2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO)
Amity University, Noida, India. June 4-5, 2020

Creation of Image Segmentation Classifiers for Sign Language Processing for Deaf and Dumb

Pushan Kumar Datta
Amity School of Engineering and Technology,
Amity University Kolkata
Kolkata, India
pkdutta@kol.amity.edu

Abhishek Biswas
Senior Consultant,
Protiviti Middle East Member Firm,
Riyadh, Saudi Arabia
abhishek.biswas@protivitiglobal.me

Ahona Ghosh
Assistant Professor
Department of Computational Science
Brainware University, Kolkata
Kolkata, India
ahg.cs@brainwareuniversity.ac.in

Nilanjana Chaudhury
Senior Consultant and Entrepreneur
Kolkata, India
chaudhuri.nil@gmail.com

***Abstract*—Recognition of sign language is an emerging area of research in the domain of gesture recognition. Research has been carried out around the world on sign language recognition, for many sign languages. The basic phase of sign language recognition systems is accurate hand segmentation. This paper used Otsu's technique of segmentation to create an improved vision-based recognition of sign language. There are about 466 million users suffering from hearing loss globally, of whom no. of kids is 34 million. Deaf people have very little or no capacity to hear. For communication, they use sign language. People in distinct areas of the globe use distinct sign languages which is very small in amount compared to spoken languages. Our goal is to create a static-gesture recognizer, a multi-class classifier that predicts the gestures of static sign language. In the proposed work, we identified the hand in the raw image and provided the static gesture recognizer (the multi-class classifier) with this section of the image. We build multi-class classifiers from the scikit learning library by first building the data set, each image being converted into a feature vector (X) and each has a label that matches the sign language alphabet denoted by (Y). Our predicted classifiers analyzed 65% of the said images with clarity.**

Keywords—Sign Language, Image segmentation, Indian Sign Language, Otsu Segmentation, Recognition, Image thresholding

I. INTRODUCTION

Sign language (SL) is a visual and gestural vocabulary applied by people who are deaf and tough-hearing. To express meanings, three-dimensional spaces and hand moving activities (and other portions of the body) are used by them. It has its own grammar and syntax that is drastically dissimilar from the verbal and written dialects. Image processing, hand detection and finger counting can help people with disabilities to interact with the technological system. For example, a student with a high degree of visual impairment may listen to the verbal test questions aloud and then show the answer to the camera with his fingers. This is the domain of gesture recognition. Gesture recognition is an open machine vision issue, a computer science discipline that allows devices to emulate human vision. Sign language recognition is one of the biggest concerns for the dumb and the deaf. Recognition of sign language is a study domain that involves pattern recognition, computer vision, processing of natural language [1]. Recognition of sign language is a comprehensive issue due to the complexity of hand gesture visual analysis and the extremely structured type of hand signals. Neha et al. [2] proposed a method that is able to identify and translate the Indian Sign Language (ISL) into a normal text. Segmentation of the skin color is done using the method of clustering k-means. Sumaira K. et al. [3] proposed a Fuzzy Sign Language Recognition (PSL) classifier. Xu Zhang et al. [4] described a framework for Chinese Sign Language (CSL) identification based on the mixture of information from active segments with a tri-axis accelerometer (ACC) and electromyography (EMG) sensors having multichannel. Cao Dong, Ming C. Leu [5] used a method of classification per pixel used to divide the hand into portions. The input to the system was the hand area's depth image, and the output was each pixel's classification label. Divya S and Kiruthika [6] had already developed a SLR (Sign Language Recognition) system to computerize a phonetics translator's work, the image is segmented using HSV, RGB and YCbCr Color space-based skin color. Gesture recognition has many apps to improve communication between people and computers, and one of them is in the sector of Sign Language Translation, where a video series of symbolic hand gestures [8-11] is transformed into natural language. In addition, in security systems, in addition to face recognition, entering password with hand gestures may be considered. Alternatively, a system that interprets sign language with finger gestures can be developed [12-14]. In the project, many filters were used for the perception of the human hand. The key now is to use a suitable approach to vector the picture and retrieve significant data in order to supply the classifier. If we plan to be using simple multi-class classifiers, simply using the raw pixel values won't operate. We use the Histogram of Oriented Gradients (HOG) method to vectorize our pictures as it has been shown to deliver good outcomes on issues like this one. First, background subtraction is performed with HOG2 algorithm. Then, using a filter specially designed to detect the color of human skin, non-human is masked as. After this stage, the non-uniform noise on the image signals is reduced as much as possible with the median filter. Then, HSV color space transformation and grayscale transformation are performed. As the last stage of filtering, Otsu algorithm is used together with the dual thresholding method to mask those below the threshold value with black color. Now we scale the test image by different factors to actually use the above classifier and then use a sliding window approach on all of them to pick the window that perfectly captures the area of interest. This is performed by choosing the region that corresponds to the maximum trust results assigned to all scales by the binary (hand / not-hand) classifier.

II. BACKGROUND SUBTRACTION AND LEATHER COLOR FILTER

Every pixel in the image has a separate probability density function. The pixels in the new image will have a different probability density function than the old image (background). Since the camera stands still, we can analyze what we call the background as a photograph. After statistical modeling analysis, variations are determined for the levels of pixel densities because the variance may vary from pixel to pixel. For this, the single Gaussian model specified in equation 1 can be used [15]. RGB color space or another position of a pixel in the color space with respect to time is indicated by x (t) notation pixels. In the background-based deletion process, we can determine the probability that a pixel is a part of the background (AP) or a part of the foreground (ÖP) by Bayesian rule (R). This process is also used in a single Gaussian model.

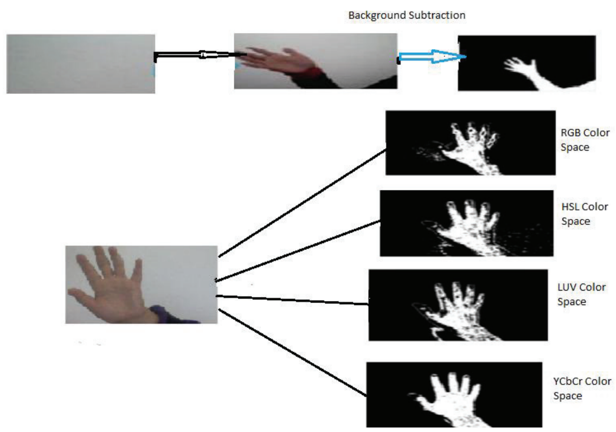

Fig. 1. Background subtraction

A method developed to detect human skin color is a hybrid method that works and decides as a combination of RGB, HSV, YCbCr color spaces. In the RGB color space, the color of human skin cannot be clearly distinguished by 3 channels (red, green, blue). However, because there are factors such as daylight, RGB color space, we will use it to reduce the effect of these factors [16]. HSV (hue, saturation, value) color extract consists of saturation and brightness components. spaces like LUV decouple the "color" (chromaticity, the UV part) and "lightness" (luminance, the L part) of color. Thus in object detection, it is common to match objects just based on the UV part, which gives invariance to changes in lighting condition. The problem of estimating the orientation of a 3D object from the flow of images is a difficult problem, especially in the context of dynamic applications where the camera is operated by the user (held in the hand or connected to the head). Under these conditions, objects in the scene undergoing any kind of movement, sometimes very fast and difficult to predict.

Fig. 2. Use of skin color filter with background subtraction

To calculate the convex hull of the given contour, the palm is perceived as the convexity defects with the fingers. The above technique can streamline the complex structure. The convex hull is the curved polygon that is encircled by the vertices that correspond to the index finger. All of the convex vertices are in motion contour. In this we apply the movement of the fingers at different time intervals to identify the sequence which can then be done using the hidden markov model approach or similar learning sequence.

III. STEPS APPLİED FOR THE STUDY

In normal isntances it has been found that the Viola-Jones HaarCascades detector developed by Juan Wachs [16] could detect easily closed fists and perform background subtraction in different color features. The other algorithm that shows the HandVu approach involves using a single sided technique to show the different sides of the hand for good tracking[17]. However both this approaches will depend on skin pigment detection and background subtraction.However with multi class variance it may so happen that we may require a weighted hull approach for gray scale image detection.

A. Hsv and grayscale conversion

Before applying the threshold value to the hand object and masking it (just before starting geometric operations), we need to apply Hsv transformation and gray transformation to the image. Otsu binarization automatically calculates a threshold value from image histogram for a bimodal image, which is an image whose histogram has two peaks. Thresholding is then performed which is a simple method of segmentation. Thresholding is used to produce a binary image from a grayscale image. Thresholding technique compares each pixel intensity value (I) to the threshold value (T). If $I < T$, the pixel is replaced by a black pixel and if $I > T$ is replaced by a white pixel. Our aim is identify the accuracy of the convexity defect measure using a plain background and with a non plain background.

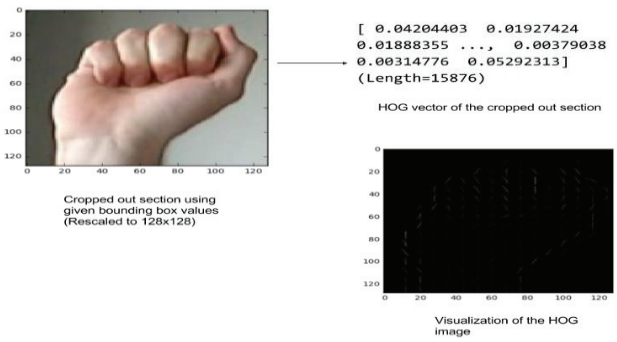

Fig. 3. Hsv color space transformation and grayscale transformation

The convex hull algorithm is an algorithm that finds the correct parts that connect the outermost points of the set of points whose contours are detected. We will also use the points where these right parts meet for our operations. These points are called extreme points. By counting the extremes, we will calculate the number of fingers that are currently residing in the image.

Otsu's method for its simplicity and efficiency is one of the most popular methods. Based on the thresholding method, it is dependent on choosing the optimum threshold value to maximize the variance of the resultant object and background classes between classes. The optimum threshold finding is done serially until a value is found which makes the difference between two or more classes maximum range. The Otsu algorithm [17] assumes that there are two classes of pixels in the image, background and foreground, using the bimodal histogram of the image. We can think of a bimodal histogram as a normal histogram. However, unlike normal, it is a histogram that reaches a significant peak at two points. Thus, the herbaceous algorithm background knows which value is closer to the average of the pixels and the approximate value of the average of the foreground pixels. The Otsu method calculates the optimal threshold value that separates these two pixel classes with the equation shown in equation 2. Thus, in-class variance is minimum, and between-class variance is maximum [18].

B. Masking

The Otsu algorithm assumes that there are two classes of pixels in the image, background and foreground, using the bimodal histogram of the image. We can think of a bimodal histogram as a normal histogram. However, unlike normal, it is a histogram that reaches a significant peak at two points. Thus, the herbaceous algorithm background know which value is closer to the average of the pixels and the approximate value of the average of the foreground pixels. The Otsu method[19] calculates the optimal threshold value that separates these two pixel classes with the equation shown in equation 2. Thus, in-class variance is minimum, and between-class variance is maximum [9]. If the intensity of a gray level image is represented in L gray levels [1,2,...,L] the number of points with gray level at i is expressed by $x\mathrm{I}$ and the total set of points can be denoted by $X = x1 + x2 + \cdots + x\mathrm{L}$. The histogram of this gray –level image is considered as an occurrence distribution of probability. Where p(i)=xi/X where xi≥0 where $\sum_{i=1}^{L} x_i = 1$

The image pixels are categorized into two categories C0 and C1 , i.e. foreground and background by a threshold t. Where C0 denotes pixels within the level [1,2,...,t] , and C1 represents pixels within the level [t+1,..., L].

$q = q(t) = \sum_{i=1}^{t} p(i)$

$\mu = \sum_{i=1}^{L} i.\, p(i)$

The divisible degree η of the class, in the discrimination analysis is

$\eta = \max_{1 \mathrm{stSL}} \sigma_B^2$

Finally, maximizing σ_B^2 to select the optimum threshold t

t = arg max σ

The between-class variance σ of the C0 and C1 is denoted by

$$\sigma^2{}_w(t) = q_1(t)\sigma^2{}_1 + q_2(t)\sigma^2{}_2$$

The algorithm step by step is as follows

1. Building the dataset by hand / not-hand.

2. Converting all pictures i.e. cropped parts to their vectorized shape with movements and hand, non-hand pictures.

3. Using these information sets, build a binary classifier to detect the part with the side and build a multi-class classifier to identify the gesture.

4. Use the above classifiers to conduct the necessary assignment one after the other.

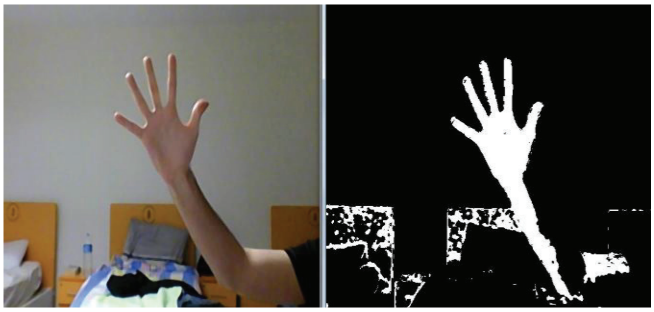

Fig. 4. Masking with Otsu algorithm and binary thresholding

Finding the largest contour and applying the convex body algorithm [20-22]. In order to be able to perform geometric operations on the image of our hand, we need to determine the limits of our hand on the filtered black and white image. For this, we should be able to perceive the contours of our hands instantly. In this way, while determining the limits of our hand, we will also be tracking our hand in software. So we don't have to put our hand on a certain area of the image, the camera will detect it wherever our hand. With the find Contours command in OpenCV, it is possible to find the edges of objects[23-26], but since this command finds all the edges in the image, the edges of this object will be detected if an object appears in the background that filters cannot block. To prevent this, we will consider the object with the largest area of the objects whose edges are detected.

IV. CONCLUSION

Segmentation of the image is frequently used to differentiate between the foreground and the background. The Otsu thresholding method for Indian sign image segmentation was proposed in this paper. This method is robust, fast and easy for implementation and creates appropriate binary images that are used in the next stages of processing including extraction and validation of attributes. This Algorithm will have faster processing time when modification of finger movement. However, the algorithm may have limited capability in identification of wrist movement and other notable features which are essential for the overall quality analysis of different representation algorithm.

REFERENCES

[1] Bojja, Abhishake Kumar, Franziska Mueller, Sri Raghu Malireddi, Markus Oberweger, Vincent Lepetit, Christian Theobalt, Kwang Moo Yi, and Andrea Tagliasacchi. "Handseg: An automatically labeled dataset for hand segmentation from depth images." In *2019 16th Conference on Computer and Robot Vision (CRV)*, pp. 151-158. IEEE, 2019.

[2] Tavari, Neha V., and A. V. Deorankar. "Indian Sign Language Recognition based on Histograms of Oriented Gradient." *International Journal of Computer Science and Information Technologies* 5, no. 3 (2014): 3657-3660.

[3] Kausar, Sumaira, M. Younus Javed, and Shaleeza Sohail. "Recognition of gestures in Pakistani sign language using fuzzy classifier." In *Proceedings of the 8th conference on Signal processing, computational geometry and artificial vision*, pp. 101-105. World Scientific and Engineering Academy and Society (WSEAS), 2008..

[4] Dong, Cao, Ming C. Leu, and Zhaozheng Yin. "American sign language alphabet recognition using microsoft kinect." In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 44-52. 2015.

[5] Zhang, Xu, Xiang Chen, Yun Li, Vuokko Lantz, Kongqiao Wang, and Jihai Yang. "A framework for hand gesture recognition based on accelerometer and EMG sensors." *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 41, no. 6 (2011): 1064-1076.

[6] Divya, S., S. Kiruthika, A. L. Nivin Anton, and S. Padmavathi. "Segmentation, tracking and feature extraction for Indian sign language recognition." *International Journal on Computational Sciences & Applications* 4, no. 2 (2014): 57-72.

[7] Abdul Rahman, Nusirwan Anwar, Kit Chong Wei, and John See. "Rgb-h-cbcr skin colour model for human face detection." *Faculty of Information Technology, Multimedia University* 4 (2007).

[8] Simons, Gary F., and Charles D. Fennig, eds. *Ethnologue: languages of Asia*. sil International, 2017.

[9] Lucas, Ceil, ed. *The sociolinguistics of sign languages*. Cambridge University Press, 2001.

[10] Joy, Jestin, Kannan Balakrishnan, and M. Sreeraj. "SignQuiz: A Quiz Based Tool for Learning Fingerspelled Signs in Indian Sign Language Using ASLR." *IEEE Access* 7 (2019): 28363-28371.

[11] Aronoff, Mark, Irit Meir, and Wendy Sandler. "The paradox of sign language morphology." *Language* 81, no. 2 (2005): 301.

[12] Mayberry, Rachel I. "First-language acquisition after childhood differs from second-language acquisition: The case of American Sign Language." *Journal of Speech, Language, and Hearing Research* 36, no. 6 (1993): 1258-1270.

[13] Schembri, Adam, and Trevor Johnston. "Sociolinguistic variation in the use of fingerspelling in Australian Sign Language: A pilot study." *Sign Language Studies* 7, no. 3 (2007): 319-347.

[14] Price, B. "Sign language costs 'too high'for some families." *Retrieved November* 30 (2017): 2017.

[15] Knoors, Harry, and Dean Hermans. "Effective instruction for deaf and hard-of-hearing students: Teaching strategies, school settings, and student characteristics." *The Oxford handbook of deaf studies, language, and education* 2 (2010): 57-71..

[16] Cheok, Ming Jin, Zaid Omar, and Mohamed Hisham Jaward. "A review of hand gesture and sign language recognition techniques." *International Journal of Machine Learning and Cybernetics* 10, no. 1 (2019): 131-153.

[17] Viola, Paul, and Michael Jones. "Rapid object detection using a boosted cascade of simple features." In Proceedings of the 2001 IEEE computer society conference on computer vision and pattern recognition. CVPR 2001, vol. 1, pp. I-I. IEEE, 2001.

[18] Kolsch, Mathias, and Matthew Turk. "Fast 2d hand tracking with flocks of features and multi-cue integration." In 2004 Conference on Computer Vision and Pattern Recognition Workshop, pp. 158-158. IEEE, 2004.

[19] Otsu, Nobuyuki. "A threshold selection method from gray-level histograms." *IEEE transactions on systems, man, and cybernetics* 9, no. 1 (1979): 62-66.

[20] Quesada, Luis, Gustavo López, and Luis Guerrero. "Automatic recognition of the American sign language fingerspelling alphabet to assist people living with speech or hearing impairments." *Journal of Ambient Intelligence and Humanized Computing* 8, no. 4 (2017): 625-635.

[21] Huang, Jie, Wengang Zhou, Qilin Zhang, Houqiang Li, and Weiping Li. "Video-based sign language recognition without temporal segmentation." In *Thirty-Second AAAI Conference on Artificial Intelligence*. 2018.

[22] Kumar, Pradeep, Partha Pratim Roy, and Debi Prosad Dogra. "Independent bayesian classifier combination based sign language recognition using facial expression." *Information Sciences* 428 (2018): 30-48.

[23] Kishore, P. V. V., D. Anil Kumar, AS Chandra Sekhara Sastry, and E. Kiran Kumar. "Motionlets matching with adaptive kernels for 3-D Indian sign language recognition." *IEEE Sensors Journal* 18, no. 8 (2018): 3327-3337.

[24] Weaver, Kimberly A., and Thad Starner. "Mobile Sign Language Learning Outside the Classroom." *Online Submission* (2012).

[25] Efthimiou, Eleni, Stavroula-Evita Fotinea, Christian Vogler, Thomas Hanke, John Glauert, Richard Bowden, Annelies Braffort, Christophe Collet, Petros Maragos, and Jérémie Segouat. "Sign language recognition, generation, and modelling: a research effort with applications in deaf communication." In *International Conference on Universal Access in Human-Computer Interaction*, pp. 21-30. Springer, Berlin, Heidelberg, 2009..

[26] Kaehler, Adrian, and Gary Bradski. *Learning OpenCV 3: computer vision in C++ with the OpenCV library*. " O'Reilly Media, Inc.", 2016.