JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Enhanced Performance Analysis of Lip Reading System with Improved SURF

Nandini M S[1], Salila Hegde[2], Ashadeepa A[3]

[1]Dept. of IS & Engg., NIE North campus, Mysuru, Karnataka, India, n
[2]Dept. of EC & Engg., NIE North campus, Mysuru, Karnataka, India, [3]Dept. of CS & Engg., Vidyavardhaka college of Engineering, Mysuru, Karnataka, India,

***Abstract**:* Lip reading is an approach for grasping spoken words/sentences .It is a methodology to grasp the spoken words by means of lip motion/gesture. The system should be robust for the variable images. Feature extraction is one of the main process in reading lips, because it plays a prominent role in classification of lip images. In Speeded Up Robust Feature extraction(SURF) for detecting accurate corners is cumbersome for more wrong prediction. In improved SURF accurate prediction of corners is possible with less wrong detection ratio. The paper summarizes improved SURF based on Harris corner detection and reduces false corner detection ratio by increasing threshold value. It effectively reduces computational aspects for descriptor component and time required to match the image.

Index terms **: SURF, Feature detection, Lip movement recognition**

1 INTRODUCTION

Machine perception is a sufficiently broader area of research now a days. It processes camera picture and it is most similar to human perception in which eye images are processed. Some important research domains under computer vision system are forensic studies, biometric, speech reading, face recognition, human recognition and lip reading.

1.2 Lip Reading

Lip reading [2,7,8,9,10,11,13,14,15,16,17,18,19] is the process of communicating verbal patterns by lip movements of a person as a source of evidence about the word/sentence. Contextual proof plays a role and may include facial, hand and body gestures of the speaker. In fact, competent lip readers need to be exceptionally good utilizers of context.

1.3 Feature Extraction

It is a significant step in lip movement recognition process. It is a significant stage in lip image analysis immediately after lip detection and tracking. We can notice feature extraction both in training phase as well as in testing phase in classification step of lip movements. Extraction of feature[3,5,6] is a method of identifying different characteristics of object such as colour, texture, shape etc. Improved SURF is modified version which is utilized in recognizing features of different parts of entities, buildings etc. The SURF technique is more moderated and utilized in the domain of speech analysis, especially in Lip movement analysis in detecting exact corner.

Section 2 provides the literature survey. In section 3, details of SURF [7,10] and procedure of improved SURF is given. Section 4 provides the view of experiments done and output obtained along with discussions. Section 5 enumerates the conclusion and future processes..

2 LITERATURE SURVEY

Since feature extraction is important process in entity prediction, many authors proposed different feature extraction methods in variant ways. Extraction of Similar scale feature technique [SIFT] proposed by D. G. Lowe et. al. [4]. Here SIFT constructs an image pyramid filtrate every layer by increasing sigma data and also the variability.

The SIFT [9] is time consuming and not so supportive in size updates. The Principle Component Analysis [PCA] combined with SIFT is an approach constructed by Y. Ke and R. suthankar[5]. It has come out to improvise the SIFT. Here complexity of analysis has decreased by reducing dimensionality. Thus, they used PCA is used to normalize the gradient path.

SURF methodology accomplished by H.Bay and T. Tuytelars [8]. Here the scripters, confirms tempo in extracting feaures built on Haris corner predition technique [1]. Fast-SIFT developed by V. Chandrashekhar et.al.[12], here feature vector is comparatively less in size than SIFT vector . F-SIFT occurs to be better and stable like SURF, as it has used K-D tree to show and number the description.

F-SIFT gains enhanced intensity against SIFT. It seems to be faster and better in many respects matched with SIFT but extended the work to predict few features for comparing. On the other hand it is suffering from predicting very less features and so the matches. F-SIFT is needed to be improvised in recognition..

Haris corner predictor[1] predicts pixels at corner .However, it conquered the disadvantage of prediction of fewer features as got in F-SIFT. Here, the work done is capable to detect not only edges but also location of the image that has more gradient rate in every direction. It is very delicate for scale similarity, so it is understood from the survey that, a high false edge recognition ratio recognized in SURF.

3 SPEEDED UP ROBUST FEATURE EXTRACTION

SURF is one of the method for shape feature extraction. It is most popularly employed approach to extract shape feature for recognition of objects such as building component. The main drawback of SURF is increased false corner detection and time consumption of feature extraction. Hence we are attempting to overcome drawbacks with improved SURF.

3.1 Enhanced Speeded Up Feature Extraction Algorithm

Step 1:

The Hessian Detector of SURF will be replaced by quick corner. To be a quick detector, threshold t has set to 25. If t is extremely small t<<0,the hurdle the amplification of the false corners will be taken into consideration.

Step 2:

Excluding maximum value (Max=25) all other suppression is set to true. Around 155 corner features per frame came out to be good balance between finding necessary features for matching and processing speed.

Step 3:

Increase the threshold t by 5 each time till the number of corners are decreased to nearest 155 , if the number is greater than 155 in the first place.

In the same way, there are paths to reduce the time consumption of produced descriptor -+ components and matching the image. Hence, effort revealing the success in getting better feature for lip movement analysis.

4 OVERVIEW OF EXPERIMENTS CARRIED OUT AND RESULTS

4.1 Dataset

Around fifty samples of video are collected as base database. A couple of bunch of words Anna, Amma, Appa, Anna, Ajja and Vana, Bana, Mana, Mara and Bara. are the input data to evalute the Improved SURFand SURF methods.

4.2 Performance Metrics Used

To demonstrate the proposed system, following metric measurements are used.

a) Recall : The Recall is expressed in the formula,

Recall=TrueP/(TrueP+FalseN) (1)

b) Precision: Precision is expressed in the formula

Precision=TrueP/(TrueP+FalseP) (2)

The TrueP,TrueN,FalseP and FalseN are:

TrueP : These are the exactly recognized positive values and the value of recognized class is also yes.

TrueN : These are the exactly recognized negative values which means that the value of original class is no and value of recognized class is also no.

FalseN : The original class is yes but recognized class in no.

FalseP: The original class is no and recognized class is yes

The samples utilized for training and testing the data are 30 and 10 .

Relative operating characteristic curve(ROC): it is an analogy of a matching two operating characteristics (FPR and TPR) as the benchmark changes. It is a most prioritized evaluation metrics for evaluating general classification model's performance. The ROC curve is designed by using TPR matched with FPR.

Area Under The Curve(AUC) provides degree or measure of separability. It says how much the model is able to distinguish between the classes. More in AUC, good the model is at recognizing 0s as 0s and 1s as 1s. By the standard criterion, Higher the AUC, good the model AUC.

4.3 Experimental Results and Discussions

The evaluation output of experimenting Improvised SURF and SURF on the couple of group of words Appa, Amma, Akka, Ajja and Anna ,Vana, Bana, Mana, Mara and Bara. are tabulated in Table1,Table2,Table3 and Table4

Table 1. Results of experimenting SURF on the words such as Amma, Appa, Ajja, Akka and Anna.

		False Positive						
		Appa	**Amma**	**Akka**	**Ajja**	**Anna**	**Precision (%)**	**Recall (%)**
True Positive	**Appa**	02	05	00	00	01	62.4	62.6
	Amma	00	02	00	06	02	65.67	60.7
	Akka	05	02	01	00	00	50.10	62.6
	Ajja	06	02	05	02	00	62.6	55.6
	Anna	05	01	01	00	01	62.6	62.6
							Avg Precision 61.84	**Avg Recall 60.73**

Table 2 The result of evaluating Improvised SURF on the words Appa, Amma, Akka, Ajja and Anna.

		False Positive						
		Amma	**Appa**	**Ajja**	**Akka**	**Anna**	**Precision (%)**	**Recall (%)**
True positive	**Amma**	06	02	00	00	00	66.66	75.00
	Appa	01	07	00	01	01	77.77	70.00
	Ajja	00	00	06	02	01	60.00	65.66
	Akka	01	00	02	05	01	62.5	62.50
	Anna	01	00	02	00	06	66.66	66.66
							Average Precision 66.71	**Avg Recall 68.16**

Table 3 The result of checking SURF on the group of words Vana, Bana, Mana, Mara and Bara.

		False Positive						
		Vana	**Bana**	**Mana**	**Mara**	**Bara**	**Precision (%)**	**Recall (%)**
True Positive	**Vana**	04	01	01	02	01	50.00	44.44
	Bana	01	05	01	01	01	62.50	55.55
	Mana	01	01	06	02	00	54.54	60.00

Mara	01	01	02	05	00	50.00	55.55
Bara	01	00	01	00	04	66.66	66.66
						Average Precision 56.74	**Average Recall 54.44**

Table 4 The evalution result of experimenting Improvised SURF on the bunch of words Vana, Bana, Mana, Mara and Bara.

		False Positive						
		Vana	**Bana**	**Mana**	**Mara**	**Bara**	**Precision (%)**	**Recall (%)**
True Positive	**Vana**	05	01	01	01	00	71.42	62.50
	Bana	00	06	01	00	01	66.66	75.00
	Mana	01	01	04	01	00	66.66	51.14
	Mara	01	00	00	07	02	63.63	70.00
	Bara	00	01	00	02	08	72.72	72.72
							Average Precision 68.21	**Average Recall 67.47**

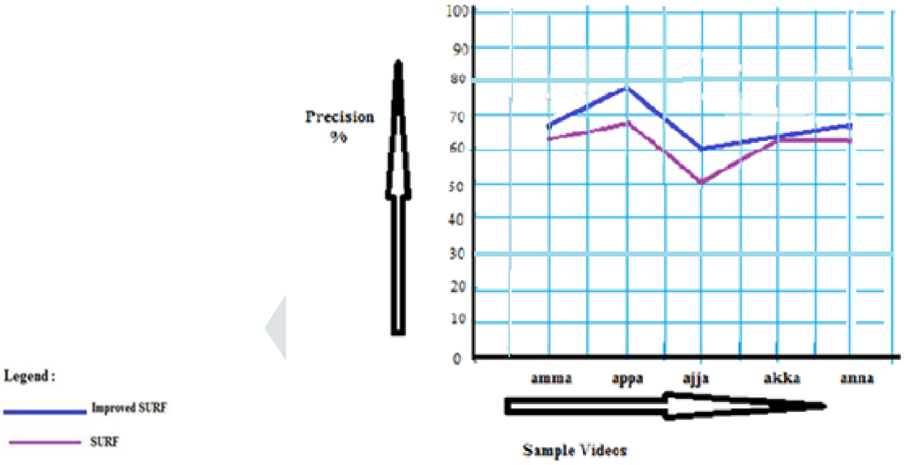

Fig. 1 Precision vs. video samples of Amma, Appa, Ajja, Akka and Anna on evaluating improvised SURF and SURF.

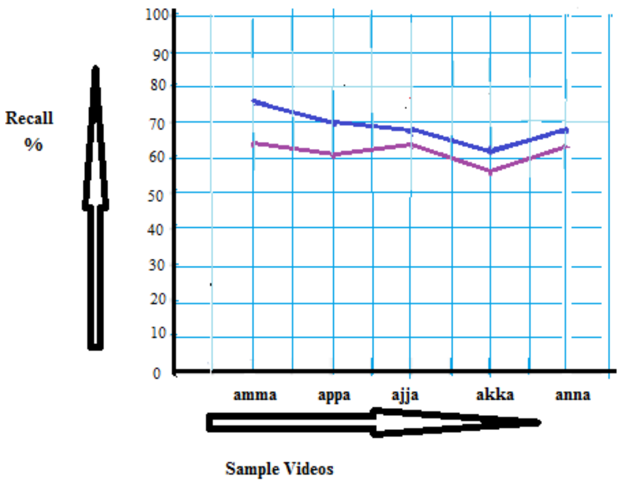

Fig. 2 Recall vs. video samples of amma, ajja, appa, anna and akka on experimenting improvised SURF and SURF.

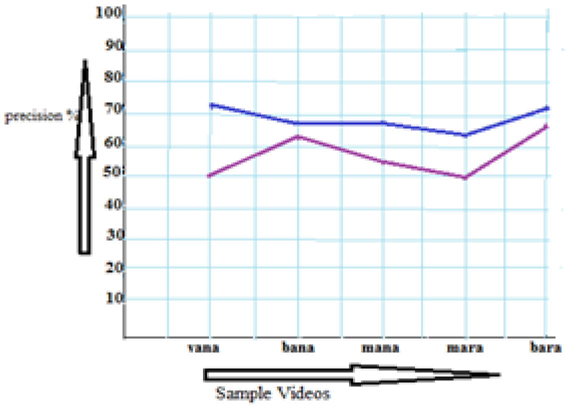

Fig. 3 Precision vs. Sample videos of Vana, Bana, Mana, Mara, Bara on applying SURF and improved SURF.

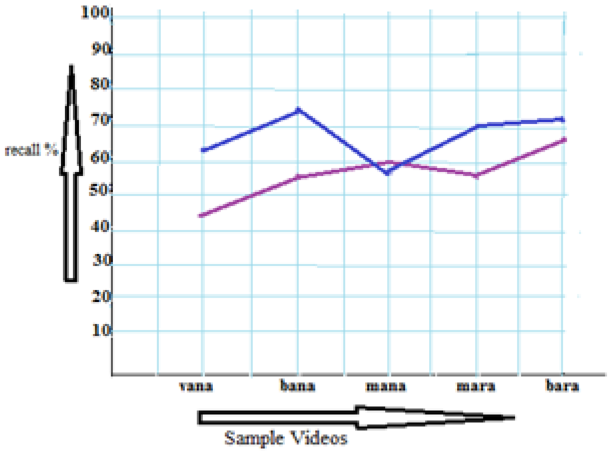

Fig. 4 Recall vs. Sample videos of Vana, Bana, Mana, Mara, Bara on experimenting SURF and improved SURF.

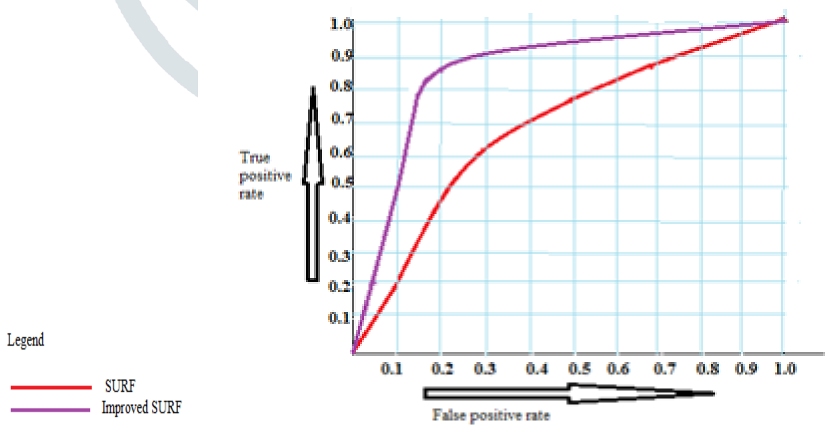

Fig. 5 ROC Curve for the video Samples of Appa, Amma, Akka, Ajja and Anna on evaluting improvised SURF and SURF.

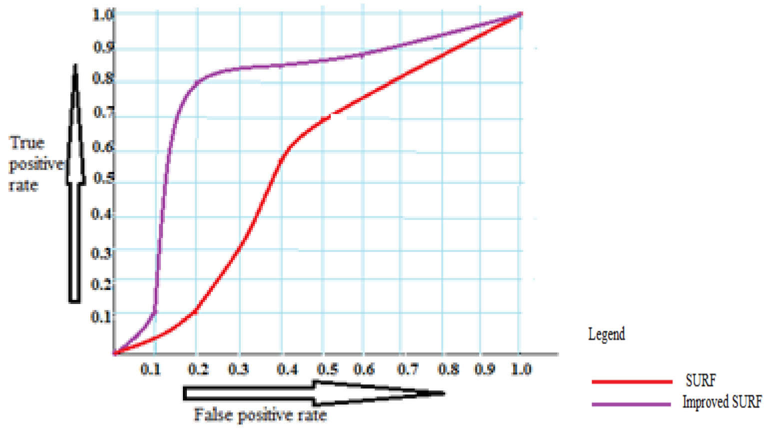

Fig. 6 ROC Curve for the Sample videos of Vana, Bana, Mana, Mara, Bara on applying SURF and improved SURF.

In SURF, we got evaluation result as rate of precision of 60.84% and rate of recall of 60.72%.Further, improved SURF enlightened the accuracy by providing performance of rate of precision of 66.73% and rate of recall of 68.17%. Recall and precision of Improvised SURF and SURF are shown in the diagram Fig. 1 and Fig. 2 .

It is recognized that there is an empirically high in rate of precision by 5.89% and rate of recall by 7.47% in improved SURF because of reduced false corner ratio.

ROC curve Fig 5 and Fig 6 plotted for words Appa,Amma,Anna,Ajja,Akka and Vana,Bana,Mana,Mara,Bara depicts that SURF has achieved good performance and improved SURF has achieved best performance.

Area under Curve(AUC) :
AUC= (Average precision - Average recall) **(3)**
Area under Curve(AUC) plotted for the words Appa,Amma,Akka,Ajja,Anna.
By the SURF technique
AUC = 61.83- 60.71
= 1.21
On Applying Improved SURF technique
AUC = 66.71- 68.16
= 1.45
Area under Curve(AUC) plotted for the words Vana,Bana,Mana,Mara,Bara:
On Applying SURF technique
AUC = 56.74- 54.44
= 2.3
On Applying Improved SURF technique
AUC = 68.21- 67.47
= 0.746
F1-score = 2* (Precision* Recall) / (Precision + Recall) (4)
F1-score for words Appa,Akka,Anna,Ajja,Amma:
By the SURF technique
F1-score = 2* (61.83 *60.71) / (61.83 +60.71)
= 61.26
On applying Improved SURF technique
F1-score = 2* (66.71 *68.16) / (66.71 +68.16)
= 67.42
F1-score for words Vana,Bana,Mana,Mara,Bara:
On applying SURF technique
F1-score = 2* (54.44 *56.74) / (54.44 +56.74)
= 55.56
On applying Improved SURF technique
F1-score = 2* (67.47 *68.21) / (67.47 +68.21)
= 67.84

It is observed from the experiment that for the words, Amma,Appa,Ajja,Akka,Anna F1-Score of SURF is 61.26 and F1-score of improved SURF is 67.42 and for words Vana,Bana,Mana,Mara,Bara F1-Score of SURF is 55.56 and F1-score of improved SURF is 67.84.From these values it is clearly visible that Improved SURF performed better than SURF

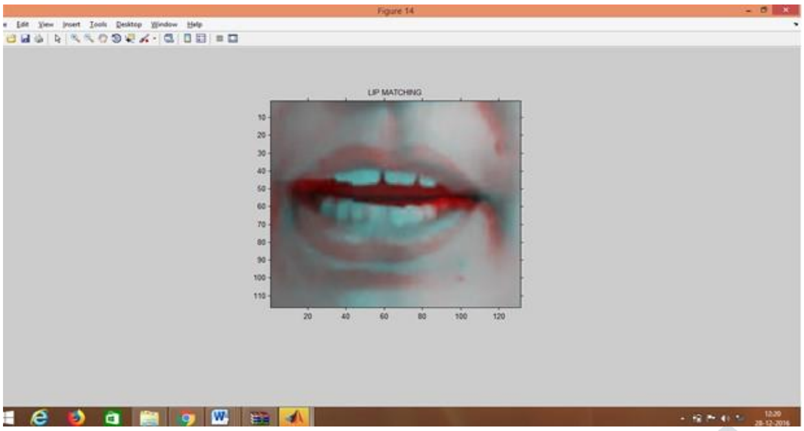

Fig. 7 Feature of lip movement prediction and lip movement matching by using SURF

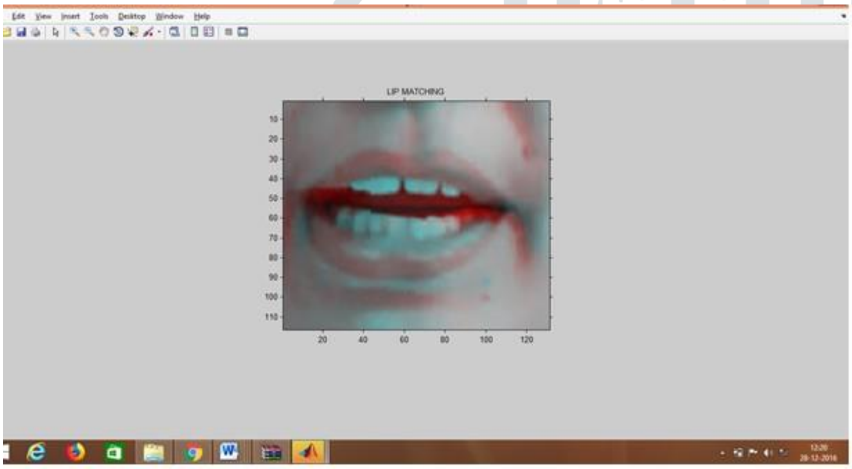

Fig. 8 Lip movement extraction and lip movement matching by applying Improvised SURF

By applying Improvised SURF and SURF technique on the word Amma displayed in Fig. 5 and Fig. 6. On the other side, the experiment tried to give importance for the eye witness which revealed the performance and efficiency of the model. The comparison of visual appearances of couple of diagrams have been supporting to gain avenues. However, the greater enhancement have recognized through edge prediction rate as noticed in Fig6.

CONCLUSION AND FUTURE ENHANCEMENT

Lip movement analysis is essential for the physically challenged or deaf person. The feature detection is most important step for classification of lip movements. The erected system with the improvement to SURF provides reduced time consumption of extracted descriptor components and matching time of the image. Hence the Improved SURF provides reduced false corner detection ratio by increasing threshold value as compared to SURF method. The Average value of rate of precision and recall are enhanced in improved SURF as matched to SURF technique. By this, we got fruitful result in extracting simillar feature for lip movement analysis.

Moreover, work can be enhanced to increase the rate of precision and recall for the same group of videos of the sentences to get increased accuracy of edge recognition ratio.

REFERENCES

[1] C. Harris and M. Stephens, "A combined corner and edge detector". Proceedings of the 4th Alvey Vision Conference., 1988, pp. 147–151.

[2] D. G. Lowe, “Object Recognition from Local Scale-Invariant Features,” in IEEE International Conference on Computer Vision. IEEE Computer Society, 1999, pp. 1150–1157.

[3] S. Agarwal and D. Roth, “Learning a Sparse Representation for Object Detection,” in European Conference on Computer Vision, vol. 4, Copenhagen, Denmark, May 2002, pp. 113–130.

[4] D.Lowe.”Distinctive Image Features from Scale-Invariant Keypoints”, IJCV, 60(2):2004,91–110.

[5] Y. Ke and R. Sukthankar.PCA-SIFT: “A More Distinctive Representation for Local Image Descriptors” ,Proc. Conf. Computer Vision and Pattern Recognition,2004, pp. 511-517.

[6] D. G. Lowe, “Distinctive Image Features from Scale-Invariant Keypoints,” in International Journal of Computer Vision. Springer Netherlands, Oct. 2004, pp. 91–110.

[7] K. Mikolajczyk and C. Schmid, “A Performance Evaluation of Local Descriptors,” IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 27, no. 10,oct.2005, pp. 1615–1630.

[8] H. Bay, T. Tuytelaars, and L. Van Gool, “SURF: Speeded Up Robust Features,” in European Conference in Computer Vision, 2006, pp. 404– 417.

[9] A. E. Abdel-Hakim and A. A. Farag, “CSIFT: A sift descriptor with color invariant characteristics,” in Computer Vision and Pattern Recognition, IEEE Computer Society Conference on, 2006,vol. 2, pp. 1978–1983.

[10] T. Tuytelaars and K. Mikolajczyk, “Local Invariant Feature Detectors: A Survey,” in Foundation and Trends in Computer Graphics and Vision, 2008, pp. 177–280.

[11] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, “Speeded-up robust features (SURF),” Computer vision and image understanding,2008, vol. 110, no. 3, pp. 346–359.

[12] V. Chandrasekhar, M. Makar, G. Takacs, D. Chen, S. S. Tsai, N.- m. Cheung, R. Grzeszczuk, Y. Reznik, and B. Girod, “Survey of SIFT Compression Schemes,” in International Workshop on Mobile Multimedia Processing, Aug. 2010.

[13] G. Wang, L. Tao, H. Di, X. Ye, and Y. Shi, “A scalable distributed architecture for intelligent vision system,” IEEE Transactions on Industrial Informatics,Feb,2012, vol. 8, pp. 91–99.

[14] H. Liu, S. Chen, and N. Kubota, “Intelligent video systems and analytics: A survey,” IEEE Transactions on Industrial Informatics,Aug 2013, vol. 9, pp. 1222–1233.

[15] M. H. Lee and I. K. Park, “Performance evaluation of local descriptors for affine invariant region detector,” in Asian Conference on Computer Vision, 2014,pp. 630–643.

[16] E. N. Salahat, H. H. M. Saleh, S. N. Salahat, A. S. Sluzek, M. AlQutayri, B. Mohammad, and M. I. Elnaggar, “Object detection and tracking using depth data,” Oct. 23 2014. US Patent App. 14/522,524.

[17] S. Tamura, H. Ninomiya, N. Kitaoka, S. Osuga, Y. Iribe, K. Takeda, and S. Hayamizu. Audio-visual speech recognition using deep bottleneck features and high-performance lipreading. In 2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA), pages 575–582. IEEE, 2015.

[18] M. Wand, J. Koutn, et al. Lipreading with long short-term memory. In 2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pages 6115– 6119. IEEE, 2016.

[19] M. Zimmermann, M. M. Ghazi, H. K. Ekenel, and J. P. Thiran. Visual speech recognition using PCA networks and LSTMs in a tandem GMM-HMM system. In Asian Conference on Computer Vision, pages 264–276. Springer, 2016.