

Vision Based Hand Gesture Recognition for Mobile Devices: A Review

Housseem Lahiani^{1,3,4}(✉), Monji Kherallah², and Mahmoud Neji^{3,4}

¹ National School of Electronics and Telecommunications,
University of Sfax, Sfax, Tunisia
lahianihousseem@gmail.com

² Faculty of Sciences, University of Sfax, Sfax, Tunisia
monji.kherallah@gmail.com

³ Faculty of Economics and Management, University of Sfax, Sfax, Tunisia
mahmoud.neji@gmail.com

⁴ Multimedia Information Systems and Advanced Computing Laboratory,
Sfax, Tunisia

Abstract. The desire to interact with a mobile device in an intuitive and natural way is growing. In fact, research in this field aims to develop systems able to model, analyze and recognize user's hand gestures to control mobiles without having the need to touch the screen. We give in this paper an overview of current research works and an analysis of comparative studies in this field. This paper focuses on the main steps of hand gesture recognition for mobile devices like detection, tracking and recognition. This work also gives an analysis of the existing literature on gesture recognition systems for human-computer interaction by classifying them under various key parameters. At the end we conclude with some reflections on future works.

Keywords: Hand gesture · Recognition · Mobile device · Human computer interaction · Android

1 Introduction

In these times, the interaction with a mobile device has become greater than ever and it has become a necessity and we couldn't live without it. This embedded technology has become so integrated in our daily life and we became obliged to use it in almost all areas of our lives. To effectively exploit those devices, we need more and more interaction with them. An interaction is a manipulation of graphic objects like icons and windows using a touchscreen or a pointing device. Even if the invention of virtual keyboard and the touchscreen represents a great progress, there are some situations in which these pointing devices are incompatible for Human-Computer Interaction. The use of hand gestures offers an interesting alternative to these cumbersome devices for human-computer interaction. Using the hand as a device can help to communicate with mobile devices in a more natural and intuitive way. Hand gestures are therefore a non-verbal means of communication. Hand Gestures can be static (pose or posture) which need less computational complexity or dynamic which could be more complex.

To recognize gestures, some gesture recognition methods, in addition to the camera, used additional hardware such as gloves and sensor to easily extract the full description of the gesture features. But there are other methods like appearance based methods that use the skin color to segment hand and then extract features. These methods are more easy, intuitive, natural and cost less compared to methods mentioned previously. The remainder of this paper is organized as follows:

Section 2 gives a literature review and explains key issues of hand gesture recognition system for mobile devices. Section 3 discusses application areas of hand gesture recognition systems for mobile devices. In Sect. 4 a summary of recent research results is shown and a possible future work is given, and finally we conclude in Sect. 5.

2 Literature Review and Key Issues of Hand Gesture Recognition for Mobile Devices

The implementation of an efficient hand gesture recognition system for mobile devices is aided through two main kinds of enabling technologies for human-machine interaction such as contact based and vision based devices. Contact based devices require physical interaction like accelerometers, data glove and multi-touch screen and they do not provide much of acceptability because their use is uncomfortable especially for an inexperienced user [1]. Thus vision based approaches have been employed for hand gesture recognition in human-computer interaction. Contact-based approach for hand gesture recognition is so embarrassing, hence vision-based approach is a comfortable experience but it is hard to deploy it in bad conditions. The main challenge of the recognition of hand gestures by a mobile device based on vision based approach is to cope with the wide variety of gestures in addition of the computational limitations of mobile devices. Note that the location of the hand in this context is a computer vision problem and hand gesture recognition is a machine learning problem which makes the fact to develop a real-time gesture recognition system by mobile device, which has a limited capacity in terms of CPU, memory and battery, challenging. Most researchers classified hand gesture recognition systems mainly in four steps after image acquisition from camera or data glove instrumented device. The captured hand gestures by the camera of the device constitute the input of the system. Systems have to work with approaches that are computationally inexpensive to compensate the weak processing capability of mobile device. These steps are mainly: hand detection, hand segmentation, feature extraction and gesture classification as illustrated in Fig. 1.

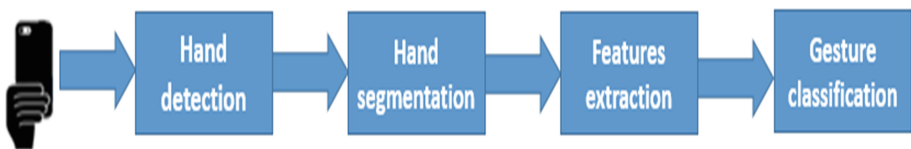


Fig. 1. Block diagram of gesture recognition system

Image acquisition is the first step in gesture recognition system. Frames are captured by the smartphone camera then the hand detection is processed. Then the segmentation of hand is done. After that hand tracking and segmentation are done, features must be extracted from the segmented hand. And finally, the classification step is processed. In this step input features are compared with features of the trained database.

2.1 Hand Detection and Image Pre-processing

For any vision-based system, Image Acquisition is the step only after which we can go forward with image processing then the detection step can be processed and we can go forward to segmentation of the hand to subtract the region of interest from the background. This segmentation is crucial because it isolates the relevant area from the rest of background image before transmitting it to tracking and recognition steps. Also the moving object could be extracted from the background using a threshold [4]. Several methods have been proposed in the literature using many types of visual features and, in several cases, their combination. Among the most used cues in hand gesture recognition systems for mobile device we find the color and the shape of the hand. The cue commonly used in gesture recognition for mobile devices, to segment the hand, is the color of the skin, because it is easy and invariant to translation and rotation changes [2] and because it is very suitable for embedded systems like mobile devices since color is computationally inexpensive, and it can give more information than a luminance-only image or an edge-segmented image which need more computational resources which make real-time systems hard to realize [3]. However some factors can obstacle segmentation process like illumination changes, complex background and low video quality. Therefore, to enhance the segmented image, there are preprocessing operations which can be applied such as subtraction and normalization. Indeed, the shape of hand can be utilized to detect it in images in several ways. Several information could be obtained by the contour extraction (Edge detection) of the desired object (hand). If it is detected correctly, the hand's shape can be presented by this contour and did not dependent on skin color and illumination.

H. Lahiani et al. in [5, 6] makes a real time hand gesture recognition system for android devices based on skin color segmentation. L^*A^*B color space was used because it makes separation between chrominance components and illumination.

In [7] V. ShirshirReddy et al. developed a system based on finger detection to control tablets. The finger is detected by putting an orange color round mark of known size on the front side of the finger of user. The round shape is chosen since it is simple to detect and the choice of the orange color is done because it is easy to differentiate it from surroundings in most of cases.

In [8] Prasuhn make a static hand gesture recognition system for the American Sign Language. To extract the hand area he applies simple binarization in HUV color space. Then binarized image has been submitted to morphological operations to reduce noise. Specifically, he applies closing first and then opening operation [9].

To recognize the shape of the hand, A. Saxena et al. [10] and J.L. Raheja et al. [11] used Sobel Edge Detector for image pre-processing for a sign recognition system using a mobile device.

To make a real time Indian Sign Language Interpreter for mobile devices S. Swamy et al. [12] used the Viola-Jones algorithm based on LBP features to recognize the posture of the hand. For this system, the pure background color is required during the image capture progress. To tackle the problem of lighting condition and background noise HSV model's threshold value was used to separate the background noise in the training images.

To make a Sign Language system based on android device for deaf people Setiawardhana et al. [13] used the Viola-Jones algorithm to detect the hand of the user and then to recognize finger alphabet. For this system, the detection of skin color, noise removal and thresholding are processed after capturing the image.

To make a Vision-based Gesture Recognition and 3D Gaming System for Android Devices controlled by hand gestures, Mahesh B. Mariappan et al. [14] used Cascaded Haar Classifiers to efficiently track object (hand) by the mobile device. After getting the image from the live video the image is converted to a grayscale image using the `cvtColor` function. Then the contrast of the image is improved because Haar classifiers use Haar features for classification purposes and those features are contrast based rectangular features.

In [16] to make a Static Hand Gesture Recognition using an Android Device, Tejashri J. Joshi et al. used thresholding to segment hand from background. They replace each pixel in the image by black pixel when the image intensity is less than threshold and a white pixel when the image intensity is greater than threshold. Then a rotation, cropping and normalization are done to the binary image.

2.2 Features Extraction

Good feature extraction is conditioned by a good segmentation and which have an important role in a good recognition process. The segmented hand image features vector can be extracted in several ways. Many methods were used to represent how features could be extracted. Some methods have used the shape of the hand, like the contour and the silhouette of the hand and some others use palm center, fingertips position, etc. For the features extraction step, in [5, 6, 19] H. Lahiani et al. used OpenCV functions to return the contour of the hand and convex hull points.

OpenCV have functions which return the contour and convex hull points of the hand from the segmented image. Using the bounding box with those elements they arrived to calculate the radius of inscribed circle of contour and the center of the palm. After that, location of Fingertips are computed using defect points and finally finger vectors are calculated and divided by the radius of the circle to obtain final feature vector.

In [8] to represent hand shape, L. Prasuhn et al. used the descriptor of Histogram of Oriented Gradients (HOG). HOG feature is robust under illumination change but its performance is vulnerable under object rotation. Although it is suitable for hand pose estimation. HOG can give a feature vector when an input image of fixed size is given. The proposed system used parameters of the original HOG paper [15].

In [12] S. Swamy et al. used LBP features. In this work LBP descriptors are experimented and they show a discriminative power in the classification of hand shape and their capability to capture enough information in order to distinguish among hand shapes in the classification phase. In [16] Tejashri J. Joshi et al. used Principal Component Analysis (PCA) in the features extraction step. At first a set “S” of “m” training images was created and each image was transferred into vector of size $1*5400$. Then a mean image was computed and the difference between each training image and the mean image was calculated to obtain the covariance matrix. Then Eigen vectors of covariance matrix was computed and finally the image was projected into space using feature of each image in the training dataset. Feature vector of all training images are computed and those feature vectors was utilized to train a classifier.

In [10] A. Saxena et al. used hand token as feature to make the image into a usable form for neuronal network. The cosine and sine angles of the shape where used to represent the criteria of a recognition pattern.

In [7] V. Shirshir Reddy et al. to track the finger which controls the tablet they extracted the area and centroid of the orange mark placed in the finger whose position is traced at any time with respect to an initial position named “O”.

Setiawardhana et al. [13] used the Viola-Jones Algorithm to detect the hand of the user and Mahesh B. Mariappan et al. [14] used Cascaded Haar Classifiers to efficiently track hand (closed fist) so as features they used Haar features. Features proposed by Viola and Jones consider detection windows delimiting adjacent rectangular areas; pixels intensities of these blocks are added, forming sums whose difference is a feature. A feature therefore is a real number which codes the pixel-wise variations of content at a given position in the detection window. The presence of edges or changes in texture are thus digitally translated by the values of Haar features.

Setiawardhana et al. [13] after detecting the hand and the preprocessing step, the hand shape image is obtained as input data and training data.

2.3 Gesture Classification and Recognition

The ultimate objective of hand gesture recognition is to interpret the semantics that the hand posture or gesture want to convey. After modeling and analyzing the image of the hand, a gesture classification method is needed to recognize gesture. The purpose of this process is usually to classify captured movements into different types of actions. Figure 2 explain the architecture of the classification system.

Two types of gesture recognition can be distinguished: the static or dynamic gesture recognition, based on one or more frames respectively. Most of the vision based systems done to recognize hand gestures by a mobile device are destined for static hand gesture recognition. Static hand gestures are classified into linear and non-linear learner. The first one is adapted to the linearly separable data and the second for the other cases. Another way to classify learning algorithms is to consider their results. Thus, it distinguishes supervised learning, unsupervised learning, semi-supervised learning, etc. The choice of the training algorithm depends mainly on the selected hand gesture representation. For example, S. Swamy et al. [12] proposed to recognize static hand gesture have choose LBP feature to apply The Adaptive Boosting Learning Algorithm

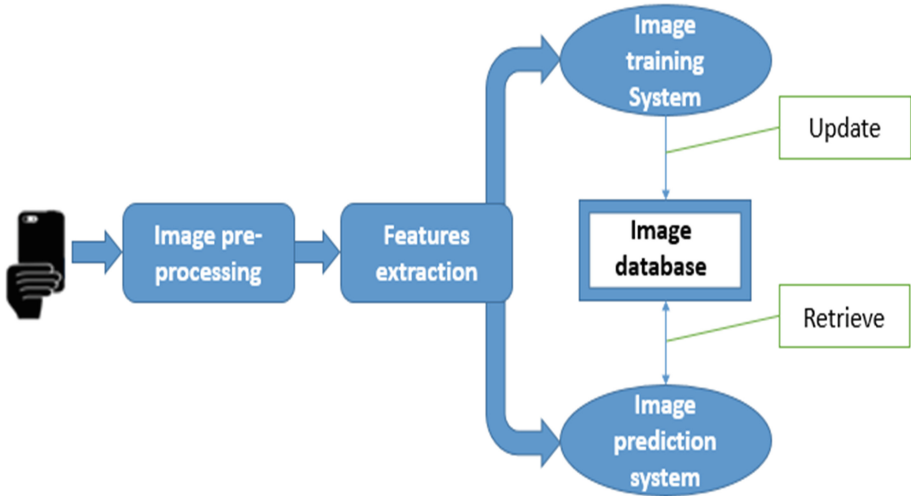


Fig. 2. Architecture of the classification system

“AdaBoost” which represents a learning algorithm that is capable of integrating the information of a category of objects. This algorithm is used by Viola-Jones algorithm to train samples’ set which includes cascade based classifier. AdaBoost combines weak classifiers that cannot give satisfactory results to become a strong classifier to obtain the better result. The Adaptive Boosting learning algorithm gets the best weak classifier from an images’ set that contains positive and negative images. After selecting the best weak classifier, Adaboost algorithm adjust the weights of the training images. At this stage, weights of classified training images decreased and unclassified images increased. After that, unclassified images are more focused by the Adaboost algorithm which tries to classify correctly the misclassified images. To recognize finger alphabets Setiawardhana et al. [13] after detecting the hand of the user by using the Viola-Jones Algorithm which use AdaBoost Algorithm cited above, they used K-nearest neighbor (k-NN) algorithm as a classifier to build the classification model which is a non-parametric lazy (no explicit training phase or it’s minimal) learning algorithm. Each data point in the set of data is considered in a known class. Then, the classification of a new data point can be predicted based on the known classifications of the observations in the database. Indeed, the database is known as the training set. The classification of a new observation is based on the classifications of the observations in the database which it’s ‘most similar’ to. Neighbors are usually calculated using Euclidean distance.

H. Lahiani et al. in [5, 6] proposed to recognize static hand gesture used Multi-class SVM (Support Vector Machine) to build the training classification model and to make predictions. SVMs are a set of supervised learning techniques aimed at tackling discrimination and regression problems. SVM are a generalization of linear classifiers. SVM can be used to solve discrimination problems, in other meaning it decides to which class a sample belongs, or regression problem, in other meaning it predicts the numerical value of a variable. The resolution of these two problems is through the construction of a function h to which an input vector x matches an output y .

$$y = h(x). \quad (1)$$

In [16] Tejashri J. Joshi et al. try to classify extracted features using minimum Euclidean Distance between the feature vectors of test image and training image. The correlation between continuous and quantitative variables are measured by Euclidean distance and it is not appropriate for ordinal data, where the preferences are listed by rank instead of depending on the actual and that reduces the accuracy. To fix this problem they tried to use K-means Clustering as simple classifier. All features of the same class are combined to form a cluster. The minimum distance between the feature vector of the test image and cluster centers and is computed and the class of the test image is predicted. However this classifier is not robust against outliers and so it reduces accuracy. All those problems were solved by using the SVM classifier to solve multi-class problem. To overcome the problem of unclassified region caused by indirect methods of the multiclass SVM which are one-against-all, one-against-one, Directed Acyclic Graph “DAG” SVM, a decision tree based multiclass SVM was used. In [10] A. Saxena et al. decided to use hand tokens cited in the section above to train a feed-forward backpropagation neural network. The network has only one input layer, hidden layer and output layer to simplify and accelerate the calculations. In [8] Prasuhn used a database storing a set of hand gesture images and to overcome the problem of labeling and classifying correctly all degrees of freedom of human’s hand image, images of database were synthesized by using an open source software named LibHand [17] which is a library for human hand articulation. Recognition method is based on finding the best-match image in the database. Each image stored in the database is pretreated and has a HOG feature. Here, the simplest brute force matching was used because the number of images in the database is low. HOG features of the test image is compared with each database image’s HOG feature. The “L2” distance of two HOG features is used as “matching metric”. Image in the database with the minimum “L2” distance is chosen as the best-matched image. In [14] Mahesh B. Mariappan et al. used Cascaded Haar Classifiers to efficiently track object (fist) by the mobile device. To collect training images, they shot a short video of a closed fist in different lighting conditions and from different angles for about two minutes at twenty five frames per second. Then a video editing program was used to extract images from video. After obtaining positive image samples and negative image samples in the training database a border around region of interest in each positive image was marked by using an object marking program. This generate an output file that contains all coordinates of positive images (the name of the image file, the width and the height of the object of interest, the coordinates (x,y) of the top-left corner of the object of interest) and the same thing was done for negative images. Finally the HaarTraining program was invoked after packing all positive images into a vector file using createsamples program which output a.vec file. Using this vector file and the file containing all data about negative images HaarTraining program generate a directory full of training data. Finally Convert_Cascade program was used to converts the training data into an XML file which represents the final output of the Haar training process and which can be used for recognition process.

3 Application Domains of Hand Gesture Recognition for Mobile Devices

This section gives a brief overview of some of the advanced application areas of vision based hand gesture recognition systems for mobile devices. Hand gesture recognition systems for mobile device are applied in different applications on several areas, it includes interpreting sign language, device control, numbers recognition, gaming, etc. An overview of some areas of hand gesture recognition application for mobile devices is given below.

3.1 Numbers Recognition

Counting numbers and digit using gestures in a mobile device can be used to place orders to access some apps or to control something. H. Lahiani et al. [5, 6] proposed a system that uses the numbering to count fingers and that could be used in different manner as controlling the smartphone or a connected device. After interpreting the hand pose sign the orders could be given to the device to perform a special task. A. Saxena et al. [10] and J.L. Raheja et al. [11] developed a system that recognize digit from one to five. In [16] Tejashri J. Joshi et al. developed also a system that recognize numbers from 1 to 5.

3.2 Sign Language Recognition

Because sign language is used for interpreting hand signs made by people with special needs it has received special attention. Many systems have been proposed to recognize hand poses using different types of sign languages. For example, Setiawardhana et al. [13] system recognized Indonesian Sign Language. In [12] S. Swamy et al. developed a system that recognize Indian Sign Language. The proposed system by [8] Prasuhn et al. is applied to recognize American Sign Language.

3.3 Gaming and Augmented Reality

Another recent application of hand gesture is control games with the 3D modeling. In [14] Mahesh B. Mariappan et al. developed a system named “PicoLife” which constitutes an augmented reality game in which 3D characters are controlled by hand gestures in an Android smartphones.

3.4 Device Control

In [7] V. Shirshir Reddy et al. developed a system based on finger detection to control tablets. It is a virtual touchscreen controlled through hand movement and it provides a suitable efficient and user friendly interface between human and tablet.

In [18] T. Marasovic et al. developed an accelerometer based (but non-vision based) gesture recognition system to control mobile devices. The system was designed to run in real-time. The application is designed for Android operating system which uses the data from a single triaxial accelerometer to recognize nine different hand gestures.

4 Summary and Prospects

The table below shows summary of some hand gesture recognition systems for mobile devices. In Table 1 a comparison between different systems is made and in which we give a summary of different extraction methods and classifiers used by system.

Table 1. Comparison between hand gesture recognition methods for mobile device

Researchers	Image processing	Extracted features	Classifier	Recognition rate
[5] H. Lahiani et al. (2015)	skin color segmentation	Finger vectors	SVM	93%
[8] L. Prasuhn et al. (2014)	binarization in HUV color space	HOG feature	Brute force matching	47%
[10] A. Saxena et al. (2014)	Sobel Edge Detector	Hand tokens	feed-forward back propagation neural network	77%
[13] Setiawardhana et al. (2015)	skin color segmentation	Haar features for detection And hand shape for recognition	Viola-Jones for detection and K-nearest neighbor (k-NN) for recognition	*100% if distance < 50 cm *25% if distance = 75 cm
[12] S. Swamy et al. (2014)	HSV model's threshold value	LBP features	AdaBoost (Viola-Jones)	62%
[16] Tejashri J. Joshi et al. (2015)	Thresholding, reorientation, cropping, normalization	Principal Component Analysis (PCA)	*Euclidian distance *K-mean cluster *SVM	*72.4% for Euclidian distance *45% for K-mean cluster * 97.6%. for SVM

Table 1 shows that vision based hand gesture recognition for mobile devices is a growing with promising results and especially with the rise of the use of smartphones and tablets. A possible future work is possible to improve existing works such as hybridization between artificial vision and the use of sensors and accelerometer which already exist in the majority of today's mobile devices. This may make systems more robust.

5 Conclusion

In this work various methods are discussed for hand gesture recognition for mobile devices, these methods include Color based segmentation, Edge detection, etc. for image processing, and PCA, LBP feature, Haar feature, etc. for features extraction and Neural Network, SVM, Viola-Jones Algorithm, etc. for classification. Also application domains for those systems are presented. A literature review and comparison of recent recognition systems are given as well. Summary of some hand gesture recognition systems are listed and a possible future work is presented as well.

References

1. Rautaray, S.S., Agrawal, A.: Vision based hand gesture recognition for human computer interaction: a survey. *Artif. Intell. Rev.* **43**, 1–54 (2012)
2. Stergiopoulou, E., Papamarkos, N.: Hand gesture recognition using a neural network shape fitting technique. *Elsevier Eng. Appl. Artif. Intell.* **22**(8), 1141–1158 (2009)
3. Wu, Y., Liu, Q.: An adaptive self-organizing color segmentation algorithm with application to robust real-time human hand localization. In: *Proceedings of Asian Conference on Computer Vision, Taiwan* (2000)
4. Yadav, D.K., Sharma, L., Bharti, S.K.: Moving object detection in real-time visual surveillance using background subtraction technique. In: *14th International Conference on Hybrid Intelligent Systems (HIS)*, pp 79–84 (2014)
5. Lahiani, H., Elleuch, M., Kherallah, M.: Real time hand gesture recognition system for android devices. In: *15th International Conference on Intelligent Systems Design and Applications (ISDA)*, pp. 592–597 (2015)
6. Lahiani, H., Elleuch, M., Kherallah, M.: Real time static hand gesture recognition system for mobile devices. *J. Inf. Assur. Secur.* **11**, 67–76 (2016). ISSN 1554-1010
7. Reddy, V.S., Raghuvver, V., Krishna, J.V., Chandralohit, K.: Finger gesture based tablet interface. In: *IEEE International Conference on Computational Intelligence and Computing Research (ICCIC)*, pp. 1–4 (2012)
8. Prasuhn, L., Oyamada, Y., Mochizuki, Y., Ishikawa, H.: A HOG-based hand gesture recognition system on a mobile device. In: *IEEE International Conference on Image Processing (ICIP)*, pp. 3973–3977 (2014)
9. Szeliski, R.: *Imageprocessing*, 1st edn. Springer, New York (2010). chapter 3, pp. 112–113
10. Saxena, A., Jain, D.K.: Hand gesture recognition using an android device. In: *Fourth International Conference on Communication Systems and Network Technologies*, pp. 819–822 (2014)
11. Raheja, J.L., Singhal, A.: *Android based portable hand sign recognition system*. Cornell University Library (2015)
12. Swamy, S., Chethan, M.P., Karnataka, M.: Indian sign language interpreter with android implementation. *Int. J. Comput. Appl.* **97**(13), 36–41 (2014)
13. Setiawardhana, S., Hakkun, R.Y., Baharuddin, A.: Sign language learning based on android for deaf and speech impaired people. In: *2015 International Electronics Symposium (IES)*, pp. 114–117 (2015)
14. Mariappan, M.B., Guo, X., Prabhakaran, B.: PicoLife: a computer vision-based gesture recognition and 3D gaming system for android mobile devices. In: *2011 IEEE International Symposium on Multimedia (ISM)*, pp. 19–26 (2011)

15. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005, vol. 1, pp. 886–893 (2005)
16. Joshi, T.J., Kumar, S., Tarapore, N.Z., Mohile, V.: Static hand gesture recognition using an android device. *Int. J. Comput. Appl.* **120**(21), 48–53 (2015). (0975–8887)
17. Saric, M.: Libhand: a library for hand articulation. Version 0.9 (2011)
18. Marasovic, T., Pasic, V.: User-dependent gesture recognition on android handheld devices. In: 22nd International Conference on Software, Telecommunications and Computer Networks (SoftCOM) (2014)
19. Lahiani, H., Kherallah, M., Neji, M.: Hand pose estimation system based on Viola-Jones algorithm for android devices. In: 13th ACS/IEEE International Conference on Computer Systems and Applications, (AICCSA) (2016)