A New Hand Gesture Recognition Approach for Robotic Assistants Based on Mobile Devices

J. S. Ochoa-Zambrano, V. E. Robles-Bykbaev, and T. E. Flores-Tapia
GI-IATA, CIDII, Universidad Politécnica Salesiana, Cuenca, Ecuador
Email: {jochoaz, vrobles, tflores}@est.ups.edu.ec

Jan Doutreloigne
Center for Microsystems Technology, Ghent University- IMEC, Ghent, Belgium
Email: Jan.Doutreloigne@elis.ugent.be

***Abstract*—Nowadays, the mobile devices (smartphones and tablets) are developed with important improvements in processing and memory capacity as well as camera resolution, massive storage or wireless connectivity. On those grounds, in developing countries these devices constitute an important alternative to develop assistive technologies to provide support in several areas as education, health care, and the elderly. Given that, in this paper we propose a new approach to handle the gesture recognition in robotic assistants that use a mobile device as main processor. In order to perform the recognition our approach uses a robust descriptor based on polygonal approximation, convex hull techniques, and the first seven HU moments. The results show 93% precision in real scenarios with different light conditions.**

***Index Terms*—mobile device, robotic assistant, computer vision, hand gesture recognition, auto-adjust thresholding**

I. Introduction

Thanks to the major progress of mobile devices (smartphones or tablets) in recent years, it has now become possible to integrate them in support tools for various tasks in everyday life. Moreover, the capacity of mobile devices converts them into powerful processing systems, capable of handling audio, video and performing diverse arithmetic operations.

On the other hand, there are several societal aspects, mainly in developing countries, that haven't improved yet. A clear example can be found in the domain of health care and education. Statistical data about Ecuador show a complex picture, where approximately 400000 persons suffer from some kind of disability, with the highest percentage in the provinces Bolívar, Cañar and Sucumbíos [1]. Therefore, it's important to rely on tools that allow improving the educational processes and providing care and assistance to the most necessitous groups in our society. Low-cost mobile devices are ideally suited for participating in those activities as support tools. Therefore, this paper presents a proposal for the implementation of hand gesture and pattern recognition functionality inside Android-based mobile devices. With this functionality, it becomes possible to develop a wide variety of human support applications, ranging from psychomotor stimulation to care and accompaniment of the elderly. Our proposal has been integrated in the Android Mobile Multipurpose Robotic Assistant (AMURA), a mobile robotic assistant whose main objective is to support various assistance applications. In order to achieve the above mentioned, we propose a new descriptor that combines several features extracted with using techniques like polygonal approximation, convex hull, and the first seven HU moments.

This paper is organized in the following way: Section 2 gives an overview of some of the most relevant work done in the area of robotic assistants and the automatic gesture recognition using mobile devices. Section 3 describes in a detailed way the procedure we propose for generating a more robust descriptor for gesture recognition. The results of the recognition process and the practical applications developed based on our approach are presented in Section 4. Finally, Section 5 formulates some conclusions and takes a look at future work.

II. Related Work

Nowadays, several robots have been developed with the aim of providing assistance for people in different areas. In the health care sector exist robots to help with the transportation of immobile patients (lifting bedridden patients) [2], provide support for rehabilitation of upper extremities in patients suffering from paresis [3], or act as support tools in laparoscopic procedures [4].

Other robots have been developed to evaluate a person's mood by looking for pathologies of depression or anxiety disorders in older adults [5], while [6] presents a semi-autonomous robot control to assist the elderly at home. By using this control, the relatives of older adults can control remotely the robot.

Likewise, in the Hand Gestures Detection (HGD) research field, some of the current approaches propose the use of hybrid environments to leverage resource constrained mobile devices, with the aim to automatically and adaptively make offloading and parallelism decisions for mobile interactive perception applications [7]. Other

Manuscript received December 20, 2014; revised April 24, 2015.

©2016 Int. J. Electron. Electr. Eng.
doi: 10.18178/ijeee.4.1.96-99

researches achieve HGD process using light sources toenhance interaction between hands and virtual musical instruments in iPAD devices [8].

III. Proposed Approach

By virtue of the above, this section presents the most important aspects of the hand gesture recognition process, utilizing a system of auto-adjustment in response to illumination changes embedded in mobile Android devices. Our proposed technique has been tested in AMURA, a robotic assistant capable of providing support to the different activities performed in the domain of psychomotor stimulation, for children as well as adults. At present, AMURA integrates an intelligent mobile device (tablet or smartphone) with an electromechanical displacement platform, by means of a robust protocol based on wireless Bluetooth technology.

To that purpose, our approach proposes a technique that allows the automatic determination of the global illumination level of a scene or setting. This approach also proposes a method to combine the mentioned techniques with various algorithms and existing descriptors, in order to recognize four hand gestures, suitable for implementation in mobile Android devices. All image processing is carried out with the aid of the OpenCV library (www.opencv.org), available for Android.

A. Auto-Adjust Thresholding

Given the fact that the user must wear a red-colored glove, segmentation by color is performed prior to carrying out the phase of gesture recognition, with the objective to subtract the background. However, changes in lighting conditions affect the way the camera perceives the color of interest. Therefore it is suggested to measure the illumination of the scene captured by the camera, and determine on the basis of this information the optimal value for the segmentation. Within an image, there may exist black or white objects or elements that affect the values of the histogram, for example, if there was a dark object in the image, then the histogram would indicate that the scene is badly illuminated, although this may not be the case in reality. In an attempt to avoid this inconvenience, we divide the global histogram of the image into two sub-histograms and assign a weight factor (from 0 to 127) to each pixel intensity, giving a higher weight factor to pixels with intermediate intensity, as indicated in Fig. 1.

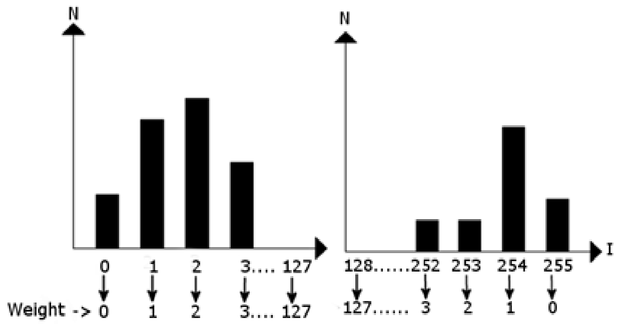

Figure 1. Histogram weighing proposal.

It is now possible to create two vectors that store the multiplication of the histogram value for each pixel intensity and the previously assigned corresponding weight, using to that purposes (1) and (2).

$$Vn = \{0 \times N_0, 1 \times N_1, \ldots .127 \times N_{127}\} \quad (1)$$

$$Vb = \{127 \times N_{128}, 126 \times N_{129}, \ldots .0 \times N_{255}\} \quad (2)$$

where N_i is the number of pixels with the indicated intensity, V_n is the vector with the values for the histogram corresponding to pixel intensities that tend to black, and V_b the similar vector for the pixel intensities that tend to white. At this point it is possible to calculate the average values of V_n as well as V_b, according to the following expressions:

$$mn = \frac{1}{128} \Sigma_{i=1}^{128} Vn_i \quad (3)$$

$$mb = \frac{1}{128} \Sigma_{i=1}^{128} Vb_i \quad (4)$$

Now a heuristic is proposed wherein these two averages are combined in a simple way into the factor f, defined as specified in (5).

$$f = \frac{mn}{mb} \quad (5)$$

This factor allows us to obtain a good global perspective of how the histogram fluctuates in different illumination conditions, irrespective whether there are totally white or black objects in the scene or image, or not. This factor tends to zero when the histogram contains mainly values close to white, whereas it tends to infinity when it contains predominantly values close to black. In this way we can estimate a trend line that enables the system to automatically adjust itself to the actual illumination level on the scene.

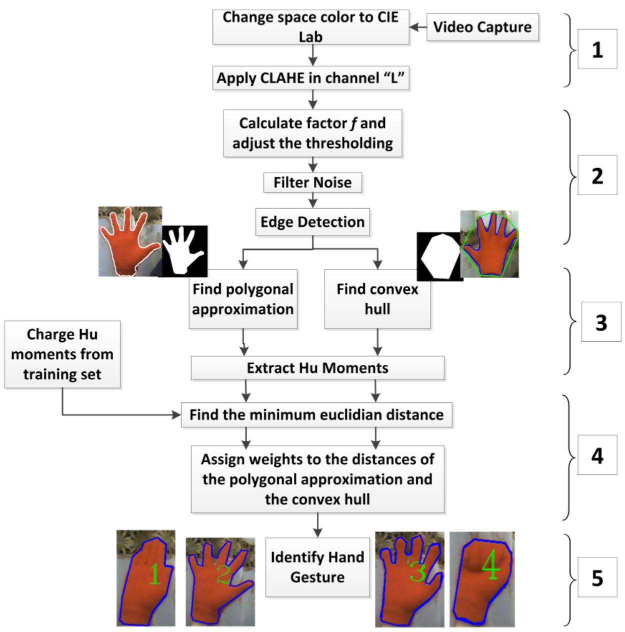

Figure 2. The flow chart for hand gesture recognition.

B. Descriptors Extraction and Recognition Phases

For the gesture recognition phase, a comparison is made between the characteristics of a new gesture

©2016 Int. J. Electron. Electr. Eng.

captured by the camera and the samples of a training set, with the aim to identify to which gesture it actually corresponds. This training set comprises a total of 800 images for 4 different gestures, i.e. 200 images per gesture. The flow chart for the hand gesture recognition process is depicted in Fig. 2.

As a first stage in the proposed process, a representation of the image in the CIE Lab color space is obtained and the Contrast Limited Adaptive Histogram Equalization (CLAHE) [9] technique is applied, with the objective to improve the image quality and to equalize its histogram.

In the second stage, three important tasks are performed:

- The previously mentioned factor f is determined, with the goal to define the thresholding value that will be used for the segmentation of the object of interest (in our case the red-colored glove).
- A filter for noise elimination is applied.
- The object edges are detected in order to be able to determine the polygonal approximation and the convex hull.

During the third stage, two different representations of the segmented glove are obtained: the polygonal approximation and the convex hull [10]. On the basis of these two representations, the 7 invariant HU moments are determined for each one of them.

As a next step, in stage four, we proceed to determine the heuristic for the weight assignment to each moment and for combining them into one single descriptor. For defining the weight factors, the following procedure is adopted: the Euclidian distances with respect to the training set are estimated for each one of the four gestures to be recognized. The equations are defined as follows:

$$dAP_i[x] = \sqrt{\sum_{j=1}^{7}(mAP[j] - gAP_i[j][x])^2} \qquad (6)$$

$$dRC_i[x] = \sqrt{\sum_{j=1}^{7}(mRC[j] - gRC_i[j][x])^2} \qquad (7)$$

where dAP_i is the matrix of distances in case of the polygonal approximation for each gesture i, and dRC_i the matrix of distances for the convex region for each gesture i. The variable x takes values from 1 to 100 and represents the index of each one of the images from the training set. $mAP[j]$ contains the 7 invariant HU moments of the polygonal approximation for the new gesture captured by the camera, whereas $mRC[j]$ contains the 7 invariant HU moments extracted from the convex region for the new gesture captured by the camera. The quantities $gAP_i[j][x]$ and $gRC_i[j][x]$ are the invariant HU moments extracted from the training set, and this for the polygonal approximation as well as for the convex hull for each gesture i. In total this would yield a vector of 100 distances, so it is necessary to extract the smallest distance from each matrix of values:

$$dmAP_i = min(dAP_i) \qquad (9)$$

$$dmRC_i = min(dRC_i) \qquad (10)$$

wherein $dmAP_i$ represents the minimal distance of the polygonal approximation for gesture, and $dmRC_i$ the minimal distance of the convex region for gesture i. Both the polygonal approximation as well as the convex hull provides important information for recognizing the hand gestures, but one may reveal more relevant information than the other. Hence, it is necessary to assign a weight to each one and to calculate a weighted average. This process is conducted during stage five according to (11) and (12).

$$res_i = p_1 \times dmAP_i + p_2 \times dmRC_i \qquad (11)$$

$$p2 = 1.0 - p1 \qquad (12)$$

herein $p1$ is the weight factor for the minimal distance of the polygonal approximation and $p2$ the weight for the minimal distance of the convex hull. Depending on the weight that is assigned to each one, the recognition rate will increase or decrease, this will depend on the quality of the samples in the training set.

Finally, in the last stage, the absolute minimal distance can be extracted as well as the corresponding position in the training set, with the objective to determine which hand gesture the captured image corresponds to. If the value of this absolute minimal distance is not within a given tolerance margin, the captured image is considered not to be a hand gesture:

$$if\{min(res) < tolerance;\ gesture = posMin(res)\} \qquad (13)$$

The tolerance is defined as:

$$tolerance = \bar{x} + \sigma \qquad (14)$$

wherein $\bar{x}$ is the average and σ the standard deviation. The appropriate values are calculated from a total of 50 samples taken from the 4 gestures we would like to recognize.

IV. RESULTS

TABLE I. HAND GESTURE RECOGNITION RATE

Weight		Recognition rate %
p1	p2	
1,0	0,0	76,25
0,9	0,1	83,75
0,8	0,2	91,26
0,7	0,3	90
0,6	0,4	93,75
0,5	0,5	85
0,4	0,6	87,5
0,3	0,7	83
0,2	0,8	83,75
0,1	0,9	75
0,0	1,0	78,75

In order to estimate the appropriate weight factors, tests were performed where p1 was varied in the range from 0 to 1.0, and where a total amount of 80 samples was introduced, i.e. 20 samples for each gesture. Table I clearly indicates how the hand gesture recognition rate changes along with variations of the weight factors. According to the conducted tests, the optimal values for this case are p1=0.6 and p2=0.4. It can also be noticed

©2016 Int. J. Electron. Electr. Eng.

that in case only the polygonal approximation descriptor is used (p1=1 and p2=0), the recognition rate becomes substantially lower than in the case of a combined descriptor. Exactly the same situation occurs when only the convex hull descriptor is employed.

Various experiments were performed to evaluate the gesture recognition and augmented reality applications on different Android architectures and devices, in the lower, medium and higher ranges, yielding a CPU usage between 36% and 67%. Fig. 3 indicates both the minimum and maximum percentages of CPU usage for each test device.

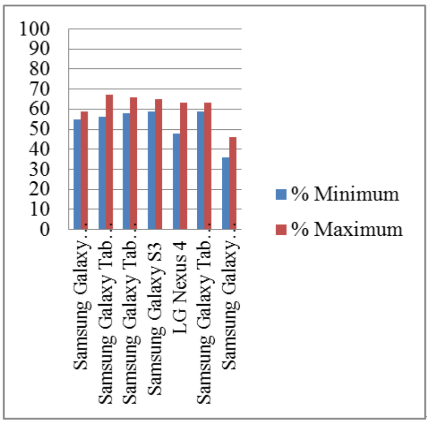

Figure 3. Percentage of CPU usage in the tested devices.

V. Conclusion

The combination of the invariant HU moments from the polygonal approximation and the convex hull, by means of assigning appropriate weight factors, allowed a significant increase of the global hand gesture recognition rate.

This proposal for hand gesture recognition can be implemented in the majority of mobile Android devices, as the CPU usage corresponding to the data processing is only moderate. All tests were conducted in real time, in different settings with variable illumination conditions, in order to guarantee that the proposed approach is viable.

Suggestions for future advanced research include the following:

- To develop a control module that enables an automatic adjustment of the weight factors starting from the illumination values on the scene.
- To analyze the performance of the application when other descriptors are being used, such as the case of Fourier.

Acknowledgment

This work was funded by the “SistemasInteligentes de Soporte a la Educación (SINSAE v2) CIDII-010213” research project of the Universidad PolitécnicaSalesiana, Sede Cuenca.

References

[1] (2014). Consejo nacional de discapacidades, Registro nacional de discapacidades. [Online]. Available: htttp://www.consejodiscapacidades.gob.ec/wp-content/uploads/downloads/2014/11/registro_nacional_discapacidades.pdf

[2] M. Ding, *et al.*, “Comfort estimation during lift-up using nursing-care robot—RIBA,” in *Proc. IEEE First International Conference on Innovative Engineering Systems (ICIES)*, 2012, pp. 225-230.

[3] M. D. Kostic, M. B. Popovic, and D. B. Popovic, “Control of robot assistant for rehabilitation of upper extremities,” in *Proc. Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2012, pp. 3918-3921.

[4] B. Estebanez, E. Bauzano, and V. F. Muñoz, “Surgical tools pose estimation for a multimodal HMI of a surgical robotic assistant,” in *Proc. IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2011, pp. 2121-2126.

[5] S. McGlynn, B. Snook, S. Kemple, T. L. Mitzner, and W. A. Rogers, “Therapeutic robots for older adults: investigating the potential of paro,” in *Proc. ACM/IEEE International Conference on Human-Robot Interaction*, 2014, pp. 246-247.

[6] M. Mast, *et al.*, “Design of the human-robot interaction for a semi-autonomous service robot to assist elderly people,” *Advanced Technologies & Societal Change*, pp. 15-29, 2015.

[7] M. R. Ra, *et al.*, “Odessa: Enabling interactive perception applications on mobile devices,” in *Proc. 9th International Conference on Mobile Systems, Applications, and Services*, 2011, pp. 43-56.

[8] O. Nieto and D. Shasha, “Hand gesture recognition in mobile devices: Enhancing the musical experience,” in *Proc. 10th International Symposium on Computer Music Multidisciplinary Research*, Marseill, France, 2013.

[9] S. J. Kim, B. S. Min, D. K. Lim, and J. H. Lee, “Determining parameters in contrast limited adaptive histogram equalization,” in *Proc. 7th International Conference on Information Security and Assurance*, 2013.

[10] G. Bradski and A. Kaehler, *Lerning OpenCV: Computer Vision with OpenCV Library*, California: O’Reilly Media, Inc., 2008.

Juan S. Ochoa-Zambrano was born in Cuenca, Ecuador in 1990. He received the engineering degree in electronics in 2015 from the Universidad Politécnica Salesiana, Cuenca, Ecuador in 2015.
From 2013 to 2014, he collaborated in the Research Group on Artificial Intelligence and inclusion technologies of the University Politécnica Salesiana en Cuenca, Ecuador, where he made his degree work in order to obtaining the title in electronical engineering.
He currently serves as innovation assistant in the Integrated Security System ECU 911 in Cuenca, Ecuador, developing monitoring technologies based on computer vision, in order to improve the security for all people in the country.

Vladimir Robles-Bykbaev was born in Azogues, Ecuador in 1980. He received the engineering degree in computer systems in 2015 from the Universidad Politécnica Salesiana, Cuenca, Ecuador in 2006.
He continued his studies in year 2008 in the area of artificial intelligence and obtained the Master degree from Universidad Politécnica de Valencia (Spain) in year 2010. Currently, he is PhD student in the University of Vigo (Spain). Nowadays, he is in charge of the research group in artificial intelligence, and assistive technologies at the Universidad Politécnica Salesiana, where collaborates as full-time teacher. Her research interests include artificial intelligence, assistive technology, data mining, and computer vision.