

Feature Extraction of Dynamic Sign Language Based on Dct

Hussein Ali Aldelfy¹, Mahmood Hamza Al-Mufraji², Thamir R. Saeed³

¹Faculty of Engineering, University of Mustansiriyah Baghdad, Iraq

^{2,3}Department of Electrical Engineering University of Technology - Iraq Baghdad, Iraq
hussienaldelfy@gmail.com, 30207@uotechnology.edu.iq, 50527@uotechnology.edu.iq

Abstract— Sign language is the basic way to communicate and understand ideas among people with disabilities, deaf and dumb, and to recognize and understand the meanings of the motions and the gestures used by them that requires knowledge of certain features such as hand position, form, motion, directions and facial expressions. Moreover, the extraction of image features is important and it plays a vital role in further analysis of image processing. Extracting the special features of the isolated Arabic sign language (ArSL) words based on Discrete Cosine Transform (DCT) was proposed in current paper. By monitoring and tracking the trajectory of the hands when single hand movement or double hands extract the required features. That is the independent sign language word features vector is incorporated as an input into the classification stage to derive the meaning of the word as voice or text. The research data were collected in cooperation with the Ministry of Labor and Social Affairs of Iraq with the assistance of the Special Needs Department and up to 40 words of isolated sign language collected. Two signs were taken as examples of feature extraction (Happy and water). Since the word happy gives an example of a sign word performed by two hands, on the other hand, the word water is performed with one hand, the Arabic sign language contains different words, part of them are performed with one hand and the other part with both hands.

Index Terms— Sign language, Feature extraction, Discrete cosine transform.

I. INTRODUCTION

Sign Language is the requisite way to communicate and exchange information and ideas among members of the deaf and dumb community to continue their normal lives. Normal people generally need an interpreter to communicate with hearing impaired. Additionally, People with hearing disabilities developed a special kind of communication culture among themselves and with healthy people using gestures. To perform the same purpose as the spoken language, the sign language fulfills that goal for people who suffer from hearing and speech impairment in a clear and understandable manner [1]. The spoken language structure uses words in linearly, one by one. Thus sign language uses many physical gestures in both spatial and time space simultaneously. Sign languages are natural languages in daily life that use various means of communication. Examples of certain sign languages include Arabic Sign Language, American Sign Language, British Sign Language, Chinese Sign Language, Korean Sign Language, Japanese Sign Language and so on. [2]. The gestures are classified into two types: **static** and **dynamic** gestures. Gestures of the static type depend on the shape of the hand to represent the desired sign in the form of a single image. While the dynamic gestures are the moving gestures represented by a sequence of images. To distinguish hand gestures and their meaning in sign language, researchers have suggested several methods, including the most used ones, for example methods based on gloves and vision. The glove based method

depends on electro mechanical apparatus .The Disabled individual wears gloves with wired connections and sensors. A digital camera such as that used in Internet applications is used to implement the second proposal which is known as vision based system for recording hand gestures, video clips for hands movement recorded represent isolated sign words [3].

The aim of this research is to propose a method for extracting the features of the group of images representing the isolated word of the Arabic sign language (ArSL) to extract the feature vector of **single hand** that represented by (**water word**) and **double hand (Happy ward)** by tracking the hands movement trajectory and extracting features founded by **DCT**. The **DCT** coefficients will be used as a feature vector and enter into the classification stage to extract the meaning of the word voice or text.

II. PREVIOUS WORKS

To make an image processing easy and expandable to handle images, the extraction of features is a high quality and essential method of reducing the dimension of the images. Additionally, to handle these images in a timely manner, the measurement of two images that entered through scouting about features is decreased. The process of alteration information to specific features is called extraction features. The gain of extracting two approaches that based totally on the direct pixel fee of the hierarchical centroid supplied in [4].Framework utilized frequent feature of area, and velocity confirmed in [5]. At the phrase degree used by the community of the deaf, successful and wonderful algorithm for the understanding of Arabic sign language was presented once in [6]. Perform a similarity matching using Euclidean space with the probability of the chain code the reference feature is very efficacious process for recognition proposed in [7]. Object detection involves finding contour of two states, one for static images and the other for video was performed in [8]. Arabic isolated sign language word feature extraction founded depending on model of chain code is proposed [9]. A fast model to extract hands form the video sequences and to generate features. Introduced in [10].

III. SIGN LANGUAGE RECOGNITION SYSTEM

The process of distinguishing the sign language used for understanding between people with hearing and speech disabilities requires the creation of a system based on vigorous image processing techniques. This system is qualified to identify a specific set of gestures and convert them to identical meaning in the form of text/audio .The system consists of three parts to carry out the process of recognition : **pre-processing, extraction of features** and **classification** for the completion of the sign language translation process . In pre-processing, important areas in the physique concerned in sign language work like palms and face are distinctive from gathered video. Diverse aspects from videos or images are extracted to characterize the feature vector that will be used as input to the classifications stage later. Part of the images representing the collected data will be used as test images and compared with the images stored in the last stage (classification) following the recognition training. A schematic of the method of sign language recognition is mentioned in FIG (1) below.



FIG (1) BLOCK DIAGRAM OF SIGN LANGUAGE RECOGNITION SYSTEM

A. Pre-processing

Using a camera device, the required data were collected from the surrounding medium and in the form of isolated words. The images are processed initially for captured gestures to detect hands and face, and then isolating the objects previously detected from the background. The images are then converted from RGB to Gray scale. Next getting a grayscale image, various techniques are applied to perform operations for image enhancement, insulating the image from the background and image manipulation to eliminate noise and reveal areas of interest such as hands shapes in the image.

B. Feature extraction

After hand segmentation from the image different features of hand that are used to identify gesture. The features extracted from the hand represent what is crucial to the classification process. By following the trajectory of the hand, many features such as direction, speed, angles, color, shape, space, length and Euclidean distance can be extracted. Input data are converted to a set of features or miniature exemplification features. When input data for the algorithm are too large to handle and apparently they are redundant (too much data, but not much information) [11]. In this case, to extract features of images a (DCT) has been applied.

C. Classification

Features can be categorized. After extracting features from the previous phase as mentioned above. Multiple classification techniques can be used to accomplish this function, for example, but not limited neural network[12], Support Vector Machine (SVM) [13] hidden markov model (HMM) [14] utilized to complete the sign language recognition process.

IV. DISCRETE COSINE TRANSFORM (DCT)

DCT is closely related to the Discrete Fourier Transform (DFT) uses only real numbers so the result is real. DCT head for concentrate information, making it useful for many sciences and engineering applications to compress images and distort audio and serve in reducing vector size parameters for various applications as well. DCTs are approximately tantamount to DFTs of twice the length, Works on real data even with symmetry, but in some variations the input and / or output data are transferred by half the sample. It proposes a simple and effective system for recognizing hand gestures in an uncomplicated background using DCT, Haar, Walsh and Kekres transformations. DCT provides a higher recognition rate than PCA. [15]

In terms of a sum of cosine functions oscillating at different frequencies, a discrete cosine transform expresses a set of data points restricted. Converts the signal or image from the spatial domain by DCT to the frequency domain. The general 2D equation (N by M image) DCT Defined by the equation below [2].

$$F(u,v) = \sqrt{\frac{2}{N}} * \sqrt{\frac{2}{M}} \sum_{i=0}^{N-1} A(i) * \cos\left(\frac{u(2i+1)\pi}{2N}\right) * \sum_{j=0}^{M-1} A(j) * \cos\left(\frac{v(2j+1)\pi}{2M}\right) * f(i,j) \quad (1)$$

Where $f(i,j)$ is the 2D input sequences.

$$A(i) = A(j) = \frac{1}{\sqrt{2}}, \text{ for } u=0, v=0$$

=0 Otherwise

V. PROPOSED METHOD

In this part of the research a description of the proposed method for extracting the features of isolated sign words is displayed. Depending on the Discrete Cosine Transform, the features from the videos that representing the sign language are extracted. Figure (2) presents the block diagram for this method. The proposed system consists of 3 stages:

- (1) Pre-processing.
- (2) Feature extraction based on DCT.
- (3) Classification.

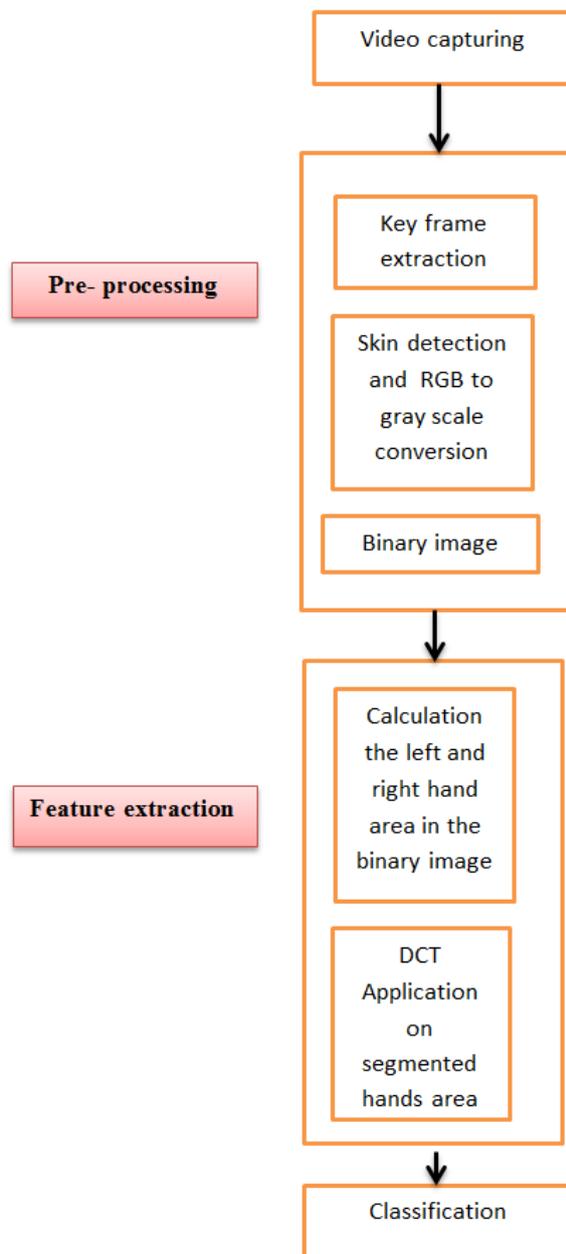


FIG (2) BLOCK DIAGRAM OF THE PROPOSED METHOD

A. Preprocessing

1. Key frame extraction.

By using the key frame, the main content of video data can be clearly represented and the amount of space available for video data analysis and processing reduced considerably. This will make processing, retrieving and identifying video information easier and more productive. This reduces the duplicate frames and leads to getting vector features which are used in the classification phase. These ten frames represent the isolated sign word from the start of the movement to the end, and therefore are processed effectively and efficiently [16]. An isolated words video (**water and happy**) words used

as real examples after reading the collected videos as an input. The first word represents the word sign, performed with one hand, and the second word is performed with two hands. Using Hamming distance, the ten most important key frames are extracted and identified from the total number of frames included in the isolated sign word. **Water words** and a **happy word** in the same way were taken into account as equivalent key frames for each video representing these signs shown in Fig(3-a&c). Gray images are obtained from converting (RGB) images from the previous step, and a binary image was obtained as shown in Fig(3-b&d).

2. Skin detection

After completion of the previous step, the skin is exposed to the important areas of treatment which are hands and face. For color division of the skin, the color space used is YCbCr. In this illustration, Y indicates to the brightness. Cb, Cr are values of the chrominance for blue light and red light. YcbCr is preferred because the reality is that the skin color can be recognized by the chrominance component and the brightness value is isolated from the coloring values. The practical values for Cb and Cr are: Cb = [77 127] and Cr = [133 173] [17]. Binary Images of Associated areas representing the mass of the palm are the result at that point.



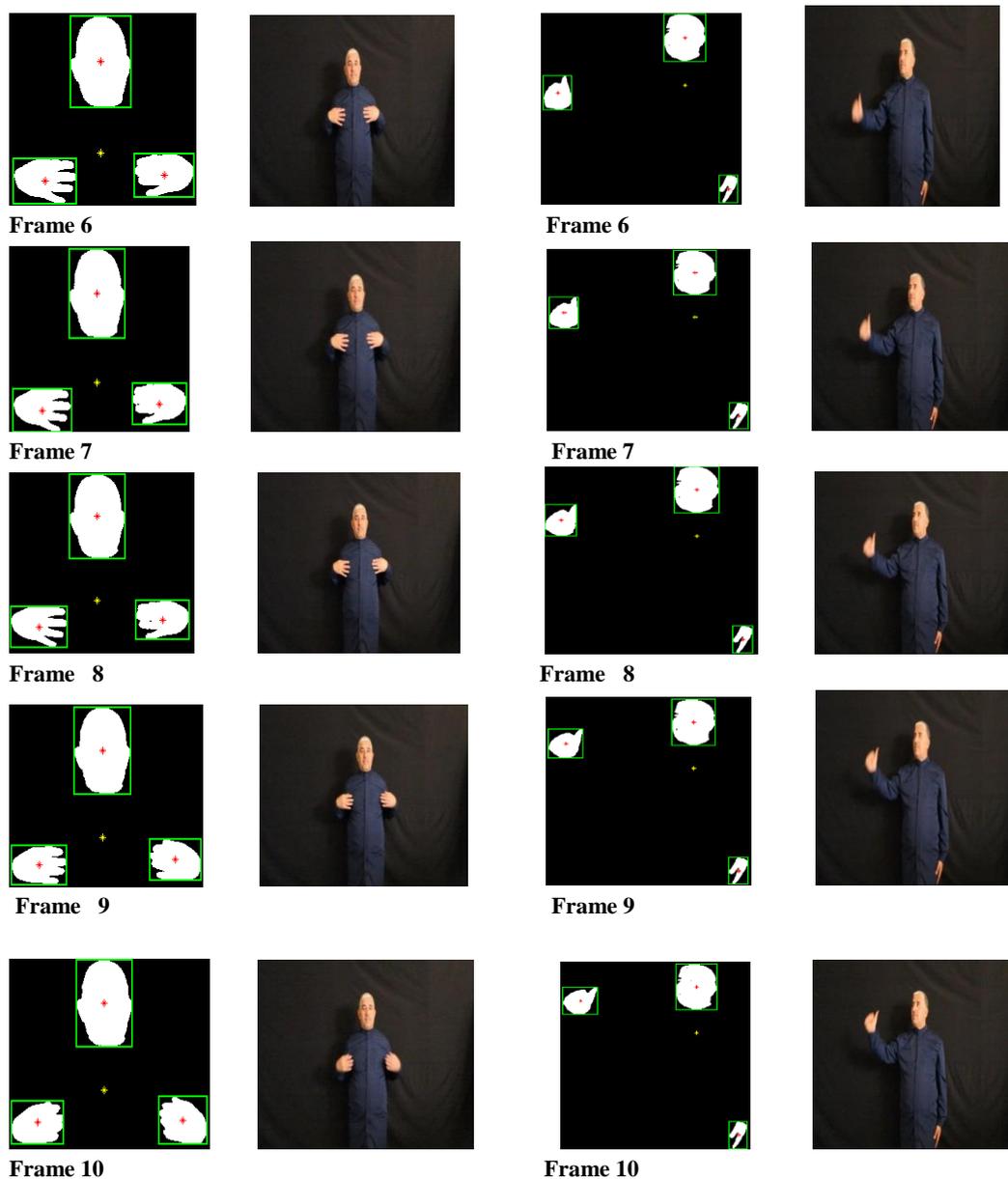


FIG (3) [A- KEY FRAMES EQUIVALENT OF WATER WORD ,B-THE BINARY IMAGES AND SEGMENTED AREAS , C- KEY FRAMES EQUIVALENT OF HAPPY WORD, D- THE BINARY IMAGES AND SEGMENTED AREAS]

B. Feature extraction based on DCT.

After Hamming distance has obtained the key frame, as described above. Body parts centers are critical for completing the follow-up and discrimination cycle such as the center of the right hand (RH) and center of the left hand (LH) as well as the center of the surrounding face, as shown in the bounding box in Figure(3). The next step is to start tracking the movement of the arms from the start to the end of the movement, which means, for example, the isolated word (water word & happy word). It has been noticed that each frame consists of three fragmented areas, namely the area of the right hand, the left hand area and the area of the face. Now treating the **right and left hands** segmented area as a discrete event, the DCT is applied on the 10 extracted key frames. Coefficients of DCT congruent to each segmented

region are known as features. A selection of DCT coefficients is used as a distinctive route for classifying the movements from hand-segmented image areas.

VI. EXPERIMENTAL RESULTS

In this part of the research, there are guiding words, for People with hearing disabilities, deaf and dumb have been addressed and studied. Such words were collected in Arabic, Iraqi dialects and character phrases in collaboration with the Iraqi Ministry of Labor and Social Affairs, with the aid of the Special Needs Department. More than 40 important words were mentioned in the deal. The performance time of the representational motion of the isolated sign word varies between (2-3sec) depending on the speed of the performance and the volunteer person. In other words , two separate terms (water.mp4, happy.mp4) are used as examples. The suggested approach was applied using the Matlab 2013b program, by incorporating the isolated words of the deaf and dumb into the framework. The word **water** and the word **happy** were taken as examples to extract the special features of these words by applying the **DCT**. **Table (1)** illustrates the variation of DCT coefficient of the (**Happy sign word**) , in the same manner the variation of DCT coefficient of the (**Water sign word**) is shown below in **Table(2)**. **Figure (4)** shows the variation of DCT coefficient of the **Happy word**. **Figure (5)** shows variation of DCT coefficient of the **Water word**.

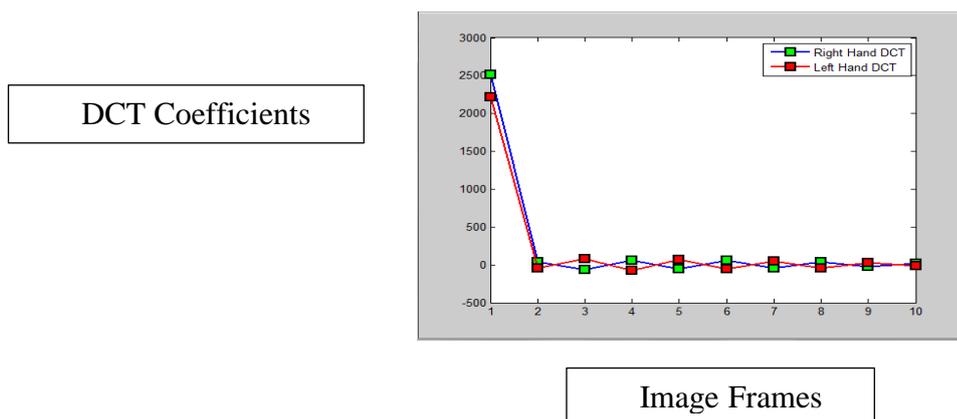


FIG (4) VARIATION OF DCT COEFFICIENT OF THE HAPPY WORD

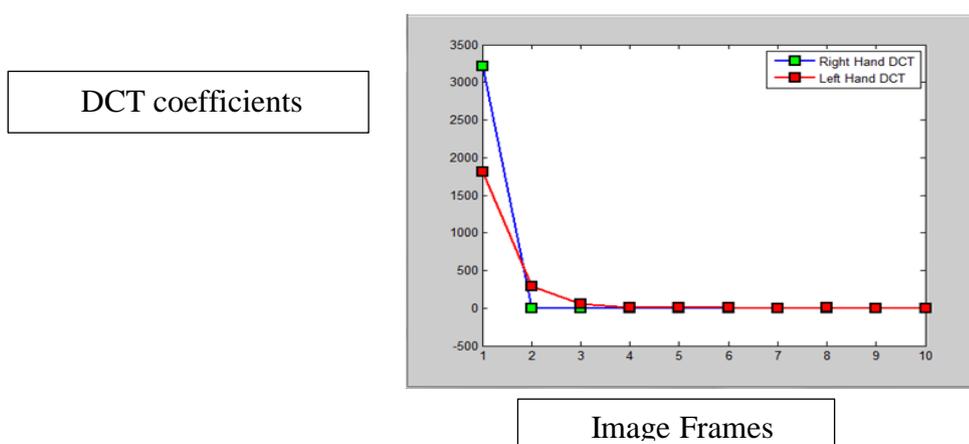


FIG (5) VARIATION OF DCT COEFFICIENT OF THE WATER WORD

Received 29 May 2019; Accepted 10 Dec 2019

TABLE (1) THE VARIATION OF DCT COEFFICIENT OF THE (HAPPY SIGN WORD)

	Frame 1	Frame 2	Frame 3	Frame 4	Frame 5	Frame 6	Frame 7	Frame 8	Frame 9	Frame 10
Area RH	792	797	800	804	810	814	818	820	833	663
Dct-RH * 10 ³	2.5143	0.0332	-0.0698	0.0622	-0.0599	0.0528	-0.0478	0.0364	-0.0266	0.0138
Area LH	698	695	692	689	686	679	675	669	662	850
Dct-LH * 10 ³	2.2120	-0.0434	0.0760	-0.0731	0.0692	-0.0591	0.0511	-0.0400	0.0268	-0.0125

TABLE (2) THE VARIATION OF DCT COEFFICIENT OF THE (WATER SIGN WORD)

	Frame 1	Frame 2	Frame 3	Frame 4	Frame 5	Frame 6	Frame 7	Frame 8	Frame 9	Frame 10
AreaRH	1013	1014	1015	1015	1015	1016	1016	1016	1016	1017
Dct-RH * 10 ³	3.2107	-0.0032	-0.0007	-0.0008	-0.0002	-0.0009	0.0002	0.0001	0.0002	-0.0002
Area LH	733	696	656	612	571	535	505	484	472	465
Dct-LH * 10 ³	1.8117	0.2863	0.0525	0.0135	0.0080	0.0047	0.0025	0.0030	0.0015	0.0009

VII. CONCLUSION AND FUTURE WORK

The dynamic sign language feature extraction of the isolated Arabic sign word is given in this paper based on the DCT. Some of the signs used in current database are presented in Table (3) representing a single hand sign examples, in the same manner Table (4) represents a double hand sign examples for different signers. Furthermore, the aim of using this method is to extract the features of the sign word represented by the movement of one hand as in the **water** word or the movement of both hands as in the **happy** word for people with **Speech** and hearing **impairment** based on DCT. Then using the features of the right and left hand to classify the term isolated sign as an input vector to the classification stage. Sign language patterns will be developed in our future work and the extracted features will be used to construct a sign recognition system for isolated sign words that are efficient and reliable in use.

TABLE (3) SINGLE HAND SIGN EXAMPLES

Sign No.	Frame-No.1	Frame-No.5	Frame-No.10	Arabic presentation of the sign
1				ماء
2				طبيب
3				السلام عليكم
4				مااسمك

5				نقود
---	---	---	--	------

TABLE (4) DOUBLE HAND SIGN EXAMPLES

Sign No.	Frame-No.1	Frame-No.5	Frame-No.10	Arabic presentation of the sign
1				صباح
2				مساء
3				سعيد
4				شارع
5				كيف حالك

ACKNOWLEDGMENT

In conclusion, thanks are extended to the Department of Electrical Engineering at the University of Technology, represented by the staff of the post graduate studies laboratory. Additionally, special thanks to the Iraqi Ministry of Labor and Social Affairs, the Department of People with Special Needs, represented by the Al-Izdihar Institute for Special Needs, for their help in collecting research data.

REFERENCES

- [1] Paulraj M P , et, al. " Extraction of Head and Hand Gesture Features for Recognition of Sign Language" Published in: 2008 International Conference on Electronic Design.
- [2] Divyashree Shinde , et,al. " Sign Language Interpreter", International Journal of Engineering Research & Technology (IJERT) , Vol. 3 Issue 5, May – 2014.
- [3] Pratibha Pandey1, Vinay Jain2, " An Efficient Algorithm for Sign Language Recognition", (IJCSIT) International Journal of Computer Science and Information Technologies, Vol. 6 (6) , 2015, 5565-5571
- [4] Madhuri Sharma, Ranjna Pal and Ashok Kumar Sahoo, " INDIAN SIGN LANGUAGE RECOGNITION USING NEURAL NETWORKS AND KNN CLASSIFIERS" ,ARPN Journal of Engineering and Applied Sciences, VOL. 9, NO. 8, AUGUST 2014
- [5] Elmezain , Mahmoud, et al. "Spatio -temporal feature extraction based hand gesture recognition for isolated American Sign Language and Arabic numbers." Image and Signal Processing and Analysis, 2009. ISPA 2009. Proceedings of 6th International Symposium on. IEEE, 2009.
- [6] Mahmoud Zaki Abdo, Sameh Abd El-Rahman Salem, Alaa Mahmoud Hamdy , Elsayed Mostafa Saad, " EMCC: Enhancement of Motion Chain Code for Arabic Sign Language Recognition", (IJACSA) International Journal of Advanced Computer Science and Applications, Vol. 6, No. 12, 2015..
- [7] Dolly Indra#, Sarifuddin Madenda*, Eri Prasetyo Wibowo*, "Recognition of Bisindo Alphabets Based on Chain Code Contour and Similarity of Euclidean Distance", International Journal on Advanced science Engineering information Technology, Vol.7 (2017) ,No,5.

Received 29 May 2019; Accepted 10 Dec 2019

- [8] R S Vaddi¹, L N P Boggavarapu¹, H D Vankayalapati², K. R. Anne¹, "CONTOUR DETECTION USING FREEMAN CHAIN CODE AND APPROXIMATION METHODS FOR THE REAL TIME OBJECT DETECTION", Asian Journal Of Computer Science And Information Technology 1:1 (2011) 15 – 17
- [9] Hussein Ali Aldelfy, Mahmood Hamza Al-Mufraji, Thamir R. Saeed" "An Efficient Feature Extraction of Isolated Word For Dynamic Sign Language Classification", Third Scientific Conference of Electrical Engineering, UOT ,IRAQ,2018.
- [10] Sunita Ravi^{1,2}, M. Suman³ and P. V. V. Kishore¹" VIDEO BASED INDIAN SIGN LANGUAGE RECOGNITION USING BLOCK ZIG-ZAG DCT FEATURES AND MAHALANOBIS DISTANCE CLASSIFIER", ARPN Journal of Engineering and Applied Sciences, VOL. 12, NO. 16, AUGUST 2017
- [11] Kavitha Jaganathan,et,al "SINGLE HAND GESTURE RECOGNITION BASED ON DWT AND DCT FEATURE EXTRACTION AND NEURO-FUZZY CLASSIFIER",E-Proceeding of the International Conference on Social Science Research, ICSSR 2015.
- [12] Oz, Cemil, and Ming C. Leu. "Linguistic properties based on American Sign Language isolated word recognition with artificial neural networks using a sensory glove and motion tracker." *Neurocomputing* 70.16 (2007): 2891-2901.
- [13] Aujeszky, Tamás, and Mohamad Eid. "A gesture recognition architecture for Arabic sign language communication system." *Multimedia Tools and Applications* (2015): 1-19.
- [14] Ding, L. and A.M. Martinez, "Modelling and recognition of the linguistic components in American Sign Language". *Image and Vision Computing*, 2009. 27(12): p. 1826-1844
- [15] Shivashankara S, Srinath S" A comparative Study of Various Techniques and Outcomes of Recognizing American Sign Language: A Review", *International Journal of Scientific Research Engineering & Technology (IJSRET)*, ISSN 2278 – 0882 ,Volume 6, Issue 9, September 2017.
- [16] Hussein Ali Aldelfy, Mahmood Hamza Al-Mufraji, Thamir R. Saeed, "Improved Key Frame Extraction Using Discrete Wavelet Transform with Modified Threshold Factor", *TELKOMNIKA*, Vol.16, No.2, April 2018, pp. 567~572.
- [17] D. Chai and K. N. Ngan, "Face segmentation using skin-color map in videophone applications," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 9, pp. 551–564, 1999.