ISSN: 0976-3104

SPECIAL ISSUE (SCMDSA)
Gupta et al.

ARTICLE OPEN ACCESS

SPEECH GUIDED FEATURE EXTRACTION AND BRAIN STORM OPTIMIZATION TO CLUSTER OBJECTS USING FUZZY LOGIC

Utkarsh Gupta*, Swarnalatha P, Prateek Chharia
**Department of SCSE, VIT University, INDIA*

ABSTRACT

In this paper a novel object recognition technique is proposed which is based on fuzzy clustering and Brain Storm Optimization Algorithm. The aim is to create classifiers based on clustered data. Object features are extracted from real time video frames guided by speech recognition. The proposed feature extraction works in two phases, first phase deals with extracting average pixel intensities of Red, Green and Blue channels respectively from the sample object image along with illuminance reading of Lux Meter and name of the object recognized by speech engine. These features are then stored as primary feature vector set. Second phase deals with extraction of keypoints using robust local feature detector algorithm called as SURF (Speeded-Up Robust Features) which will be stored as secondary feature vector set. FREAK (Fast Retina Key-point) descriptor has been combined with SURF detector algorithm for comparison with SIFT (Scale Invariant Feature Transform). Brain Storm Optimization helps in optimization and minimization of cluster distances. In our proposed technique we perform clustering using fuzzy C-Means and BSO only on primary feature vector set. The aim is to reduce keypoints matching time complexity. Computing distances like mahalanobis distance between primary feature vector and test object features will reduce the candidate rows of feature set. Applying keypoints mapping on fewer records will reduce the complexity of recognition algorithm. 65.8% reduction in time has been observed using this strategy over the conventional method of mapping keypoints of complete dataset with test object. Clustering algorithm has 86.9 per cent accuracy for the primary feature vector set consisting of 56 real time object data points.

Received on: 30th-Nov-2015
Revised on: 11th-March-2016
Accepted on: 29h – March-2016
Published on: 10th–June-2016

KEY WORDS

Object Recognition, Image Processing, Speech Recognition, Brain Storming Optimization

***Corresponding author:** **Email:** utkarsh.satishg2014@vit.ac.in **Tel:** +91-9944702033

INTRODUCTION

Multimodal recognition is one of the most important fields of robotics. Multimodal feature extraction will increase the accuracy of recognition. Suppose an image of pen is captured, then image processing is done to extract features. In this case system may get confused with other cylindrical objects. Similarly if a user says “It is pen”, and speech engine detects words, there is a possibility that acoustic model of system recognizes the word “pan” and not actually “pen”. Therefore multiple modalities are required to increase the accuracy of the system. Reconsidering the previous example using multimodal inputs where system captures image and user’s speech input, “It is pen”. In this case, system can parallely know that the object is cylindrical in shape and its name/category is pen, so it can filter the data set, and map cylindrical shape with keyword “pen”. Thus the accuracy will be really improved if we fuse the speech with real time video frame to classify object [1]. The proposed technique splits feature set into primary vector of features and secondary vector of features. Primary vector contains generic and less complex numerical data. Whereas secondary vector contains more complex data. Example: primary vector set contains average pixel intensities of various channels of image and secondary vector set contains more complex keypoints extracted from image (using SURF with FREAK). Separating feature vectors helps in reducing the overall matching time. A comparative study between SURF with FREAK and SIFT has also been conducted and results have been included in the experiments and discussions section.

MULTITHREADED SYSTEM ARCHITECTURE

Multithreaded architecture ensures parallel feature extraction and thus reduces overall running time. One thread uses speech engine for object’s name/category/type recognition and other thread does the image processing over real time video frames. This will decrease the computation time of the system. **Figure- 1** shows the flow of information representing multithreaded architecture. Speech recognition engine and video capture engine are working parallel. Speech engine converts speech to text using voce library. In case of video processing engine, camera continuously captures video frames, from which features are extracted. Features are extracted using the openCV libraries.

THE IIOAB JOURNAL
ISSN: 0976-3104

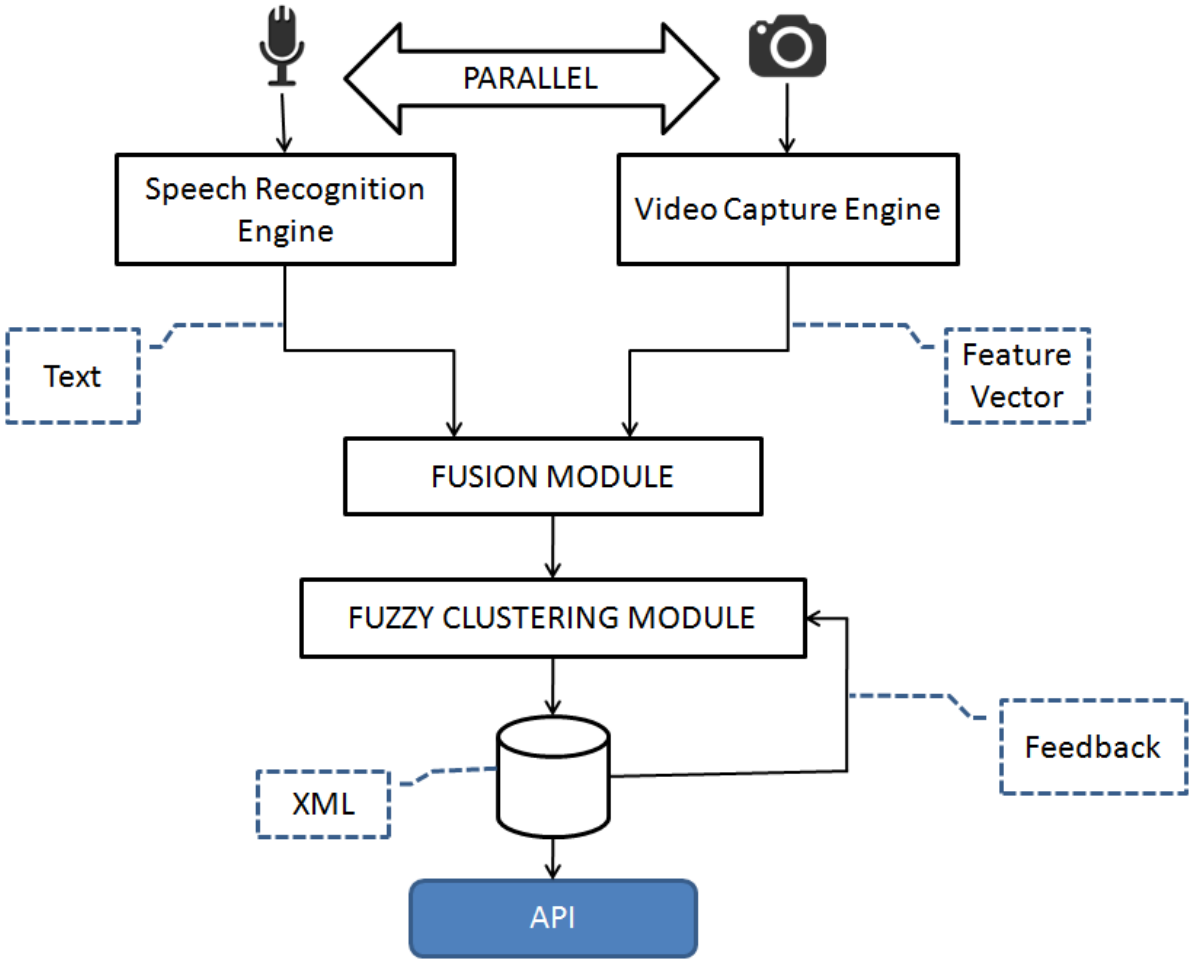

Fig: 1. Architecture showing multithreaded processes and modules

BRAIN STROM OPTIMIZATION

In order to solve a very difficult problem, different people from different backgrounds get together to brain storm. Such methodology helps in generating a large number of ideas because of collaborative thinking. Great ideas originate because of interactive sharing of information. Brain storming focuses initially on bulk ideas generation then eliminating ideas of less importance. Here in our application brain storming is applied to generate new feature vectors and adding them as new objects as per their fitness (calculated by computing distance values).

The BSO algorithm [2, 3] first finds out random cluster centers then applies FCM (Fuzzy C-Means), then it finds best data points (best ideas) in each cluster formed. Next the algorithm generates new data points (new ideas) on the basis of experimentally derived probability attribute. Then for the complete set of data points, the algorithm either selects single cluster center or two cluster centers (again on the basis of experimentally derived probability values). Then for each selected cluster, it finds new data points on the basis of activation function like sigmoid function. Finally the newly generated data points are checked for the fitness among several clusters. If their fitness is as per threshold then the new data points are added or replaced in the data set. Brain Storming Process is illustrated in the following steps in **[Table-1]**.

THE IIOAB JOURNAL
ISSN: 0976-3104

Table: 1. BRAIN STORMING PROCESS STEPS

STEP I	Assemble a group of people from different backgrounds and disciplines.
STEP II	Produce several ideas as per rules in **[Table-2]**.
STEP III	Select certain number of owners of the problems to generate better ideas to solve the problem.
STEP IV	Follow the ideas generated in Step III with greater probabilities to engender new ideas as per the rules in **[Table-2]**.
STEP V	Again owners must select certain nearer ideas as done in Step III.
STEP VI	In this step random selection of objects is made. The looks and functionalities of the objects can be used to generate further new ideas as per rules in **[Table-2]**.
STEP VII	Inform the owners to again select better ideas.
STEP VIII	Final step deals with merging or replacement of newly generated ideas with old ideas.
STEP I	Assemble a group of people from different backgrounds and disciplines.

Table: 2. IDEA GENERATION RULES GIVEN BY OSBORN

Rule or Pattern 1	Suspend Judgement.
Rule or Pattern 2	Anything Works.
Rule or Pattern 3	Cross-Fertilize (Piggyback)
Rule or Pattern 4	Achieve Quantity

In a process of brain storming, generally there is enabler, a group of problem members (people) the brain storming of ideas, and various owners of the problems. The function of enabler is to enable the generation of idea by imposing the group to adapt the Osborn's 4 rules of generation of ideas in a process of brainstorming [4]. The Osborn's 4 rules are presented in **Table- 2** below. The enabler is not be implied in generation of ideas, but alleviating the process of brain storming only. The road map for choosing enabler is to have a good enabler who has prior experience but has less expertise on the background knowledge related to the problem to be solved and who can help in alleviation. The aim of this is that generated ideas should have less, if not zero, biases from the enabler.

EXPERIMENTS AND DISCUSSIONS

We carried out experiments using standard IRIS dataset. We used single iteration of Fuzzy C-Means Algorithm [11, 12] to create initial clusters. Then applied the BSO on created clusters to generate new data points. IRIS dataset consists of 3 classes called as Setosa, Verginica and Versicolor. The blue color (cluster_0) in below given figures represents Setosa, red color (cluster_1) indicates Verginica and yellow color (cluster_2) indicates Versicolor. Figure 2 represents clusters created after applying BSO. **Figure- 2** is plotting of membership values on the graph of PW (Petal Width) vs PL (Petal Length). The DUNN's index was observed to be 2.908. **Figure- 3** is plotted on PW vs SW (Sepal Width). **Figure- 4** is plotted on PL vs SW.

Yellow color cluster (cluster_1) is Setosa; Blue color cluster (cluster_0) is Verginica; and Red color cluster (cluster_2) is Versicolor. Points with black color boundary represent cluster centers. Fuzzy index used in the algorithm was considered to be 2. New data point generation is computed using Equation 1.

$$X_i^d = X_{selected} + \xi * \mathcal{N}(\mu,\sigma) \quad (1)$$

$X_i{}^d$ is the dth dimension of the individual data point chosen to generate new individual data point, $X_{selected}$ is is the dth dimension of the newly generated data point. $\mathcal{N}(\mu,\sigma)$ represents the Gaussian random function with μ as mean and σ as variance.

THE IIOAB JOURNAL
ISSN: 0976-3104

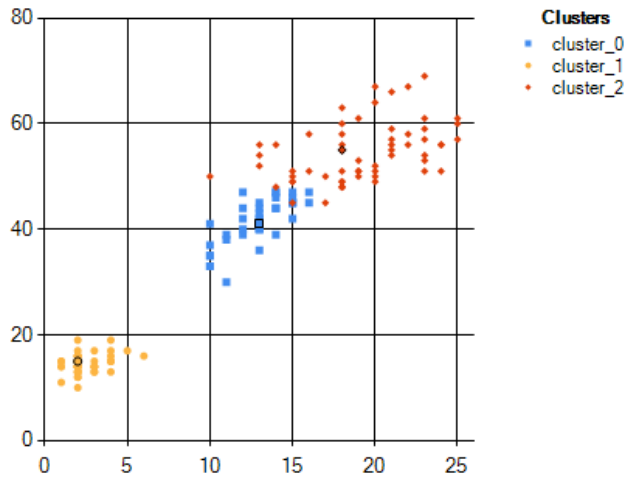

Fig: 2. Plot of membership values on graph of PW vs PL

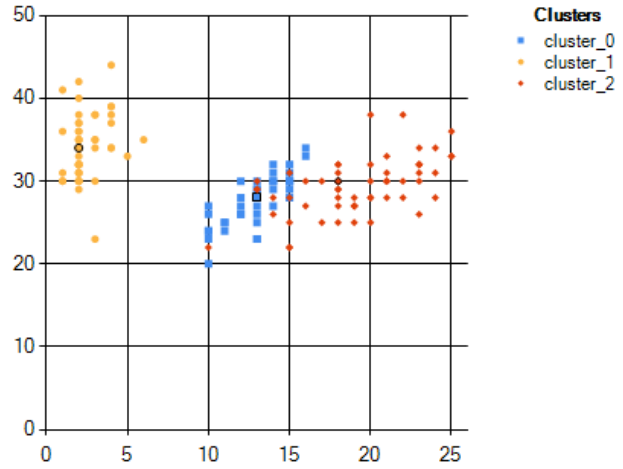

Fig: 3. Plot of membership values on graph of PW vs SW

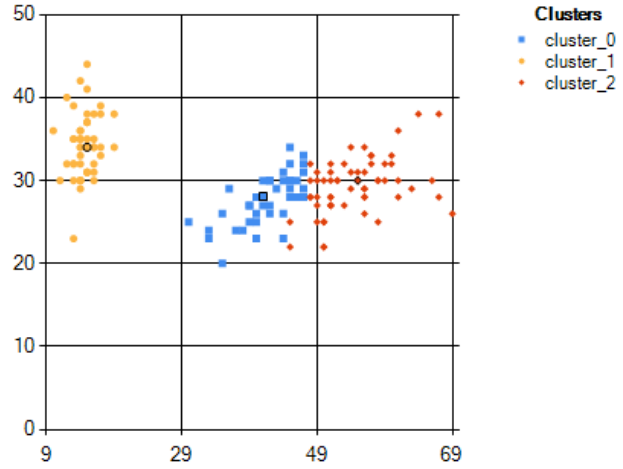

Fig: 4. Plot of membership values on graph of PL vs SW

THE IIOAB JOURNAL
ISSN: 0976-3104

$$\xi = \text{logsigmoid}\,(\,0.5 * \text{max\_i} - \text{curr\_i} / k) * \text{rand}(0,1) \quad (2)$$

k represents change factor for logsigmoid() function's slope. In our experimentation we considered k as 20. max_i is the maximum number of iterations and curr_i represents current iteration number.

Our experimental setup consists of a quad core processor Intel(R) Core(TM) i7-4700MQ having capacity to reach 2.34GHz (each core), an integrated web camera and integrated microphone. We created a multithreaded application to train objects' images through camera and speech through microphone simultaneously. We created a dataset of objects containing 56 tuples, each tuple represents a particular object given as input to system via camera and microphone. First 20 tuples of the dataset consist of "faces", next 20 tuples consist of "hands" and remaining 16 were of "watches". We extracted following features as primary feature vector set: (1) Name of the object obtained from speech recognition engine (2) Average pixel intensity of red plane (3) Average pixel intensity of green plane (4) Average pixel intensity of blue plane (5) Average pixel intensity of canny edge plane (6) Number of keypoints extracted using SURF (Speeded Up Robust Feature) extraction algorithm and FREAK descriptor of OpenCV (7) Capacity of SURF keypoints vector (8) Illuminance reading abtained from LUX Meter. Apart from this primary feature vector set, we stored keypoints extracted from video frames of object using SURF as secondary feature vector set. **Figure- 5** shows the dataset in chronological ordering i.e. the numbering is done from left to right in each line of the figure. We applied Brain Storm Optimization algorithm over all attributes except the first attribute i.e. name of the object. **Figure- 6** represents the graph of membership values plotted on "average pixel intensity of red plane" vs "average pixel intensity of green plane". **Figure- 7** represents the graph of membership values plotted on "average pixel intensity of red plane" vs "average pixel intensity of blue plane". **Figure- 8** represents the graph of membership values plotted on "average pixel intensity of green plane" vs "average pixel intensity of canny edge plane". **Figure- 9** represents the final clusters of the objects (data points). Blue color in the figures represent "Faces", yellow color represent "Hands" and red color represent "Watches". There is 1 outlier in cluster 1 i.e. "Faces", there are 4 outliers in cluster 2 i.e. "Hands" and 4 outliers in cluster 3 i.e. "Watches". Remaining points lie in correct clusters. The aim of separation of primary feature vector set from secondary feature vector set is to minimize the computational complexity required to match a large number of keypoints of every data point with every other data points.

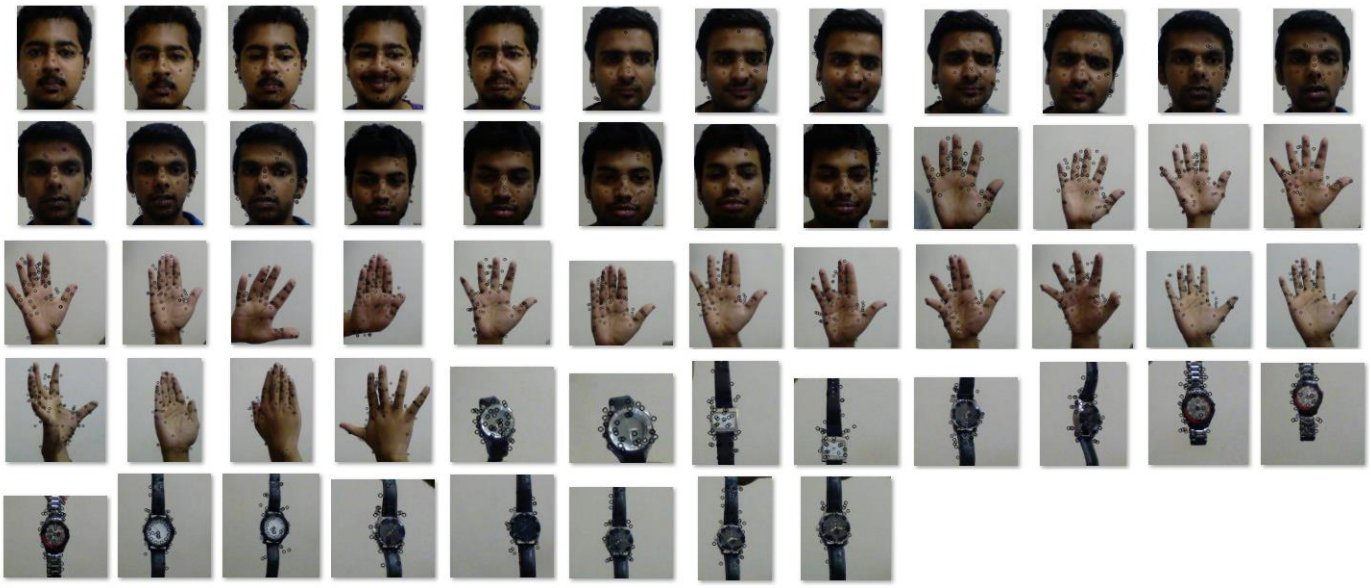

Fig: 5. Chronological Ordering of Data Set representing keypoints extracted from SURF algorithm

THE IIOAB JOURNAL
ISSN: 0976-3104

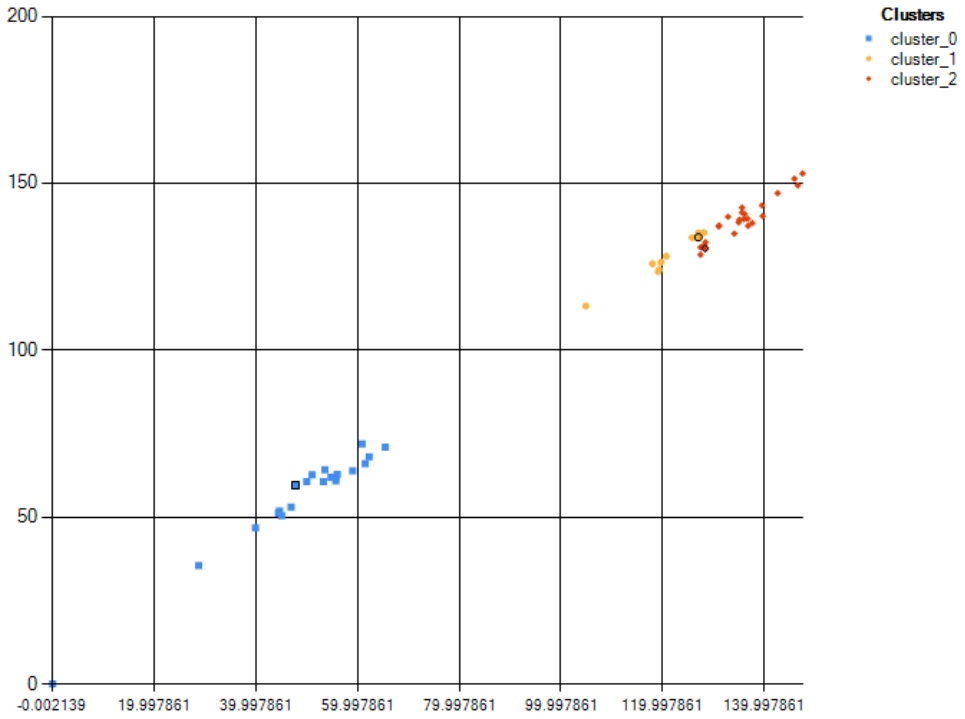

Fig: 6. Plot of membership values on graph of average pixel intensity of red plane vs average pixel intensity of green plane

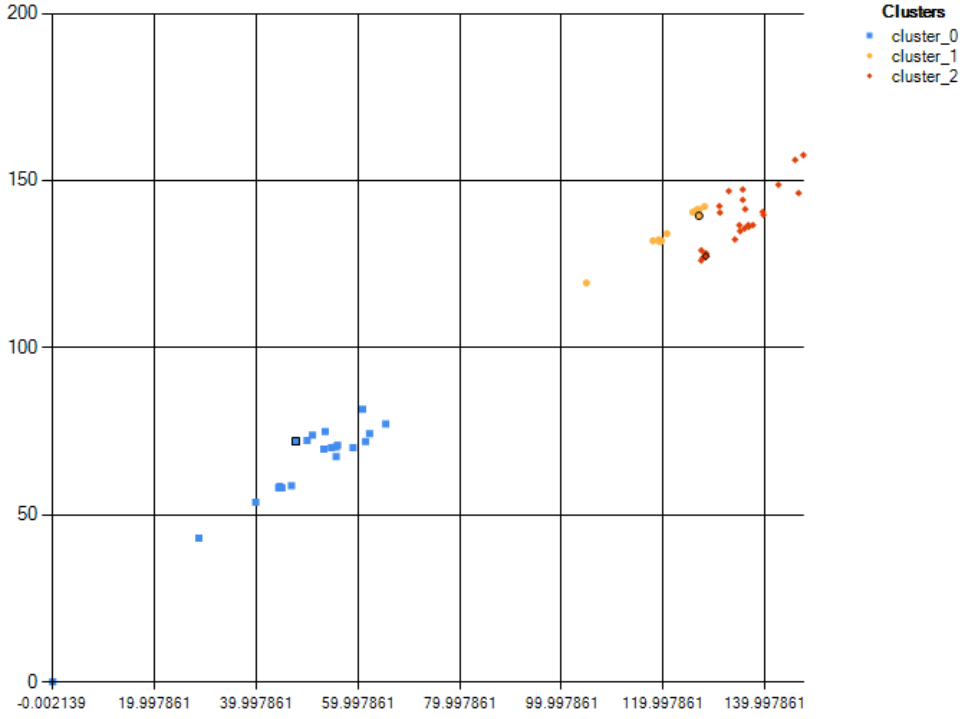

Fig: 7. Plot of membership values on graph of average pixel intensity of red plane vs average pixel intensity of blue plane

THE IIOAB JOURNAL
ISSN: 0976-3104

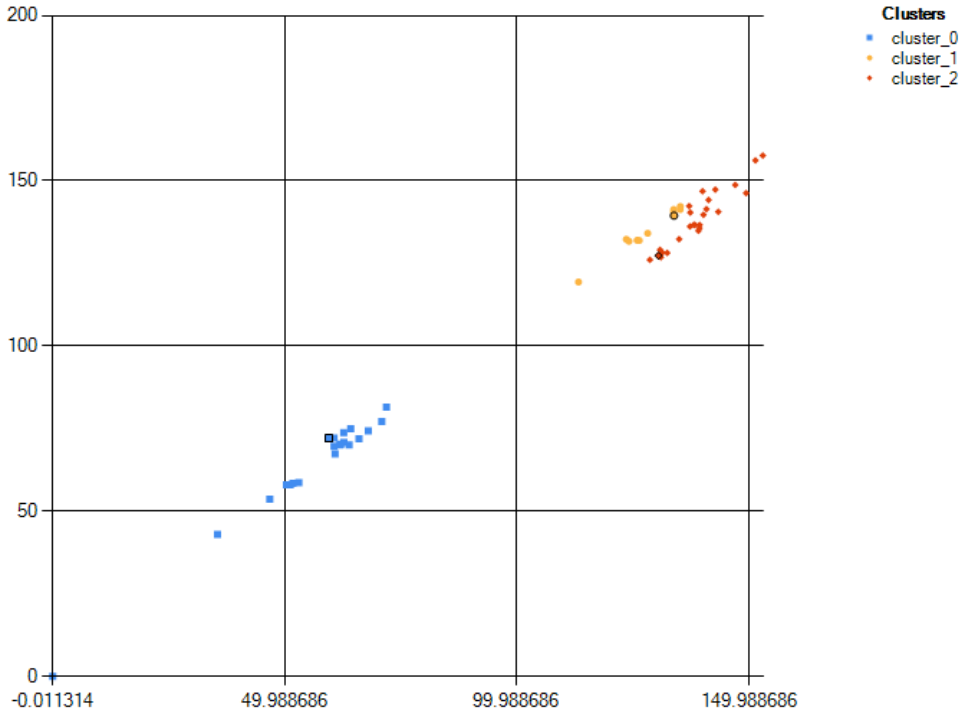

Fig: 8. Plot of membership values on graph of average pixel intensity of blue plane vs average pixel intensity of canny edge plane

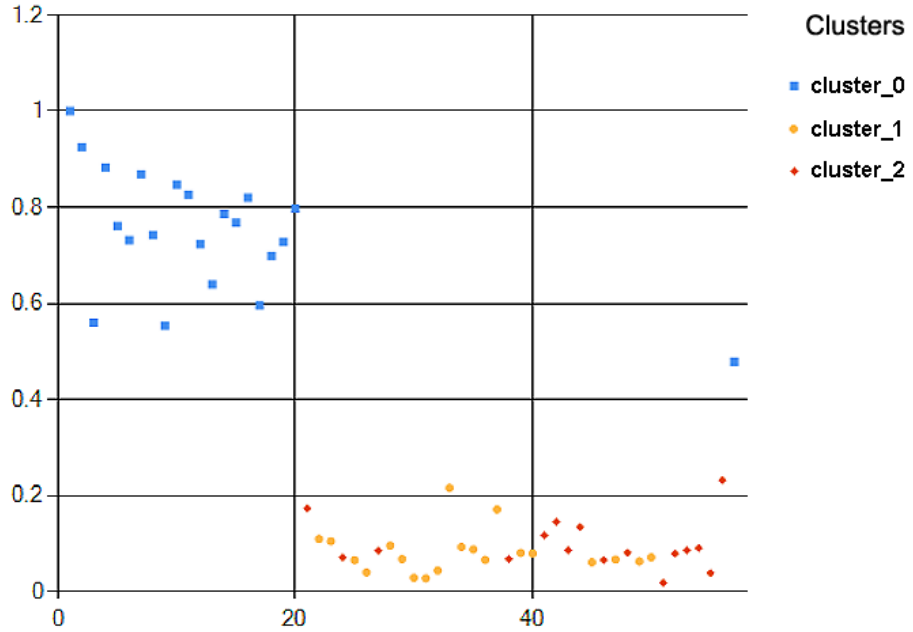

Fig: 9. Final Clusters - X Axis represents data point number, Y Axis represents membership value of data point for first cluster.

Since there are total 9 outliers among the dataset of 56 objects, the accuracy of the clustering system is computed to be 83.92%. Following table contains the data of 30 objects out of 56 objects, 10 from each category. In the following **table- 3**, Type 0 indicates “face”, Type 1 indicates “hand” and Type 2 indicates “watch”.

THE IIOAB JOURNAL
ISSN: 0976-3104

Table: 3. Features extracted from 3 types of objects

Type	Avg red intensity	Avg green intensity	Avg blue intensity	Avg edge intensity	Num keypoints	capacity	lux
0	47.86284	59.492586	72.006016	11.103232	46	63	43
0	51.042227	62.684196	73.82555	10.616559	45	63	43
0	60.908019	71.887839	81.503608	9.881328	66	94	43
0	50.014296	60.59837	72.156533	11.147912	54	63	43
0	53.570218	64.173977	74.944223	11.456442	63	63	43
0	54.747319	61.883129	70.098111	9.620038	38	42	43
0	55.986561	62.738898	70.834437	9.007336	44	63	43
0	53.298458	60.609684	69.636722	9.697591	40	42	43
0	65.459998	70.869158	77.104781	9.7257	67	94	43
0	55.705207	62.016665	70.288868	10.167026	53	63	43
1	104.872065	113.241252	119.362969	10.864501	60	63	43
1	147.529959	152.940015	157.625289	7.561342	72	94	43
1	145.91001	151.363548	156.174345	7.327393	74	94	43
1	126.614514	133.733009	141.312034	8.491996	63	63	43
1	132.854051	139.959366	146.840013	7.80492	67	94	43
1	128.094414	135.174938	142.229197	7.048164	69	94	43
1	119.110055	123.521824	132.29523	8.331642	53	63	43
1	119.293145	124.118101	131.703569	6.810675	87	94	43
1	135.600186	142.708154	147.335056	7.691119	76	94	43
1	127.035292	135.138251	141.312565	7.727407	79	94	43
2	146.597138	149.358333	146.258618	7.885305	44	63	43
2	135.942973	139.288143	135.654857	13.260759	33	42	43
2	136.67944	139.34492	136.652582	6.887209	46	63	43
2	139.538909	143.370271	140.620213	5.783088	40	42	43
2	137.617936	138.075293	136.660811	7.148363	71	94	43
2	136.799375	137.246696	136.17469	6.865798	62	63	43
2	139.680335	140.177728	139.716264	8.818973	73	94	43
2	134.933946	138.309244	136.6316	7.511284	47	63	43
2	127.459702	130.817212	129.102416	9.615451	78	94	43
2	128.349193	130.488091	127.352004	6.542601	69	94	43

Feature vector matching in FREAK descriptor [13] involves following steps:

1. The descriptor uses varied scales of Difference of Gaussians (DoG) which extract the object information. It contains binary symbols set.
2. FREAK descriptor simulates the topology of retina [13, 14].
3. Gaussian is used to smooth sampling points which are distributed on concentric circles where Gaussian kernel size is proportional to the radii of current sampling point's concentric circle.
4. Hamming distance is used a measure of similarity between sampling points.

Our algorithm using SURF with FREAK descriptor gives 22.3% more accuracy than using SIFT.

THE IIOAB JOURNAL
ISSN: 0976-3104

CONCLUSION

Proposed algorithm of division of features into primary and secondary feature sets help in reduction of complexity of mapping large number of keypoints of all objects. Our experiments show that BSO based clustering over the dataset yields in 83.92% accuracy, which indicates that the system can strongly eliminate the problem of large number of keypoints mapping. As per the experiments conducted over a quad core processor with total 8 logical processors having capability of reaching 2.34GHz each, the total running time for matching SURF keypoints of a test object with that of complete dataset (56 objects) takes 4.2 minutes. Whereas our algorithm takes 1.43 minutes which corresponds to 65.8% reduction in time.

CONFLICT OF INTERESTS

Authors declare no conflict of interest.

ACKNOWLEDGEMENT

We would like to thank our dean SCSE Department, Vice Chancellor and our parents because of whom, we had the opportunity to perform research in this field.

FINANCIAL DISCLOSURE

No financial support was received to carry out this project.

REFERENCES

[1] Kate Saenko, Trevor Darrell,"Object Category Recognition Using Probabilistic Fusion of Speech and Image Classifiers" Springer Berlin Heidelberg vol. 4892 ISSN 0302-9743 pp 36-47 2008

[2] Yuhui Shi, Brain Storm Optimization Algorithm, ICSI 2011, Part I, LNCS 6728, pp. 303–309, 2011.

[3] Zhi-hui Zhan; Jun Zhang; Yu-hui Shi; Hai-lin Liu.[2012] A modified brain storm optimization, Evolutionary Computation (CEC), 2012 *IEEE Congress* on , 1(8): 10-15

[4] Smith R. The 7 Levels of Change, 2nd edn. Tapeslry Press (2002)

[5] Tyler Streeter. Open Source Speech Interaction with the Voce Library".

[6] Tsontzos G, Orglmeister R. CMU Sphinx4 speech recognizer in a Service-oriented Computing style," Service-Oriented Computing and Applications (SOCA), 2011 *IEEE International Conference on* , pp.1,4, 12-14 Dec. 2011.

[7] Hong Kook Kim Cox, RV,Rose RC. [2002] Performance improvement of a bitstream-based front-end for wireless speech recognition in adverse environments," Speech and Audio Processing, *IEEE Transactions* on , .10(8):591,604, Nov.

[8] Sheikhzadeh H, Li Deng, [1994]Waveform-based speech recognition using hidden filter models: parameter selection and sensitivity to power normalization," Speech and Audio Processing, *IEEE Transactions on* , 2(1):80-89, Jan. 1994.

[9] Anderson S, Kewley-Port D. [1995] Evaluation of speech recognizers for speech training applications, Speech and Audio Processing, *IEEE Transactions on* , 3(.4):229-241, Jul 1995.

[10] Yao X, Liu Y, Lin G. [1999] Evolutionary Programming Made Faster. *IEEE Transactions on Evolutionary Computation* 3: 82-102

[11] Swarnalatha Purushotham, BK Tripathy. [2015]A Comparative Analysis of Depth Computation of Leukaemia Images using a Refined Bit Plane and Uncertainty Based Clustering Techniques", *Cybernetics and Information Technologies*, ISSN: 1314-4081 15(1):.126-146.

[12] Tripathy BK, P Swarnalatha, et.al.[2013] Rough Intuitionistic Fuzzy C-Means Algorithm and a Comparative Analysis, Proceedings of the 6th ACM India Computing Convention, COMPUTE '13, Aug 22-24, 2013 ACM 978-1-4503-2545-5/13/08.

[13] Wu Yanhai, Zhang Cheng, Wang Jing, Wu Nan.[2015] Image registration method based on SURF and FREAK, in Signal Processing, Communications and Computing (ICSPCC), 2015 *IEEE International Conference* on , 1-4, 19-22 Sept.

[14] Križaj J,Štruc V, Dobrišek S, Marčetić D, Ribarić S. SIFT vs. FREAK.[2014] Assessing the usefulness of two keypoint descriptors for 3D face verification, in Information and Communication Technology, Electronics and Microelectronics (MIPRO), 2014 37th International Convention on ,.1336-1341, 26-30 May 2014

ABOUT AUTHORS

Utkarsh Gupta *is a Student of Master of Technology at VIT University in Vellore, Tamil Nadu, India. He has worked in the field of image processing and has published research work in the subject of Face Recognition and Multimodal Biometric System.*

THE IIOAB JOURNAL
ISSN: 0976-3104

Prof.Swarnalatha Purushotham *is an Associate Professor, in the School of Computer Science and Engineering, VIT University, at Vellore, India. She pursued her Ph.D. degree in Image Processing and Intelligent Systems. She has published more than 50 papers in International Journals/International Conference Proceedings/National Conferences. She is having 14+ years of teaching experiences. She is a member of IACSIT, CSI, ACM, IACSIT, IEEE (WIE), ACEEE. She is an Editorial board member/reviewer of International/ National Journals and Conferences. Her current research interest includes Image Processing, Remote Sensing, Artificial Intelligence and Software Engineering.*

Prateek Chharia *is a Student of Master of Technology at VIT University in Vellore, India.*