

# Human Body Tracking and Pose Estimation Using Modified Camshift Algorithm

Seung-Jun Hwang, Jae-Hong Min, In-Gyu Kim, Seung-Jae Park, Gwang-Pyo Ahn,  
Joong-Hwan Baek

Department of Information and Telecommunication Engineering, Korea Aerospace University, Go-yang, South Korea.  
Email: sj.fogfog@gmail.com

Received 2013

## ABSTRACT

In this paper, we propose multiple CAMShift Algorithm based on Kalman filter and weighted search windows that extracts skin color area and tracks several human body parts for real-time human tracking system. The CAMShift Algorithm we propose searches the skin color region by detecting the skin color area from background model. Kalman filter stabilizes the floated search area of CAMShift Algorithm. Each occlusion areas are avoided by using weighted window of non-search areas and main-search area. And shadows are eliminated from background model and intensity of shadow. The proposed modified Camshaft algorithm can estimate human pose in real-time and achieves 96.82% accuracy even in the case of occlusions.

**Keywords:** Body Tracking; CAMShift; Pose Estimation; Kalman Filter; Weighted Search Windows

## 1. Introduction

Recently, with the spread and development of 3D display, the 3D(three-dimensional) content has been developing. In order to control 3D content, there is a need to develop more convenient and intuitive interface. Therefore, in order to implement an interface that matches with these devices, it is necessary to recognize technology to control the objects on the 3D space. 3D gesture recognition hardware has been developed such as a TOF camera and Kinect. However there is a disadvantage that the price is high compared with webcam[1].

In this paper, we propose an algorithm to track and recognize the body in the picture based RGB. It requires a precise tracking to estimate the posture of the body based on the hands and face. However, the hands and face colors are similar to each other. In addition, in the case of the background color with the color of the skin, the tracking error occurs. In order to solve this problem, we propose CAMShift Algorithm that searches the skin color region by detecting the skin color area from background model[2,3]. At this time, we used a Kalman filter to stabilize the detection area of CAMShift. In addition, to prevent the loss of the detection area, we add the weights of the main-detection area and the non-detection area. For example, as hands and face overlap with each other, we propose an algorithm to avoid each other obstruction area.

This paper is organized as follows. In Chapter 2, we

describe how to remove the shadow and the Gaussian background model for detection of the body. Chapter 3 describes how to recognize different body parts such as hands, face, elbows, and feet. Chapter 4 explains how the CAMShift algorithm can be used to avoid obstruction during between tracking regions, and applying Kalman filter algorithm for stabilization. Proposed experimental results are provided in Chapter 5, and concluding remarks in Chapter 6.

## 2. Body Detection

### 2.1. Gaussian Background Model

In order to keep track of each part of the body, it is necessary to distinct the body part from the background accurately. We use an adaptive Gaussian background model that can respond adaptively to changes in the background. We set the background for a period of time using a Gaussian probability density function and weights based on a plurality of color model. Applying the Gaussian background model, **Figure 1** shows an example of extraction of the foreground image of the original image.

### 2.2. Shadow Elimination

Data that has passed through the background separation process by the Gaussian background model, has the moving object and shadow. In the normalized RGB color model is capable of comparison of the color without

brightness of pixel. Moreover, in this model, it is possible to calculate the similarity of the color of each pixel in the shadow using the equation (1).

$$r = \frac{R}{T}, g = \frac{G}{T}, b = \frac{B}{T}, T = R + G + B \quad (1)$$

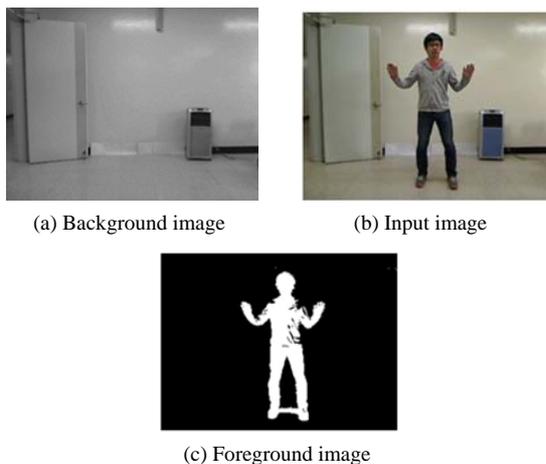
If any pixel from the background is above the threshold, is judged as the area of the shadow, thus the pixel is removed. Through shadow removal, **Figure 2** shows that the shadow around the lower body portion of the foreground image is removed, making the toes exactly separated.

### 3. Recognition of Body Parts

#### 3.1. Hand and Face Recognition

In this paper, in order to detect the skin area, we use the method to extract the skin color and the center of gravity. To detect the skin color from the incoming video, apply certain rules to calculate a numerical distance of skin and non-skin color. Because it defines the boundary value of a certain skin area, fast detection is possible. The input video format is set to RGB, if value for each pixel of the image meets the following equation (2), then it is detected as the skin color[4,5].

$$\begin{aligned} R > 95 & \quad G > 40 & \quad B > 20 \\ \max\{R, G, B\} - \min\{R, G, B\} > 15 & & \\ |R - G| > 15 & \quad R > G & \quad R > B \end{aligned} \quad (2)$$



**Figure 1.** Foreground extraction using Gaussian background model.



**Figure 2.** Result of shadow elimination.

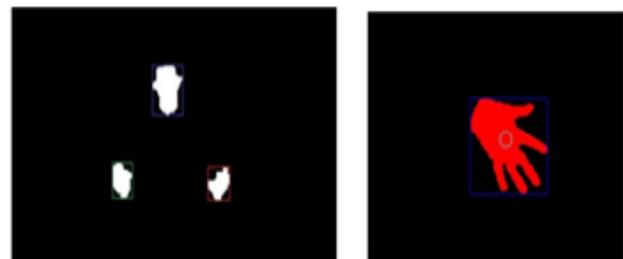
However, in addition to the area of skin to be detected, the noise will appear in other parts. Through morphology operation, these noise components are removed. Set the region with more than a certain size as the region of interest. The coordinates of the detected area is recognized as the center of gravity of the region of interest. **Figure 3** shows the extracted skin color region and the center of gravity.

#### 3.2. Elbow Recognition

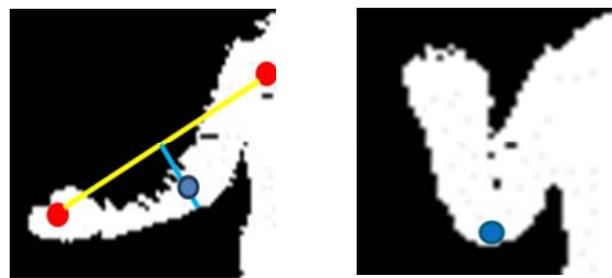
When the lower arm and upper arm are overlapped, the position of the elbow is ambiguous. So calculate separately. At this time, using a Kalman filter makes the position of the elbow to have more stable value. With the arms extended as shown in **Figure 4(a)**, the elbow is present in the normal direction relative to the center of the distance between the hand and shoulder. As shown in **Figure 4(b)** when the upper arm and the lower arm are folded together, the distance between hand and shoulder are relatively shorter than other poses. When the distance is smaller than a certain value, we assume that the arm is folded and the farthest point is considered as the elbow.

#### 3.3. Toe tip Recognition

We have assumed that the foot is in the lowest position in the body to recognize toe. As shown in the Figure, we track the foot from the lower part of shadow-removed foreground image. Like **Figure 5**, in the case of left foot, scan from left to right, and the right foot right to left. When the amount of pixels exceeds a threshold value, then the center of the pixels is the toe tip.



**Figure 3.** Skin color area extraction and center of gravity.



**Figure 4.** Result of elbow tracking.

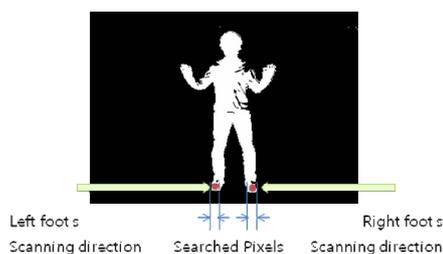
## 4. Body Tracking

**Figure 6** shows KWMCAMShift algorithm's block diagram that applying the weight to avoid obstruction among tracking regions, and adapting Kalman filter algorithm for stabilization.

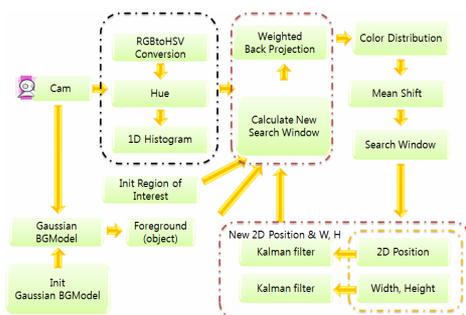
### 4.1. CAMShift Applying the Weighted Search Window

Initial region of the face and hand is specified by the skin color extraction. The color histograms extracted in the initial region are similar to each other, thus diffusion of tracking area by obstruction can occur. In addition, there is no division among the regions and the center point of the tracking area appears in almost the same place. **Figure 7(b)** is a case where the searching area is overlapped and the color distribution of the image area is expanded. We can see that the center point of the track in the original searching area is getting expand to another searching area. Therefore, it is necessary for each other's searching area to prevent expanding if the searching areas are overlapped.

In this paper, As in Equation (3), the filter which adds the weight was designed for tracking region of the previous frame.  $M_g^a$  represents the mean motion vector of the object  $a$ ,  $y^a$  is the next position of tracking object,  $n$  is the number of pixels in the tracking area,  $w$  is the weight of color,  $g$  is the profile of the kernel and  $h$  indicates the window size. In addition, we can do robust tracking in obstruction with eliminating other tracking areas within current tracking area in distribution function.



**Figure 5. Finding toes' ends.**



**Figure 6. Block diagram of KWMCAMShift algorithm.**

$$M_g^a(y_i^a) = \frac{\sum_{i=1}^n x_i^a w_i g_i \left( \left| \frac{y^a - x_i^a}{h} \right|^2 \right)}{\sum_{i=1}^n w_i g_i \left( \left| \frac{y^a - x_i^a}{h} \right|^2 \right)} \quad (3)$$

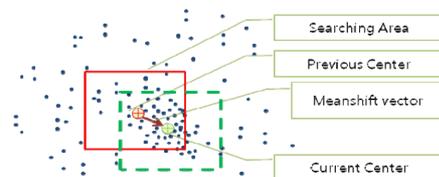
In this case, we apply a filter of equation (4), (5) for the overlapped areas to estimate a new motion mean vector.

$$T^a = k_i g_i \left( \left| \frac{y^a - x_i^a}{h} \right|^2 \right) \quad (4)$$

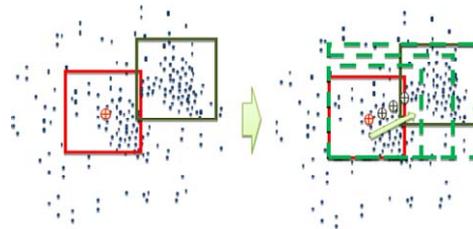
$$k_i = \begin{cases} a_m + b_m w_i \\ a_n^j + b_n^j w_i \end{cases} \quad (5)$$

In this case, the value of  $b_m$ ,  $b_n^j$  have 0 to 1 and  $a_m$ ,  $a_n^j$  are the weights of each searching area.  $a_m$ ,  $b_m$  indicate weight-added function in the main searching area,  $a_n^j$ ,  $b_n^j$  are weight-added function of the non-searching area and  $j$  is the number of non-searching area. We highlighted the main search area by adding the high-weighted value and add the low-weighted value in the non-search area. It changes the variables of the weights of the histogram and estimates the new center point by computing repeatedly.

By adding weights to the non-searching area such the dotted area in **Figure 8**, we prevented the spread of the searching area and the center point. Therefore, to reduce the weight of the area of the hands during the search of the face area as shown in **Figure 8**, since the value is reduced in formula (4), the value of the mean motion vector by  $M_g^a(y_i^a)$  is decreased. We could calculate repeatedly by adjusting the variables of the weight of the histogram, prevented the spread of the search area, and maintained the original search area.



**(a) Non-occlusion case**



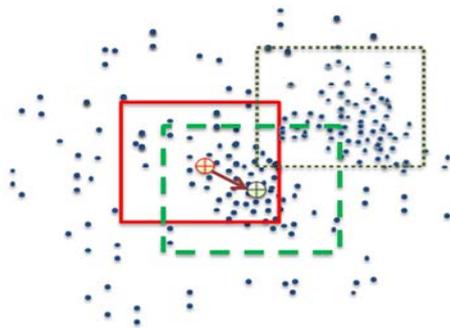
**(b) Occlusion case**

**Figure 7. Non-occlusion and occlusion cases.**

**Figure 9(a)** is the video that used only the CAMShift algorithm to recalculate area in the range of two times of the current-searching region. Therefore, it can be seen that the area is extended if the similar color histogram exists in current searching area. The results of maintaining the search area per each part of **Figure 10** with weight adjustment of histogram show that the extension was not occurred as shown in **Figure 9(b)**.

#### 4.2. Stabilization Algorithm by Using Kalman Filter

Searching area obtained from the mean motion does not hold a stable value because the shape and the intensity of the hand's region in the searching area obtained in each frame are not constant.



**Figure 8. Mass center when occlusion.**



(a) Only CAMShift (b) Modified CAMShift

**Figure 9. Result of weighted window CAMShift.**



(a) Face Tracking Color Distribution Image



(b) Left and Right Tracking Color Distribution Image

**Figure 10. Hue distribution images for search areas.**

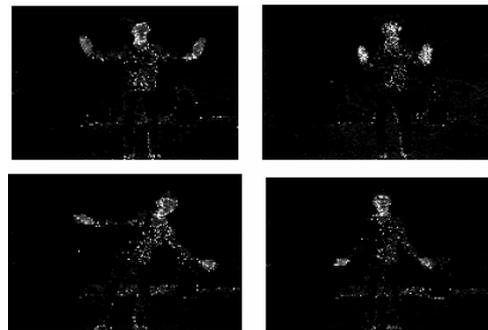
In **Figure 11**, Change of color values occurs due to irregular illumination, even though the shape in skin area is similar. Since the CAMShift algorithm is applied in each frame, each searching area changes in color distribution images obtained by color histogram back-projection. In order to stabilize this change, we should track the center point and the size of the searching area with Kalman filter. **Figure 12** is a block of CAMShift using a Kalman filter.

In this paper, the state equation and the measurement equation were defined in formula (6) and (7).

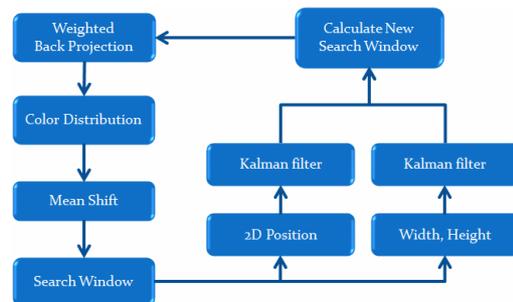
$$\begin{bmatrix} x_{k+1} \\ y_{k+1} \\ v_{x_{k+1}} \\ v_{y_{k+1}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} x_k \\ y_k \\ v_{x_k} \\ v_{y_k} \end{bmatrix} + W_k \quad (6)$$

$$\begin{bmatrix} x_{ck} \\ y_{ck} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} * \begin{bmatrix} x_k \\ y_k \\ v_{x_k} \\ v_{y_k} \end{bmatrix} + V_k \quad (7)$$

$W_k$ ,  $V_k$  are Gaussian noise,  $x_k$ ,  $y_k$  are the center points of the searching area, and  $x_{ck}$ ,  $y_{ck}$  are the center points of the current measurement. Also,  $v_{x_k}$ ,  $v_{y_k}$  are the speed of the object. Furthermore, we defined the state and measurement equations of  $w_k$ ,  $h_k$  (8), (9) which mean the width and height of the searching area.



**Figure 11. Histogram back-projection image in various poses.**



**Figure 12. Block Diagram of CAMShift using Kalman filter.**

$$\begin{bmatrix} w_{k+1} \\ h_{k+1} \\ r_{x_{k+1}} \\ r_{y_{k+1}} \end{bmatrix} = \begin{bmatrix} 1 & 0 & T & 0 \\ 0 & 1 & 0 & T \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} * \begin{bmatrix} w_k \\ h_k \\ r_{x_k} \\ r_{y_k} \end{bmatrix} + U_k \quad (8)$$

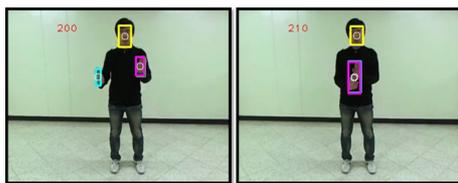
$$\begin{bmatrix} w_{ck} \\ h_{ck} \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} * \begin{bmatrix} w_k \\ h_k \\ r_{w_k} \\ v_{h_k} \end{bmatrix} + Z_k \quad (9)$$

$w_{ck}$ ,  $h_{ck}$  are the center points of the current measurement,  $U_k$ ,  $Z_k$  are Gaussian noise,  $r_{w_k}$ ,  $v_{h_k}$  are the ratio of the size of the searching area.

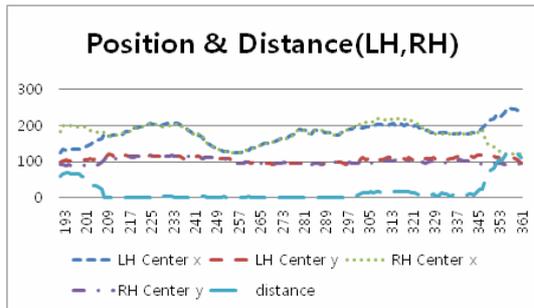
### 5. Applying the CAMShift Algorithm using the Kalman Filter and Weighted Search Windows and the Result

#### 5.1. Experimental Comparison of the Proposed Algorithm with the Obstruction

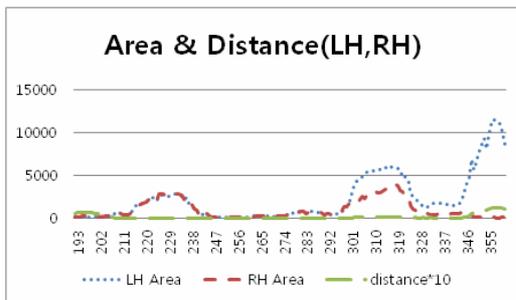
Figure 13(a) shows the images applying multiple CAMShift algorithm that do not add the Kalman filter and the



(a) Result image



(b) The distance between the center points of hands



(c) Size and distance of the hands

Figure 13. Result of CAMShift without weighted search window.

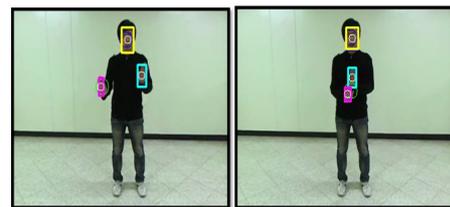
weighted searching window areas of skin color. We found that the region of the both hands with a similar color value caused overlap with each mean shift algorithm and around of center's value. We could know that we failed tracking due to the extension and the overlap of the size of two hands' area near the 205 frame with Figures 13(b), (c).

Figure 14(a) shows the images applying multiple CAMShift algorithm that add the Kalman filter and the weighted search window areas of skin color. Since the trace is avoided when area of hands are obstructed, we could know that it is possible to maintain the original search area.

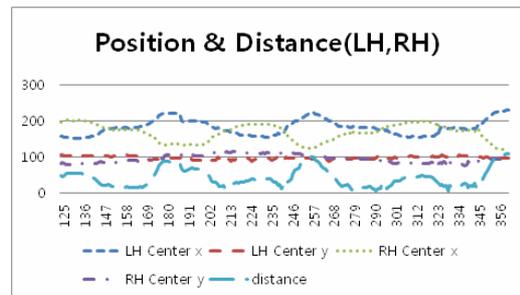
We deserved that searching area of both hands avoided each other in occasion of both hands' overlap by using Figures 14(b), (c). When both hands overlapped, the size of searching area extended and return to original size.

#### 5.2. Experiments the Change of Weights with the Obstruction in Hands' Overlapped.

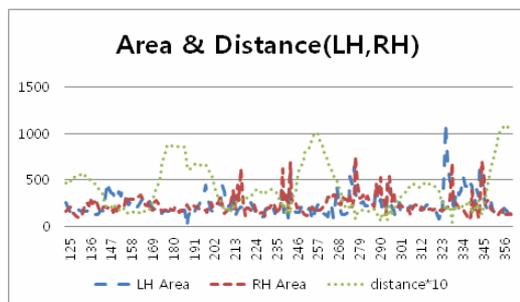
In order to investigate the recognition rate corresponding to the weight in obstruction, Figures 15, 16 has changed



(a) Result image



(b) The distance between the center points of hands



(c) Size and distance of the hands

Figure 14. Result of proposed KWMCAMShift.

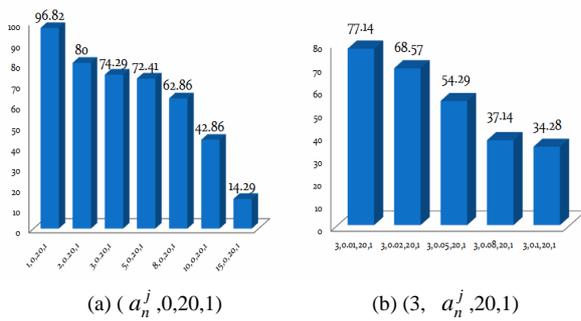


Figure 15. Result of various non-search window weights.

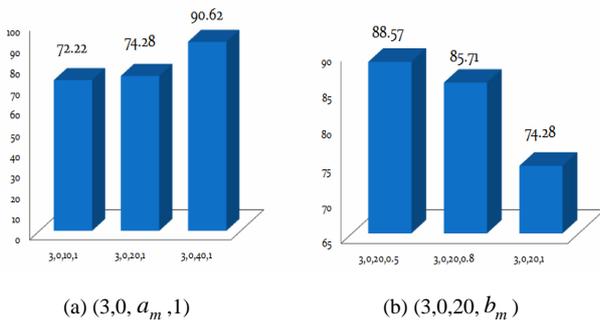


Figure 16. Result of various main-search window weights.

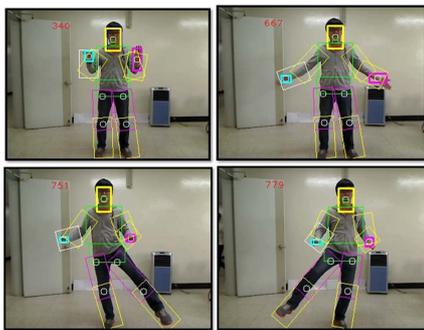


Figure 17. Result of pose estimation using proposed algorithm.

the value of the weight variable  $(a_n^j, b_n^j, a_m, b_m)$  of  $k_i$  in equation (5). As a result, we concluded the recognition rate of 96.82% in weight  $(1,0,20,1)$ .

### 5.3. Result of Pose Estimation

Figure 17 is a video that estimates the posture of the body when the proposed algorithm is applied.

## 6. Conclusions

In this paper, we propose Multi CAMShift Algorithm based on Kalman filter and weighted search windows that extracts skin color area and tracks several human body parts for real-time human tracking system.

We estimated the width, the height, and the position in searching area of CAMShift algorithm with the Kalman filter, and we made accurate searching in occasion of obstruction adding the weight to main-searching area and non-searching area in mean motion vector. We found that the recognition rate of 96.82% when we applied modified CAMShift algorithm proposed in this paper even with the obstruction.

## 7. Acknowledgements

This study was conducted with the assistance of the Korea Aerospace University Technical Research Center of the next generation broadcast media by the GRRC (Gyeonggi-do Regional Research Center) program.

## REFERENCES

- [1] J. Shotton, *et al.*, "Real-time Human Pose Recognition in Parts from Single Depth Images," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 20-25 June, 2011, pp. 1297-1304.
- [2] G. R. Bradski, "Computer Vision Face Tracking for Use in a Perceptual User Interface," *Intel Technology Journal*, 2nd Quarter, 1998.
- [3] Xun Cai, Long Jiang, *et al.*, "A New Region Gaussian Background Model for Video Surveillance," *Natural Computation*, 2008, Vol. 6, pp. 123-127.
- [4] V. Vezhnevets, V. Sazonov and A. Andreeva, "A Survey on Pixel-based Skin Color Detection Techniques," *Graphicon03*, 2003, pp. 85-92.
- [5] P. Peer, J. Kovac and F. Solina, "Human Skin Colour Clustering for Face Detection," *Eurocon 2003*.