Arabic Sign Language Alphabet Recognition Methods Comparison, Combination and implementation

Mohamed Youness Ftichi[1], Abderrahim Benabbou[1], Khalid Abbad[1]
[1] Dept. of Intelligent Systems and Applications
Faculty of Sciences and Technologies
FES, Morocco

ABSTRACT: *Sign language can be defined as a combination of hand motion mainly used for communication purposes, especially for the deaf-mute community. More than 5% of the worldwide population (320 million) are concerned by the use of it. Through our work, we aim to provide a mean to automate the process of translation from Arabic sign language to written Arabic, in the static context. As a first step, we will produce a three-level process, allowing the recognition of static Arabic sign.*

Keywords: Arabic Sign Language, Recognition, Skin Color Segmentation, Hull Convex, Classification, Hand Pose

Received: 18 June 2017, Revised 28 July 2017, Accepted 4 August 2017

1. Introduction

In literature, most static sign language recognition systems (based on images), and generally static hand gesture recognition systems, based on vision, are composed of three basic phases: segmentation, features extraction and recognition. Our work will be a combination of several methods, following the same schema to get a translation from Arabic sign language alphabet to written Arabic.

First step will be the skin color segmentation within a static image. This step will be crucial to get the correct segmented image for the rest of the process by generating a skin color model to work with.

The features extraction will allow us to get the characteristics best describing the sign (hand pose), and pass it on to the classifier. A number of classifiers will be tested and one will be chosen as the one with best results.

2. Skin Color Segmentation

Skin color segmentation is one of the widely used techniques to get a robust hand segmentation, it main aims to build a decision rule allowing as to separate skin from non-skin pixels. A metric measuring the distance between the pixels values and skin

tone is generally used, this metric is defined according to the method chosen to model the skin color.

2.1 Skin Modeling Method

Segmentation through a predefined set of rules is among the many efficient methods available to get skin clusters in images. Many researchers, such as [1], [2], [3], [4], have tempted to use this method in their works, essentially because of how easy it is to construct a fast skin color classifier. Nevertheless, achieving a high recognition rate lies on the color space choice and the tests to get the adequate set of rules.

Recently, there have been proposed a method that uses machine learning algorithms to find both suitable color space and a simple decision rule that achieve high recognition rates [5]. The authors start with a normalized RGB space and then apply a constructive induction algorithm to create a number of new sets of three attributes being a superposition of *r*, *g*, *b* and *a* constant 1/3, constructed by basic arithmetic operations. A decision rule, is estimated for each set of attributes. The authors prohibit construction of too complex rules, which helps avoiding data over-fitting, that is possible in case of lack of training set representativeness. They have achieved results that outperform Bayes skin probability map classifier in RGB space for their dataset.

In our work, and through experiment and comparison of the different available techniques, we have concluded that the following criteria, presented in the [1], is the best suited for our usage, and is described as follow:

$$\begin{aligned} &\text{If} \quad \frac{3br^2}{(r+g+b)^3} > 0.1276 \\ &\text{And} \quad \frac{r+g+b}{3r} + \frac{r-g}{r+g+b} \leq 0.9498 \\ &\text{And} \quad \frac{rb+g^2}{gb} \leq 2.7775 \\ &\text{Then} \quad \text{Class} = \text{Skin} \end{aligned} \tag{1}$$

2.2 Application

A standard hand gesture dataset, provided in [6] [7] [8], is used to test our process as in Fig. 1.

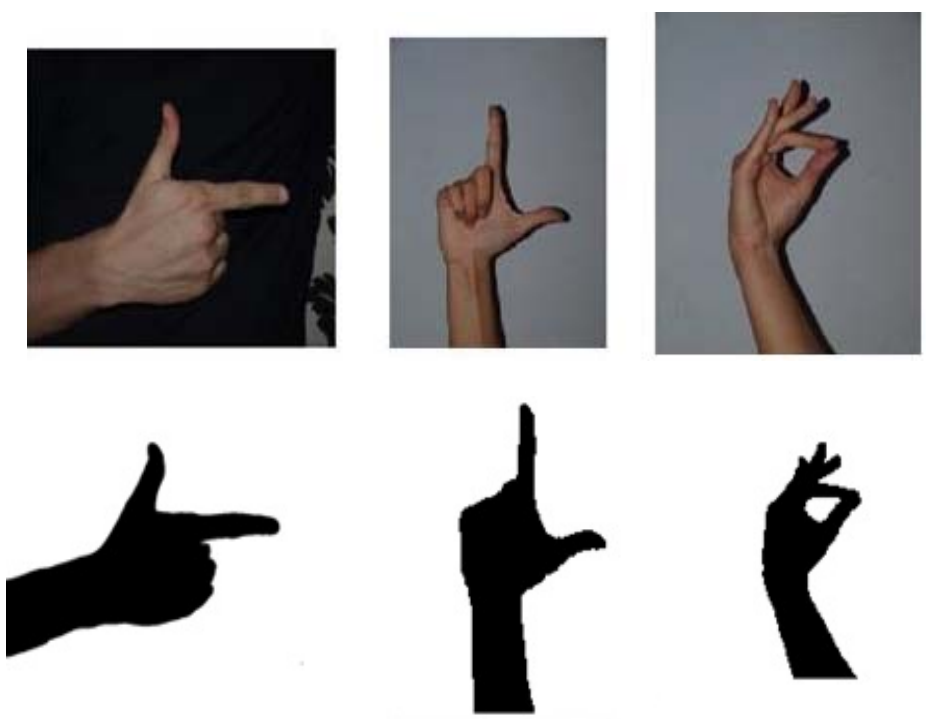

Figure 1. Hand segmentation based on the chosen method, on top the original image, on the bottom the segmented one

3. Features

3.1 In Literature

Features extraction is the major step toward the recognition of the Arabic sign language alphabet. This step is the mean to reduce the dimensionality of the problem, and give the minimum amount of data to the classifier (see section 4), so we can get an identification of the hand pose correspondent in the sign language vocabulary.

In literature, many different methods have been used to characterize the main key elements of an image, in the hand pose recognition context. Among the various methods, SIFT [9], Hu Moments [10], dimensional Gabor Wavelets [11] and Fourier descriptors [12] have been used with more or less success to extract the features to use, efficiently, to recognize the sign.

3.2 Our Approach

For the purpose of work, we choose to use the hull convex [13] as a mean to detect the defection point that are the base on the features extracted for the recognition part (see section 4).

Based on the results of the skin color segmentation, we apply the Hull Convex algorithm, following the extraction of the hand's contour to get the convexity defect points in those images as shown Fig 2.

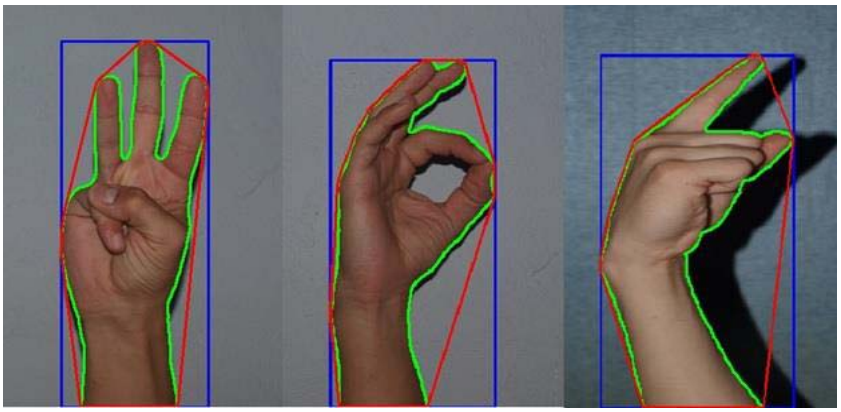

Figure 2. Hull convex method applied to segmented images

The Hull convex step is followed by the drawing of the convexity defect points for each sign as shown Fig. 3.

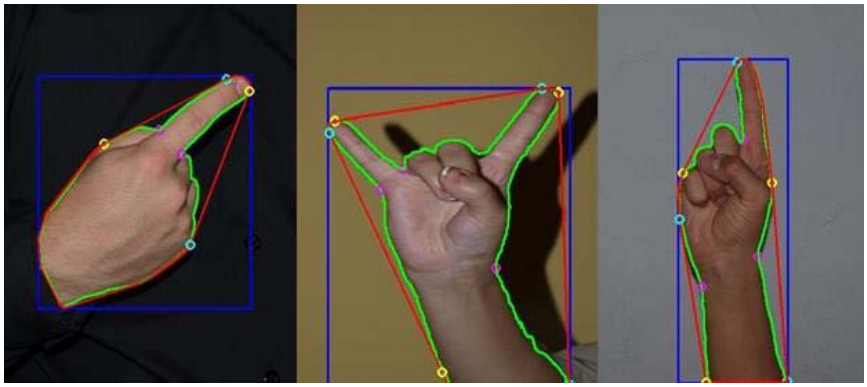

Figure 3. Defect points identification

The points in yellow, blue and purple represent respectively the beginning of the defection (Bg), its end (En) and the defection point (Df). The next step is to calculate the following distances:

• A = Distance from the defect point to the beginning of the defection;

• B = Distance from the defect point to the end of the defection;

• C = The angle $(B\widehat{gDf}En)$.

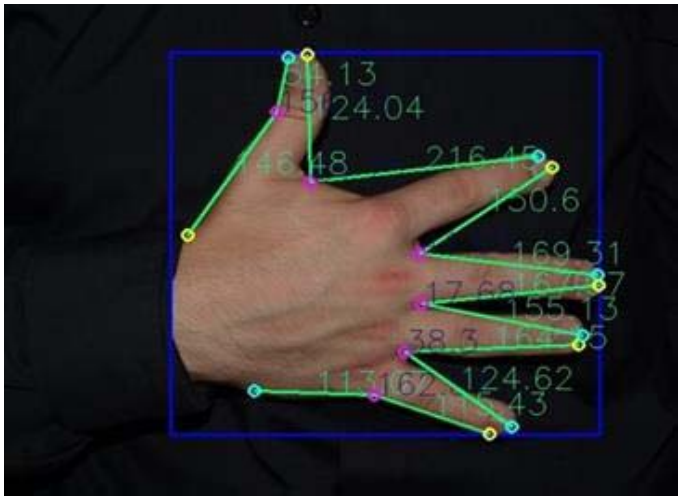

Figure 4. Numbers in blue are the angles, the green ones are the distances

3.3 Classification

Classification is the last, but most important step of the recognition process. In order to get the best result, we experimented different classification methods available in the literature, most specifically Stochastic Gradient Descent [14], Random Forest [15], Logistic Regression [16], KNN [17], Decision Tree [18], SVC and Linear SVC [19]. The vector used for classification contains the number of defect points, the angle C, the distances A and B, and the different locations of the defects point.

We use the dataset provided by [6] [7] [8]. The learning process contains 10 images for each sign, for a total of 240 images. The next step is to apply those methods to the rest of the dataset which consists of a total of 899 images, approximately 33 images for each sign. The results of the recognition process are Fig. 5.

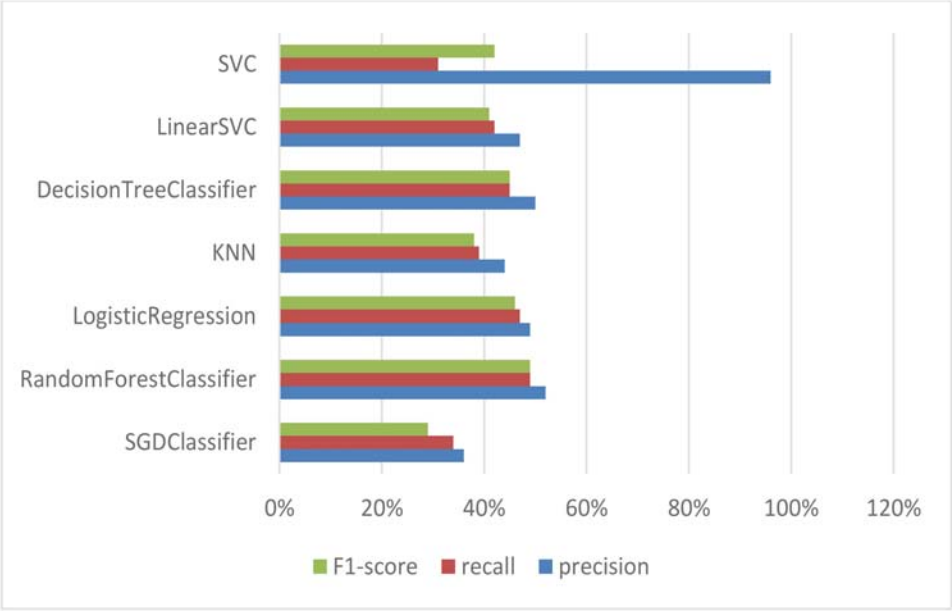

Figure 5. Classification methods comparative results

4. Conclusion

Through our work, we were able to address the matter of translation from Arabic sign language based on static images provided by the standard hand pose dataset on [6][7][8]. Test were realized on an intel core I7-4710MQ CPU 2.50 GHZ 2.50 GHZ computer and using Scikit image and OpenCV frameworks on python.

5. Future work

We are currently working on the establishment of a translation process based on dynamic hand poses (videos), then, later, hand gestures in real time. Such objectives are realizable based on the work presented in this paper and are our main focus.

References

[1] Peer, P., Kovac, J., Solina, F. (2003). Human skin colour clustering for face detection. Submitted to EUROCON 2003 – International Conference on Computer as a Tool.

[2] Ahlberg, J. (1999). A system for face localization and facial feature extraction. Tech. Rep. LiTH-ISY-R-2172, Linkoping University.

[3] Fleck, M., Forsyth, D. A., Bregler, C. (1996). Finding naked people. *In*: Proceedings of the ECCV, 2, 592–602.

[4] Jordao, L., Perrone, M., Costeira, J., Santos-Victor, J. (1999). Active face and feature tracking. *In:* Proceedings of the 10th International Conference on Image Analysis and Processing, 572–577.

[5] GOMEZ, G., MORALES, E. (2002). Automatic feature construction and a simple rule induction algorithm for skin detection. *In:* Proceedings of the ICML Workshop on Machine Learning in Computer Vision, 31–38.

[6] Kawulok, M., Kawulok, J., Nalepa, J., Smolka, B. (2014). Self-adaptive algorithm for segmenting skin regions, EURASIP *Journal on Advances in Signal Processing* (170)

[7] Nalepa, J., Kawulok, M. (2014). Fast and accurate hand shape classification, *In*: Beyond Databases, Architectures, and Structures, S. Kozielski, D. Mrozek, P. Kasprowski, B. Malysiak-Mrozek, and D. Kostrzewa, Eds., 424 of Communications in Computer and Information Science, p. 364-373. Springer.

[8] Grzejszczak, T., Kawulok, M., Galuszka, A. (2016). Hand landmarks detection and localization in color images, *Multimedia Tools and Applications*, 75 (23) 16363-1638.

[9]Carson, C., Belongie, S., Greenspan, H., Malik, J (2002). Blobworld: Image segmentation using expectation-maximization and its application to image querying, *IEEE Trans*. PAMI, 2002.

[10] Hu, M (1962). Visual pattern recognition by moment invariants, *Ire Trans Inf Theory*, (8) 179-187.

[11] Chang, T., Kou, C. (1993). Texture analysis and classification with treestructured wavelet transform, *IEEE Trans. On Image Processing* , 2 (4) 429-441.

[12] Wang, Tao., Liu, Wen-Yin., Sun, Jia-Guang., Zhang, Hong-Jiang. (2002). Using Fourier Descriptors to Recognize Object's Shape, *Journal of Computer Research and Development*, 39 (12) December.

[13] Youssef, M. (2011). Hull Convexity Defect Features For Human Action Recognition, Ohio: Day-ton University.

[14] Bottou, L. (2012). Stochastic Gradient Descent Tricks. *In:* Montavon G.,Orr G.B., Müller KR. (eds) Neural Networks: Tricks of the Trade. *Lecture Notes in Computer Science*, 7700. Springer, Berlin, Heidelberg

[15] Breiman, L. (2001) Machine Learning. 45. 5.

[16] Yu, HF., Huang, FL., Lin, CJ. (2011) *Mach Learn*. 85: 41.

[17] Andoni, A., Indyk, P. (2006). Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions, *In:* Proceedings 47th Annu. IEEE Symp. Found. Comput. Sci. (FOCS), October, 459–468.

[18] Breiman, L., Friedman, J., Olshen, R., Stone, C. (1984). Classification and Regression Trees, Wadsworth, Belmont, CA.

[19] Cortes, C., Vapnik, V., Mach Learn. (1995). 20. 273.