

HD: Efficient Hand Detection and Tracking

C. Rouge, S. Shaikh, and J.I. Olszewska
School of Computing and Technology
University of Gloucestershire
Cheltenham, United Kingdom
Email: joanna.olszewska@ieee.org

Abstract— Automated hand detection is useful for applications requiring reliable hand posture and hand gesture processing. Such applications include human-computer/human-robot interfaces for rehabilitation, serious games, or non-invasive medical diagnosis. Hence, in this paper, we focus on the design and development of a robust and fast hand detection and tracking (HD) system. The design of our HD system involved the study of the human skin colour and people’s foreground properties, in order to merge these information for an efficient hand detection and tracking. Experiments have been carried out in real-world environment and have demonstrated the excellent performance of our HD system.

I. INTRODUCTION

Hand detection and tracking are the actions of automatically locating and following human hands to extract information useful for various applications, which involve interactions of humans with computers (HCI) [1] through interfaces such as mouses [2], gloves [3], interactive tables [4], or natural gestures [5].

Automated hand detection could be applied for multimodal interactions of humans within augmented-reality frameworks, e.g. game control [6], character animation [7], etc., or it could be used for secure and safe interactions of humans within real-world environment. Indeed, hand detection could be used in biometrics [8], since palms of the human hands contain unique patterns of ridges and valleys, in the same way that fingerprints, whereas presenting much larger areas and thus being even more distinctive than the fingerprints [9]; in surveillance for abnormal behaviour detection [10] or in crime prevention [11].

Medical applications involving the monitoring of human interaction with robots (HRI) [12] can embed hand detection, e.g. for functional rehabilitation [13] such as exercises of patient’s interaction with objects s/he grasps [14] to improve stroke recovery [15].

Moreover, hand detection and tracking is promising in diagnosis of neurologic disorders such as Parkinson’s disease. Indeed, this progressively degenerative movement disorder [16] implies quantitative evaluations of hand movement as well as hand tremor for neurological examination [17]. These measurements could be based on the visual tracking of patient’s hand motion [18] or on the paced finger tapping test [19] assessed with visual rating scales such as the Unified Parkinson’s Disease Rating Scale (UPDRS), rather than on the traditional invasive methods such as electroencephalograms (EEGs) [20].

Other applications of hand detection and gesture recognition could improve the communication and interaction of people presenting difficulties [21]. For example, hand detection in context of sign language could help deaf people naturally interacting with machines or non-deaf people interacting with deaf ones [22].

The main challenge of vision-based hand detection is to cope with the large variability of human hand’s appearance due to a huge number of degrees of freedom (DoFs) of the hand’s movements, to different skin-colour possibilities as well as to the variations in view points, scales, and speed of the camera capturing the scene [21].

Hence, hands have been represented as 2D [23] or 3D models [24], and various methods have been proposed, involving eigen-based hand model [25], hand/not-hand SVM classifier [26] or AdaBoost-based hand detector [27], to automatically recognize human hands in images or videos. Despite their effectiveness, these methods lack of detection accuracy in some real situations.

In this work, the adopted approach for hand detection is focused on robustness, and it combines threshold-based colour detection with background subtraction. Moreover, enhanced Adaboost face detection [28] is used to differentiate hands from the human face region, while hand tracking is achieved by adding the foreground extraction obtained from one frame to another.

Hands could be visually detected or tracked by a range of sensors [29], such as ego-centric/self-mounted cameras [30], stereo camera [31], or Kinect [32]. In this paper, we focus on hand detection and tracking solely based on a single, static camera feed, which is a non-invasive and low-cost solution, leading to a convenient, computer-vision-based system.

The product developed in this work is capable of both effectively detecting human hands without building a hand model, but instead applying pixel thresholding based on a skin-colour model, and efficiently tracking an individual’s hand movement based on the foreground blob and background subtraction information. This HD system could be used for numerous real-world purposes, and in particular for medical applications.

The original contribution of our work is twofold and consists in the design and development of a new hand detection and tracking system based on features such as colour and motion as well as in a study of the colour skin appearance within YCbCr colour space.

The paper is structured as follows. In Section II, we present our hand detection and tracking (HD) system, while in Section III we report and discuss the carried out experiments which results show the developed HD system has excellent performance on real-world video datasets. Conclusions are drawn up in Section IV.

II. OUR METHOD

The developed HD system as depicted in Fig. 1 uses skin-colour thresholding as explained in Section II-A. Foreground detection and background subtraction are described in Section II-B. Head detection is computed in order to discard that region to improve hand detection accuracy in case of full-body tracking as mentioned in Section II-C. Moreover, the fusion of all the computed information to detect the hands as well as the tracking process of the hand are presented in Section II-D.

A. Skin-Colour Detection

Let us consider a colour image or video frame $I(x, y)$ with M and N , its width and height, respectively.

At first, our system performs skin thresholding [33]. Among major colour spaces such as RGB [34], HSV [35], or YIQ [36], our system uses YCbCr colour space for thresholding to distinguish between skin and non-skin colours. Indeed, the YCbCr colour space is the most popular choice in skin-colour detection methods, because its luminance component Y as well as chrominance components Cb and Cr are separated and could be easily computed from RGB values [37]. This RGB-YCbCr transformation possibility as well as the explicit separation of luminance and chrominance components, which furthermore brings some degree of robustness to illumination variations [38], make this colour space attractive for skin-colour modeling [39].

Moreover, YCbCr colour space is used in our work due to the fact it has a good skin feature clustering as different human skin colours from different races fall in a compact region in the YCbCr colour space [40]. On the opposite, the RGB colour space, which is the basic colour space of image processing, cannot be segmented by simple thresholding [41], because RGB colour space has perceived non-uniformity [42]. Although it may be possible to detect one skin colour using other colour spaces, it would be not possible to detect with the RGB colour space the perceived colour of human skin which varies greatly across human races or even between individuals of the same race. So, to increase invariance against illumination variability, our HD system operates in the YCbCr colour space in order to approximate the chromaticity of skin rather than its apparent colour value.

Indeed, in our system, the distribution of skin pixels values is highlighted based on Cb and Cr components [40] as an additional threshold conditions reported in Fig. 2, whereas the luminance component is eliminated to remove the effect of shadows, illumination changes, and modulations of orientation of the skin surface relative to the light source(s) [38].

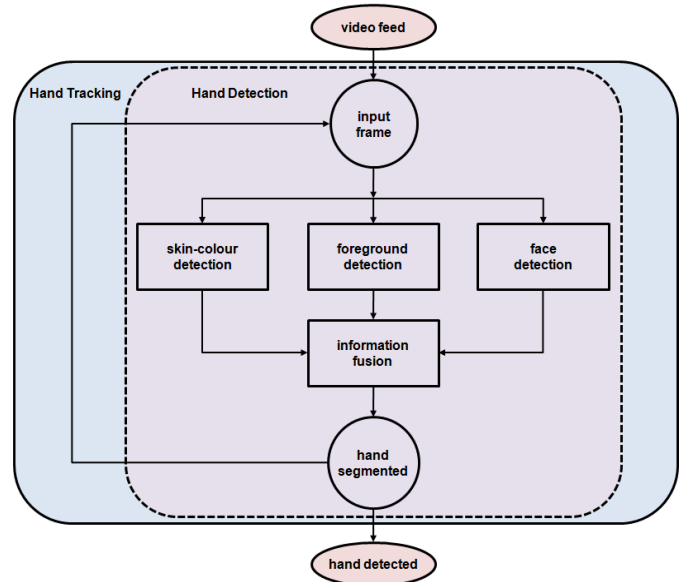


Fig. 1. Overview of our hand detection and tracking (HD) system.

So, the adopted YCbCr colour threshold detection values are as follows:

$$Cb = 0.148 \times R - 0.291 \times G + 0.439 \times B + 128, \quad (1)$$

$$Cr = 0.439 \times R - 0.368 \times G - 0.071 \times B + 128, \quad (2)$$

where the skin region is segmented based on the thresholding values [43] defined as $140 \leq Cb_{skin} \leq 195$ and $140 \leq Cr_{skin} \leq 165$.

B. Foreground Detection

To extract the foreground blob, we combine frame difference and background subtraction techniques [44]. This consists in computing in parallel, on one hand, the difference between a current frame $I_k(x, y)$ and the previous one $I_{k-1}(x, y)$, and on the other hand, the difference between the current frame $I_k(x, y)$ and a background model of the scene, and afterwards, to combine both results in order to extract the foreground in the corresponding view [44]. It is worth noting that background subtraction is based on the assumption that the camera does not move with respect to the static background.

To model the background, we adopt the running Gaussian average (RGA) [45], which is suitable for real-time tracking, and which is characterized by the mean μ_b and the variance $(\sigma_b)^2$.

Hence, the foreground is determined by

$$F(x, y) = \begin{cases} 1 & \text{if } |F_f(x, y) \cup F_b(x, y)| = 1, \\ 0 & \text{otherwise,} \end{cases} \quad (3)$$

with

$$F_f(x, y) = \begin{cases} 1 & \text{if } |I_k(x, y) - I_{k-1}(x, y)| > tf, \\ 0 & \text{otherwise,} \end{cases} \quad (4)$$

and

$$F_b(x, y) = \begin{cases} 1 & \text{if } |I_k(x, y) - \mu_b| > n \cdot \sigma_b, \\ 0 & \text{otherwise,} \end{cases} \quad (5)$$

where tf and $n \in \mathbb{N}_0$ are the thresholds for frame difference and background subtraction processes, respectively [44].

Finally, to compute a blob defined by labeled connected regions, morphological operations [4], [46], such as dilation, filling, etc. are applied to the extracted foreground F , in order to exploit the existing information on the neighboring pixels, in the frame:

$$f(x, y) = \text{Morph}(F(x, y)). \quad (6)$$

C. Face Detection

Face detection is performed using the lighting-variable Adaboost algorithm [28], relying on global image intensity [28] and Haar-like features [47] rather than face skin-colour as in [48], in order to robustify the HD system against illumination changes.

Hence, based on the gray image $I_g(x, y) = 0.299 R(x, y) + 0.587 G(x, y) + 0.114 B(x, y)$, the average value I_{gAVG} of the global image intensity is defined as

$$I_{gAVG} = \frac{1}{MN} \sum_{x=1}^M \sum_{y=1}^N I_g(x, y), \quad (7)$$

while local Haar-like features f [47] encode the existence of oriented contrasts between regions in the processed image, and are computed by subtracting the sum of all the pixels of a subregion of the local feature from the sum of the remaining region of the local feature using the integral image representation $II(i, j)$ which is defined as follows:

$$II(i, j) = II(i-1, j) + II(i, j-1) - II(i-1, j-1) + I(i, j), \quad (8)$$

where $I(i, j)$ is the pixel value of the original image at the position (i, j) .

This detection system requires a training phase during which it builds strong classifiers based on cascades of weak classifiers [28]. In particular, to form a T -stage cascade, T weak classifiers are selected using the AdaBoost algorithm [47]. In fact at a t stage of this cascade, a sub-window u of an image from the training set is computed by Eq. (8) and it is passed to the corresponding t^{th} weak classifier. If the region is classified as a non-face, the sub-window is rejected. If not, it is passed to the $t+1$ stage, and so forth. Consequently, more stages the cascade owns, more selective it is, i.e. less false positive detections occur. However, that could lead to the increase in the number of false negative detection.

In order to select at each t level (with $1 < t < T$) the best weak classifier, an optimum threshold θ_t is computed by minimizing the classification error due to the selection of a particular Haar-like feature value $f_t(u)$. The resulting weak classifier k_t is thus obtained as follows:

$$k_t(u, f_t, p_t, \theta_t) = \begin{cases} 1 & \text{if } p_t f_t(u) < p_t \theta_t, \\ 0 & \text{otherwise,} \end{cases} \quad (9)$$

where p_t is the polarity indicating the direction of the inequality.

Then, a strong classifier $K_T(u)$ is constructed by taking a weighted combination of the selected weak classifiers k_t according to

$$K_T(u) = \begin{cases} 1 & \text{if } \sum_{t=1}^T \alpha_t k_t(u) \geq \frac{1}{2} \sum_{t=1}^T \alpha_t, \\ 0 & \text{otherwise,} \end{cases} \quad (10)$$

where $\frac{1}{2} \sum_{t=1}^T \alpha_t$ is the AdaBoost threshold and α_t is a voting coefficient computed based on the classification error in each stage t of the T stages of the cascade [47].

Next, the lighting-adaptable strong super-classifier $\mathcal{K}(u)$ is defined as

$$\mathcal{K}(u) = \begin{cases} K_L(u) & \text{if } I_{gAVG} > I_{g_{th}}, \\ K_D(u) & \text{if } I_{gAVG} \leq I_{g_{th}}, \end{cases} \quad (11)$$

where $I_{g_{th}}$ is the global image intensity threshold and where D and L (with $D \geq L$) are the numbers of the weak classifiers for dark and light images, respectively.

In this way, the system allows to automatically select the number of stages of the AdaBoost cascade accordingly to the lighting conditions expressed in Eq. (11) by I_{gAVG} and could be applied on any new input frame [28].

D. Hand Detection and Tracking

Information fusion between the skin-colour detection, foreground detection, and additional face detection is performed in order to detect the hand(s) as illustrated in Fig. 1. Indeed, the skin detection (Section II-A) allows the HD system to remove the forearm from the detected foreground (Section II-B), while the hand blob complements the skin-detected region in order to have a precise detection of the human hand(s) as shown in Fig. 3 (d). In case a face is detected (Section II-C) in a frame containing the full body, this head region is discarded to avoid confusion between regions containing skin colour and to extract only hands' regions as in Fig. 4 (d).

As the detection method is fast enough to operate at image acquisition frame rate, it can be used for hand tracking. Tracking hands is difficult since hands can move very fast and their appearance can change vastly within a few frames. Tracking hands is defined in the HD system as the frame-to-frame correspondence of the segmented hand regions (see Fig. 1). Tracking provides the inter-frame linking of hand/finger appearances. This provides trajectories of features in time. These trajectories convey essential information regarding the gesture [38] and might be used either in a raw form (e.g. in hand-guided control applications) or after further analysis (e.g. in hand gesture recognition).

III. EXPERIMENTS AND DISCUSSION

Our HD system and its components have been tested on real-world videos captured at 25f/s and with a resolution of 640x360 pixels. All the experiments have been carried out on a computer with an Intel Core (TM) 2 Duo CPU @2.5GHz processor, with a 2Gb RAM, and running MatLab software.





















| SKIN-COLOUR DETECTION | | | |
|-----------------------|---|--|--|
| data | skin colour palette | mathematical formula | examples of original frames |
| input |  | $\begin{pmatrix} Cb \\ Cr \end{pmatrix} = \begin{pmatrix} c_1 & c_2 & c_3 \\ c_4 & c_5 & c_6 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix} + \begin{pmatrix} d_b \\ d_r \end{pmatrix}$ |  |
| tests | segmentation result | coefficient values | skin detection results |
| test 1 output |  | $c_1 = 0.048; c_2 = -0.191; c_3 = 0.339; d_b = 128;$ $c_4 = 0.339; c_5 = -0.268; c_6 = -0.071; d_r = 128;$ |  |
| test 2 output |  | $c_1 = 0.048; c_2 = -0.191; c_3 = 0.339; d_b = 128;$ $c_4 = 0.439; c_5 = -0.368; c_6 = -0.071; d_r = 128;$ |  |
| test 3 output |  | $c_1 = 0.148; c_2 = -0.291; c_3 = 0.439; d_b = 123;$ $c_4 = 0.439; c_5 = -0.368; c_6 = -0.071; d_r = 123;$ |  |
| test 4 output |  | $c_1 = 0.148; c_2 = -0.291; c_3 = 0.439; d_b = 128;$ $c_4 = 0.339; c_5 = -0.268; c_6 = -0.071; d_r = 128;$ |  |
| test 5 output |  | $c_1 = 0.148; c_2 = -0.291; c_3 = 0.439; d_b = 128;$ $c_4 = 0.439; c_5 = -0.368; c_6 = -0.071; d_r = 128;$ |  |
| test 6 output |  | $c_1 = 0.148; c_2 = -0.291; c_3 = 0.439; d_b = 133;$ $c_4 = 0.439; c_5 = -0.368; c_6 = -0.071; d_r = 133;$ |  |
| test 7 output |  | $c_1 = 0.148; c_2 = -0.291; c_3 = 0.439; d_b = 128;$ $c_4 = 0.539; c_5 = -0.468; c_6 = -0.171; d_r = 128;$ |  |
| test 8 output |  | $c_1 = 0.248; c_2 = -0.391; c_3 = 0.539; d_b = 128;$ $c_4 = 0.439; c_5 = -0.368; c_6 = -0.071; d_r = 128;$ |  |
| test 9 output |  | $c_1 = 0.248; c_2 = -0.391; c_3 = 0.539; d_b = 128;$ $c_4 = 0.539; c_5 = -0.468; c_6 = -0.171; d_r = 128;$ |  |

Fig. 2. Results of the skin-colour modeling and detection tests with different coefficient values.

A. Experiment 1

The first experiment consists in skin-colour detection and it evaluates the colour distributions from the hand in order to improve overall detection performance as illustrated in Fig. 2. Indeed, in the YCbCr colour space, Cb and Cr channels are employed owing to their relative insensitivity to the lighting variations [49].

In test 1, decreasing both $c_1 - c_3$ and $c_4 - c_6$ coefficient values came to no positive effect, showing a decrease in lighter and darker skin tones compared to the original palette, and resulting in poor hand detection. In test 2, decreasing $c_1 - c_3$ threshold values reduces system's capability to detect darker skin regions, and thus is not recommended to use with different skin tones and environmental changes e.g. light, shading and backgrounds. The purpose of the test 3 is to determine if changing the end values for both $c_1 - c_3$ and $c_4 - c_6$ thresholds would have any effect on results. The difference is easy to see between this test and the original, as the darker skin region is not significantly detected. In test 4, decreasing the $c_4 - c_6$

values demonstrated no positive changes to that of the original one. Lighter skin tones are not as visible and darker ones have not particularly changed. Once again, this can be demonstrated within frame results as, in this case, half of the hand seems missing in some frames. In test 5, this model accepts a wide range of skin colours, although it can be seen within the skin colour palette that it has had issues with detecting darker skin regions. In test 6, most of skin regions of the palette have been detected. However, within the frames tested, one can see that less of the hand is detected. Within test 7, darker regions have had minimal change between this and the original palette, but it is the lighter skin tones that have lost parts. This is demonstrated within the frame results as the hand in some frames becomes barely visible. In test 8, it is noticeable that the detection of darker region has increased, although it has also decreased detected regions within lighter skin tones. This can also be highlighted within the same test on the running frames, as in many parts of the hands, detection is missed. In test 9, by increasing both $c_1 - c_3$ and $c_4 - c_6$ coefficient

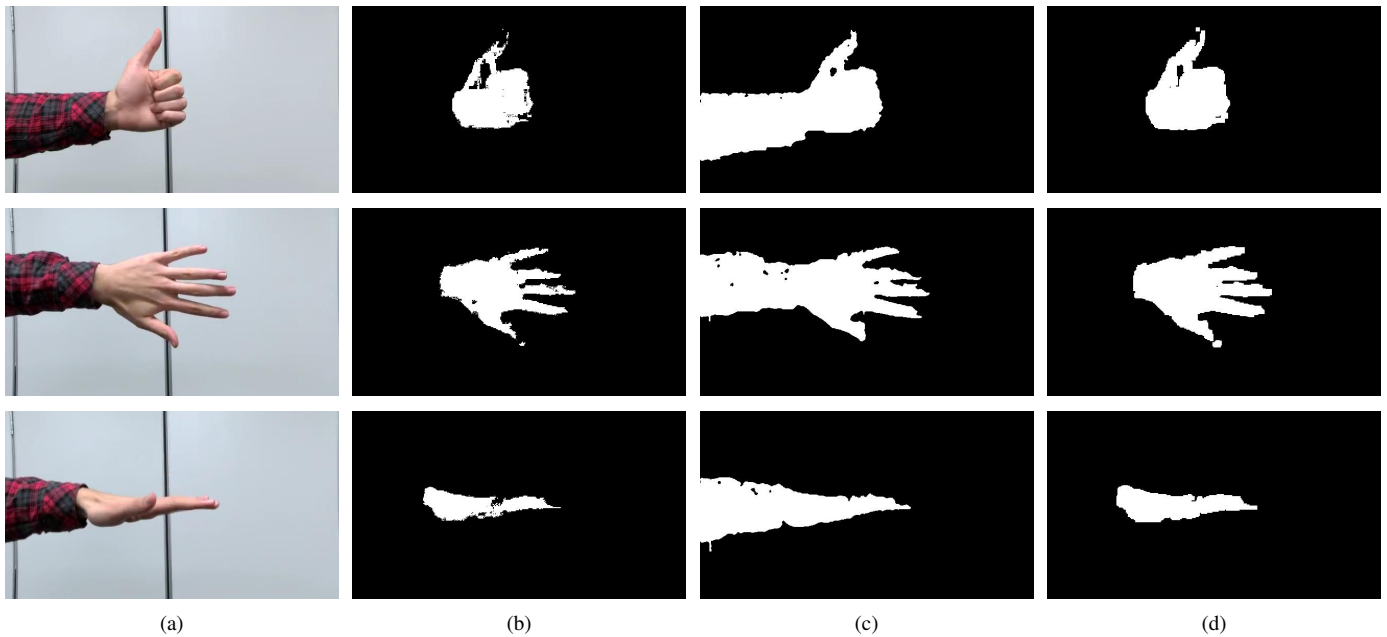


Fig. 3. Samples of our HD system for (a) input frame (1st row: frame 91, 2nd row: frame 288, 3rd row: frame 374); (b) skin-colour detection; (c) foreground detection; (d) hand detection.

values at the same time, there is a decrease within the detection of lighter skin tones and an increase in detection of darker skin tones. However, it is very prone to noise as demonstrated within each of the tested frames.

Finally, the adopted model is the one from the test 5, since the palette has been well respected and the hand has been the most clearly detected in all frames.

B. Experiment 2

Experiment 2 considers the hand detection in sequences without visible face. So, it tests the skin-colour detection and the foreground detection as well as the fusion of the related information for the hand detection.

Moreover, the background noise removal value was tested between the range of 0 and 1000, and morphological methods were also attempted. The skin colour was set to range of values in proportion to standard skin colours as described in Section III-A, the challenge being to detect darker shades of skin, nails, red tints within skin and shadows.

Results demonstrate the effectiveness of our HD system, since at all time the hand (Fig. 3(a)) is correctly detected (Fig. 3(d)), leading to 100% accuracy. Figure 3 clearly shows the blobs of the skin-detected wrist (Fig. 3(b)) and the foreground forearm (Fig. 3(c)) as well as the combined final result providng the detected hand (Fig. 3(d)) in each frame.

C. Experiment 3

The third experiments are testing the entire HD system for scenarios where other commonly visible, skin-colour-like body parts appear in the frames (Fig.4(a)), the face for instance. Hence, the HD system computes the skin colour objects (Fig. 4(b)), the foreground (Fig. 4(c)) as well as the face (Fig.

4(d)) which is detected within the frame by the lighting-variation-robust cascade object detector. Then, the HD system uses this information to remove the detected face blob, while the remaining skin-colour objects and the foreground are combined together in order to allow accurate hand detection (Fig. 4(e)).

The hand itself has been detected at all times, which gives a 100% detection rate, which is an excellent results compared to typical values for hand detection using blob approach (96.77%) [50], Histogram of Gradient (HOG) method (94.40%) [51], and Skin Colour Histogram of Gradient (SCHOG) technique (97.80%) [51]. Moreover, the running time of our overall HD system has been assessed, and it allows real-time hand tracking.

IV. CONCLUSIONS

The human hand has crucial importance to many actions that may occur in a person's day-to-day life, e.g. person-to-person interactions or person-to-object interactions. Hence, this paper focuses on the automatic detection of hands, with the use of background subtraction, foreground detection, and skin-colour thresholding methods. With these techniques combined, our HD system is capable of detecting human hands in a number of different scenarios such as person hand detection and tracking for Parkinson's disease diagnosis or for stroke recovery monitoring. Other potential applications could be 3D hand tracking [52] and gesture control for serious game [53], HRI and rehabilitation interactions [54] as well as individual finger processing for sign-language communication between deaf/non-deaf people [55].

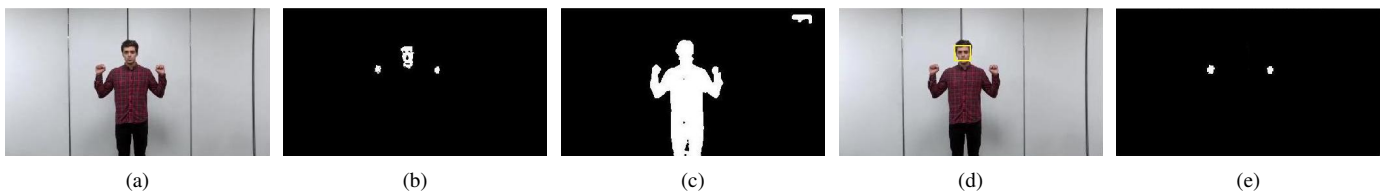


Fig. 4. Samples of our HD system for (a) input frame; (b) skin-colour detection; (c) foreground detection; (d) face detection; (e) hand detection.

REFERENCES

- [1] V. I. Pavlovic, R. Sharma, and T. S. Huang, "Visual interpretation of hand gestures for human-computer interaction: A review," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 677–695, 1997.
- [2] R. Suriya and V. Vijayachamundeswari, "A survey on hand gesture recognition for simple mouse control," in *Proceedings of the IEEE International Conference on Information Communication and Embedded Systems*, 2014, pp. 1–5.
- [3] A. Aristidou and J. Lasenby, "Motion capture with constrained inverse kinematics for real-time hand tracking," in *Proceedings of the IEEE International Symposium on Communications, Control and Signal Processing*, 2010, pp. 1–5.
- [4] N. J. Kim, S. Suh, and C. Choi, "Robust finger contact detection with majority quadrant search for interactive tabletop displays," in *Proceedings of the IEEE International Conference on Consumer Electronics*, 2014, pp. 518–519.
- [5] M. Refice, M. Savino, M. Adduci, and M. Caccia, "Automatic classification of gestures: A context-dependent approach," in *Proceedings of the IEEE Federated Conference on Computer Science and Information Systems (FedCSIS)*, 2011, pp. 743–750.
- [6] A. T. S. Chan, H. V. Leong, and S. H. Kong, "Real-time tracking of hand gestures for interactive game design," in *Proceedings of the IEEE International Symposium on Industrial Electronics*, 2009, pp. 98–103.
- [7] A. H. J. Moreira, S. Queiros, J. Fonseca, P. L. Rodrigues, N. F. Rodrigues, and J. L. Vilaca, "Real-time hand tracking for rehabilitation and character animation," in *Proceedings of the IEEE International Conference on Serious Games and Applications for Health*, 2014, pp. 1–8.
- [8] S. Ben Jemaa, M. Hammami, and H. Ben-Abdallah, "Contactless hand detection in complex image based on data-mining process," in *Proceedings of the IEEE Conference on Computer Systems and Applications*, 2013, pp. 1–4.
- [9] A. N. Kataria, D. M. Adhyaru, A. K. Sharma, and T. H. Zaveri, "A survey of automated biometric authentication techniques," in *Proceedings of the IEEE Nirma University International Conference on Engineering*, 2013, pp. 1–6.
- [10] W. Kong, A. Hussain, M. H. M. Saad, and N. M. Tahir, "Hand detection from silhouette for video surveillance application," in *Proceedings of the IEEE International Colloquium on Signal Processing and its Applications*, 2012, pp. 514–518.
- [11] E. Dente, J. Ng, A. Vrij, S. Mann, A. Bull, and A. Bharath, "Tracking small hand movements in interview situations," in *Proceedings of the IEEE International Symposium on Imaging for Crime Detection and Prevention*, 2005, pp. 55–60.
- [12] C. D. Lim, C.-M. Wang, C.-Y. Cheng, Y. Chao, S.-H. Tseng, and L.-C. Fu, "Sensory cues guided rehabilitation robotic walker realized by depth image-based gait analysis," *IEEE Transactions on Automation Science and Engineering*, vol. 13, no. 1, pp. 171–180, January 2016.
- [13] N. Nordin, M. R. Arshad, U. Soori, and N. M. Yin, "Virtual input using skin color model for robotic platform control," in *Proceedings of the IEEE International Conference on Signal and Image Processing Applications*, 2009, pp. 305–311.
- [14] G. P. Rosati Papini, M. Fontana, and M. Bergamasco, "Desktop haptic interface for simulation of hand-tremor," *IEEE Transactions on Haptics*, vol. 9, no. 1, pp. 33–42, 2016.
- [15] L. Shires, S. Battersby, J. Lewis, D. Brown, N. Sherkat, and P. Standen, "Enhancing the tracking capabilities of the Microsoft Kinect for stroke rehabilitation," in *Proceedings of the IEEE International Conference on Serious Games and Applications for Health*, 2013, pp. 1–8.
- [16] G. V. Kondraske and R. M. Stewart, "Web-based evaluation of Parkinson's disease subjects: Objective performance capacity measurements and subjective characterization profiles," in *Proceedings of the IEEE Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 799–802.
- [17] R. LeMoyné, C. Coroian, and T. Mastroianni, "Quantification of Parkinson's disease characteristics using wireless accelerometers," in *Proceedings of the IEEE International Conference on Complex Medical Engineering*, 2009, pp. 1–5.
- [18] J. Ide, T. Sugi, N. Murakami, F. Shima, H. Shibasaki, and M. Nakamura, "Quantitative evaluation of hand movement on visual target tracking for patients with Parkinson's disease," in *Proceedings of the IEEE International Conference on Complex Medical Engineering*, 2007, pp. 1896–1900.
- [19] S. Das, L. Trutoiu, A. Murai, D. Alcindor, M. Oh, F. De la Torre, and J. Hodgins, "Quantitative measurement of motor symptoms in Parkinson's disease: A study with full-body motion capture data," in *Proceedings of the IEEE International Conference on the Engineering in Medicine and Biology Society*, 2011, pp. 6789–6792.
- [20] J. Chiang, Z. J. Wang, and M. J. McKeown, "A generalized multivariate autoregressive (GmAR)-based approach for EEG source connectivity analysis," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 453–465, 2012.
- [21] S. S. Rautaray and A. Agrawal, "Design of gesture recognition system for dynamic user interface," in *Proceedings of the IEEE International Conference on Technology Enhanced Education*, 2012, pp. 1–6.
- [22] V. Viitaniemi, M. Karppa, and J. Laaksonen, "Experiments on recognising the handshape in blobs extracted from sign language videos," in *Proceedings of the IEEE International Conference on Pattern Recognition (ICPR'14)*, 2014, pp. 2584–2589.
- [23] R. Z. Khan and N. A. Ibraheem, "Survey on gesture recognition for hand image postures," *Computer and Information Science*, vol. 5, no. 3, pp. 110, 2012.
- [24] A. El-Sawah, C. Joslin, N. D. Georganas, and E. M. Petriu, "A framework for 3D hand tracking and gesture recognition using elements of genetic programming," in *Proceedings of the IEEE Canadian Conference on Computer and Robot Vision*, 2007, pp. 495–502.
- [25] S. Lu, G. Tsechpenakis, D. N. Metaxas, M. L. Jensen, and J. Kruse, "Blob analysis of the head and hands: A method for deception detection," in *Proceedings of the IEEE Annual Hawaii International Conference on System Sciences*, 2005, pp. 20c–20c.
- [26] A. Thangali and S. Sclaroff, "An alignment based similarity measure for hand detection in cluttered sign language video," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, 2009, pp. 89–96.
- [27] E.-J. Ong and R. Bowden, "A boosted classifier tree for hand shape detection," in *Proceedings of the IEEE International Conference on Automatic Face and Gesture Recognition*, 2004, pp. 889–894.
- [28] R. Wood and J. I. Olszewska, "Lighting-variable AdaBoost based-on system for robust face detection," in *Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*, 2012, pp. 494–497.
- [29] S. Berman and H. Stern, "Sensors for gesture recognition systems," *IEEE Transactions on Systems, Man, and Cybernetics - Part C: Applications and Reviews*, vol. 42, no. 3, pp. 277–290, 2012.
- [30] J. Kumar, Q. Li, S. Kyal, E. A. Bernal, and R. Bala, "On-the-fly hand detection training with application in egocentric action recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW'15)*, 2015, pp. 18–27.
- [31] D. D. Nguyen, T. C. Pham, X. D. Pham, S. H. Jin, and J. W. Jeon, "Finger extraction from scene with grayscale morphology and BLOB analysis," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics*, 2009, pp. 324–329.

- [32] J. Suarez and R. R. Murphy, "Hand gesture recognition with depth images: A review," in *Proceedings of the IEEE International Symposium on Robot and Human Interactive Communication*, 2012, pp. 441–417.
- [33] S. Rungruangbaiyok, R. Duangsoithong, and K. Chetpattananondh, "Ensemble threshold segmentation for hand detection," in *Proceedings of the IEEE International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology*, 2015, pp. 1–5.
- [34] E. Marilly, A. Gonguet, O. Martinot, and F. Pain, "Gesture interactions with video: From algorithms to user evaluation," *Bell Labs Technical Journal*, vol. 17, no. 4, pp. 103–118, 2013.
- [35] Y. R. Wang, J. L. Syu, H. T. Li, and L. Yang, "Fast hand detection and gesture recognition," in *Proceedings of the IEEE International Conference on Machine Learning and Cybernetics*, 2015, vol. 1, pp. 408–413.
- [36] Z. Musa, K. Jumari, and N. Zainal, "A method of human skin detection base on background subtraction and color enhancement," in *Proceedings of the IEEE Symposium on Business, Engineering and Industrial Applications*, 2011, pp. 498–502.
- [37] D. Xu, Y. L. Chen, X. Wu, Y. Ou, and Y. Xu, "Integrated approach of skin-color detection and depth information for hand and face localization," in *Proceedings of the IEEE International Conference on Robotics and Biomimetics*, 2011, pp. 952–956.
- [38] S. Rautaray, S. Siddharth, and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: A survey," *Artificial Intelligence Review*, vol. 43, no. 1, pp. 1–54, 2015.
- [39] V. Vezhnevets, V. Sazonov, and A. Andreeva, "A survey on pixel-based skin color detection techniques," in *Proceedings of the IEEE Graphicon*, 2003, vol. 3, pp. 85–92.
- [40] S. Bilal, R. Akmeliawati, M. J. E. Salami, A. A. Shafie, and E. M. Bouhabba, "An hybrid method using Haar-like and skin-color algorithm for hand posture detection, recognition and tracking," in *Proceedings of the IEEE International Conference on Mechatronics and Automation*, 2010, pp. 934–939.
- [41] S. Xie and J. Pan, "Hand detection using robust color correction and Gaussian mixture model," in *Proceedings of the IEEE International Conference on Image and Graphics*, 2011, pp. 553–557.
- [42] B. Junxia, Y. Jianqin, W. Jun, and Z. Ling, "Hand detection based on depth information and color information of the Kinect," in *Proceedings of the IEEE Chinese Control and Decision Conference*, 2015, pp. 4205–4210.
- [43] J. Rajan, "Pantech Solution," 2013, Available online at: <https://www.pantechsolutions.net/blog/matlab-code-for-background-subtraction/>.
- [44] J. I. Olszewska, "Multi-camera video object recognition using active contours," in *Proceedings of the International Conference on Bio-Inspired Systems and Signal Processing*, 2015, pp. 379–384.
- [45] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Real-time tracking of the human body," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780–785, 1997.
- [46] J. Sonkusare, N.B. Chopade, R. Sor, and S. L. Tade, "A review on hand gesture recognition system," in *Proceedings of the IEEE International Conference on Computing, Communication, Control, and Automation*, 2015, pp. 790–794.
- [47] P. Viola and M. J. Jones, "Robust real-time face detection," *International Journal of Computer Vision*, vol. 57, no. 2, pp. 137–154, January 2004.
- [48] N. Soontranon, S. Aramvith, and T. H. Chalidabhongse, "Face and hands localization and tracking for sign language recognition," in *Proceedings of the IEEE International Symposium on Communications and Information Technology*, 2004, vol. 2, pp. 1246–1251 vol.2.
- [49] S. Kang, J. Oh, and H. Hong, "Human gesture detection based on 3D blobs and skeleton model," in *Proceedings of the IEEE International Conference on Information Science and Applications*, 2013, pp. 1–4.
- [50] H. S. Park and K. H. Jo, "Real-time hand gesture recognition for augmented screen using average background and Camshift," in *Proceedings of the IEEE Korea-Japan Joint Workshop on Frontiers of Computer Vision*, 2013, pp. 18–21.
- [51] X. Meng, J. Lin, and Y. Ding, "An extended HOG model: SCHOG for human hand detection," in *Proceedings of the IEEE International Conference on Systems and Informatics*, 2012, pp. 2593–2596.
- [52] H. Cheng, L. Yang, and Z. Liu, "A survey on 3D hand gesture recognition," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 9, pp. 1659–1673, 2016.
- [53] L. Evett, A. Burton, S. Battersby, D. Brown, N. Sherkat, G. Ford, H. Liu, and P. Standen, "Dual camera motion capture for serious games in stroke rehabilitation," in *Proceedings of the IEEE International Conference on Serious Games and Applications for Health*, 2011, pp. 1–4.
- [54] W.-H. Chen, Y.-H. Lin, and S.-J. Yang, "A generic framework for the design of visual-based gesture control interface," in *Proceedings of the IEEE Conference on Industrial Electronics and Applications*, 2010, pp. 1522–1525.
- [55] H. V. Verma, E. Aggarwal, and S. Chandra, "Gesture recognition using Kinect for sign language translation," in *Proceedings of the IEEE International Conference on Image Information Processing*, 2013, pp. 96–100.