Vision-Based Remote Control System by Motion Detection and Open Finger Counting

Daeho Lee[1] and Youngtae Park[2]

Abstract — *In this paper, we present a universal remote control system based on computer vision. The method is composed of two stages of detecting visual evidences. Motion and skin color information is utilized to detect waving hands requesting control commands. Upon the request, the camera is controlled to zoom in on local region of the hand. The number of open fingers is counted by the shape analysis on the segmented hand region image. Control command is issued when a predetermined sequence of gesture state transition is produced. Experimental results show that the shape of open fingers exhibits strong features for determining correct gesture states. The use of stable features on consecutive frames yields robust and accurate performance regardless of operating conditions*[1].

Index Terms — Remote control, computer vision, gesture, open finger counting, human-machine interaction.

I. INTRODUCTION

Infrared remote controllers have been widely used in most of home electronics. Most of users have difficulties of finding the appropriate controller among the increasing number of such controllers. An integrated and more convenient means has been devised to replace the existing remote control system (RCS) [1], [2]. Voice or gesture recognition has been recently adopted to implement the universal RCS. Voice recognition [3]-[5], however, may fail to provide robust performance because of the speaker variation and surrounding noise. RCSs using gloves or markers [6]-[8] are based on the gesture recognition. They may provide accurate control performance, but the use of the gloves or markers makes the users uncomfortable. More natural way to use human actions is required for commercially worthy RCS. Vision-based approaches to recognize the hand gestures or postures [1], [9]-[11], may be exemplified.

A remote control mode is set by detecting a hand pointing toward an appliance to be controlled in [1], [9]-[12]. Then the local hand region is tracked to recognize unique hand gestures. To detect pointing hand, 3-dimensional coordinate is computed in a framework of stereo vision using two or more cameras. In stereo vision, correct correspondences may not be found for occluded regions or for the regions without textures [13]-[15]. Also, it is well known that the 3-dimensional acquisition using the stereo vision may cover only a limited angle of view.

Waving hands are detected by frequency analysis to set the remote control mode in [16]. After the appliance is selected, hand gestures are recognized by an HMM (or neural network) using region information such as shape features. Control commands are driven by pointing menus displayed on the television screen in [17]. Most of the methods assume that the user is wearing shirts with long sleeves [1], [2], [11], [12].

This paper is aimed at developing a new scheme that can overcome most of the drawbacks of existing methods, and thereby can be used for implementing commercial RCSs with high performance and low-cost. We use one zoom camera which is mounted below the ceiling. When local motion regions with skin color are found, the remote control mode is invoked, and then the camera zooms in on the hand region. The open fingers are detected using the boundary analysis on hand regions. The gesture state is determined when the number of open fingers is found to be stable at consecutive frames. Control commands can be defined by a set of state transitions. The processing speed is fast enough for real-time operation.

II. ORGANIZATION OF THE PROPOSED RCS

Vision-based RCS is organized as illustrated in Fig. 1. If users request remote control commands by waving hands, the request is recognized by analyzing the image sequences acquired from the camera in the control box. Then the interactive menus of current state (remote control mode, TV control mode, volume control mode, channel control mode, etc) are displayed on the menu display. Remote control signals of recognized control gestures are transmitted to the designated appliance. A computer monitor, a color LED (light-emitting diode) array or an overlaid television may be used for the menu display.

Overall scheme of our method is shown in Fig. 2. Initially the camera is zoomed out to acquire images of wide view angle. And the system initially begins with r_0 = false, where r_t denotes a remote control mode at time t. When an appropriate control motion is detected, the system enters into the control mode, i.e., r_t = true. Both motion and color information is utilized to detect correct hand gestures. Details are described in the next section. Once the control motion is confirmed, the camera zooms in on the detected local region of interest. An example of zooming control is shown in Fig. 3.

When r_t = true, a sequence of zoomed images is processed to detect the number of open fingers. The number is used as a gesture state. Combinations of the gesture state transitions

[1]Daeho Lee is with College of Liberal Arts, Kyung Hee University, Korea (e-mail: nize@khu.ac.kr).

[2]Youngate Park, the corresponding author, is with College of Electronics and Information, Kyung Hee University, Korea (e-mail: .ytpark@khu.ac.kr).

Contributed Paper
Manuscript received October 13, 2009

provide control commands as exemplified in Table 1. The control box transmits the recognized control command to an electronic appliance. If the Exit command is obtained, the system enters into r_0 = false, and the camera zooms out.

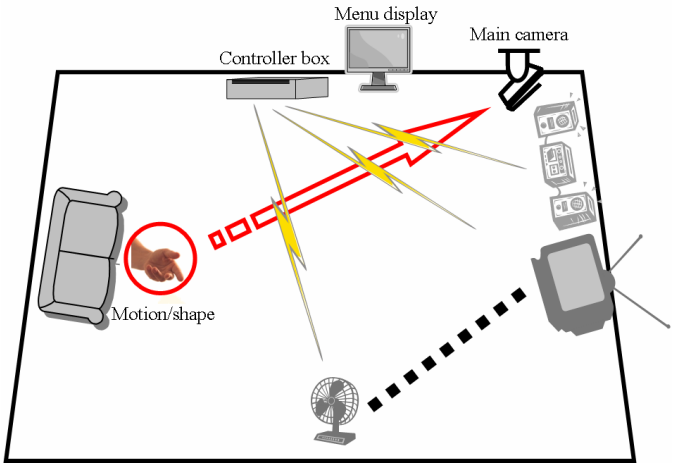


Fig. 1. Organization of the proposed RCS.

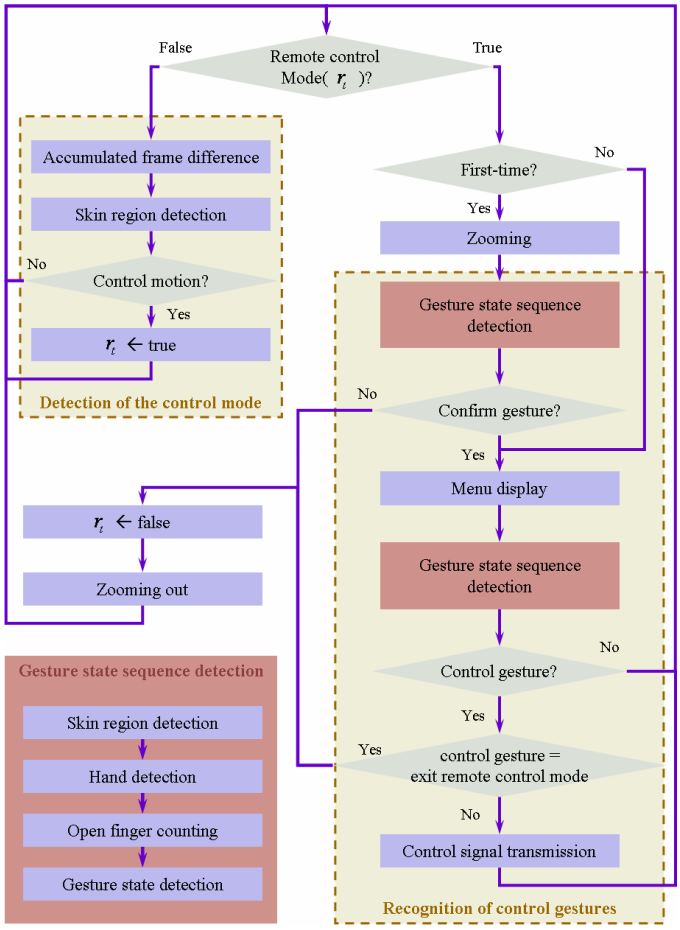


Fig. 2. Overall scheme of the proposed RCS.

An example of detecting control motion and gesture states are shown in Fig. 3. We describe the details in the following sections.

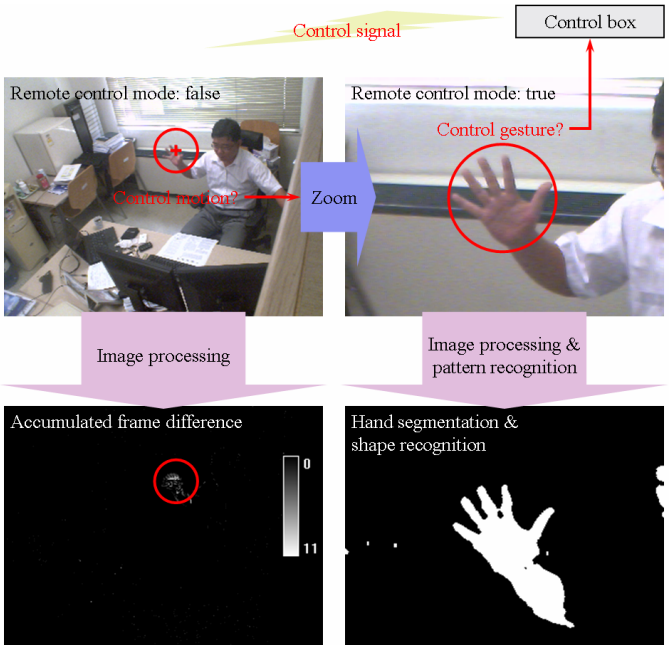


Fig. 3. Brief processing illustration of the proposed RCS.

III. Detection of the Control Mode

In the non remote control mode (r_t = false), waving hands requesting the remote control mode is detected. Movement information obtained from the frame difference may be too sensitive to user's motion. Thus, we use accumulated frame difference defined by

$$AD_t(x,y) = \begin{cases} AD_{t-1}(x,y)+1 & \text{if } |I_t(x,y) - I_{t-1}(x,y)| > \tau_d \\ AD_{t-1}(x,y)-1 & \text{otherwise} \end{cases}, \quad (1)$$

where I_t denotes the intensity of the image at time t, and τ_d a threshold. Pixels associated with movement are characterized by large values of AD_t. Candidate region of movement M_t can be found by selecting pixels having sufficiently large value of AD_t.

$$M_t(x,y) = \begin{cases} 1 & \text{if } AD_t(x,y) > \tau_m \\ 0 & \text{otherwise} \end{cases}, \quad (2)$$

where τ_m denotes a threshold. To suppress clutter movement due to abrupt change of illumination, we additionally use unique color information of skin.

Skin regions S_t can be detected by color processing [18], [19], as follows;

$$S_t(x,y) = \begin{cases} 1 & \text{if } \min\left(R_t(x,y) - G_t(x,y), R_t(x,y) - B_t(x,y)\right) > \tau_s \\ 0 & \text{otherwise} \end{cases}, \quad (3)$$

where R_t, G_t and B_t denote red, green and blue components, respectively, and τ_s a threshold.

As shown in Fig. 4, motion regions detected by the accumulated frame difference are compared to the skin regions. The motion regions associated with the skin regions are determined as the control motion region. Specifically, we

use the following condition:

$$\left(B_S^i \cap B_M^k\right) / B_S^i > \gamma\,, \tag{4}$$

where B_S^i denotes the i th label of S_t, B_M^k denotes the k th label of M_t, and γ denotes a threshold.

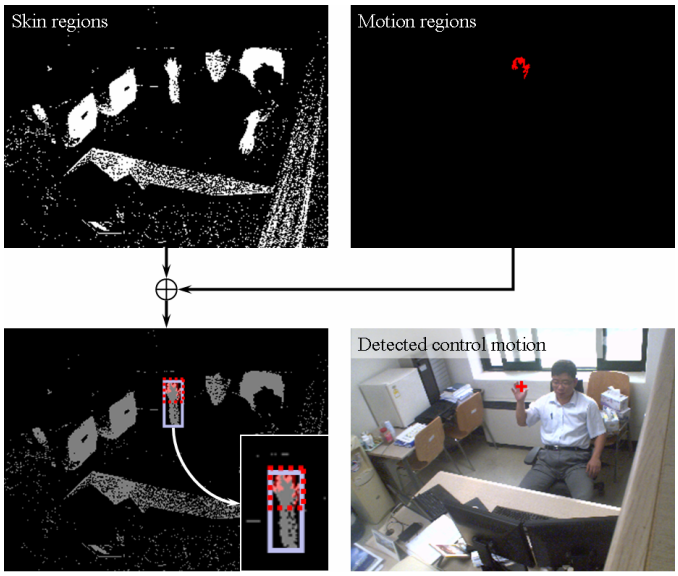


Fig. 4. Control motion detection.

If there exists a labeled region satisfying (4), the system enters into the remote control mode, i.e. r_t = true, and the gesture states are detected to issue control commands. If desired hand gestures are not obtained in a few second, r_t is set to false, and the camera zooms out.

IV. Recognition of Control Gestures

In the remote control mode (r_t = true), the camera zooms in on the detected control motion, and the number of open fingers is counted to determine the gesture states. The hand region image is segmented using the skin color information given by (3). The hand region can be obtained easily by applying an image labeling procedure. The isolated hand region is located at the center of the image, since the camera zooms in on the hand control motion. An example is shown in Fig. 5.

Once the hand shape is obtained, open fingers may be detected by finding finger tips, which are usually characterized by high value of curvature. We utilize k-cosine [20], [21] functions to measure the curvature. Let C be a boundary point set, $p_0,\cdots,p_{i-1},p_i=(x_i,y_i),\cdots,p_n$, of the hand shape. The k-vectors at the point p_i are define as

$$\begin{aligned}\mathbf{a}_{ik} &= (x_i - x_{i+k}, y_i - y_{i+k}) \\ \mathbf{b}_{ik} &= (x_i - x_{i-k}, y_i - y_{i-k})\end{aligned}. \tag{5}$$

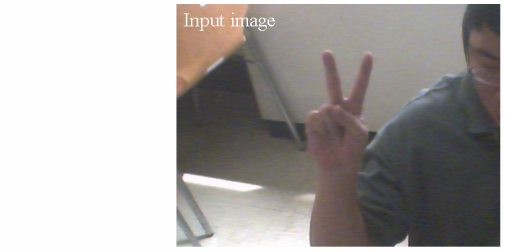
The k-cosine, which is the cosine of the angle between $\mathbf{a}_{ik}$ and $\mathbf{b}_{ik}$, is given by

$$\cos p_i^k = \frac{\mathbf{a}_{ik}\cdot\mathbf{b}_{ik}}{|\,\mathbf{a}_{ik}\,||\,\mathbf{b}_{ik}\,|}\,. \tag{6}$$

The convex points associated with the local maxima of k-cosine represent finger tips. Finger tips can be discriminated from the valleys by comparing the distance from the centroid. Let p_i denote a boundary point associated with the local maxima, and d_i the distance from the centroid of the hand region. The finger tip condition is

$$d_i > (d_{i-k} + d_{i+k})/2\ . \tag{7}$$

An example of detecting the finger tips is shown in Fig. 6. The number of open fingers is used as the gesture states. Depending on operating conditions like illumination variation, the gesture states might suffer from noisy random variations. To enhance the robustness, only stable states in several consecutive frames are used as confirmed states.



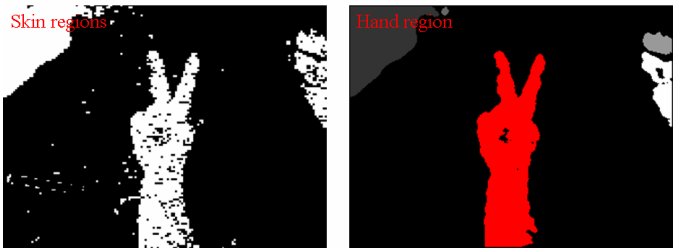
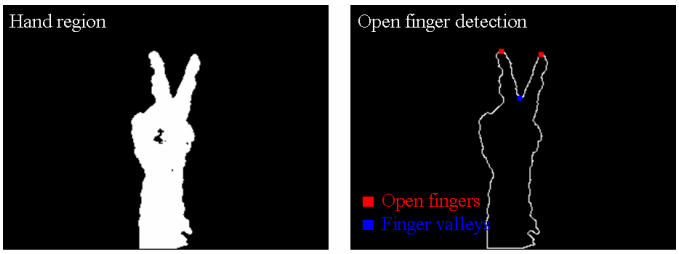


Fig. 5. Hand region detection.



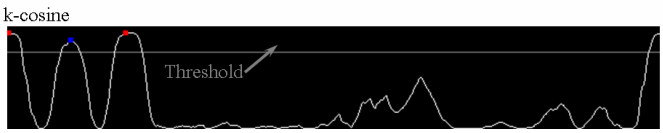


Fig. 6. Open finger detection.

Control commands can be obtained by combinations of gesture states as shown in Fig. 7. Gesture states are stored into a gesture queue, and the control commands are generated by the state sequences such as exemplified in Table 1, where control commands required can be defined by variable-length sequences. A hierarchical menu structure can be constructed depending on user's needs as in Fig. 8

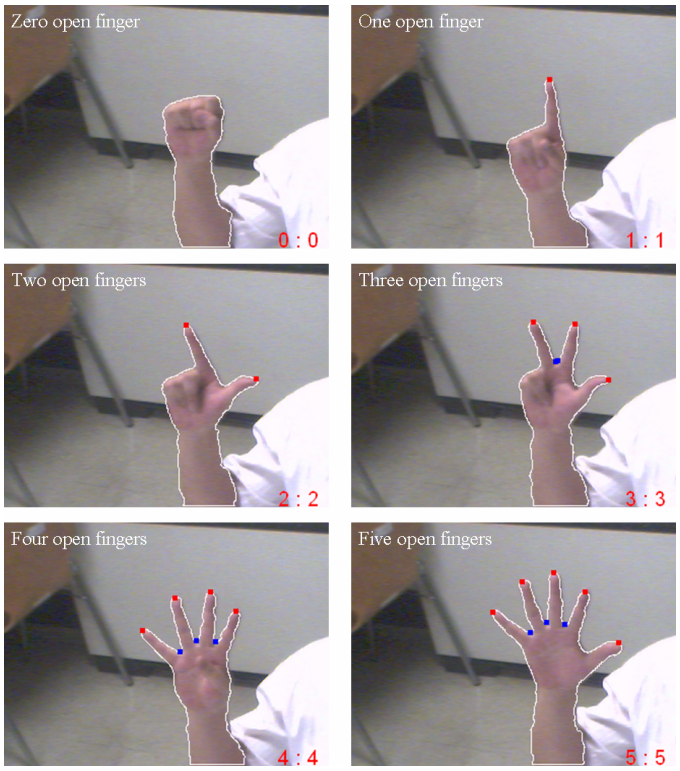


Fig. 7. Example of finger detection.

Computer monitors, color LED arrays or overlaid televisions can be used for the menu display. Fig. 9 shows an example of the menu display on the overlaid television.

TABLE I
EXAMPLE OF GESTURE STATE SEQUENCES AND CONTROL COMMANDS

Gesture State Sequences (Number of Open Fingers)	Control Gestures
1→ 0	Select 1
2→ 0	Select 2
3→ 0	Select 3
4→ 0	Select 4
1→ 5→ 0	Previous menu
2→ 5→ 0	Exit remote control mode
…	…

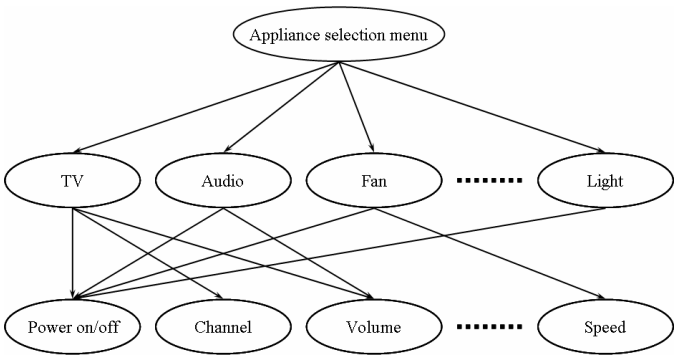


Fig. 8. Example of a menu hierarchy.

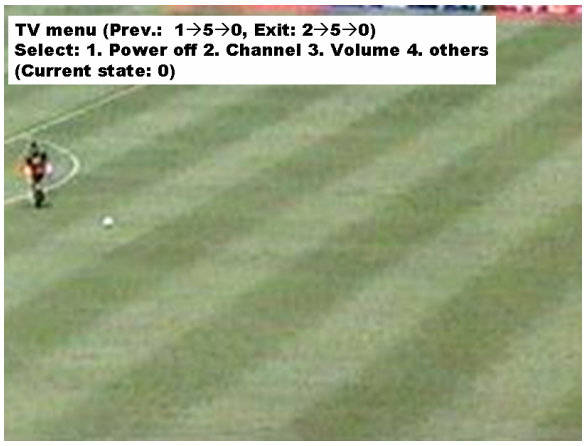


Fig. 9. Example scene of a menu display on the overlaid television.

V. EXPERIMENTAL RESULTS

The proposed RCS was implemented in Visual C++, and tested on a Pentium PC (Core™ 2 Duo, 2.40GHz). A Web camera (LifeCam VX-6000) mounted below the ceiling was used. The image resolution is 320×240.

Fig. 10 shows an example of detecting the remote control mode, where the left-top numbers of the accumulated frame difference images denote the frame numbers and the red cross of the right-bottom image indicates the maximum *AD* point of the detected motion region.

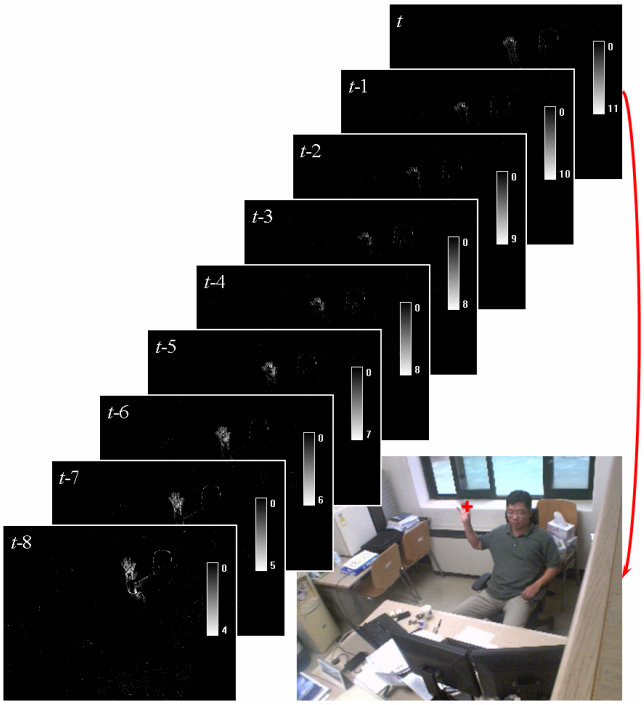


Fig. 10. Detection example of the remote control mode.

The use of color information is illustrated in Fig. 11, where the maximum *AD* points (indicated by yellow crosses) are suppressed because corresponding skin color is not found.

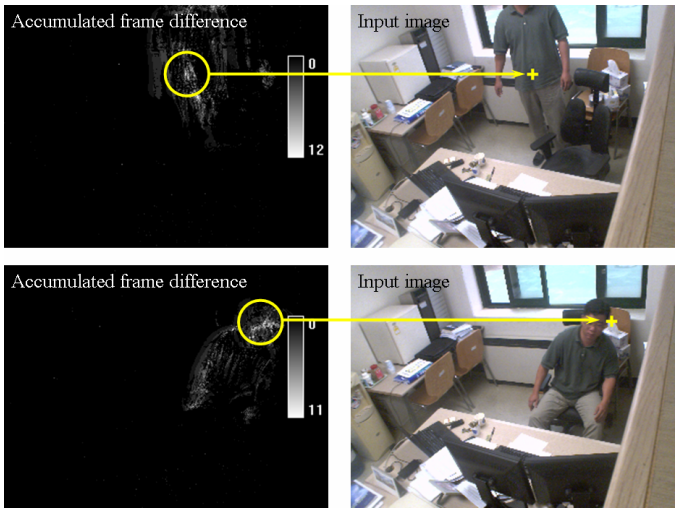


Fig. 11. Accumulated frame differences by big movements.

An example of the gesture state transition (2→5→0) is shown in Fig. 12. Numbers at the top of images denote the frame number, and bottom two numbers denote the number of open fingers and the gesture state, respectively. Only stable states over a number of consecutive frames (say τ_a frames) are validated to suppress noisy states. We set $\tau_a = 4$ in the experiments. Note that the gesture state at $t+11$ th frame is suppressed since consecutively stable features are not found.

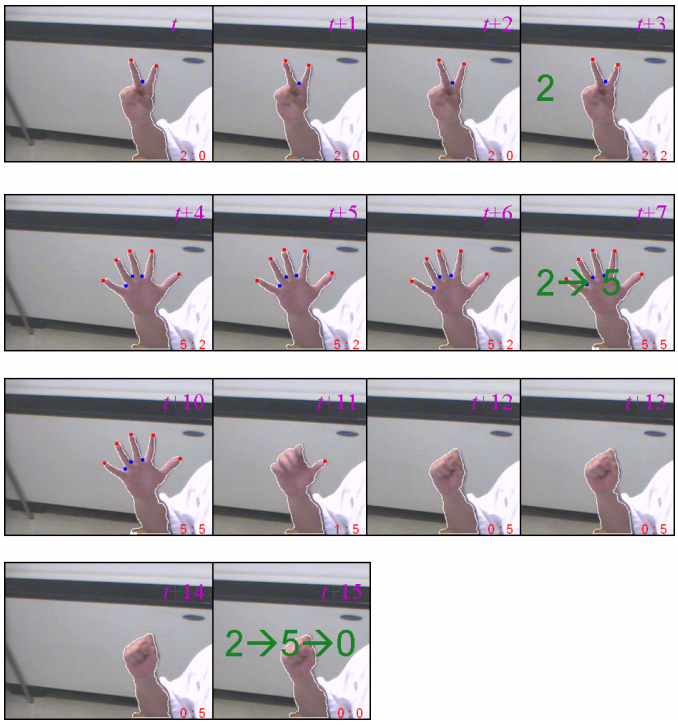


Fig. 12. Recognition result of a control gesture (2→5→0).

False detections of the remote control mode were observed at a few times for two hour tests. These errors, however, were corrected automatically by the test of the gesture confirmation. Recognition results of finger counting are shown in Table 2. The false classifications did not lead to wrong control commands, since the gesture sequence transitions did not match to one of the registered commands, and thereby were rejected. Registered sequences given in Table 2 were tested ten times for each, where incorrect command was not found to be issued. This result shows the strong capability of error correction.

TABLE II
RECOGNITION RESULT OF FINGER COUNTING

Number of Open Fingers	Correct/Total (Accuracy rates)
0	349/350 (99.71 %)
1	73/73 (100 %)
2	218/220 (99.09 %)
3	76/76 (100 %)
4	28/31 (90.32 %)
5	85/96 (88.54 %)

It is difficult to compare the proposed method with others because of the different test environments. Our method, however, has advantages over other methods as follows; a low-cost single camera can be used, users may not wear the shirts with long sleeves, and the capability of error handling results in robust performance.

VI. CONCLUSION

We presented a new vision based RCS for controlling electronic appliances. Remote control mode is initiated by detecting waving hands. To remove false initiation, movement regions associated with skin color are selected. Once accepting the request, the camera is controlled to zoom in on local region of the hand. A corner detection algorithm based on the k-cosine function is utilized to find the finger tips in the zoomed local image. The gesture state is determined by the number of open fingers. If a predetermined sequence of gesture state transition is found in the state transition queue, the corresponding control command is issued. Since stable features on consecutive frames are retrieved, robust and accurate performance may be guaranteed regardless of operating conditions. Using the proposed method, universal RCS's with high performance and low-cost may be realized.

REFERENCES

[1] J. Do, J. J. Jung, S. H. Jung, H. Jang and Z. Bien, "Advanced soft remote control system using hand gestures," *MICAI (Advances in Artificial Intelligence) 2006, LNAI*, vol. 4293, pp. 745-755, 2006.

[2] P. Premaratne and Q. Nguyen, "Consumer electronics control system based on hand gesture moment invariants," *IET Computer Vision*, vol. 1, no. 1, pp. 35-41, Mar. 2007.

[3] K. M. Fludd, W. Pruehsner and J. D. Enderie, "Multi remote appliance controller," *Proc. IEEE 27th Annual Northeast Bioengineering Conf.*, pp. 87-88, 2001.

[4] K. Fujita, H. Kuwano, T. Tsuzuki, Y. Ono and T. Ishihara, "A new digital TV interface employing speech recognition," *IEEE Trans. Consumer Electron.*, vol. 49, no. 3, pp. 765-769, Aug. 2003.

[5] H. Jiang, Z. Han, P. Scucces, R. Robidoux and Y. Sun, "Voice-activated environmental control system for persons with disabilities," *Proc. IEEE 26th Annual Northeast Bioengineering Conf.*, pp. 167-168, 2000.

[6] N. X. Tran et al., "Wireless data glove for gesture-based robotic control," *Human-Computer Interaction, Part II, HCII 2009, LNCS*, vol. 5611, pp. 271-280, 2009
[7] D.J. Sturman and D. Zeltzer, "A survey of glove-based input," *IEEE Comp. Graph. Appl.*, vol. 14, no. 1, pp. 30-39, Jan. 1994.
[8] T. Baudel and M. Baudouin-Lafon, "Charade: remote control of objects using free-hand gestures," *Communications of the ACM*, vol. 36, no. 7, pp. 28-35, Jul. 1993.
[9] C. Colombo, A. D. Bimbo and A. Valli, "Visual capture and understanding of hand pointing actions in a 3-D environment," *IEEE Trans. Syst. Man Cybern. B*, vol. 33, no. 4, pp. 677-686, Aug. 2003.
[10] S. Sato and S. Sakane, "A human-robot interface an interactive hand pointer that projects a mark in the real work space," *Proc. 2000 IEEE int. Conf. Robotics & Automation*, pp. 589-595, 2000.
[11] M. Kohler, "Vision based remote control in intelligent home environments," *3D Image Analysis and Synthesis*, pp. 147-154, 1996.
[12] L. Bretzner, I. Laptev, T. Lindeberg, S. Lenman and Y. Sundblad, "A prototype system for computer vision based human computer interaction," Technical report ISRN KTH/NA/P-01/09-SE, 2001.
[13] A. D. Sappa, F. Dornaika, D. Ponsa, D. Geronimo and A. Lopez, "An efficient approach to onboard stereo vision system pose estimation," *IEEE Trans. Intell. Transp. Syst.*, vol. 9, no. 3, pp. 476-490, 2008.
[14] A. Saxena, J. Schulte and A. Y. Ng, "Depth estimation using monocular and stereo cues," *Int. Joint Conf. on Artificial Intelligence*, 2007.
[15] W. Jiang and Jian Lu, "Panoramic 3D Reconstruction by fusing color intensity and laser range data," *Proc. of the 2006 IEEE Int. Conf. on Robotics and Biomimetics*, pp. 947-953, 2006.
[16] K. Irie and N. Wakamura, "Construction of an intelligent room based on gesture recognition," *Proc. 2004 Int. Conf. Intelligent Robots and Systems*, pp. 193-198, 2004.
[17] F. Shafait, M. Grimm and R. Grigat, "Evaluation of a vision based 2-button remote control for interactive television," *Int. Worksh. Systems, Signals and Image Processeing*, pp. 67-70, 2004.
[18] S. K. Singh, D. S. Chauhan, M. Vatsa and R. Singh, "A robust skin color based face detection algorithm," *Tamkang J. of Science and Engineering*, vol. 6, no. 4, pp. 227-234, 2003.
[19] Y. J. Lee and D. H. Lee, "Research on detecting face and hands for motion-based game using Web camera," *2008 Int. Conf. on Security Technology*, pp. 7-12, 2008.
[20] A. Rosenfeld and E. Johnston, "Angle detection on digital curves," *IEEE Trans. Comp.*, vol. 22, no. 9, pp. 875-878, 1973.
[21] T. Sun, "K-cosine corner detection," *J. of Comp.*, vol. 3, no. 7, pp. 16-22, 2008.

Daeho Lee received M.S. and Ph.D. degrees in Electronic Engineering from Kyung Hee University, Seoul, Korea, in 2001 and 2005, respectively. He is an Assistant Professor in the College of Liberal Arts at Kyung Hee University. His research interests include computer vision, computer games, ITS, and signal processing.

Youngtae Park received the B.S. degree in Electronic Engineering from Seoul National University, Seoul, Korea in 1979, the M.S. degree in Electrical Science from Korea Advanced Institute of Science and Technology, Seoul, Korea in 1981, and the Ph.D. degree in Electrical and Computer Engineering from the University of California, Irvine in 1990. He has been a Professor of Electronics Engineering at Kyung Hee University, Korea, since 1992. His research subjects include computer vision, pattern classification, image processing, intelligent transportation systems, and vision-based human identification.