Convert Gestures of Arabic Words into Voice

Shaker K .Ali, Ali Al-Sherbaz, Zahoor M. Aydam

Abstract: Gestures are one of the best ways of communication between dumbs and other people using the expression of signs language. In this paper, we suggest an algorithm for recognizing hand gestures of Arabic words (اتمنى لك حياة سعيدة-اقتباس) to by using dumb (through signs) and convert the sings into voice corresponding to sings words. The proposed algorithm for Convert Gestures of Arabic Words into Voice , record video of gesture (of the dumb person) then convert the video into frames (images), preprocessing for the resulted image must done by remove the noise, resize the images and increase the contrast, then calculate the distance to clustering the words by using (C4.5 , k-mean , k- medoid and artificial neural network), calculate the distance (or features) by using Euclidean distance and slope where ,there are eighteen features (eight features from Euclidean distance, eight features from slop, Area, and perimeter). The results in the training stage were (C4.5 gave 100%, k-mean gave 95.2% k-medoid gave 91.9% and ANN gave 91.27%). While in the testing stage we used three classifiers (Euclidian Distance, Modify of the Standardize Euclidian Distance and Correlation) and the results show that (Euclidian Distance gave 94.4%,Modify of the Standardize Euclidian Distance gave 100% and Correlation gave 94.4%) We create our database (three videos with 250 frames) for training and one video for testing.

Keywords: Gestures, Feature Extraction, C4.5, K-Mean , K-Medoid and ANN

1. Introduction

Communication is the way for expression about thoughts, opinions, information, or messages between the people by writing, speaking, or signs. Communication is usually oral expression between people by talking to each other while people dumb cannot communicate with others as ordinary people do, they can't speaking people who are deaf are able to speak, but they unable to hear. While the blind are unable to see but they can speaking and listen [1]. The gesture is a kind of nonverbal communication with a part of the body, which used together with verbal communication. The gestures are obscuring not totally specific. Like the talk and handwriting, gestures change from individual to individual, even to the same person in different cases [2].A gesture is a language used by dumb people. Dumb people use signs to show their ideas. Gesture language is different from each country to another country with its special vocabulary and grammarian. In fact, gesture language can vary in one country from one place to another, as Languages spoken [3]. The gesture is the movement of any part of the body such as the face and hands a kind of motion [4]. There are two methods for recognizing the gesture; the first way is based glove and the second was based on computer. The first way depends on the hardware and gets information from the joints of the hand by using sensors to know the classification of hand gesture. This way use video and convert the video into frames to identify the pattern they know the hand gestures [5].

Recognize of gesture language at present, by the token gesture of humans using video camera such as a mobile, tablet, special camera, or laptop camera [6] then convert the video into the image and extract the features then classify each number into voice, this paper focuses on the how the gesture language translate into voice to make the dumb communicate with other people through voice . The Arabic words gesture as shown in Figure 1.

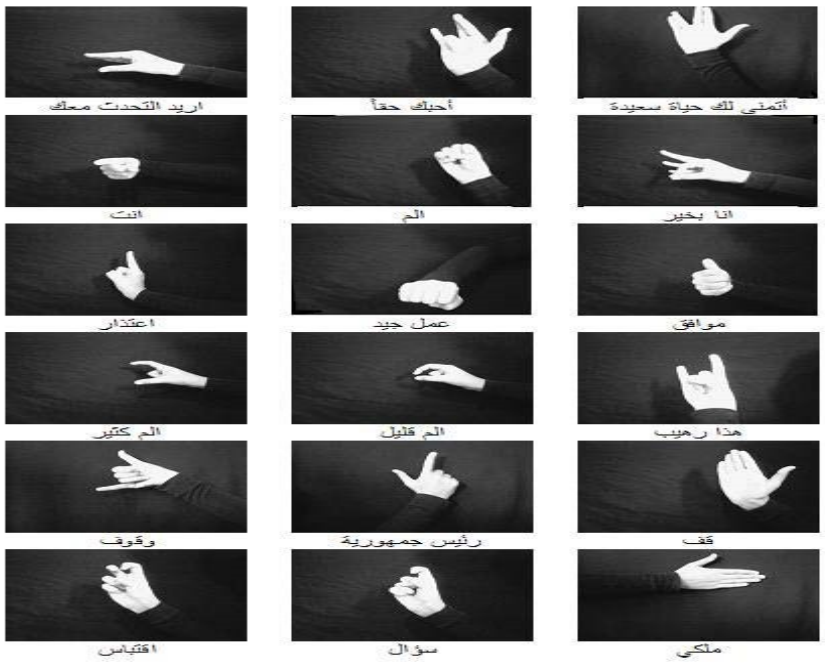

Figure 1. Gesture of Arabic Words

2.Clustering Algorithms and Classification Algorithms

There are many algorithms for clustering and classification, in our algorithm we tried to use the C4.5, K-mean, K- Medoid algorithms and ANN the result from our experiments shows that C4.5 is the best one and high accuracy. **C4.5 algorithm**

C4.5 is a standard algorithm for inducing classification rules in the form of the decision tree. As an extension of ID3, the default criteria of choosing splitting attributes in C4.5 is information gain ratio instead of using information gain as that in ID3, information gain ratio avoids the bias of selecting attributes with many values[7]. C4.5. Algorithm steps [7]: Check the basic cases.

For each calculate features :(Acquire the normalized information from the division on an attribute X).

Select the best features that have the highest gain for information.

Create a node is divided by the best decision point, such as the root node.

Repeated the sub-menus obtained by splitting on the best a and adding those nodes as the children node.

3.Proposed Algorithm

The proposed algorithm consists of four steps (images acquisition, preprocessing step, features extraction, classify (in training) or comparison (in testing) and convert into voice). as show in Figure 2.

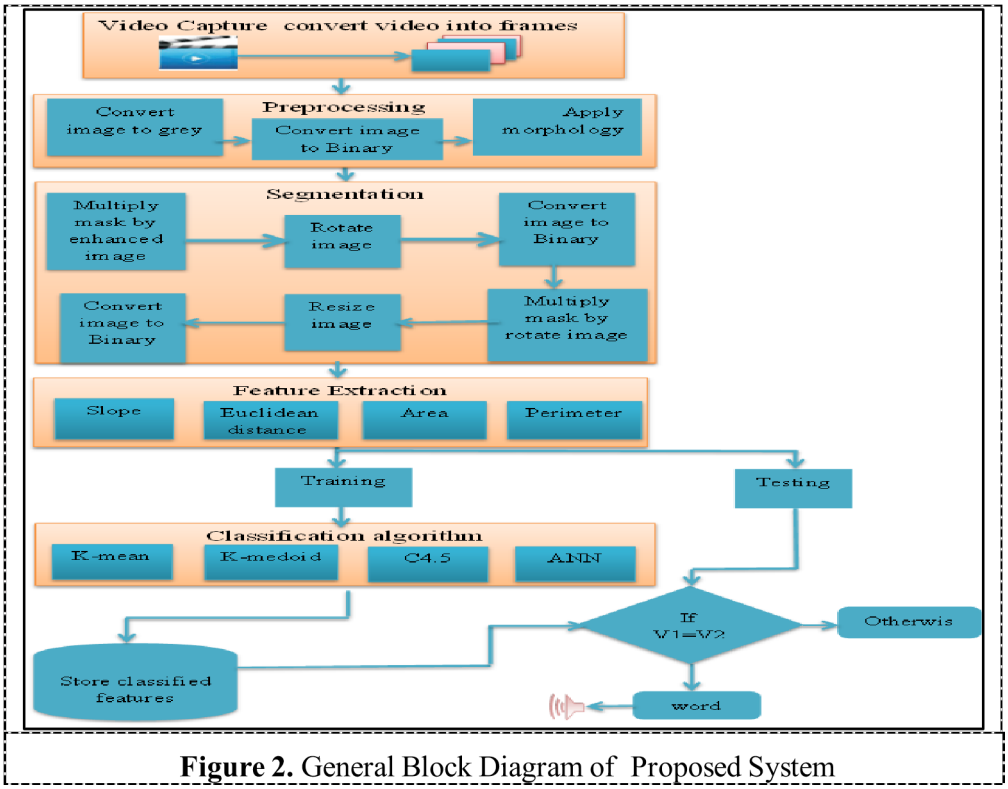

Figure 2. General Block Diagram of Proposed System

3.1 Dataset Aquisyion

We created our dataset by using an external camera in the laboratory by using different cameras to create our dataset. The background of images must be black color to be easy for classify objects (the image must contain only sign part). The position of the camera is also important issues to remove the background and keep the sing only.

Here we took eighteen words in Arabic words of three different persons, where these words will be as different templates. The videos are contains a series of frames (images) with size (720*1280 pixels) and (640 * 480 pixels). In this paper we used three videos contains 123 frames (images) with (720*1280 pixels) and 127 frames (images) with size (640 * 480 pixels) (for training stages).

3.2. Pre- Processingre

The Pre-processing includes the following steps:

Transform the videos into the required frames (images $_k$)

Convert the image (images k) to gray scale format then convert the resulted image into a double image the resulting image is an enhanced image.

Transform the enhanced image (images $_{k)}$ into a binary image.

Remove little objects from the binary image using morphological operations (Dilation and closing).

3.3 Segmentation

Segment the palm area from the resulting image of the morphological operations (Dilation and closing).

Use the resulted image as a mask

Multiply the enhanced image by mask.

Calculate the angle according to the following Equation (1) [8].

$$\theta = \tan^{-1}((t_1 - t_2)/(1 + t_1 * t_2)) \qquad (1)$$

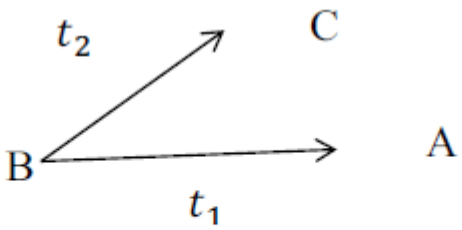

Figure 3. Slope

$t_1$is the slope between B and A, t_2 is the slope between B and C.

Rotate (image $_k$) according to the following equation (2) [9].

$$\left.\begin{aligned}\hat{z} &= z * cos(\theta) - w * sin(\theta)\\ \hat{w} &= z * sin(\theta) - w * \cos(\theta)\end{aligned}\right\} \qquad (2)$$

(z, w) and (z,w) : are pixel coordinates before and after rotation, respectively,

θ : is the counter clockwise angle of rotation.

Convert the image (image $_k$) to a binary image and segment the palm area (used as mask). Multiply rotate (image $_k$)with a palm-sized mask. Resize the image [any * 100].

Convert the image (Image $_k$) to a binary image.

3.4 Feature Extraction

In our proposed algorithm there are 18 geometrics features for each frame (image) these 18 features divided in (8-features) calculated from the distance between the 8 points within the center of hand, (8features) calculated from the slop of the same 8 points in first (8-features), one feature calculated from area and one feature calculated from perimeter. We can calculate the 18-features by using as following steps:

Extract 8 points of palm.

Calculating the 8-features include the distance of 8 points by using Euclidean distance from the center of palm to the 8- points according to Equation (3) [10], and calculate the length of the palm and divide the distance by the length of the palm

$$DE_{zo} = \sqrt{\sum_{i=1}^{n} (x_{zi} - x_{oi})^2} \quad (3)$$

Where as:

(.

n: Number of properties

DE_{zo}: Distance between points and center of palm

x_{zi} : The coordinates of the i property for Z (where Z: points)

x_{oi}: The coordinates of the i property for o (where o: center point of palm)

The second 8-features include the slop from center of palm to the 8-points as according to Equation (4) [8].Then calculate (tan) of the slope.

$$Slop = \frac{yz - yo}{xz - xo} \quad (4)$$

Where as:

yz: the y- axis points

x_z: the x- axis points

y_o: the y- axis of center value point of palm

x_o: the x- axis of center value point of palm

Calculate center point of palm as in Equation (5) [11].

$$x_o = \frac{\sum x_{oi} A_i}{\sum A_i}, \qquad y_o = \frac{\sum y_{oi} A_i}{\sum A_i} \quad (5)$$

Where:

x_o: the x- axis of center value point of palm

y_o: the y- axis of center value point of palm

x_{oi}: The distance at which the center of the shape moves away from the junction point of the axes on the axis (x)

y_{oi}: The distance at which the center of the shape moves away from the junction point of the axes on the axis (y)

A_i: Area the shape

Calculate area of palm as according to Equation (6) [12], and divided the results by 10000 (to reduce the big numbers).

$$A = \frac{1}{2}\sum_{i=0}^{n-1}(x_i \times y_{i+1}) - (x_{i+1} \times y_i) \qquad (6)$$

n: Number of points

x_i : x- axis coordinates points

y_i : y- axis coordinates points

Calculate Perimeter of palm as according to Equation (7) [13] and divided the results by 500.

$$Per = \sum_{i=0}^{n-1} x_i \qquad (7)$$

n: Number of ribs

x: length of the rib

Calculate the feature vector for each word (18 words) by using two steps:

Calculate the average of features for the same words from different images of the same word.

Calculate the feature vector for the first image of the word and ignore the rest images of the same word.

4. Result

The proposed algorithm contains two parts; one for training with 250 images while the second part is for testing by using 18 images as shown in Figure 2. In training part we need to calculating 18 features using C4.5 algorithm for classify the 18 types of words for each image as shown in Table 1 for distance, slop, area and perimeter respectively, then calculate the features for each words 18 types words as shown in Table 2 (the average of features for the same words from different images of the same word) and Table 3 shows the feature vector for the first image of the word and ignore the rest images of the same word. When we used three videos with 250 frames (images) we found that the results from four clustering algorithms; K-mean, K-mediod, C4.5, and ANN, for 18 words which gave different results for recognition as shown in Table 4 as following:

When we Implement of the k-mean cluster algorithm on the extracted features we found the accuracy of K-mean is 95.2000% , K-medoid is 91.9111% , C4.5 is 100% and ANN is 91.2727%for training stage when the dataset is 250, as shown in Table 4 and Figure 3 respectively.

Table 1 .Features vector of word

Feature Geometry																	
Distance								Slope								Area	Perimeter
D1	D2	D3	D4	D5	D6	D7	D8	S1	S2	S3	S4	S5	S6	S7	S8		
0.1610	0.1808	0.1383	0.1376	0.2304	0.2325	0.2108	0.2205	-0.6508	-0.7832	-1.4638	-1.5361	-0.1267	-0.1836	1.1912	1.0927	0.4891	0.7902

Table 2 .Features vector of average 18 words

word	Number of image	Feature Geometry																	
		Distance								Slope								Area	Perimeter
		D1	D2	D3	D4	D5	D6	D7	D8	S1	S2	S3	S4	S5	S6	S7	S8		
وقوف	24	0.2361	0.2339	0.1676	0.1595	0.2491	0.2548	0.2377	0.2542	0.3185	0.2368	-1.1806	-1.0451	-0.3602	-0.4138	0.9003	0.8236	0.5866	1.0113
ملكي	17	0.1652	0.1683	0.1930	0.1967	0.1359	0.1250	0.1524	0.1573	-0.1445	-0.2421	1.3325	1.2636	0.6734	0.2277	-0.5935	1.2345	0.4259	0.6585
قف	16	0.2506	0.2499	0.2213	0.2201	0.2248	0.2295	0.1393	0.1180	0.0753	-0.0249	-1.4027	-1.0462	-0.2291	-0.3086	-0.9949	1.0926	0.7258	0.8587

Table 3. Features vector for the first image of the word

word	Number of image	Feature Geometry																	
		Distance								Slope								Area	Perimeter
		D 1	D2	D3	D4	D5	D6	D7	D8	S1	S2	S3	S4	S5	S6	S7	S8		
وقوف	24	0.2596	0.2485	0.1807	0.1739	0.2561	0.2643	0.2455	0.2594	0.7512	0.7022	-1.0083	-1.0732	-0.3891	-0.4592	0.8261	0.7694	0.5951	1.0999
ملكي	17	0.1704	0.1750	0.2005	0.2041	0.1399	0.1379	0.1492	0.1617	-0.2302	-0.3230	1.2279	1.1818	0.6577	0.6385	-1.3173	1.1052	0.4411	0.6786
قف	16	0.3339	0.3333	0.2114	0.2072	0.3048	0.3082	0.1678	0.1436	0.0608	-0.0091	-1.1139	-1.1573	-0.1768	-0.2301	-1.0274	-1.5702	1.2197	1.0646

Table 4 .shown the difference between four algorithms rate

word	No image	C4.5	K-mean	K- mediod	ANN
	250	100%	95.2000%	91.9111%	91.2727%

As shown in Table 4 the C4.5 algorithm is the best algorithm in the classification and accuracy

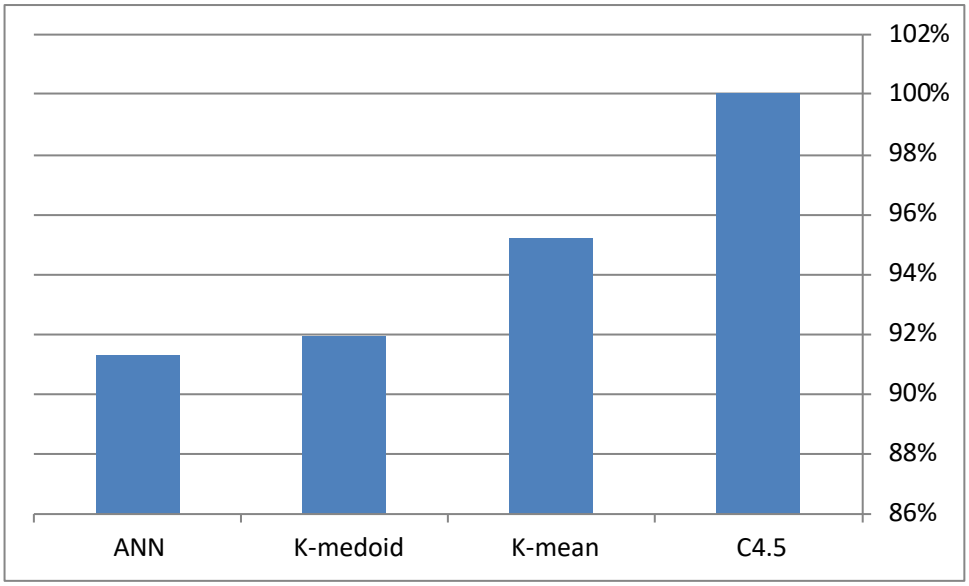

Figure 3. shown the difference between Four algorithms rate

In testing stage the features will classify where the input image will be in cluster or class of 18 types of words by comparing the feature vector with the 250 vectors stored in the dataset then the result will convert the class type or cluster type into corresponding voice (words) as shown in Figure 2. The testing of our algorithm is done by using Equation (8)[14] which is the modify of the Standardized Euclidean distance , by using Equation (3) of the Euclidean distance and also by using Equation (9) [15] of the correlation to compare the new features of input image with the classified features database of images.The results shows the accuracy of three ways (Modify of the Standardized Euclidean distance, Euclidean distance and Correlation) in Tables 5 and Figure 4

$$Dst = \sqrt{\sum_{i=1}^{n} \frac{(x_i - y_i)^2}{\sqrt{\frac{1}{(n-1)}\sum_{j=1}^{n}(x_j - \bar{x})^2 + (y_j - \bar{x})^2}}} \qquad (8)$$

Where :

x_i :is the i^{th} valuse of first vector value.

y_i :is the i^{th} valuse of second vector value.

n : is the number of elements in vector.

$\bar{x}$: is the mean value of first and second vector

$$R1 = \frac{\sum_i (x_i - x_m)(y_i - y_m)}{\sqrt{\sum_i (x_i - x_m)^2}\sqrt{\sum_i (y_i - y_m)^2}} \qquad (9)$$

Where xi is the intensity of the ith value in vector 1, yi is the intensity of the ith value in vector 2, xm is the mean intensity of vector 1, and my is the mean intensity of vector2.

Table 5. the result accuracy of by using three ways

No. of tested images	No. of dataset images	Accuracy by using		
		Euclidean distance	correlation	Modify of the Standardized Euclidean distance
18	250	94.4444%	94.4444%	100%

The result shows that Modify of the Standardized Euclidean distance is the best accuracy from Euclidian distance and the correlation.

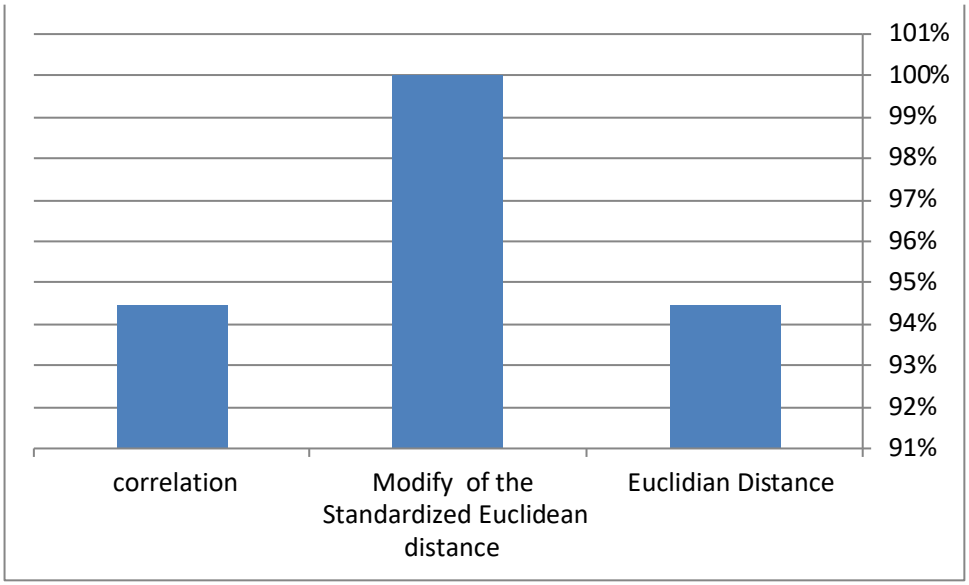

Figure 4.shows the accuracy rate using (Euclidian Distance ,Correlation and Modify of the Standardized Euclidean distance

5.Conclusion

In this paper, we have designed a system for recognition of Arabic words for gesture language based on clustering methods. In our experiments, we found that the geometric features (distance, area, perimeter, and slope) are the good eatures rather than the others features such as (shape, texture, color,…..etc). There are many clustering and classify algorithms used in our proposed algorithm such as (K-mean, K-medoid, C4.5, and ANN) where the experiments found that C4.5 algorithm is the best one in clustering or classify with the percentage of(100%) in training and (100%) in testing. In the testing stage, the results found that the modify of the Standardized Euclidean distance is best metric for calculating the corresponding features vector of the tested image to know the type of which words (18 words) while others metrics such as (Euclidean distance and Correlation) is lower accuracy than the Modify of the Standardized Euclidean distance.

References

[1] Ch. V.N. Syam Babu, V.J.K. Kishor Sonti and Y. Varthamanan 2016 *Design and Simulation of Communication Aid for Disabled Using Threshold Based Segmentation* (I J C T A Vol 9,No7) pp. 3275-3281.

[2] Mohamed S. Abdalla and Elsayed E. Hemayed 2013 *Dynamic Hand Gesture Recognition of Arabic Sign Language using Hand Motion Trajectory Features*(Global Journal of Computer Science and Technology Graphics & Vision Vol 13, Issue 5)pp26-33.

[3] Hemina Bhavsar and Dr. Jeegar Trivedi 2017 *Review on Classification Methods used in Image based Sign Language Recognition System*(International Journal on Recent and Innovation Trends in Computing and Communication Vol5 ,Issue 5)pp 949 – 959.

[4] Kumud Tripathi ,Neha Baranwal and G. C. Nandi *2015Continuous Indian Sign Language Gesture Recognition and Sentence Formation*(Procedia Computer Science Vol 54)pp 523 – 531.

[5] Liu Yun, Zhang Lifeng and Zhang Shujun 2012 *A Hand Gesture Recognition Method Based on Multi-Feature Fusion and Template Matching* (Procedia Engineering Vol 29)pp 1678 – 1684.

[6] Pablo Barros, Nestor T. Maciel-Junior , BrunoJ.T. Fernandes , ByronL.D. Bezerra and Sergio M.M. Fernandes *2017A dynamic gesture recognition and prediction system using the convexity approach*(Computer Vision and Image Understanding Vol 155)pp139-149,2017.

[7] Wei Dai and Wei Ji 2014 *A Map Reduce Implementation of C4.5 Decision Tree Algorithm*(international journal of database theory and application Vol 7,No 1)pp.49-60..

[8] Christopher Clapham and James Nicholson 2009 *Oxford Concise Dictionary of Mathematics*(OUP oxford).

[9] Hermann K. 2011 *Real-Time Systems Design Principles for Distributed Embedded Applications*(Springer) Second edition.

[10] Michel Marie Deza and Elena Deza 2009 *Encyclopedia of Distances*(Springer)pp 94.

[11] Dan B. Marghitu and Mihai Dupac 2012 *Advanced Dynamics*(Springer)Chapter 2 pp 73-141.

[12] H. Stroud 1971 *Approximate calculation of multiple integrals*(Prentice-Hall Inc., Englewood Cli_s, N. J.).

[13] Dr Yeap Ban Har,Dr Joseph Yeo,Teh Keng Seng,Loh Cheng Yee,Ivy Chow,Neo Chai Meng and Jacinth Liew 2018 *NEW SYLLABUS MATHEMATICS TEACHER'S RESOURCE BOOK1*(OXFORD UNIVERSITY Press) 7th edition.

[14] Pavol OR.ANSKÝ 2009 *Fundamentals of Mathematical Statistics*(Slovakia: Statistics Faculty of Management, University of Presov)

[15] A. Miranda Neto, A. Correa Victorino, I. Fantoni, D. E. Zampieri, J. V. Ferreira and D. A. Lima 2013 *Image Processing Using Pearson's Correlation Coefficient: Applications on Autonomous Robotics*(IEEE).