

Real Time Gesture Recognition with Heuristic-Based Classification

Omar Lopez-Rincon and Oleg Starostenko^(✉)

CENTIA, Department of Computing, Electronics, and Mechatronics,
Universidad de Las Americas Puebla, 72820 Cholula, Mexico
{omar.lopezrn, oleg.starostenko}@udlap.mx

Abstract. The recognition of human gestures in real time is still open problem due to low success rate of systems recently reported in scientific literature. This paper presents the proposed method and designed prototype for motion analysis and classification for human-computer interaction. The method is based on pattern recognition techniques of artificial vision without applying any markers or special sensors as well as utilizing low resolution cameras and simple hardware specifications. The proposed method provides interaction of user with computer via gestures in habitual and normal manner in order to activate system events (up, left and right) in real time. The proposed heuristic classifier recognizes specified gestures with an appropriate system context precision of 91.25 %. Comparing the obtained results with recent reports, the proposed approach provides satisfactory gesture recognition in real time with low resolution cameras.

Keywords: Pattern recognition · Gesture recognition · Artificial vision · Heuristic classifier

1 Introduction

Nowadays, the most common interfaces between human and computer systems still are the keyboard and the mouse, but the tendency in a short term has been focused on devices with touchscreen and gesture recognition with movements executed by the user [1]. The gestures are generated from body movements such as arms, hands, fingers, head, face or body [2].

Karam [3] reported that hands are the most used to execute gestures compared with any other body parts, as it is a natural part of human's communication, either for sentiments as for intentions. Therefore they are the most adequate for the natural interaction with computers too. The research concerning pattern recognition is directed to systems that can identify human gestures as entries and process them to control devices, mapping those gestures as commands. The main technologies at the present moment are based on artificial vision and contact [4]. We will focus in the artificial vision (AV) recognition. The capacity to detect gestures using AV and pattern recognition, allows to explore a variety of interaction techniques to control different environments, for example: changing the music volume or manipulating the thermostat without approaching to it [5]. In devices based on touchscreens for gesture recognition, it is necessary to detect the

beginning of the movement, called *gesture localization* [6], which is recognized at the moment of making contact with the surface or the sensible part of the device.

By keeping record of the executed movement over the surface or tactile sensor, the registered sequences are verified to evaluate if they match with the established classifications. If a system matches with any of the gestures to which the system responds, it is considered as an action by the user and then the system triggers an event in response. This kind of feedback doesn't exist in touchless systems. Several research groups are "on the run" in developing the standard scope for pattern recognition. There are several alternatives for pattern recognition, from complex devices, such as complete body suits, to non-invasive devices such as infrared depth cameras as Kinect early known as PrimeSense of Israeli 3D sensing company [2]. There are also complex methods such as those which by using wireless network signal they detect the corporal movements and recognize the human gestures [5].

Through movement analysis using a web cam and user interface for a simple computing tasks system, new technologies are within reach of everyone. Using AI would be a practical way of solving gesture recognition [7]. Thus, the main purpose of this paper is gesture interpretation, using a static camera as well as to present a novel and fast method to classify gestures.

The rest of paper is organized as follows. Section 2 presents an analysis of well-known relevant methods of real time gestures recognition systems based in artificial vision. Section 3 describes the proposed method and Sect. 4 presents the heuristic and classification approaches. Section 5 shows experimental results and performance evaluation of the proposed method. Finally, Sect. 6 presents conclusions and future work.

2 Related Works for Real Time Gesture Recognition

According to Mitra [2], gesture recognition is the process where the user acts a gesture out and the receptor recognizes it as an input. According to this, we could interact with machines, sending them messages as a signal relating them with the environment and the system's syntaxes. In order to achieve this, the image processing and furthermore, feature extraction are required. Most of vision based systems, comprehend three stages: detection, following, and classification or recognition [8]. At the first stage the challenges are the hands recognition and segmentation of the desire region within the image. This process is imperative to eliminate irrelevant information from the background and then follow the movement as a sequence. Several characteristics had been taken into consideration in different methods to achieve this like color, shape, movement or templates [8]. Due to space-time variations, desired segmentation of the hand and correct movement tracking are still a higher challenge. Errors at this early stage of the process cause deviation of the real trajectory during movement tracking [9].

Coming up next, some of the most used methods from the past 5 years for gesture recognition with AI are summarized according to Athavale [4] (see Table 1). All of them work in real time, searching and detecting skin color, which makes them sensible to image color and lighting as well as users must have their hands uncovered.

Table 1. Summary of recent relevant methods for human gesture recognition.

Source:	Bhuyan 2014 [9]	Kshirsagar 2013 [10]	Palochkin 2014 [11]	Rios-Soria 2013 [12]	Lee 2013 [13]
Advantage	robust	easy	easy	easy	Fast/robust
Disadvantage	short gestures	weak	static gestures	static gestures	expensive
Features	TV ^a	TV	C-H ^b	C-H	TV
Extraction	CRF ^c	FSM ^d	Viola-Jones	heuristic	SVM ^e
Face detection	Yes	No	No	No	Yes
Motion tracking	local features	states	skin detection Lucas-Kanade	skin detection	Kanade-Lucas-Tomasi
Training	Yes	NO	Yes	NO	Yes
Classification	classifier	templates	parameter analysis	parameter analysis	templates
Background	simple	complex	simple	simple	complex

^aTrajectory vector; ^bConvex hull; ^cConditional random fields; ^dFinite state machine; ^eSupport vector machine

The overall success rate of the recent and the most relevant systems for gesture recognitions lie in the range of 70–96 % [1, 9, 11, 13]. As mentioned before, the gesture analysis methods as usual use skin recognition, which in case of presence of gloves would disable most of them [14]. Some methods process only static gestures applying complex algorithms that frequently do not provide fast recognition required for real-time applications. In this paper the proposed approach has been developed for high speed short gestures recognition in real-time using image acquisition and processing tool without any high quality requirements.

3 The Proposed Method for Gesture Detection

Commonly, an object detection process includes a differences frame implementation, background elimination and a method for the movement tracking. To achieve these several processes to work at real time we start making two copies of each frame in reduced different sizes: one with 50×50 pixels and another one with 100×80 . The biggest image (the one with 130×95) is scanned for face detection using the Viola-Jones method [15]. When a face is detected the system grabs the frame t and the frame $t-1$ of the 50×50 frames and starts subtracting them in a loop until there is no face detection.

In this proposal we used a frame differences method due to its high speed to detect the motion in a video sequence. This is done with a dyadic pixel by pixel comparison and the difference calculation on both spatial axes x, y of the image [16]. Then a color reduction is done on each frame f on the pixel P at the t moment on the x, y position and its V value is obtained with the luminance calculation with its RGB values using the following equation (Eq. 1).

$$V = 0.21 * \text{red} + 0.72 * \text{green} + 0.07 * \text{blue} \quad (1)$$

Each of the frames $f(x, y, t - 1)$ and $f(x, y, t)$ are gray scale images and continuous from the real time input video sequence. Their difference is expressed as follows (Eq. 2) drawing desired pixels of detected objects in motion in black.

$$D_t(x, y) = -1|V(f(x, y, t - 1) - V(f(x, y, t))| \quad (2)$$

Then the difference obtained from the matrices D_t is used to create a new binary image B_t by taking each D_t value and evaluate it at different thresholds. If the D_t value

lies between $uMax$ and $uMin$, then corresponding pixel in B_t would be 0, otherwise the pixel is depicted as 1 (Eq. 3).

$$\begin{aligned} B_t(x, y) &= val; val = 0, uMin > val < uMax \\ B_t(x, y) &= val; val = 1, uMin < val > uMax \end{aligned} \quad (3)$$

At the same time all of the gray values are evaluated to create the binary image. For this two histograms are established H_x and H_y each of them with 50 values. The x histogram (H_x) will have all the values of the columns of the binary image and the y histogram (H_y) all the ones that correspond to each row (see Fig. 1).

The motion tracking is achieved by crossing the maximum values in each of the histograms and the final classification is done by tracking the intersection of these two values of the histograms and comparing them to the heuristics (see Fig. 2).

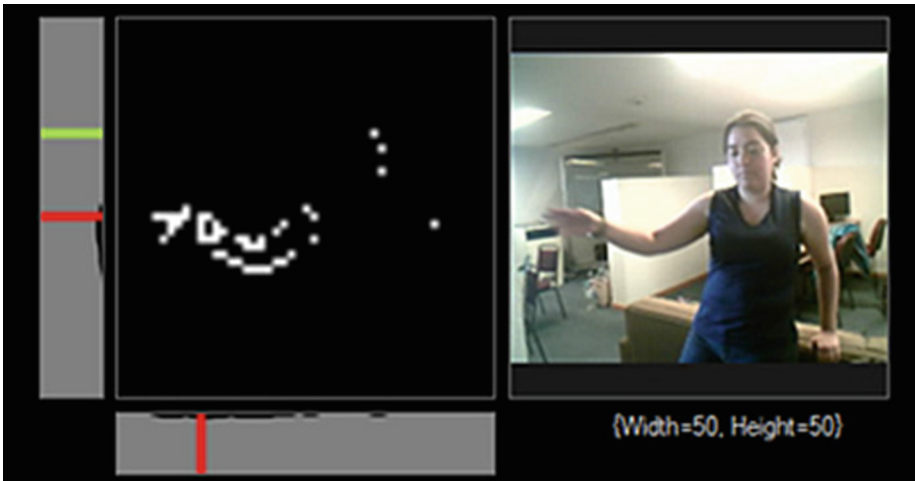


Fig. 1. Object in motion and its histogram. Despite of blurred object in motion due to low resolution camera, the approach satisfactory detects the region of interest.

The second part of the classification involves the face detection. It is used as a local reference parameter. The gesture detection and motion tracking start with a recognized face. At this moment we have new dynamic thresholds inside the image which are based on the position of the detected face.

We looked up for proportions of thresholds, where three gestures (*up*, *right*, *left*) could be recognized without collision reducing in this way false positives and true negatives. If both of the histograms are empty then it means there is no detected motion from the user. The block diagram of the procedure for fast gesture recognition and classification is shown in Fig. 3.

Applying fast feature extraction approaches with quite low computational complexity, the gesture detection is done effectively in real time as well as the used procedures are simple and easy to implement with low cost hardware.

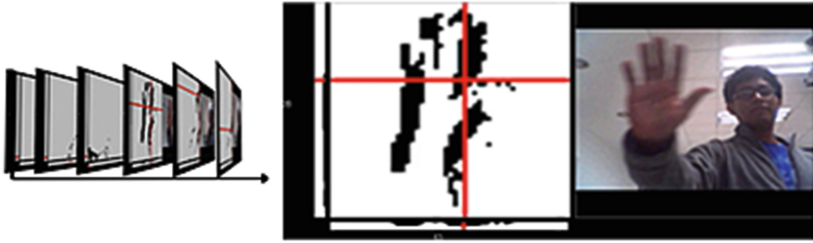


Fig. 2. The motion tracking by crossing the maximum values in each of the histograms detecting the maximum motion in the image sequence.

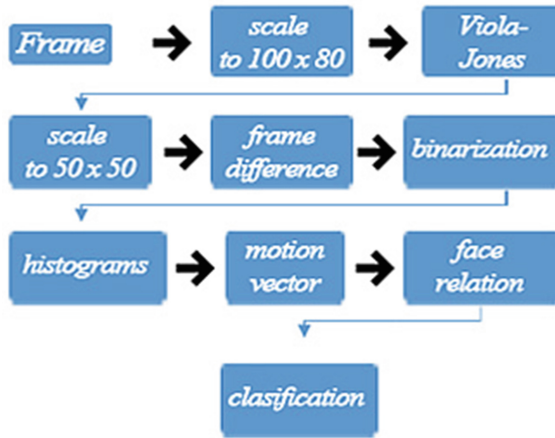


Fig. 3. Block diagram of the algorithm for gesture detection and recognition

4 The Proposed Algorithms for Recognition and Classification

In skeleton based classification, the gesture is determined by comparing the movement and the position of the wrist, elbow and shoulders from the detected body [17]. We start face recognition taking into account that it is important not only to determine if there is a user in motion, but also a face position must be detected. A face will be used as the reference of hand motion in relation to a human body without restricting its position.

This is used to determine three thresholds to create three different rules in order to find if the detected movement is *left*, *right* or *up* gesture. All the detected movements above the face are ignored. This allows the user to move in a natural way filtering several true negatives.

Three final thresholds related to a face have been proposed to use. The first one is measured as a center point of the height of the recognized face and it is used to detect the *up* gesture. If the face is detected at the starting point of the superior left side on the image $P(x, y)$ with a height H and a width W , then the center point would be calculated as it follows (see Eq. 4)

$$U_1 = P(x + \frac{W}{2}, y + \frac{H}{2}) \quad (4)$$

The second threshold is set to the right of the face at a distance localized in the face width multiplied by 2.5. This is the approximation of the hand position pointing towards right above its shoulder and elbow (see Eq. 5).

$$U_2 = P(x - W * 2.5, y + \frac{H}{2}) \quad (5)$$

The last threshold is at the end of the right side of the image, given by the same point $P(x, y)$ and the addition of the face width W (see Eq. 6).

$$U_3 = P(x + W, y) \quad (6)$$

On the Fig. 4 we can see the image with each of the mentioned thresholds. Additionally, there are two moments considered for gesture activation $M(t)$ and $M(t - 1)$. If a motion at the point $P(H_x(Max), H_y(Max))$ crosses one of the thresholds with two moments at the same time then the gesture is activated. With the histogram crossing the position of a movement is stored as a moment $M_{t-1}(P(H_x(Max), H_y(Max)))$ from the previous frame and one of the actual frame $M_t(P(H_x(Max), H_y(Max)))$. This generates a directional vector with the motion direction.

On Fig. 5 we can see the yellow cross of the center of face detection, done on a 50×50 image. The red lines indicate where the 3 thresholds are placed according to the position of a face. The white pixels are part of motion detection on the image and finally, with color circles we can see the motion vector direction. The red circle (closest to cross) represents where the motion came from and the green one (right most) shows the final position and defines the direction.

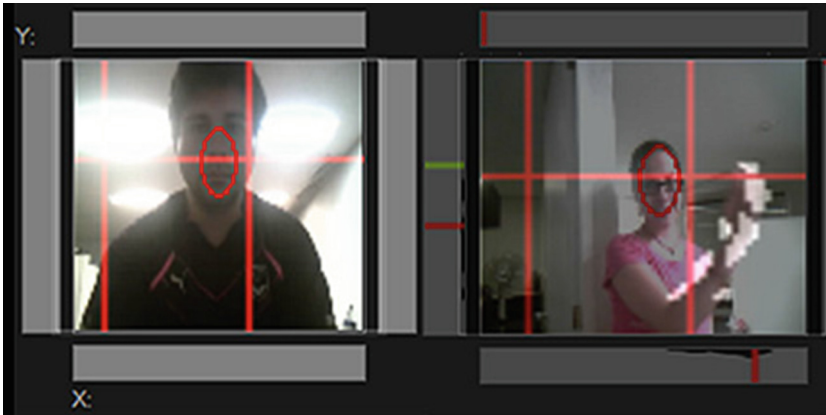


Fig. 4. Two examples of three thresholds related to a face used for gesture detection

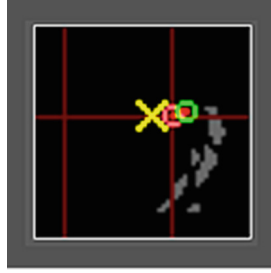


Fig. 5. Motion vector defined by red starting (left) circle and green final (right) circles. (Color figure online)

When the vector v completely crosses one of the thresholds, a gesture is detected and classified given by the direction of the vector $G(v)$ (see Eq. 7).

$$G(v) \begin{cases} UP, v > U_1 \\ LEFT, v < U_2 \\ RIGHT, v > U_3 \end{cases} \quad (7)$$

Additionally, the particular priority of gestures for classification is applied. The highest priority has the *up* gesture, then the outside threshold is analyzed specifying the *right gesture* and finally, the *left gesture* is evaluated analyzing presence of motion from center of the face to the left direction.

On the Fig. 6 we can see the histograms, the movement found by the histograms, the detected face (yellow cross) and the motion vector with the thresholds drawn in red accordingly with the face position.

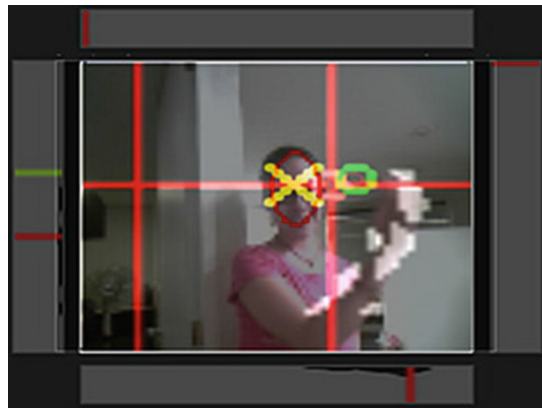


Fig. 6. Final frame composed by the histograms for motion detection, the detected face (yellow cross) and the motion vector with the thresholds drawn in red (Color figure online)

The method compared with other proposals is the simplest in implementation and the easiest one from computational complexity point of view. It is not necessary to recognize skin color and the searching ace position is used only as the reference parameter to establish the thresholds for detection of the motion of user hands.

It is important to mention that the simplicity of each part contributes to low requirements for used hardware and implementation of the whole procedure.

5 Experimental Results and Evaluation

In Table 2 the results of several experiments are presented. The tests have been carried out with volunteers using the designed system based on the proposed method for gesture recognition. Every user had a session of interaction with a system making different gestures in a random sequence. All of the volunteers were sitting down, so the individual height of each user wouldn't varies too much to become an issue of adjustment every time at each test. Table 2 resumes the total number of attempts of each user for every produced gesture (up, right, left) and precision of recognition of gesture of this particular user by system. The recognition rates of each gesture for all users are resumed in the columns.

Table 2. Results of gesture recognition by different users

Type of gesture	Attempts	<i>up</i> /%	<i>right</i> /%	<i>left</i> /%	Success rate
User 1	13	5	4	4	92.3 %
User 2	28	4	11	13	92.8 %
User 3	19	5	8	6	93.1 %
User 4	17	4	8	5	86.5 %
User 5	38	4	18	16	88.8 %
User 6	12	4	4	4	98.1 %
Total attempts:	127	26	53	48	
Total precision:		98.7 %	95.1 %	83.0 %	91.25 %

The average success recognition rate computed taking into account the number of attempts proportional to contribution of each user is about 91.25 %. That is a quite acceptable result for system that has no any particular requirements for high quality tool. For instance, the tests were made using webcam of low resolution (800 × 600) at the approximated distance of 1.5 m in ambient with soften indoor light.

The precision of recognition varies from 86.5 % (the case when user 4 has been illuminated with diffuse ambient light only) to 98.1 % (the case when user 6 has been illuminated with additional directional light from lamp directly to his face). Better illumination facilitates the face detection and increases precision of recognition.

Analyzing columns of Table 2, it is important to mention that the gestures *up* and *right* have been recognized with the highest precision (98.7 % and 95.1 %, respectively),

while the recognition of left gesture (83 %) had errors due to more complex background of hand in motion to left for right handed users.

In Table 3 the final comparison of success rate of the proposed and recently used methods discussed in Sect. 2 is presented. Unfortunately, the most of reports use their proper non-standard video sequences or databases for performance evaluation of proposed and designed systems for gesture recognition. Therefore, the presented in Table 3 recognition rate may be considered only as related possibly achieved efficiency of systems for gesture analysis in very specific controlled environment described into each particular report.

Table 3. Recognition rate of well-known and the proposed systems for gesture recognition

	[9]	[10]	[11]	[12]	[13]	Proposed
Success rate	96 %	65.7 %	80 %	93.17 %	80 %	91.25 %

One of the main disadvantages of the proposed method is its sensitivity to the lighting conditions. Since the method was thought to work in airport information modules, the lighting of that ambient should be sufficient for satisfactory operation of the gesture recognition system. Besides this, the proposed low cost and simple procedure for gesture detection and recognition could be used for design several interactive systems with ease navigation. The real time gesture recognition during natural human-computer interaction may be considered as significant advantage of the proposal comparing it with other well-known systems.

6 Conclusions

We found that most recent methods for gesture recognition using artificial vision as whole or as part of their method depend on the skin recognition and some of them use specialized hardware. We proposed a simple and fast method to detect, track and recognize short gestures with high precision. It works with simple heuristics classifying three basic gestures in real time during natural human interaction with computers.

Compared to the high percentage of correct recognition of the gestures *up* and *right* we can assume that the heuristics of the lateral gesture *left* might not be the optimal still yet. In the future research we need to improve the performance of the approach, make adjustments to find new heuristics for gesture detection and recognition as well as increment the discrimination ability of the method in case of a the complex background and low or variable illumination.

References

1. Bondre, M.H.S., Pimple, J.: Survey on touch less computer control system using hand gesture recognition. *Int. J. Recent Innov. Trends Comp. Commun.* **3**(2), 209–213 (2015)
2. Mitra, S., Acharya, T.: Gesture recognition: a survey. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* **37**(3), 311–324 (2007)

3. Karam, M.: Ph.D. thesis: A framework for research and design of gesture-based human-computer interactions, University of Southampton (2006)
4. Athavale, S., Deshmukh, M.: Dynamic hand gesture recognition for human computer interaction, a comparative study. *Int. J. Eng. Res. Gen. Sci.* **2**(2), 38–55 (2015)
5. Pu, Q., Gupta, S., Gollakota, S., Patel, S.: Whole-home gesture recognition using wireless signals. In: *MobiCom 2013 Conference*, Miami, FL, USA, pp. 1–12, 30 September 2013
6. Elmezain, M., Al-Hamadi, A., Michaelis, B.: A novel system for automatic hand gesture spotting and recognition in stereo color image sequences. *J. WSCG* **17**, 89–96 (2009)
7. Chaudhary, A., Raheja, J.L., Raheja, S.: A vision based geometrical method to find fingers positions in real time hand gesture recognition. *J. Softw.* **7**(4), 861–869 (2012)
8. Zabulis, X., Baltzakis, H., Argyros, A.: Vision-based hand gesture recognition for human-computer interaction. In: *Universal Access Handbook*, Chap. 34. Lawrence Erlbaum Associates, Inc. (LEA), pp. 1–30 (2009)
9. Bhuyan, M.K., Kumar, D.A., MacDorman, K.F., Iwahori, Y.: A novel set of features for continuous hand gesture recognition. *J. Multimodal User Interfaces* **8**(4), 333–343 (2014)
10. Kshirsagar, K.P., Sahu, R.M., Bankar, S.M., Moje, R.K., Doye, D.D.: K one hand gesture recognition. *Int. J. Innov. Res. Electr. Electron. Instrum. Control Eng.* **1**(7), 330–334 (2013)
11. Palochkin, V., Demidov, P.G., Alexander, M.J., Priorov, A.: Recognition of hand gestures on the video stream based on a statistical algorithm with pre-treatment. In: *Open Innovations Association FRUCT Conference*, St. Petersburg, Russia, pp. 105–111 (2014)
12. Rios-Soria, D.J., Schaeffer, S.E., Garza-Villarreal, S.E.: Hand-gesture recognition using computer-vision techniques (2013)
13. Lee, S.-H., Sohn, M.-K., Kim, D.-J., Kim, B., Kim, H.: Smart TV interaction system using face and hand gesture recognition. In: *IEEE International Conference on Consumer Electronics*, Mexico, pp. 173–174 (2013)
14. Brethes, L., Menezes, P., Lerasle, F., Hayet, J.: Face tracking and hand gesture recognition for human robot interaction. In: *Proceedings of International Conference on Robotics and Automation*, New Orleans, USA, vol. 2, pp. 1901–1906 (2004)
15. Viola, P., Jones, M.J.: Robust real-time face detection. *Int. J. Comp. Vis.* **57**(2), 137–154 (2004)
16. Jun-Qin, W.: Moving object detection using the edge-prefetch frame difference method. *Int. J. Adv. Comp. Technol.* **5**(5), 1139–1145 (2013)
17. Celebi, S., Aydin, A.S., Temiz, T.T., Arici, T.: Gesture recognition using skeleton data with weighted dynamic time warping. In: *VISAPP*, pp. 620–625 (2013)