Skin Color Weighted Disparity Competition for Hand Segmentation from Stereo Camera

Qi Wang[1]
wangqi@jdl.ac.cn
Xilin Chen[1]
xlchen@ict.ac.cn
Wen Gao[2]
wgao@pku.edu.cn

[1] Key Lab of Intelligent Information Processing, Institute of Computing Technology
Chinese Academy of Sciences
Beijing, China

[2] Institute of Digital Media, School of EE & CS
Peking University
Beijing, China

Abstract

Hand segmentation is the basis of many vision-based hand gesture applications in human computer interaction (HCI). This paper proposes a novel method of skin color weighted disparity competition to incorporate the skin color and depth cues for hand segmentation in clutter environment from stereo camera. The basic idea is that the hand disparity will be apparently different from most of the background disparities. Note that we refer to the scene except the user as the background. The proposed method validates a hand pixel by verifying whether the pixel satisfies the disparity limit of the hand more than of the background. Its novelty lies in incorporating the skin color probability to weight the disparity competition to improve the robustness of validation and the accuracy of boundary location. Experiments demonstrate the efficiency of the proposed method.

1 Introduction

Hand gesture gets more and more attention for its convenience and nature in HCI [9, 13]. Hand segmentation, as the first step towards most hand gesture applications, plays an important role in the robustness, accuracy and efficiency of a HCI system. However, it suffers from clutter background, lighting variation, the flexibility or concave character of the hand, etc.

Up to now, many HCI systems based on hand gesture are built with a single camera [9, 13]. The popular monocular methods for hand segmentation include skin color detection [3, 7, 14], background subtraction [11, 16], active contour [21], and model-based method [5, 17, 18, 19]. Skin color detection is very common for its simpleness and easy implementation, and is very efficient against simple background or in the scene of hand being the only skin-colored object. However, skin color detection is sensitive to confusing pixels, so it is not expected to work efficiently against clutter background or in the scene of more than one skin-colored object. Background subtraction can deal with the clutter background to

© 2010. The copyright of this document resides with its authors.
It may be distributed unchanged freely in print or electronic forms.

some degree, but suffers from the shadow problem. Active contour starts from an initial contour and dynamically moves toward the true contour by minimizing some predefined energy function iteratively. Some researchers apply active contour to hand segmentation against clutter background and reports accurate result [21]. However, locating the hand in clutter environment and selecting the appropriate initial contour are not trivial. Model-based methods receive attention from many researchers since the prior knowledge can be easily introduced . On one hand, model-based methods can present excellent performance against clutter background [18] and moreover directly obtain the model parameters of the hand. On the other hand, model-based methods suffer from some problems such as the exponential growth over time with the larger search space, the local minima and singularities. It is worth noting some interactive segmentation methods such as GraphCut and its variations [4, 10, 15]. With simple user's interaction such as marking some object pixels or some background pixels, GraphCut is able to generate a very good segmentation. GraphCut is essentially a global optimization method and able to approach the global optimal solution to some degree for the given energy function. However, our experiments show that GraphCut seems to fail in segmenting the concave detail of the hand for the possible reason of the serious concave or consequent image blur.

Rather than monocular camera, some researchers study hand gesture by using stereo camera [8, 12]. Stereo camera can provide the disparity map more than the optical image. The disparity cue, i.e. the depth cue, helps to many challenging problems in monocular methods such as the clutter background, the lighting variation, the occlusion and the singularity. In the case of stereo camera, researchers employ generally the simple thresholding operation for hand segmentation by assuming the disparity map is dense. However, it is difficult to compute the dense disparity map for the hand due to lack of texture. Although sophisticated stereo algorithms may be employed for the dense disparity map [6], the time consuming would weaken their advantage for the HCI application where the speed is an important factor for real interaction.

Is the non-dense disparity map enough for efficient hand segmentation? This paper focuses on this problem and proposes the novel method of skin color weighted disparity competition. The proposed method incorporates both the skin color and depth cues. The main idea is to validate a hand pixel by verifying whether the pixel satisfies the disparity limit of the hand more than of the background and furthermore incorporate the skin probability to weight the disparity competition to improve the robustness of validation and the accuracy of boundary location.

Manders et al.'s work [12] is close to our method and employs also both the disparity cue and the skin cue. Their method produces the output by taking the pointwise multiplicative product of the disparity map and the skin probability map. So, their method would be efficient if the disparity map is dense and would neglect those hand pixels with uncertain disparity if the disparity map is not dense.

The rest of this paper is organized as follows. Section 2 presents the proposed method and Experimental results in section 3 demonstrates the efficiency of the proposed method. Section 4 concludes this paper and gives the future directions.

2 The proposed method

The framework of the proposed method is shown in Figure 1. The proposed method contains two main stages of marking the confident hand pixels following by performing skin

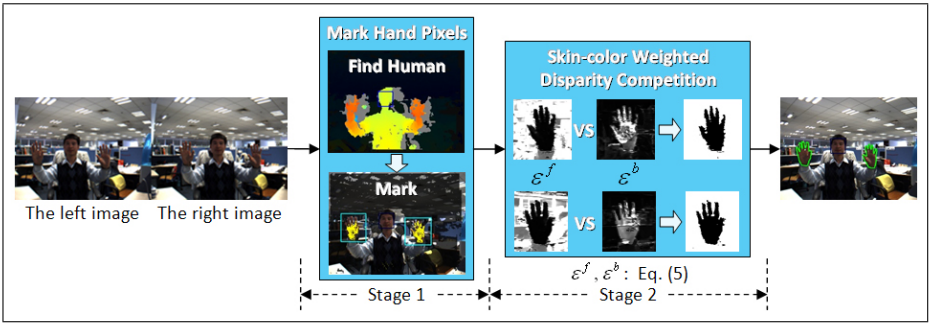


Figure 1: The framework of the proposed method.

color weighted disparity competition. The first stage marks the confident hand pixels and obtains all regions of interest (ROI) in the left image. The second stage employs skin color weighted disparity competition to validate the hand pixels for each ROI. Thereafter, the proposed method segments the hand in each ROI by performing the morphological operation of opening following by searching the maximal connected region in the related ROI.

2.1 Marking the confident hand pixels

How to mark some confident hand pixels in clutter scene is not trivia since the hand is possibly not the only skin object or not the closest object to the camera. In the proposed method, we mark the confident hand pixels by employing the following flow: first computing the disparity map and detecting face in the left image, then tracking down by following the disparity cue of the detected face to find human body regions, finally incorporating the disparity cue and the skin color cue from the detected face to mark the hand pixels in the found human body regions, as illustrated in Figure 1. It is feasible since face detection is very robust and the disparity cue of the face helps to distinguish the possible part of human body.

Computing the disparity map We employ the simple block matching stereo algorithm based on the gray images related to the left and right image to produce the disparity map, which is typically not dense for the hand due to lack of texture. The zero value in the disparity map means uncertain disparity for the related pixel.

Detecting face We employ the face detector [20] to detect face in the left image. After the face is detected, we use the average disparity of the detected face region to threshold the disparity map and then obtain both human body candidate regions that we refer to as the foreground (FG) map and the confident background regions that we refer to the background (BG) map. It is reasonable since the hands are in front of the face in most HCI cases. Due to the clutter, the FG map typically includes some non-human-body regions.

Finding human body regions We obtain the human mask from the union image of the zero pixels in the disparity map and the non-zero pixels in the FG map by finding the maximal connected region containing the detected face and mask the FG map to obtain the valid human body regions. This operation is reasonable: first, the zero pixels

in the disparity map contains those human pixels that don't appear in the FG map, consequently, these zero pixels can lead to all human body regions in the FG map being a whole connected region; second, the non-zero pixels in the BG map separate the human body regions from non-human-body regions in most cases, consequently, the obtained human mask less likely contains the non-human-body regions.

Marking the confident hand pixels We mark those pixels of the FG map, whose color is close to the detected face and whose disparity fall in the small neighbor of the local maxima of the FG map, as the hand pixels. The marked hand pixels will be confident since the hand color is close to the face color and each hand is usually the foremost part of each arm in HCI scene. About the skin model for detecting skin-colored pixel, we employ the Hue-Saturation Histogram learned from the detected face region.

Moreover, we obtain the ROI for each appear hand by considering the zero pixels in the disparity map as non-zero in the FG map and performing the camshift algorithm based on the marked hand pixels. Of course, some allowance is made on the camshift result for few unexpected cases. The obtained ROI would contains the hand region since the zero pixels in the disparity map contain those hand pixels with uncertain disparity. The obtained ROIs enable a selective analysis for the following skin color weighted disparity competition and thus contributions to the efficiency of the proposed method.

2.2 Segmenting the Hand with skin color weighted disparity competition

The non-dense disparity map prevents the simple depth threshold method from the efficient hand segmentation. Although, three facts stand by the feasibility of accurate hand segmentation with the non-dense disparity:

1. The disparity of each hand fluctuates very slightly around the mean disparity of the marking hand pixels since the depth scale of hand is far less than the distance between the hand and the stereo camera;

2. The hand disparity is apparently different from the background disparity since the user must lie in some space for the HCI application at most cases;

3. The hand is an enough simple object against clutter background from color aspect.

Inspired by the three facts, we propose the novel method of skin color weighted disparity competition for hand segmentation.

2.2.1 The skin color probability

We use the HSV color space and utilize a three-dimensional Gaussian model for the skin color probability as defined in Eq. (1).

$$\begin{cases} P(I) = \frac{1}{(2\pi)^{3/2}|\Sigma|^{1/2}} \exp\left[-\frac{1}{2}(I-\mu_c)^t \Sigma^{-1}(I-\mu_c)\right] \\ \Sigma = \begin{pmatrix} {\sigma_h}^2 & & \\ & {\sigma_s}^2 & \\ & & {\sigma_v}^2 \end{pmatrix} \end{cases}, \tag{1}$$

where I is the probe pixel, μ_c is the mean HSV vector of all marked hand pixels in the current ROI, σ_h, σ_s or σ_v encodes respectively the standard deviation of the H, S or V component for the skin color in the current ROI.

Although many researchers argue that the intensity component of V is sensitive to the lighting variation and employ only H and S components for skin color detection, we argue that the intensity component of V also has some discrimination ability to distinguish the hand pixel from the background pixel because the mean color vector μ_c and the probed color vector I come from the identical small local region, i.e. the current ROI in the current frame. But, we utilize greater standard deviations for the V component than the H or S component by taking into account the fact that the V component is actually more sensitive to lighting variation than the H or S component.

2.2.2 The two competing disparity limits

The two competing disparity limits include the disparity limit of the hand and of the background as defined as follows:

$$\begin{cases} \chi^h = \left[u_d^h - e^h, u_d^h + e^h\right] \\ \chi^b(i,j) = \left[D^b(i,j) - e^b, D^b(i,j) + e^b\right] \end{cases}, \tag{2}$$

where the superscript of h or b represents the hand or background respectively, e^h or e^b encodes the maximal bias(typically set to 2 or 1 respectively), u_d^h is the mean disparity of all marked confident hand pixels in the current ROI and D^b is the result image of the WaterShed algorithm based on the BG map as the seed image and the zero image as the origin image.

We set the identical disparity limit of the hand χ^h for all pixels in the current ROI since the hand disparity fluctuates very slightly around u_d^h.

The disparity limit of the background is not as simple as of the hand due to the clutter. Firstly, different background pixels may lie in different disparity levels. Secondly, many background pixels in the left image have no corresponding pixels in the right image because of occlusion from the foreground objects, which results in many zero pixels i.e. pixels with uncertain disparity in the disparity map. The only available cue for the background disparity limit is the BG map where the non-zero pixels indicate the confident background pixels. So we consider performing the WaterShed algorithm based on the BG map as the seed image and the zero image as the origin image to fill in all zero pixels in the BG map. This operation encourages the zero pixels in the BG map to have the same disparity as the closest non-zero pixel, which is reasonable if no more evidences. Different from the disparity limit of the hand being identical for all probe pixels, the disparity limit of the background for different probe pixels may be different.

2.2.3 Skin color weighted disparity competition

The hand pixel will satisfy the hand disparity more than the background disparities. This fact inspires the idea of disparity competition.

First, Let us introduce two definitions for convenient formation of the proposed skin color weighted disparity competition. The first one is the disparity verification V defined as

$$V(L(i,j),d,N) = \frac{1}{2\times N+1}\sum_{n=-N}^{n=N} \left\| L(i,j+n), R(i,j+d+n) \right\|, \tag{3}$$

where $L(\cdot)$ or $R(\cdot)$ denotes the HSV color vector of the considering pixel in the left or right image respectively, d is the verifying disparity, N is the neighborhood size, and $\|\cdot\|$ represents the L_2 Norm. The value of V implies the degree of $L(i,j)$ satisfying the disparity d.

The other one is the skin probability of the local neighborhood of $I(i,j)$ defined as

$$\psi_I(i,j,N) \propto P\left(\frac{1}{2\times N+1}\sum_{n=-N}^{n=N} I(i,j+n)\right), I \in \{L,R\}\ , \tag{4}$$

where P is defined in Eq. (1) and N denotes the neighborhood size. The range of ψ is $[0,1]$.

By using V and ψ, we format the proposed skin color weighted disparity competition for the pixel $L(i,j)$ as

$$\begin{cases} \tilde{H}(i,j) = \begin{cases} 0, & if\ \varepsilon^h < \varepsilon^b\ and\ E(i,j) == 0 \\ 1, & otherwise \end{cases} \\ \\ where \\ \varepsilon^h(i,j) = \min\limits_{d\in[u_d^h - e^h, u_d^h + e^h]} \exp^{\frac{\left(1-\min(\psi_L(i,j,N^h),\psi_R(i,j+d,N^h)\right)-(1-T_s)}{0.5\times(1-T_s)}} \times \left(V\left(L(i,j),d,N^h\right)+\alpha\right) \\ \varepsilon^b(i,j) = \min\limits_{d\in[D^b(i,j)-e^b, D^b(i,j)+e^b]} \exp^{\frac{\psi_L(i,j,N^b)-T_s}{0.5\times T_s}} \times \left(V\left(L(i,j),d,N^b\right)+\alpha\right) \\ \\ E(i,j) = \begin{cases} 0, & if\ \psi_L(i,j,N^h) \geq T_s\ and\ \varepsilon^h \leq T^h \\ 1, & otherwise \end{cases} \end{cases} , \tag{5}$$

where

1. The verification error $\varepsilon^h(i,j)$ or $\varepsilon^b(i,j)$ is employed for verifying the degree that $L(i,j)$ satisfies the disparity limit of the hand or the background respectively:
   - Less value means more satisfying,
   - V is defined in Eq. (3) and ψ is defined in Eq. (4) where the neighborhood size of N^h or N^b is typically set to 2,
   - T_s is a probability threshold for distinguishing the skin color from the non-skin color and α is a small tuning constant;
2. $E(i,j)$ is employed for excluding the non-hand pixels ($E(i,j)=1$) and consequently contributes to the low false acceptance rate:
   - $L(i,j)$ is probably not a hand pixel if it is impossibly a skin-colored pixel (indicated by $\psi_L(i,j,N^h) < T_s$),
   - $L(i,j)$ is probably not a hand pixel if it satisfies less the disparity limit of the hand (indicated by $\varepsilon^h > T^h$, T^h is a threshold for disparity validation);
3. $\tilde{H}(i,j)$ is employed to validate a hand pixel by competition:
   - $\tilde{H}(i,j)=0$ means that $L(i,j)$ is a hand pixel since it satisfies the disparity limit of the hand more than of the background.

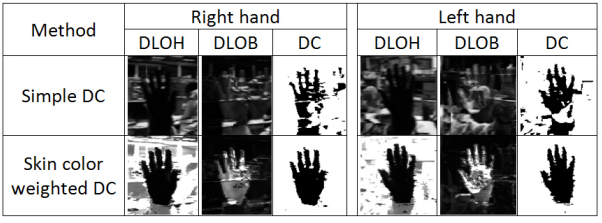


Figure 2: An illustrated example of skin color weighted disparity competition over simple disparity competition without weighting: DC means disparity competition, DLOH means the disparity limit of the hand while DLOB means the disparity limit of the background.

The exponential terms in $\varepsilon^h(i,j)$ and $\varepsilon^b(i,j)$ are used as the weight function. The weight function is inclined to minify the value of V for a skin-colored pixel but magnify the value of V for a non-skin-colored pixel in $\varepsilon^h(i,j)$, and the contrary in $\varepsilon^b(i,j)$. Consequently, ε^h affirms the hand pixels as well as reversely relieves most of the background pixels while ε^b affirms part of the background pixels as well as reversely relieves most of the hand pixels, as illustrated in Figure 2. This weight strategy contributes greatly to the robustness of the disparity competition $\tilde{H}(i,j)$ validating a hand pixel.

After all pixels in the current ROI are validated, we obtain the final hand segmentation by performing the morphological operation of opening following by searching the maximal connected region in the resulted binary image.

3 Experimental results

We implemented the skin color weighted disparity competition method in C++ by using OpenCV [1] and the Triclops Stereo Vision SDK [2]. The test sequence is captured by using the Bumblebee 2 stereo camera from Point Grey Research Inc. and refers to a chinese gesture sequence of numerals from 1 to 10 in a clutter lab environment. The test sequence contains 225 frames with the resolution of 1024×384, i.e. 512×384 for each view image.

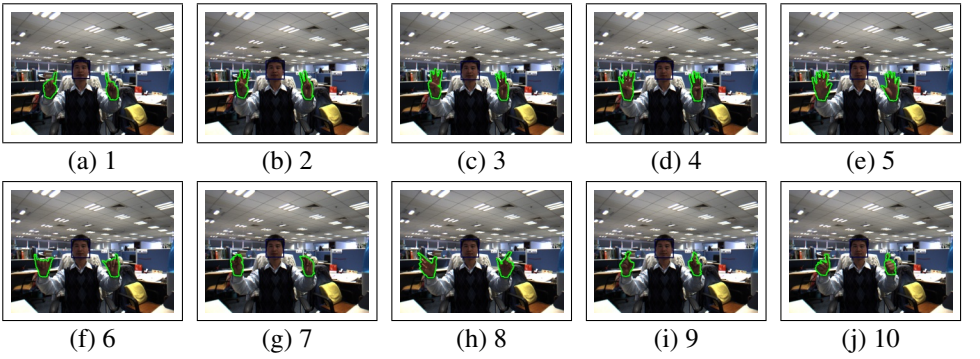
About parameter settings, our experiments show that the proposed method works well when $(\sigma_h = \sigma_0, \sigma_s = \sigma_0, \sigma_v = \sigma_0/0.6, \sigma_0 \in [55,80])$ for the skin color probability(Eq. (1)) and $(T_s \in [0.55,0.75], T^h \in [13,36])$ for the skin color weighted disparity competition(Eq. (5)). Note that we set $(\sigma_0 = 64)$ and $(T_s = 0.6, T^h = 32)$ for all tests in this paper. Figure 3 shows some segmentation examples which demonstrate the efficiency of the proposed method.

3.1 The segmentation accuracy

We compare the performance of the proposed skin color weighted disparity competition method with Manders et al.'s joint probability model [12], GraphCut [4][1] and the simple disparity competition which is not weighted by the skin color probability.

An illustrative example is shown in Figure 4. Manders et al.'s joint probability model produces the output by taking the pointwise multiplicative product of the disparity map and the skin probability map, So its result would be confined to the disparity map and neglect

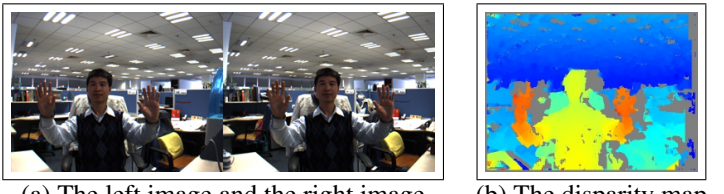
[1] We use the maxflow code by V. Kolmogorov (http://www.cs.ucl.ac.uk/staff/V.Kolmogorov/software.html).



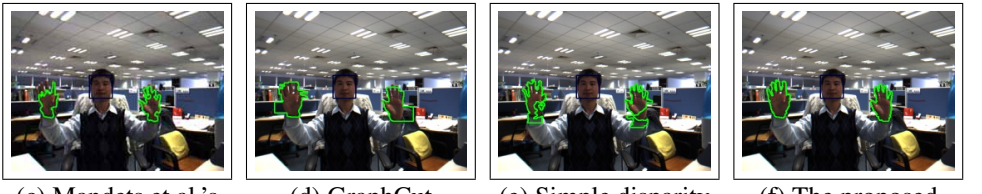
(a) 1 (b) 2 (c) 3 (d) 4 (e) 5

(f) 6 (g) 7 (h) 8 (i) 9 (j) 10

Figure 3: Segmentation examples of the proposed method for chinese numeral gesture from 1 to 10



(a) The left image and the right image (b) The disparity map



(c) Mandets et al.'s method (d) GraphCut (e) Simple disparity competition (f) The proposed method

Figure 4: Performance comparison of Manders et al's joint probability method [12], GraphCut [4], simple disparity competition without weighting and the proposed skin color weighted disparity competition.

those hand pixels with uncertain disparity, as illustrated by Figure 4(b) and Figure 4(c). It is plausible to employ GraphCut to achieve the accurate hand segmentation by considering the marked confident hand pixels as the foreground seeds and those non-zero pixels in the BG map as the background seeds. But our experiments show that the results are not satisfied, as shown in Figure 4(d). The possible reason includes two aspects. Firstly, the background color model learned from few background feeds can not enough characterize the background and thus mislead GraphCut to judge some background pixels as the hand pixels. Secondly, GraphCut seems to fail in segmenting the concave detail of the hand for the possible reason of the serious concave or consequent image blue. The simple disparity competition without weighting suffers from the problem of confusion pixels, i.e. some background pixels satisfy the hand disparity very good or some hand pixels satisfy the background disparity very good, which leads to a high false acceptation or rejection rate, as shown in Figure 2 and Figure 4(e). Owing to incorporating the skin probability to weight the disparity competition,

Step	Computing disparity	Detecting face	Marking hand pixels	Performing disparity competition	Total
Time	0.041 sec	0.104 sec	0.011 sec	0.052 sec	0.208 sec

Table 1: The time detail of the proposed method. Note that only the fourth step is necessary when incorporated with ROIs tracking

the proposed method validates a hand pixel in a low false acceptation and a low false rejection rate. Figure 4(f) shows the improvement of the proposed method over the other three methods.

Furthermore, we also give the quantitive comparison by comparing the segmentation result of each method with the ground truth which has been obtained manually with the magic wand tool of PhotoShop. We use $(\mathbf{A} \cap \mathbf{B})/(\mathbf{A} \cup \mathbf{B})$ for the accuracy calculation, where $\mathbf{A}$ represents the segmentation result and $\mathbf{B}$ represents the ground truth. The average segmentation accuracy for the proposed method is 83.8%, which is better than 63.9% for Manders et al.'s joint probability model, 69.8% for GraphCut and 59.2% for the simple disparity competition.

3.2 The Frame Rate

The main steps of the proposed disparity-competition method contains computing the disparity map, detecting face, marking the confident hand pixels, and performing skin color weighted disparity competition. For the first frame, all the four steps are needed and require a total of 0.208 sec on a Lenovo ThinkPad T400 with 2.5 GHz Dual Intel CPU and 2GB RAM. Note that the time of 0.208 sec contains no capture time. In detail, the proposed method requires 0.041 sec, 0.104 sec, 0.011 sec and 0.052 sec for the four steps in order, as listed in Table 1.

The frame rate will be the inefficient 5 fps if all four steps are also necessary for the successive frames. Fortunately, with ROIs tracking, only the last step of performing skin color weighted disparity competition is necessary for the successive frames because the predicted ROIs in the tracking can be directly considered as the ROIs in the current frame and all needed models for skin color weighted disparity competition, including the skin color probability model, the disparity limit of the hand and of the background, can be directly transferred from the previous frame. Of course, all models need to be updated to the current frame after disparity competition. In fact, the average frame rate for the test sequences is 19.2 fps that is acceptable for the common HCI application.

4 Conclusions

This paper has presented the novel skin color weighted disparity competition method for hand segmentation in clutter environment from stereo camera. Its efficiency lies in the ability of robustly validating the hand pixel owing to utilizing the skin probability to weight disparity competition. Although the proposed method can segment the detail of the hand to some degree, there is room for further improvement in the segmentation accuracy. Besides, the proposed method doesn't work well when hand-hand occlusion or hand-face occlusion happens in both the left image and the right image. We plan to work on the two aspects in the coming future. Moreover, we also plan to reconstruct the 3D appearance of each appearing hand since its silhouette/contour from either view of stereo camera is available now.

Acknowledgment This work is partially sponsored by National Key Technology R&D Program under contract No.2008BAH26B03, Natural Science Foundation of China under contract No.60973067, and QUALCOMM Incorporated.

References

[1] http://www.sourceforge.net/projects/opencvlibrary.

[2] http://www.ptgrey.com/products/triclopsSDK/index.asp.

[3] A. Argyros and M. Lourakis. Real-time tracking of multiple skin-colored objects with a possibly moving camera. In *European Conference on Computer Vision*, pages 368–379. 2004.

[4] Y. Y. Boykov and M. P. Jolly. Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images. In *IEEE International Conference on Computer Vision*, volume 1, pages 105–112, 2001.

[5] M. Bray, E. Koller-Meier, and L. Van Gool. Smart particle filtering for 3d hand tracking. In *IEEE International Conference on Automatic Face and Gesture Recognition*, pages 675–680, 2004.

[6] M. Z. Brown, D. Burschka, and G. D. Hager. Advances in computational stereo. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 25(8):993–1008, 2003.

[7] W.Y. Chang, C.S. Chen, and Y.P. Hung. Appearance-guided particle filtering for articulated hand tracking. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 1, pages 235–242, 2005.

[8] Q. Delamarre and O. Faugeras. 3d articulated models and multiview tracking with physical forces. *Computer Vision and Image Understanding*, 81(3):328–357, 2001.

[9] A. Erol, G. Bebis, M. Nicolescu, R. D. Boyle, and X. Twombly. Vision-based hand pose estimation: A review. *Computer Vision and Image Understanding*, 108(1-2):52–73, 2007.

[10] V. Lempitsky, P. Kohli, C. Rother, and T. Sharp. Image segmentation with a bounding box prior. In *IEEE International Conference on Computer Vision*, pages 277–284, 2009.

[11] J. Letessier and F. Berard. Visual tracking of bare fingers for interactive surfaces. In *UIST'04: 17th Annual ACM symposium on User Interface Software and Technology*, pages 119–122, New York, NY, USA, 2004. ACM.

[12] C. Manders, F. Farbiz, J. H. Chong, K. Y. Tang, G. G. Chua, M. H. Loke, and M. L. Yuan. Robust hand tracking using a skin tone and depth joint probability model. In *IEEE International Conference on Automatic Face & Gesture Recognition*, pages 1–6, 2008.

[13] S. Mitra and T. Acharya. Gesture recognition: A survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C: Applications and Reviews*, 37(3):311–324, 2007.

[14] Z. Mo, J.P. Lewis, and U. Neumann. Smartcanvas: a gesture-driven intelligent drawing desk system. In *IUI q́05: 10th International Conference on Intelligent User Interfaces, ACM*, pages 239–243, 2005.

[15] C. Rother, V. Kolmogorov, and A. Blake. "grabcut": interactive foreground extraction using iterated graph cuts. In *ACM SIGGRAPH 2004*, pages 309–314, Los Angeles, California, 2004. ACM.

[16] M. Shahzad and L. Joe. Visual touchpad: a two-handed gestural input device. In *Proceedings of the 6th international conference on Multimodal interfaces*, pages 289–296, State College, PA, USA, 2004. ACM.

[17] B. Stenger, P. R. S. Mendonca, and R. Cipolla. Model-based 3d tracking of an articulated hand. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, volume 2, pages 310–315, 2001.

[18] B. Stenger, A. Thayananthan, P. H. S. Torr, and R. Cipolla. Model-based hand tracking using a hierarchical bayesian filter. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 28(9):1372–1384, 2006.

[19] Y. Wu, J. Lin, and T. S. Huang. Analyzing and capturing articulated hand motion in image sequences. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 27(12):1910–1922, 2005.

[20] S. Yan, S. Shan, X. Chen, and W. Gao. Locally assembled binary (lab) feature with feature-centric cascade for fast and accurate face detection. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7, 2008.

[21] L.G. Zhang, X. Chen, C. Wang, and W. Gao. Robust automatic tracking of skin-colored objects with level set based occlusion handling. In *Proceedings of The 8th Int'l Gesture Workshop, GW'09*, Bielefeld, Germany, 2009.