Scientific Research Publishing

Journal of Signal and Information Processing, 2018, 9, 258-265
http://www.scirp.org/journal/jsip
ISSN Online: 2159-4481
ISSN Print: 2159-4465

Low-Cost Posture Recognition of Moving Hands by Profile-Mold Construction in Cluttered Background and Occlusion

Din-Yuen Chan*, Guan-Hong Lin, Xi-Wen Wu

Department of Computer Science and Information Engineering, National Chiayi University, Taiwan
Email: *dychan@mail.ncyu.edu.tw

How to cite this paper: Chan, D.-Y., Lin, G.-H. and Wu, X.-W. (2018) Low-Cost Posture Recognition of Moving Hands by Profile-Mold Construction in Cluttered Background and Occlusion. *Journal of Signal and Information Processing*, **9**, 258-265.
https://doi.org/10.4236/jsip.2018.94016

Received: September 18, 2018
Accepted: October 30, 2018
Published: November 2, 2018

Copyright © 2018 by authors and Scientific Research Publishing Inc.
This work is licensed under the Creative Commons Attribution International License (CC BY 4.0).
http://creativecommons.org/licenses/by/4.0/

Open Access

Abstract

In this paper, we propose a low-cost posture recognition scheme using a single webcam for the signaling hand with nature sways and possible occlusions. It goes for developing the untouchable low-complexity utility based on friendly hand-posture signaling. The scheme integrates the dominant temporal-difference detection, skin color detection and morphological filtering for efficient cooperation in constructing the hand profile molds. Those molds provide representative hand profiles for more stable posture recognition than accurate hand shapes with in effect trivial details. The resultant bounding box of tracking the signaling molds can be treated as a regular-type object-matched ROI to facilitate the stable extraction of robust HOG features. With such commonly applied features on hand, the prototype SVM is adequately capable of obtaining fast and stable hand postures recognition under natural hand movement and non-hand object occlusion. Experimental results demonstrate that our scheme can achieve hand-posture recognition with enough accuracy under background clutters that the targeted hand can be allowed with medium movement and palm-grasped object. Hence, the proposed method can be easily embedded in the mobile phone as application software.

Keywords

Bounding Box, Hand Profile Mold, Motion-Hand Posture Recognition

1. Introduction

Nowadays, the virtual reality will be expected under friendly contactless control where the recognition of hand gesture/posture is required [1] [2] for no need of wearable equipment. In recent, such methods are frequently Kinect-based [1]

[2]. In [1], the smart color glove of special design facilitates the achievement of robust method. In [2], for the heavy-noise problem in use of Kinect sensor, a distance metric named Finger-Earth Mover's Distance is given to effectively compare the dissimilarity between hand shapes extracted with salient distortions. The similar technology issue in similar works such as [1] and [2] cannot help but need the features extraction from the smoothed entire hand contour for complex hand-gesture recognitions. For pursuing the accuracy increment in applicative specialties, the methodology of object segmentation consecutively advances. Up to now, the deep learning network (DLN) [3] can attain well semantic object segmentation at the expense of laborious training task which inevitably causes the impropriety of DLN on specified semantic object segmentation. The early work [4] successfully realized a learning-based strategy using a nearest-neighbor deciding predictor and a refining vs. discarding verifier to segment the diverse objects deformed from complex backgrounds. In [5], the segmentation purely depends on the employing the depth information to facilitate the natural gestural interface development, but the applicability will be bounded on well-controlled settings. The work in [6] achieved a well geometric combination of processes including the searching of reliable center in the targeted palm, the mean shift algorithm and Distance Transform (DT) algorithm for moving hand segmentation. Recenth and segmentation approaches are focused on how to acquire robustness in real time or at least just-in-time [7]. In [7], the method tries to effectively suppress the skin color detection mistakes, remove the complex background, and handle the double-handed gestures in complex background with photometric change. The segmentation strategy of methods addressed in [7] [8] is separately taking care of statistic and dynamic hands. The work in [8] is to recognize the natural hand gesture. The initial segmentation is for the hand localization by only using Kinect depths. Then, in the resultant boxing window, the pixels of hand border are tracked by some preferred search directions and referring the previous pixel to obtain the closed contour. A backtrack process is added in the case that the unknown valid configuration is encountered. Observe existing hand-gesture recognizers, the extracted hand contours with the aid of Kinect have inevitable stair, saw-tooth and peak noises due to the intrinsic handicaps of sensed Kinect depth including low resolution, instability, high distortion, sensibility to radiant, specular and metal objects. And, the Kinect device is hardly applied to general (high-light) outdoor conditions in practical. Hence, in this work, we develop a webcam-based hand-gesture recognition scheme to timely figure out signals of hand with national movement at very low cost. It could directly benefit the popularization of applying human computer interfaces (HCI) [9]. The proposed scheme shown as **Figure 1** is composed of two phases: the tracking of best-fit bounding box and the HOG-based SVM. The novelty of best-fit bounding box purely including the bi-level hand mold is based on the efficient mutual adoption of low-level hand cues. This can greatly promote the effectiveness of common features. Hence, instead of heterogeneous features

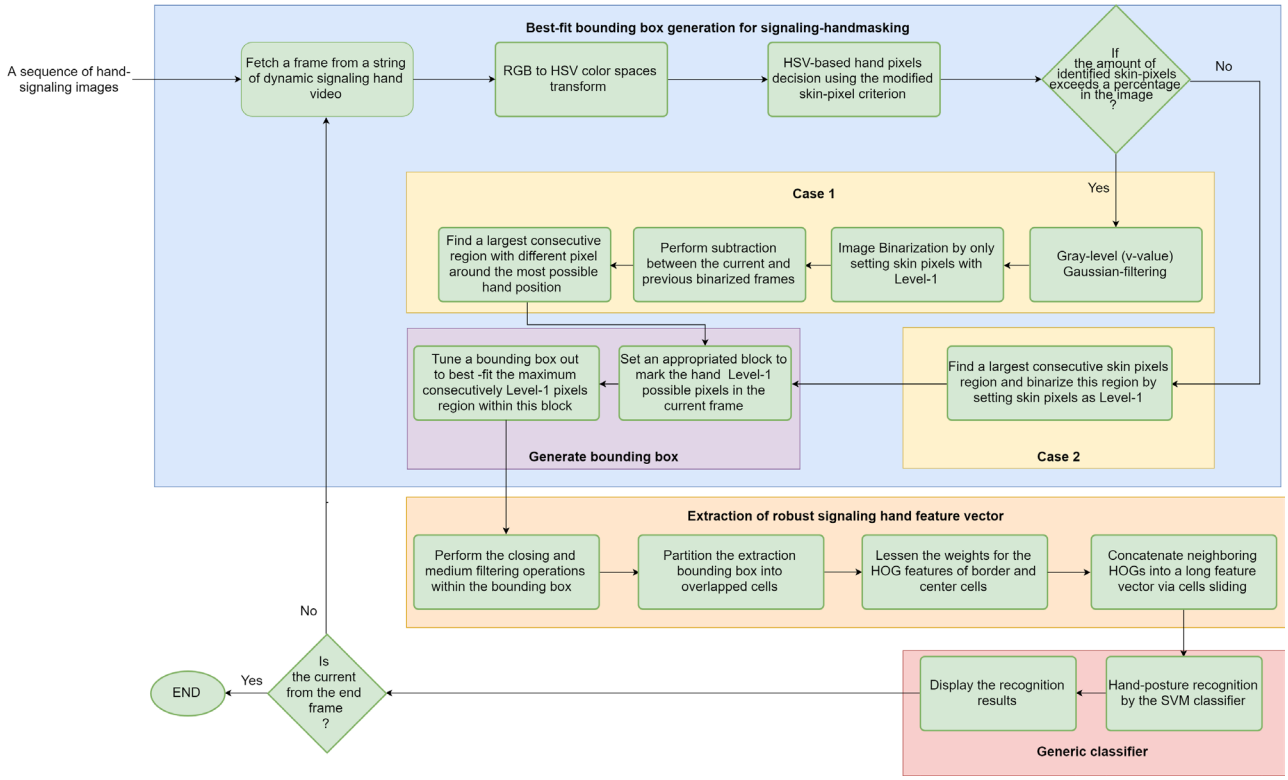


Figure 1. Processing flow diagram of the proposed hand-posture recognizer.

combined descriptor and time-consuming classifier such as AdaBoost classifier [10], the use of pure HOG feature and prototype SVM can offer enough recognition rates to motion hand postures. In short, the advantage of proposed scheme can support robust hand-posture recognition without constraints of monotonic background, occlusion-free palm, steady hand shape, fixed signaling location.

2. Best-Fit Bounding Box Generation for Signaling-Hand Masking

By only adopting a low-cost webcam, this phase is to instantly track the location of signaling hand and offer its bounding box. Since the skin-pixels decision criterion is majorly designed for human faces for most existing papers, we find the HSV-based criterion can be more easily modified to capture hand pixels relative to other colorific domains in respect of tuning of threshold parameters. So, the simplified criterion to detect whether the pixel with vector (h, s, v) in HSV color space has the skin color or not is given by

$$\left(\left(0 \le h \le \Delta_C\right) \vee \left(T_C \le h \le T_C + \Delta_C\right)\right) \& \left(T_I \le s \le 3T_I\right) \& \left(2T_I \le v\right) \tag{1}$$

For dedicating the criterion in (1) to capture hand pixels as accurate as possible, its parameters can be ranged as $\Delta_C = 10 \sim 15$, $T_C = 165 \sim 170$ and $T_I = 50 \sim 52$ according to the operating person and environment through our experiments. In general, those parameters can be tried from the empirical means with $\Delta_C = 12.5$, $T_C = 167.5$ and $T_I = 51$ or directly selected within afore-

mentioned ranges according to the operating scene cataloged in advance. In initial, the detection is applied to the current global image. If the skin-pixel criterion given by Equation (1) holds, the pixel is considered as a skin pixel of hand.

In this processing stage, Equation (1) is primarily applied to extract the all of possible skin-pixels including targeted hand, other body portions (such as the human face and upper limbs), and inevitable background fake skin points. However, tuning parameters in (1) to keep few fake skin points is not difficult. Thus, a global threshold ratio is applied to divide the images into two types under the existence of skin-pixel that the first case is the amount of detected skin pixels beyond this ratio and another is below it. This implies that the first type of image has too many other body portions, and another can almost only contain the signaling hand pixels, the desired dominant ones. Then, after the detection, the fetched image is binarized by denoting skin-pixels with grey level 1 and others with grey-level 0. Within the attention block, the hand depth pixels are simply quantized to 1's and the others to 0's. Particularly, the rate of level-1 pixels denoted by for the this case can be used in the status-decision criterion given by

$$R_t^{1's} = \frac{N_{1's}\left(M_{t+1}^b(x,y) - M_t^b(x,y) \middle| (x,y) \in M_{t+1}^b \cap M_t^b\right)}{N_{1's}\left(M_t^b(x,y)\right)} < T_S \tag{2}$$

where M_t^b expresses the binarized attention block captured at time t that its point at (x, y) is $M_t^b(x,y)$, and T_S where M_t^b expresses the binarized attention block captured at time t that its point at (x, y) is $M_t^b(x,y)$, and T_S is the status-checking threshold. If the criterion in (2) holds, the signaling hand is acknowledged in a low-motion or steady state; otherwise, in a high-activity state.

Basically, the first type of image (*i.e.*, the first case) can be regarded as the signaling hand in cluttered backgrounds. Rather, the use of temporal relation is effective because the normal user will more intentionally vary the ROI hand from an initial status than the other body portions for the aim of signaling-by-hand. The logical exclusive-OR of two consecutive binarized frames can easily capture the maximum area of spatial consecutive 1's (temporal different pixels) set as the region of dynamic ROI hand by an attention block. On the contrary, the second case, a simpler one, implies that the image could have already majorly be occupied by the signaling hand, so the maximum area of consecutive 1's within the current framebinarized is masked with an appropriated attention block. Then, a best fit bounding box is generated by interlacing the shrinking and shifting on the attention block as a minimum rectangle to mask the signaling hand.

3. Extraction of Robust Signaling Hand Feature Vector

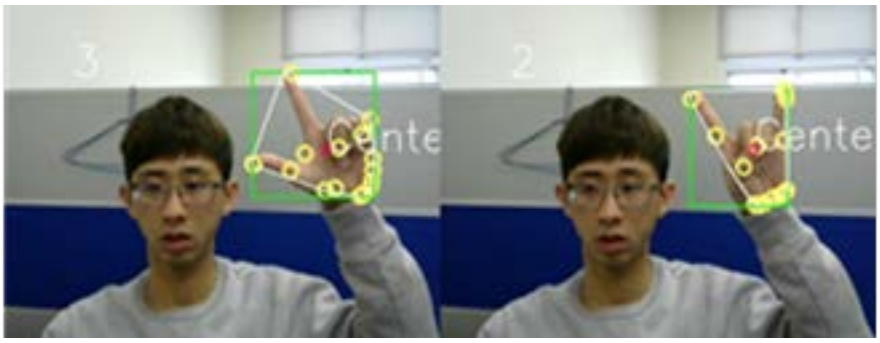
The procedure phase is primarily to reconstruct a prototype shape of moving hand as the hand posture mold rather than recover the actual one because the strength of recognizing the hand postures against random noise, occlusion,

hand-shape variation, and hand swinging is really wanted. For this goal, the closing and medium filtering are first performed on the bi-level pixels within the best fit bounding box in order. They can efficiently smooth out over-segmentations and distortions to leave a stabilized shape of segmented motion hand as the hand-profile prototype. Instead of recovering the actual of signing hand shape, the system aims to construct the prototype mold of motion hand profile.

For simple feature extraction, each best fit bounding box in the consecutive images is divided into a constant number of overlapped cells. The histogram of oriented gradient (HOG) vector of cell is figured out through dropping out too small gradient magnitudes in 8-direction histogram bins. The above two stages are dedicated to suppress the noises. Secondly, the HOG of cells are concatenated in the raster-scan order with different weights that the HOG vector of cells which are located near the border corner and the center of bounding box are weighted down. Thus, the linked long HOG vector acting as the signaling hand feature vector can be more robust against the occlusion from the other object to the signaling hand. With the feature vector on hand, the generic SVM can fast supply adequate and robust hand posture recognitions for the moving hand in this study.

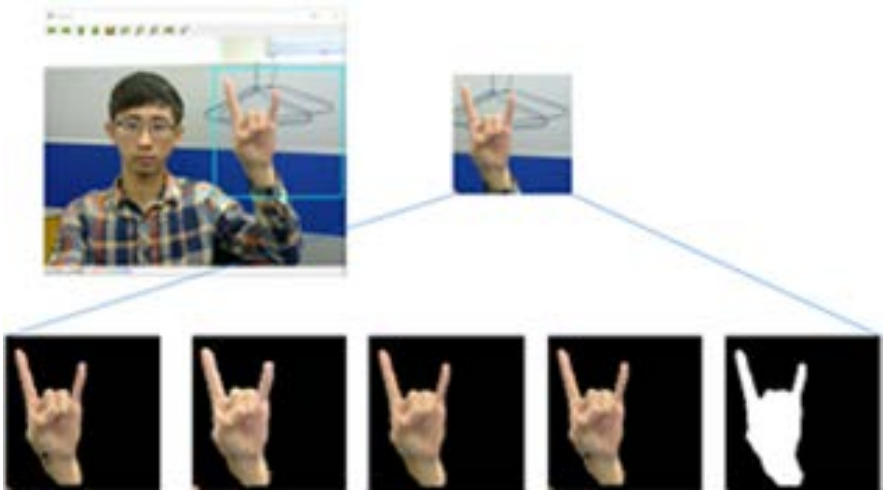
4. Simulation Results

The proposed method is implemented by student-level programming with the aid of Open CV functions in Python, Windows 10 and Intel Core i7-6700 CPU. In our experiments, 400 video sequences are collected for ten hand postures. There are 40 sequences for each hand posture that a half of them are randomly selected as training set, and the remaining ones are treated as the testing set for the recognition examination. As **Figure 2** shown, the experiment related to the first type of video demonstrates that the first phase of proposed algorithm can obtain the well best-fit bounding boxes marked by the green-rim rectangles. **Figure 3** exhibits the inter-medium outcomes of step-by-step generating the profile mold of binarized signaling hand from left to right images for the second-type video in the second row. In **Figure 3**, the final resulted image exhibits a smooth-profile mold for easily offering correct recognition. In **Figure 4**, we give an occlusion-simulated example, where a pen held in the hand causes occlusion through the palm of moving hand displaying a signified symbol "4". As **Figure 4(a)** shown, tracking and segmenting the moving hand are somewhat interfered with skin-like background part, *i.e.*, the yellow door. Rather, the proposed algorithm can still identify the production of an acceptable profile-mold, as **Figure 4(b)** shown. The performance of proposed scheme only using a web-cam can compete with the similar method in [11], which utilizes Leap Motion and Kinect, both, to obtain about 91.3% accuracy while recognizing the moving hand signaling under background clutter, hand pixel detecting-miss, hand occlusion. It is worthwhile to note that the proposed method can have effectiveness



(a) (b)

Figure 2. Example of the first type of images with moving dynamic signaling hand (a Case-1 video string) that the resulted bounding box is marked by the green-rim rectangle.



(a) (b) (c) (d) (e)

Figure 3. Profile moldgenerationofan image sampled from the second-type signaling hand video (a Case-2 video string), (a) skin-pixel detected image with background substraction, (b) dilation result of image in (a), (c) erosion result of image in (b), (d) Gaussion smoothingresultof image in (c), (e) binariztion result (*i.e.*, profile mold) of image in (d).

(a) (b)

Figure 4. An occlusion-simulated example, (a) a sampled color image from a video string exhibiting that a pen is held inside and lies through the palm of moving hand while signaling a signified symbol "4", (b) generated profile-mold from image in (a) which can successfully lead hand-posture to indicate "4" through SVM.

and concise structure, both. Hence, not only the hand-posture recognition by the proposed method can be obtained real time or just-in-time, but also the training period can be short. The training period for the ten hand postures using 200 sequences only requires about 2.5 seconds. Consequentially, the proposed scheme can be realized as an easily-installed friendly system for robust hand posture recognition at low cost.

5. Conclusion

In this paper, an efficient low-cost yet robust recognizer of moving signaling hand is proposed in only use of a cheap webcam that the implementation can be regular. The extracted hand mold can pretty benefit discrimination and robustness of extracted shape features applied to hand posture/gesture recognitions, providing a non-trivial hand shape. Hence, the proposed method can obtain robust recognition with only generic HOG features and SVM, with no need of gloves, wearable devices and Kinect. It also allows the skin-color background pixels around the signaling hand and the object occlusion on it. Experimental results verify that our method can achieve moving-hand postures recognition with low-cost, low complexity and high accuracy for common or unsophisticated cases. Hence, the proposed scheme could be a very low cost contactless non-wearing hand-control device. Basically, with such a concise structure, the proposed algorithm can be easily realized as a real-time processor.

Acknowledgements

This investigation was supported by Ministry of Science and Technology 107-2221-E-415-016-MY2, Taiwan.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Yuan, Y. and Fu, Y. (2014) Contour Model-Based Hand-Gesture Recognition Using the Kinect Sensor. *IEEE Transactions on Circuits and Systems for Video Technology*, **24**, 1935-1944. https://doi.org/10.1109/TCSVT.2014.2302538

[2] Ren, Z., Yuan, J., Meng, J. and Zhang, Z. (2013) Robust Part-Based Hand Gesture Recognition Using Kinect Sensor. *IEEE Transactions on Multimedia*, **15**, 1110-1120. https://doi.org/10.1109/TMM.2013.2246148

[3] Badrinarayanan, V., Kendal, A. and Cipolla, R. (2017) SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence,* **39**, 2481-2495. https://doi.org/10.1109/TPAMI.2016.2644615

[4] Cui, Y. and Weng, J. (1999) A Learning-Based Prediction-And-Verification Segmentation Scheme for Hand Sign Image Sequence. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **21**, 798-804. https://doi.org/10.1109/34.784311

[5] Tara, R.Y., Santosa, P.I. and Adji, T.B. (2012) Hand Segmentation from Depth Image Using Anthropometric Approach in Natural Interface Development. *International Journal of Scientific & Engineering Research,* **3**, 1-4.

[6] Wang, B. and Xu, J. (2012) Accurate and Fast Hand-Forearm Segmentation Algorithm Based on Silhouette. 2012 *IEEE 2nd International Conference on Cloud Computing and Intelligence Systems*, **2**, 976-979.

[7] Choudhury, A., Talukdar, A.K. and Sarma, K.K. (2014) A Novel Hand Segmentation Method for Multiple-Hand Gesture Recognition System under Complex Background. 2014 *International Conference on Signal Processing and Integrated Networks*, 136-140. https://doi.org/10.1109/SPIN.2014.6776936

[8] Plouffe, G. and Cretu, A.-M. (2016) Static and Dynamic Hand Gesture Recognition in Depth Data Using Dynamic Time Warping. *IEEE Transactions on Instrumentation and Measurement,* **65**, 305-316. https://doi.org/10.1109/TIM.2015.2498560

[9] Hamza, A., Anand, R., Shivhare, P. and Gaurav, A. (2017) Hand Gesture Recognition Applications. *International Journal of Interdisciplinary Research*, **13**, 2073-2075.

[10] Misral, S. and Laskar, R.H. (2017) Multi-Factor Analysis of Texture and Color-Texture Features for Robust Hand Detection in Non-Ideal Conditions. *IEEE Region Ten Conference* (*TENCON*), 1165-1170.

[11] Marin, G., Dominio, F. and Zanuttigh, P. (2014) Hand Gesture Recognition with Leap Motion and Kinect Devices. *IEEE International Conference on Image Processing*, 1565-1569. https://doi.org/10.1109/ICIP.2014.7025313