Depth-based real-time hand tracking with occlusion handling using Kalman Filter and DAM-Shift

Kisang Kim, Hyung-Il Choi

School of Media, Soongsil University, Seoul, Korea

Abstract. In this paper, we propose real-time hand tracking with a depth camera by using a Kalman Filter and an improved DAM-Shift(Depth-based adaptive mean shift) algorithm for occlusion handling. DAM-Shift is a useful algorithm for hand tracking, but difficult to track when occlusion occurs. To detect the hand region, we use a classifier that combines a boosting and a cascade structure. To verify occlusion, we predict in real time the center position of the hand region using Kalman Filter and calculate the major axis using the central moment of the preceding depth image. Using these factors, we measure real-time hand thickness through a projection and the threshold value of the thickness using a 2nd linear model. If the hand region is partially occluded, we cut the useless region. Experimental results show that the proposed approach outperforms the existing method.

1 Introduction

In the last few decades, various studies have been conducted on automatic analysis of human behavior. The most sophisticated research on the subject is being carried out in HCI(human-computer interaction). Human gesture recognition is an important area in this field. A gesture is a simple and effective nonverbal communication tool that assists complex human interactions. In many fields such as sign language, hand gesture recognition is a primary method for those with hearing impairment to smart devices for effective interactions. Several methods have been proposed for gesture recognition, which includes hand region detection and hand feature extraction. Existing research on the subject includes analyzing hand images using a data glove[1,2,3], color data [4, 5], a combination of color and depth data [6,7,8], and depth data alone [9,10,11,12,13]. It is difficult to devise an interface for a data glove because it requires a line to connect to the entire system. Various studies that integrate depth and color data seek to mitigate the sensitivity of the color data method to environmental changes. Under the assumption that the hand lies before the body. Park et al. [6] generated a histogram from a depth image of the Kinect motion sensing input device to detect candidate hand regions, and located a final hand region by using Bayes rule and skin color to find the precise hand region. This method executes significantly better than the sole use of the color and depth data, however performance

decreases in darkness due to its basic assumption; that is, the hand always lies before the body and the use of color. Van den Bergh and Van Gool [7] suggested a combined method that locates a face in a color image, removes the background by using the threshold value along with the distance of the detected face, and searches the hand region in the remaining region. Furthermore, this method is more accurate than ones that use either a depth or a color image, but requires more processing time and is more difficult to use in dim lighting. Trindade et al. [8] proposed skin color filtering by using an RGB-D sensor prior of detecting the face, body, and hand regions, distributing the histogram with the depth axis, and filtering out the hand region based on a threshold value. The outliers are thus removed by k-means clustering to find the center of the hand region, which becomes the base point for detecting the hand region and for pose recognition. This method, with a mixed use of color and depth data, can improve detection accuracy by deleting outliers during filtering and applying a segmentation technique. However, this method is difficult to use in varying lighting conditions and is sensitive to errors because it undergoes several processes prior to hand region detection.

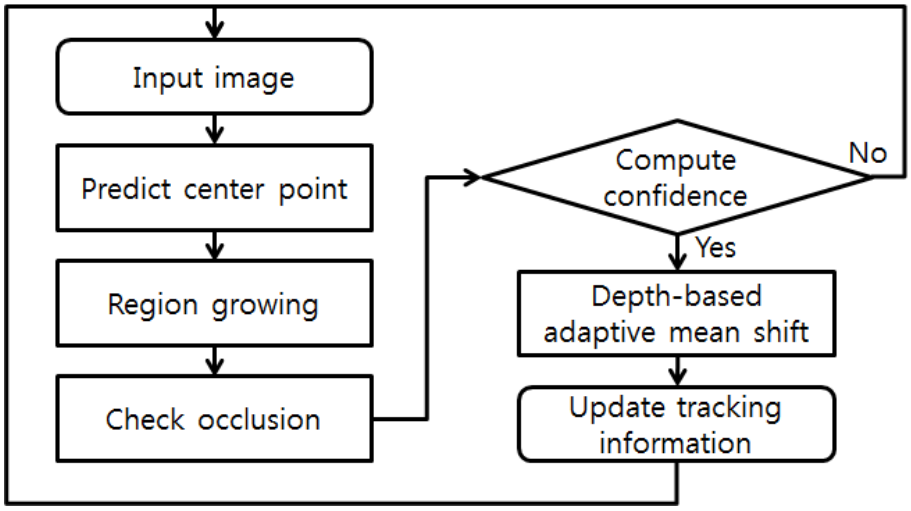

Fig. 1. Structure of occlusion handling tracking system.

To overcome these problems, we only use depth data from the Kinect camera sensor. To track the hand region, we use a boosting and cascading algorithm to detect region. Figure 1 shows the structure of our proposed system which consists of main two steps: prediction and verification of occlusion. The remainder of this paper is structured as follows: in section 2, we explain hand region prediction using Kalman Filter. Section 3 explains improved DAM-Shift[14], comparing with traditional DAM-Shift. Section 4 describes our testing environment along with

experimental results that confirm the effectiveness of the proposed algorithm. We present our conclusions in section 5.

2 Prediction of Hand Region

Prior to the region growing process, we need to find the point, which is the one of hand region. Traditional method locates the nearest point from the center of the previous hand region. However, if the hand moves too fast or if the background environment is too complex, this method causes an error in tracking the hand region. Figure 2 shows the problem with the traditional method.

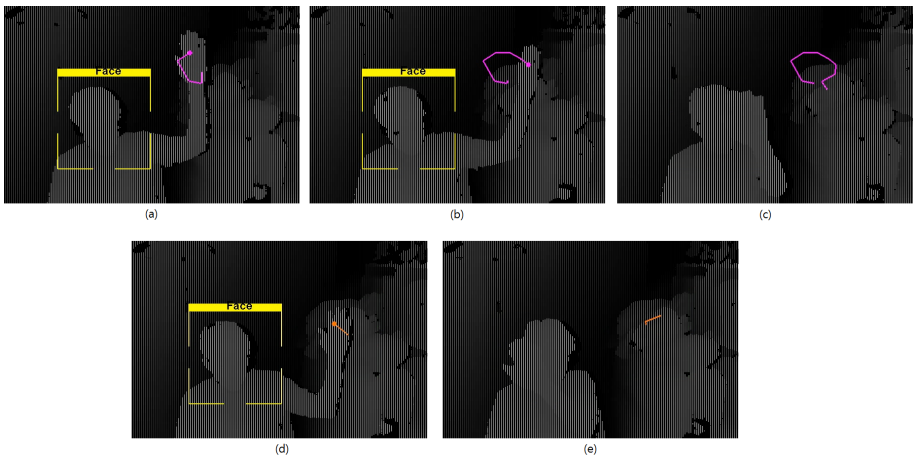

Fig. 2. Problem with the traditional tracking method. (a),(b),(c) is a set that fail to track. (d),(e) is also a set.

To solve this problem, we use a prediction method that combines Kalman Filter, previous hand moving velocity and 2nd polynomial model. Using only velocity, it is too risky because in except situation, this method occur the error due to the prediction point moves too much. Thus, in order to reduce the error rate, we use 3D Kalman Filter. Furthermore, when the distance between the hand and the camera is small, hand tends to move considerably in a corresponding image; otherwise, it only moves slightly. We use this observation in a 2nd polynomial model in order to predict hand region size and use 2nd polynomial model to predict hand region movement. [15]

$$r = \begin{bmatrix} 1 \; x \; x^2 \end{bmatrix} \cdot \begin{bmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{bmatrix} \tag{1}$$

$$\alpha = (P^T \cdot P)^{-1} \cdot P^T \cdot y \tag{2}$$

$$P = \begin{bmatrix} 1 & x_1 & x_1^2 \\ 1 & x_2 & x_2^2 \\ \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 \end{bmatrix}, y = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{bmatrix} \tag{3}$$

Equation (1) is to presume the radius of the including circle of a hand region. x represents the depth of a candidate hand region in the current image. To make the value r, we need the coefficients $\alpha = \begin{bmatrix} \alpha_1 & \alpha_2 & \alpha_3 \end{bmatrix}$. These coefficients can be calculated by (2) and (3). The value of x_i and r_i are manually collected during the learning phase. We check the size of a hand region by varying the depth values of the hand region. We assume that the size of the hand could be represented as the 2nd polynomial function of depth. Therefore, if α is determined by the learning data, the hand region size can be predicted by using (1).

$$P_i(x, y, z) = KF\{C_{i-1} + \frac{\beta \cdot r}{\sqrt{w^2 + h^2}}(2 \times C_{i-1} - 3 \times C_{i-2} + C_{i-3})\} \tag{4}$$

Equation (4) estimates of the hand region with the predicted hand region size. $KF(\cdot)$ represents the Kalman Filter. $P_i(x, y, z)$ is the predicted hand center. C_i is the hand center of i-th image. w and h represent the image width and height, respectively and β is the weight value, mostly it use from 0.3 to 0.7. Figure 3 shows the predicted point and the previous center point of the hand region.

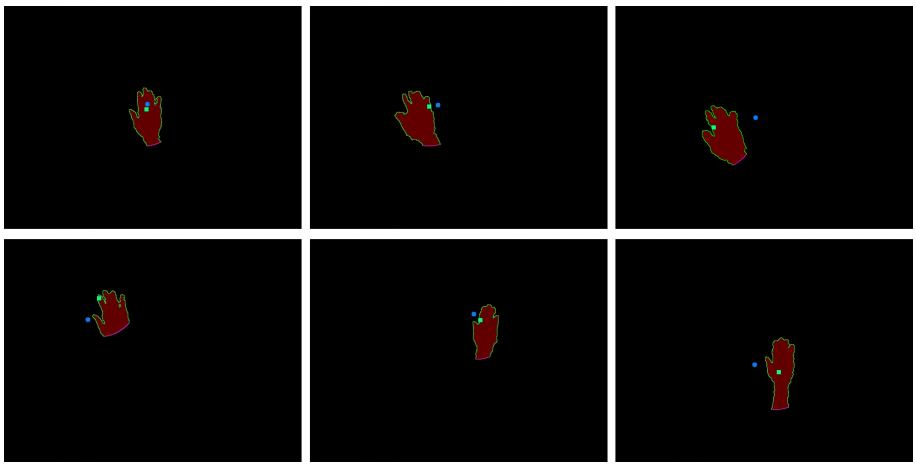

Fig. 3. Prediction of center point. Rectangle : Predicted center, Circle : Previous center.

3 DAM-Shift with Occlusion Handling

3.1 Detecting and Handling Occlusion

The major assumptions underlying hand occlusion prevention are that the tracking arm region always lies below the hand region, and that the thickness of the arm is always less than the thickness of hand. Given these assumptions, the occluded image can be found and the hand region can be revised. After the occlusion frames, this method is easy to re-track the hand region. To calculate thickness, the major axis of the hand region is required. We use central moment to measure this axis.

$$\theta = \frac{1}{2} tan^{-1}(\frac{2\mu_{11}^{'}}{\mu_{20}^{'} - \mu_{02}^{'}}) \tag{5}$$

$$C_x = \frac{M_{10}}{M_{00}}, C_y = \frac{M_{01}}{M_{00}} \tag{6}$$

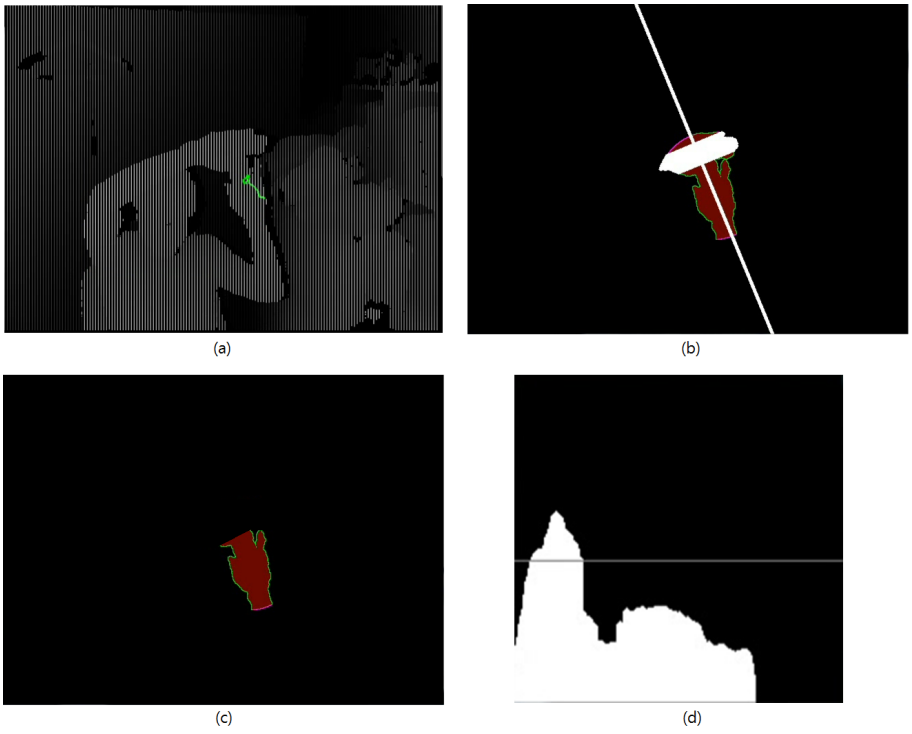

Fig. 4. Elimination of occluded region. (a) Input image, (b) Found occluded region, (c) Hand region, (d) Thickness projection.

In equation (5), θ is the angle of the major axis. $\mu_{11}^{'}$ is the value of μ_{11}/M_{00}, $\mu_{20}^{'}$ is that of μ_{20}/M_{00} and $\mu_{02}^{'}$ is that of μ_{02}/M_{00}. In equation (6), C_x and C_y are the center point of each coordinate of this region. Following this measurement, we make a projection histogram using this axis. Figure 4 (b) shows a projected image with the major axis.

In figure 4 (d), the gray line in the projection histogram is the value of r, the radius of the hand region in equation (1). If the thickness value is greater than the radius, it means the region is occluded by other objects. Therefore, if we detect a region as occluded, we remove it , as it is useless. Figure 4 (c) shows an example of the removal of the occluded region.

3.2 DAM-Shift

DAM-Shift is defined similarly to Mean Shift [16], but its kernel size changes according to the depth values and the iteration time. Equation (7) represents the DAM-Shift algorithm.

$$TP_{i+1}^{mean} = \frac{\sum_{p\in\Omega} p \cdot K(p, TP_i^{mean}, DTP_{i-1}^{mean}, i)}{\sum_{p\in\Omega} K(p, TP_i^{mean}, DTP_{t-1}^{mean}, i)} \tag{7}$$

$$K(p, s, d, i) = \begin{cases} 1 & ||p - s|| < R(d, i) \\ 0 & otherwise, \end{cases} \tag{8}$$

$$R(d, i) = \begin{cases} SPM(d) & 2SPM(d) - i \cdot T_{rc} < SPM(d) \\ 2SPM(d) - i \cdot T_{rc} & otherwise, \end{cases} \tag{9}$$

$$SPM(d) = \alpha_1 + \alpha_2 d + \alpha_3 d^2 \tag{10}$$

TP_{i+1}^{mean} in the above represents the tracking point coordinates at the $i+1$ iteration. These coordinates are updated as the iteration continues. $K(\cdot)$ presents a kernel function whose size changes depending on the depth of the tracking point. DTP_{t-1}^{mean} depicts the depth value of the tracking point in the $t-1$th frame. p represents a point that belongs to set Ω and denotes the set of valid borderlines VB obtained during the region growing. The coordinates of the nearest point TP_t^{seed} is acquired as in region growing, and it is substituted for TP_0^{mean} . As the iteration continues, TP_{t+1}^{mean} is alternated with TP_t^{mean} for the next iteration. The process is repeated until the point of convergence. In Equation (8), The radius $R(d, i)$ of the kernel function changes according to the depth value and the number of repetition. Equation (9) extracts the radius according to depth and repetition. For this purpose, the function $SPM(d)$ is used, which corresponds to the 2nd polynomial model defined in Section 2. Figure 6 shows the process of determining the tracking point using the DAM-Shift algorithm.

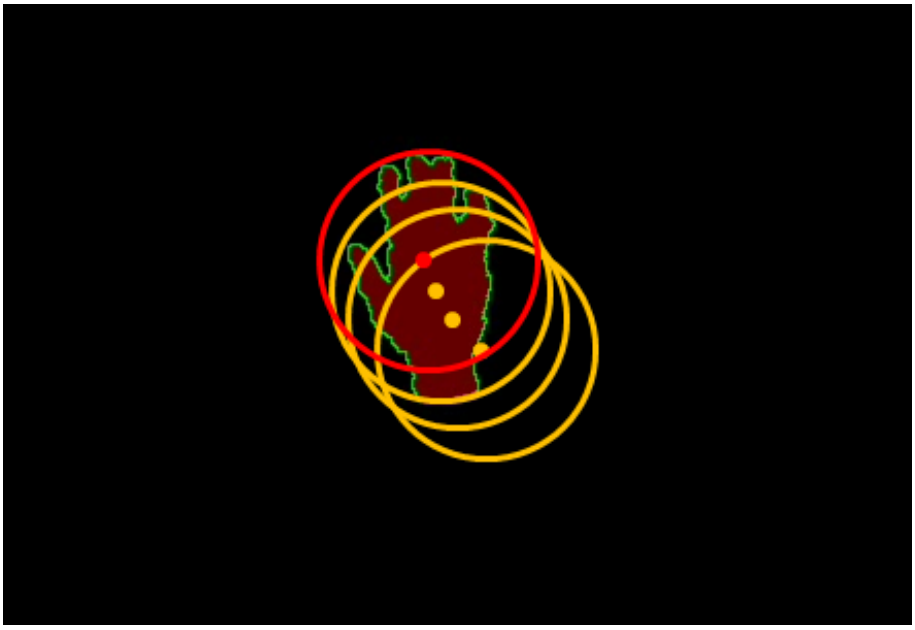

Fig. 5. Example of the movement of DAM-Shift.

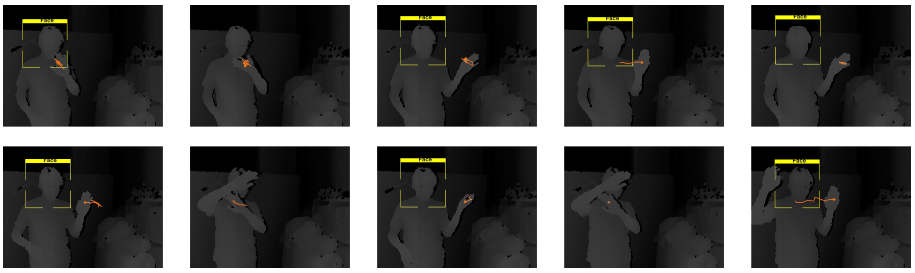

Fig. 6. Results of hand tracking.

4 Experimental Results

For experimental evaluation, we used a computer with an Intel(R) Core(TM) i5-3470 CPU and an 8GBbyte memory. We have used Microsoft Kinect camera, (320 × 240 pixels) at 30fps to acquire depth images.

In figure 6 shows the results of hand tracking in input frames at one second interval. The point shows the center point of the hand region and the tail represents the tracked line.

Figure 7 shows the features of the hand region. The white line shows the axis of the hand region and the rectangle is its predicted center point. This shows that

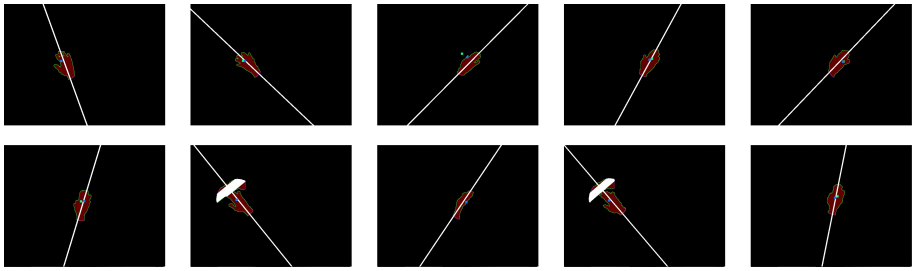

Fig. 7. Results of hand features.

it solved the occlusion problem to track. Further, occluded region can checked at Figure 8, shows the results of hand projection.

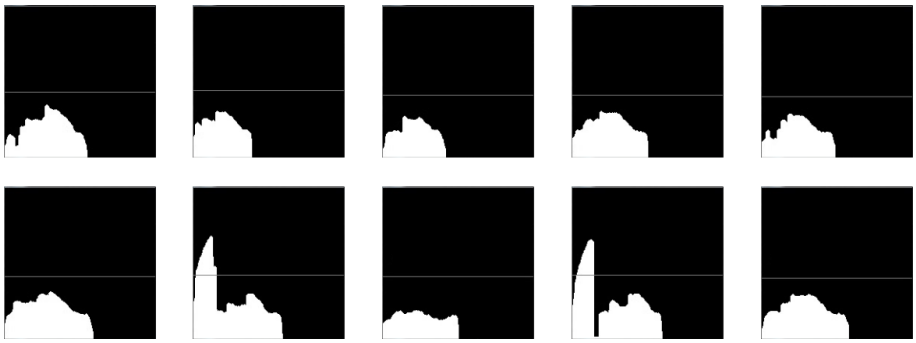

Fig. 8. Results of hand projection with checking occlusion.

5 Conclusions

We proposed a hand tracking method that works well in the real world environment. For tracking a hand, we have developed improved DAM-Shift to handle the occlusion. Our 2nd polynomial model and Kalman Filter work well to predict the center of the hand, which plays an important role in confining a search area. To handle occlusion, we developed an improved DAM-Shift, which consists of axis extraction and make hand projection to verify and remove occlusion region. As the experimental results, our method eliminates the occlusion region well and reduce problem encountered in hand tracking.

Acknowledgement. This study was supported by the Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT and Future Planning (2013R1A1A2012012).

References

1. D. L. Quam.: *Gesture recognition with a DataGlove.* in Proceedings of the IEEE National Aerospace and Electronics Conference (1990) 755760
2. R. Y. Wang and J. Popovic.: *Real-time hand-tracking with a color glove.* ACMTransactions onGraphics (2009)
3. L. Lamberti and F. Camastra.: *Handy: a real-time three color glove-based gesture recognizer with learning vector quantization.* Expert SystemswithApplications (2012) 10489-10494
4. H. I. Suk and B. H. Sin.: *Dynamic Bayesian network based two-hand gesture recognition.* Journal of KIISE: Software and Applications (2008)
5. M. K. Bhuyan, D. R. Neog, and M. K. Kar.: *Fingertip detection for hand pose recognition.* International Journal on Computer Science and Engineering (2012) 501-511
6. M. S. Park, M. Md. Hasan, J. M. Kim, and O. S. Chae.: *Hand detection and tracking using depth and color information.* in Proceedings of the International Conference on Image Processing, Computer Vision, and Pattern Recognition (2012) 779-785
7. M. Van den Bergh and L. Van Gool.: *Combining RGB and ToF cameras for real-time 3D hand gesture interaction.* in Proceedings of the IEEE Workshop on Applications of Computer Vision (2011) 66-72
8. P. Trindade, J. Lobo, and J. P. Barreto.: *Hand gesture recognition using color and depth images enhanced with hand angular pose data.* in Proceedings of the IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (2012) 71-76
9. Z. Mo and U. Neumann.: *Real-time hand pose recognition using low-resolution depth images.* in Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2006) 1499-1505
10. X. Liu and K. Fujimura.: *Hand gesture recognition using depth data.* in Proceedings of the 6th IEEE International Conference on Automatic Face andGesture Recognition (2004) 529-534
11. S. Malassiotis and M. G. Strintzis.: *Real-time hand posture recognition using range data.* Image and Vision Computing (2008) 1027-1037
12. P. Suryanarayan, A. Subramanian, and D. Mandalapu.: *Dynamic hand pose recognition using depth data.* in Proceedings of the 20th International Conference on Pattern Recognition (2010) 3105-3108
13. I. Oikonomidis, N. Kyriazis, and A. A. Argyros.: *Efficient model-based 3Dtracking of hand articulations using Kinect.* in Proceedings of the British Machine Vision Conference (2011)
14. Joo, Sung-Il, Sun-Hee Weon, and Hyung-Il Choi.: *Real-Time Depth-Based Hand Detection and Tracking.* The Scientific World Journal (2014)
15. S. Park, S.Yu, J.Kim, S.Kim, and S. Lee.: *3D hand tracking using Kalman filter in depth space.* EURASIP Journal on Advances in Signal Processing (2012)
16. G. R. Bradski.: *Computer vision face tracking for use in a perceptual user interface.* Intel Technology Journal (1998)