 3D coordinate of fingertip. The extracted 3D coordinate is used to generate the new elemental images for tracking of real 3D image.



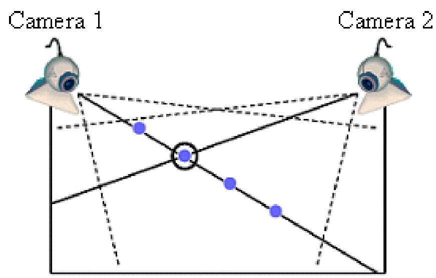
(a)

(b)

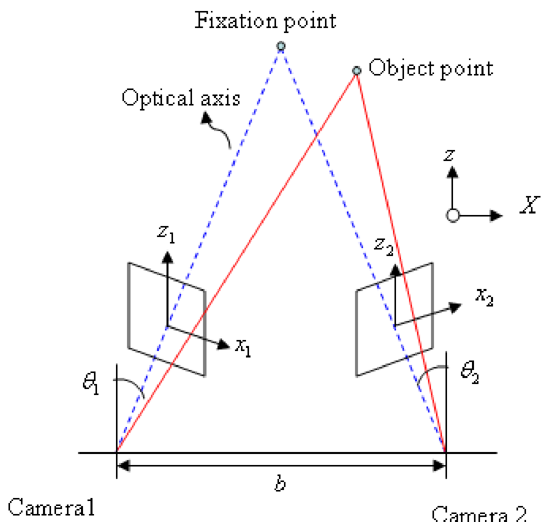
Figure 7. (a) Original image (b) Segmented images.

4. Experiments and results

To prove the usefulness of the proposed system, the preliminary experiments were performed. The whole implemented system for the proposed system is shown in Fig. 9. The integral imaging display system was implemented based on the DPII structure. Lenslet array has 150×150 elemental lenses whose focal length is 3 mm and diameter is 1.08 mm. The gab between the LCD panel and the lenslet array was 3 mm. The elemental images were displayed in a LCD panel whose pixel number is 1024 by 768. The initial elemental images were synthesized to display real 3D image in both the center position of LCD panel and 21 mm at the depth. When 3D object and background image are 'boat' located at *z*= 18 mm and 'beach' at *z*=-15 mm, the elemental images are shown in Fig. 10, respectively.



(a)



(b)

Figure 8. Convergent stereo camera geometry

After displaying the real 3D image at *z*=18 mm, we performed experiments for fingertip detection using two cameras as shown in Fig. 9. We use two common webcams with 640×480 image resolution and the frame-rate of 30 fps. According to the algorithm shown in Fig. 6, the fingertip detection process was carried out. Figure 11 shows our implemented user interface based on Visual C++ and a picture for fingertip detection using the implemented hardware system. The position of fingertip was well tracked and thus the 'boat' 3D image was moved according to the corresponding information.

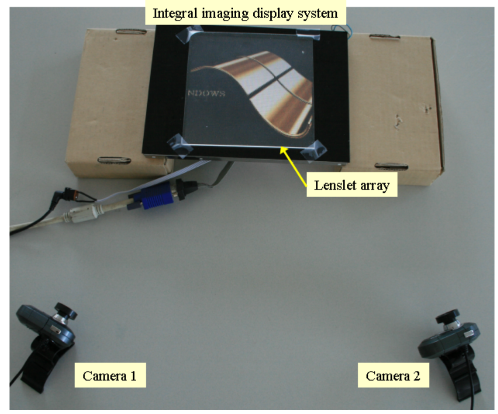
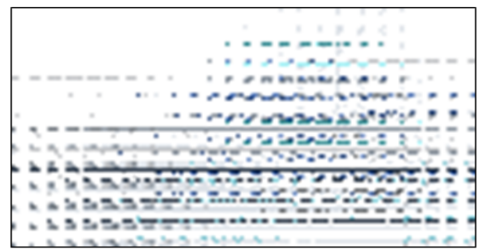
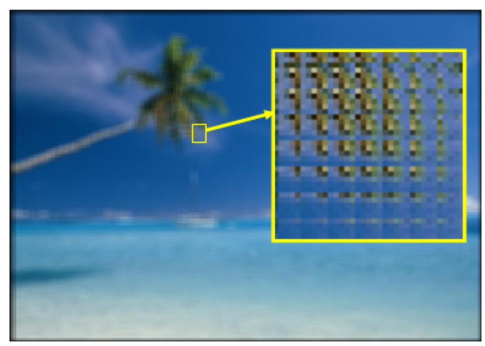


Figure 9. Experimental system structure



(a)



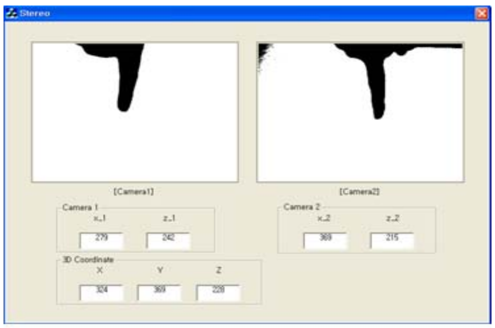
(b)

Figure 10. Elemental images used in the experiment (a)'boat' (b) 'beach'

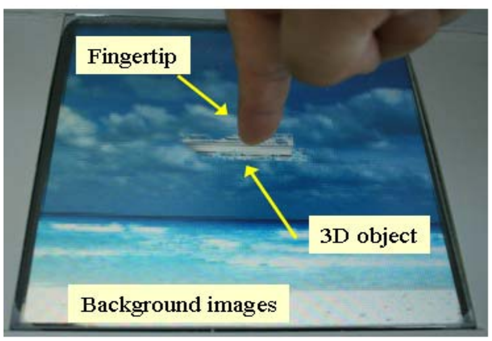
5. Discussion and conclusion

Although the preliminary results are quite satisfactory for our interactive integral imaging system, several aspects of our interface need further improvement. For examples, the quality of the fingertip tracking relies heavily on the vision-based segmentation algorithm. Form this point of view, the fingertip are only of limited range of the implemented system. And the multiple fingertips may create new problem of image detection. Due to the difficulty of these problems, a lot of research needs to be done for effective HCI interface.

In conclusion, a novel interactive integral imaging system using vision-based 3D fingertip interface was proposed and demonstrated. The proposed system consisted of the real 3D image generation system based on integral imaging technique and the interaction device using real-time fingertip detection interface. To demonstrate our proposed system, we performed the preliminary experiments and the results were presented. We expect that the proposed system can be used in effective human computer interaction method for real 3D images.



(a)



(b)

Figure 11. (a) Implemented user interface (b) Experiments for interaction

Acknowledgements

This research was supported by the MKE (Ministry of Knowledge Economy), Korea, under the ITRC (Information Technology Research Center) support program supervised by the IITA (Institute of Information Technology Assessment) (IITA-2008-C1090-0801-0018).

References

[1] Pavlovic, V.I., Sharma, R., Huang, T.S. “Visual Interpretation of Hand Gestures for Human-Computer Interaction: a Review,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 19, no. 7, pp. 677-695, 1997.

[2] Ebert, D., Bedwell, E., Maher, S., Smoliar, L., and Downing, E. “Realizing 3D visualization using crossed-beam volumetric displays,” Communications of the ACM, vol. 42, no. 8, pp. 101-107, 1999.

[3] Adcock, M., Hutchins, M., and Gunn, C. “Augmented Reality Haptics: Using ARToolKit for Display of Haptic Application,” 2nd International Augmented Reality Toolkit Workshop, Tokyo, Japan. 2003.

[4] Balakrishnan, R., Fitzmaurice, G., & Kurtenbach, G. “User interfaces for volumetric displays,” IEEE Computer. March 2001. p. 37-45. 2001.

[5] S. Pastoor and M. Wopking “3D displays: A review of current technologies,” Displays, vol. 17, no 2, pp. 100-110, 1997.

[6] M. Halle, “Autostereoscopic displays and computer graphics,” Computer Graphics, ACM SIGGRAPH, vol. 31, no. 2, pp. 58-62, 1997.

[7] G. Lippmann, “La photographic integrale,” C.R. Acad. Sci. Vol. 146, pp. 446-451, 1908.

[8] F. Okano, H. Hoshino, J. Arai, and I. Yuyama, “Three-dimensional video system based on integral photography,” Opt. Eng. vol. 38, pp. 1072-1077, 1999.

[9] B. Lee, S. Y. Jung, S.-W. Min, and J.-H. Park, “Three-dimensional display by use of integral photography with dynamically variable image planes,” Opt. Lett. vol. 26, pp. 1481-1482, 2001.

[10] J.-S. Jang and B. Javidi, "Improved viewing resolution of three- dimensional integral imaging by use of nonstationary micro-optics," Opt. Lett., vol. 27, pp. 324-326, 2002.

[11] A. Stern and B. Javidi, “3D image sensing, visualization, and processing using integral imaging,” Proc.of IEEE vol. 94, pp. 591-608, 2006.

[12] D.-H. Shin, B. Lee, and E.-S. Kim, "Multidirectional curved integral imaging with large depth by additional use of a large-aperture lens," Appl. Opt., vol. 45, pp. 7375-7381, 2006.

[13] J. Jang, F. Jin, and B. Javidi, "Three-dimensional integral imaging with large depth of focus by use of real and virtual image fields," Opt. Lett., vol. 28, pp. 1421-1423, 2003.

[14] D. Chai, K. Ngan, "Face segmentation using skin-color map in videophone applications," IEEE Trans. On Circuit and Systems for Video Technology, vol. 9, pp. 551-564, 1999.

[15] E. Krotkov, K. Henriksen and R. Kories, "Stereo raning with verging cameras," IEEE Trans. Pattern Anal. Machine Intell., vol. 12, pp. 1200-1205, 1990.