4D AUTOMATIC LIP-READING FOR SPEAKER'S FACE IDENTIFCATION

Adil AbdUlhur AboShana

Department of Information System,
University of eötvös loránd, Budapest, Hungary

ABSTRACT

A novel based a trajectory-guided, concatenating approach for synthesizing high-quality image real sample renders video is proposed . The lips reading automated is seeking for modeled the closest real image sample sequence preserve in the library under the data video to the HMM predicted trajectory. The object trajectory is modeled obtained by projecting the face patterns into an KDA feature space is estimated. The approach for speaker's face identification by using synthesise the identity surface of a subject face from a small sample of patterns which sparsely each the view sphere. An KDA algorithm use to the Lip-reading image is discrimination, after that work consisted of in the low dimensional for the fundamental lip features vector is reduced by using the 2D-DCT.The mouth of the set area dimensionality is ordered by a normally reduction base on the PCA to obtain the Eigen lips approach, their proposed approach by[33]. The subjective performance results of the cost function under the automatic lips reading modeled , which wasn't illustrate the superior performance of the method.

KEYWORDS

Lip Segmentation, Discrete cosine transform algorithm, kernel Discriminant Analysis , Discrete Hartley Transform, hidden Markov Model.

1. INTRODUCTION

It represented once for any viseme image visible and model 'visual silence'. The hidden Markov model(HMM) is trained from the visual features with three states and a diagonal covariance Gaussian Mixture Model (GMM) associated with each task are sequences of embedded trained and test with view angle dependence . The approach was visual speech synthesis and the visual parameters were generated from HMM by using the dynamic ("delta") constraints of the features. The mouth motion under the video can be rendered from the predicted visual parameter trajectories. The drawback of the HMM-based visual speech synthesis method is generated blurring due to feature dimension reduction in statistical modeling, i.e. PCA and the maximum likelihood(MLL). Proposed by [5][6].

Dhinaharan Nagamalai et al. (Eds) : ACSIT, ICITE, SIPM - 2018
pp. 115–126, 2018. © CS & IT-CSCP 2018 DOI : 10.5121/csit.2018.80809

2. RELATED WORK

The employed lip-reading systems uses developed features, i.e. csamandhilda, that are consistent across a view, which indicates improved robustness to viewpoint updated for all the types of the primitive feature . The goal of this experiment is to obtain the best viewing angle for computing lip-reading and active appearance model (AAM) features that are extracted from each view, respectively by using their second derivatives .They use a linear predictor based on the tracking and it's has more robust lip-contourthan the AAM That was introduced by [9].The audio-visual speech recognition system, visual features obtained from Discrete Cosine Transform(DCT) and active appearance model. (AAM) were projected onto a 41 dimensional feature space using the LDA, proposed by[34]The systems reduce the dimensionality for Linear Discriminant Analysis (LDA) or Fisher's Linear Discriminant (FLD) as introduced by[57]. The project pixel and the colour information pixels both are using a lower dimensional space. The threshold operation based on the lower dimensional space is represented for the lip segmentation. [58] located the face region with a skin-colour model. The mouth region was localized by involving in the skin region.

The lips region segmentation was using the G and B components of the RGB colour based contented for Fisher transform vectors. Adaptive thresholds representation as an operation of the grey scale histogram of the image were then employed to segment the lip pixels. [59] used RGB values from training images to learn the Fisher discriminant axis. The mouth region colour onto their axes used to enhance the lip-skin boundary then a threshold was applied to segment the lip pixels. [60] used an identical approach to lip segmentation. [41] used FLD by visible for an edge detection, dynamic based on the split ability of multi-dimensional distributions. The edge of the lip contour extraction was described as the point at which there is a maximized distributions of lip and non-lip pixels. Principal Components Analysis (PCA), as suggested by [57] has been reduced to dimensionality automatic for the identification of lip pixels.

2.1 Segmentation

Pixel-based of any lip segmentation have been using a specific colour space appearance to compare involve for pixel colour combined with a set of thresholds. They changed image from the binary separation into a part of lip pixels and non-lip pixels. They used colour spaces by choice and the operation threshold base by histogram selection. [50] worked a normalized RGB colour which is created based on a process to satisfy the maximum intensity normalization. They had represented colour components based on thresholding was beneficial to classify the image into lip and non-lip pixels. [51] represented a mouth region by using three different colour appearance in YCbCr, RGB and HSV. Lip pixels segmentation obtained by thresholding the YCbCr, Green and Hue components of the colour appearance and associated the results by using a logical and operator.

Region based lip Segmentation systems use developed cost-functions to constrain the subset of pixels being chosen. In region based lip segmentation these cost functions usually include the shape constraints or the locality of lip pixels. The aim of this section is to use some criterion functions to choose chromatically and homogeneous pixels in an image to be the lip region. The approach to region-based lip segmentation is the Markov Random Fields technique (MRF) , proposed by [51] .Each lip pixel is processed as a stochastic variable sensitive by a result of exactly the neighborhood is connected. The identification of an objects in an image formula where each object referred to as a set is the like to the a single MRF. In[53],use a spatiotemporal neighborhood compute to form a region lip segmentation using MRF. The temporal input is the difference between the binary labels in two consecutive images, respectively [54] use primitive pixel dependent on thresholding operation under the HSV mouth region image with edge information to create a label the MRF set as part of the lip or non-lip regions. They weren't using spatial homogeneity, pixel-based algorithms attempted to be faster than region-based approaches from time to time generating coarse segmentation results and no re- correctly classify for pixels is noisy. The pixel appearance is relative to the colour as define and the local neighborhoods of pixels was included. Lip pixels segmentation were practiced by a Bayesian classifier ,this introduced by [56] which uses Gaussian Mixture Models (GMM) that are estimated by using the Expectation Maximization algorithm (EM) , refer to [55] and their normalized the R and G components of the RGB space by using the intensity of the pixel. The normalization was an illumination invariant to RGB. Lip pixels segmentation were used by a thresholdin colourspace.

2.2 Data Acquisition

The dataset was used to record and study by [7].The datasets contain discrete English phonemes correspond to the visemes visible in The face . The face model of MPEG-4 standards is used two Facial Animation Parameters and Facial Definition Parameters. The visemes connected to form words and sentences is due the specification of used visemes as the recognition unit. The calculated number of visemes is less than phonemes , due speech is partially visible, refer to [8].Video data was recorded by a movie camera in a typical mobile environment. The camera view was on the speaker's lips-reading and it was also kept fixed throughout the recordings. Factors, examples window size (240x320 pixels), view angle of the camera, background and illumination were kept constant for each speaker. To validate the proposed method, 12 subjects (6males 6 females) were used. Each speaker, recorded 6 phonemes at a sampling rate of 30 frames/sec and every phoneme was recorded five more times to give a sufficient variability.

3. PROPOSED METHOD

They used automated lip-reading, consist of 2D DCT and Eigen-lips. The lip shape of an AAM algorithm defined by, **s**, is associated with coordinates (x,y) of the set N vertices that determined the features on an object: **s** $=(x_{1,}y_1, \dots \dots x_{n,}y_n.)^T$ .A model that appeared a linear variation in the shape is explain in below equation,

$$S{=}s_0 + \sum_{i=1}^{m} p_i \, s_i \tag{1}$$

Where s_0 is a term called mean shape and **si** are defined by the eigenvectors corresponding to the largest number (m)of the covariance matrix consist of the eigenvectors. The coefficients (pi)are defined for the shape parameters that associated with each eigenvector in the shape **(s)** is appeared. The model is always calculated by using Principal Component Analysis (PCA) to a set of shapes handle in a corresponding for each image. To get the shape vertices **s**, required in Equation (1) and to compute the shape parameters, it was proposed by [12]. The mouth of the set area dimensionality is ordered by a normally reduction base on the PCA to obtain the Eigenlips approach, their proposed approach by[33].

The area (A),a term of an AAM is represented by the pixels that stretch inside the mesh s_0. AAMs represented as linear a variation visible, so is appeared as term base on A0 plus a linear associated withlto display the images A_i.

$$A=A_0 + \sum_{i=1}^{l} \lambda_i \; A_i \tag{2}$$

where λi are represented as parameters. As well, shapes s, represented the base on $A_0$is the mean shape normalized image and vectors Ai are the reconstructed the shaped eigenvectors corresponding to the largest eigen values and both are always calculated by using PCA to the shape training images, it is normalized by [11].The scenario, established a lipin contained on the speaker's face identification system based on optimal performance of the control. A lip is necessary for discrimination, after that the their work consisted of in the low dimensional for the fundamental lip feature vector and reduced by using the 2D-DCT .They found that subjects have a two-stage discrimination analysis for speaker identify, such as to exploit two pair correlations temporal and spatial correlations by [16].The eigen face approaches [18][17]are using the principal component analysis(PCA),or Karhunen-Loeve transforms (KLT). It obtained to account the statistical base on the measure between the pixel values of images to be visual in a training update to create an orthogonal for representing images. The eigenvectors of the covariance matrix of the training update of face images are calculated and they are retrained for describing the images are called eigen faces. The ones corresponding to the largest eigen values of the covariance matrix for each training and test face is characterized by its projection on the eigen faces, and the comparison of two faces is obtained by comparing two sets of projections.

3.1 Lips Transformation

The maximum flexibility in deforming to movie shapes of body likely lips shapes .so that, then applying for complex shapes must have control points to describe them, this is usually implied, to instability in tracking [24][23]. A movie shapes, provided the lips are not flexing, is an approximately rigid, planar shape under orthographic projection [25].They show that form a vector Q for the lips a represented body with the affine transformation that is applies to the template to get the shape. No-rigid motion can be handled in the same way, by providing that represents both the rigid and non-rigid degrees of freedom of the shape. This is used, for instance, to figure out the movements in hand tracking, or to automatic lip movements. It is crucial point therefore to allow the right degrees of freedom for non-rigidity, neither too few resulting in excessive stiffness, nor leading to instability and then the snake-based methods for handling non-rigid motion by[22] allowed for the high degrees of freedom which leads to instability, well unusable for real-time. Their scenario, the framework of localized color active contour model (LCACM) expand from scheme ,introduced by [19] given that the foreground and background regions with variation in color space. They utilize a 16-point deformable model by [20]with geometric constraints to achieve lip contour extraction. They used deform modelled location region base to approach lip tracking combined with the extraction of lips contour in the color image [21].

The dimensionality is reduced for colour-space transforms, generally hue-based transforms are used in pixel based lip segmentation systems. [54] were changing the red hue, value base on a threshold to identify the lip pixels. The hue transforms originally, proposed by [61].Their purpose is to transform the colour space to maximize the chromatic difference between skin and lips that led to use based on the connected between the R and G components of the RGB system.

The transform is combined with pixel-based classification by [62].In 63], used to transform the information on the RGB colour under the CIELUV space. The adaptive used histogram-based operation thresholds later resulted in a binary classifier for lip segmentation.

Discrete Hartley Transform(DHT) is subjected to wavelet multi-scale edge detection to obtain the lip contour which is smooth by using morphological operations, refered[42] obtain in the mouth region by using generally directed camera and subject of the region hue colour transform. hybrid of representing edge which are used edge enhancement exactly a polynomial curve , suggested by [49] for the mouth region to the YCbCr colour space transformation. A parametric model using cubic curves is used to detect the lip contour. [48] approach with normalised RGB values as the colour features. A geometric model is used to detect the lip contour. In [47],use a cost function that the boundary of the mouth region is closed or open combined with a parabolic curve to extract the lip contour.[46] employed lip contour extraction using cubic B-Splines to boundary the lip-pixels extracted, however, the classification is a binary processed . Snakes for segmentation were proposed by [45].Active contours give a deformable model of a contour and it is an object under image by using internal and external cost functions or energy functions to lead the model to satisfy the object boundary. The idea of using “hybrid edges” was extended with a snakes formulation is “jumping” defined to extract the lip contour by[43]. [44] use an active contour style energy design to detect the inner lip contour. In[42], use active contours to build a geometric template with a mouth region image, proceeded that the keypoints of the lip contour are using pixel extracted. B-Spline based active contour is proposed by [41] . [40] use an active contour based codebook to synthesize similarly lip contours under any image .The convergence of active contour created is defined in the Gradient Vector Flow (GVF) , introduced by Xu and Prince .(1997) uses edge represented and an edge-based vector-field to formulate that can be external energy for snakes. GVF snakes are used for lip contour detection are proposed by [64].Active Shape Models (ASM) [38] and [28] used Active Appearance Models (AAM) and they have been used for contour-based lip segmentation. The active contours mod-based approach to lip segmentation uses level sets. It provides an energy minimization by curve created called B-Spline technique that has been created based on Widely model shapes and object boundaries in the computer vision presented by [37].

3.2 Synthesizing Identity Surfaces

In [27][28][29], use analysis base on synthesis. They approach by using the synthesise identity surface of a subject face from a small sample of patterns which sparsely fill the view sphere. The base, approximate of the identity surface using a set of $\mathrm{N_p}$ planes separated by $\mathrm{N_p}$ multiple views. They used PQl tilt and yaw are the **z** discriminating feature vector of a face pattern.

KDA vector.$(\mathrm{x}_{01,}y_{01}), (\mathrm{x}_{02,}y_{02}), \dots\dots(\mathrm{x}_{0\mathrm{N},}y_{0\mathrm{N}})$ are define views which separate the view plane into $\mathrm{N_p}$spieces . On each of these $\mathrm{N_p}$ spieces, the identity surface is approximated by a plane suppose the $\mathrm{M_i}$ sample patterns filled by the plane are $(\mathrm{x}_{01},\mathrm{y}_{01},\mathrm{z}_{01})$, $(\mathrm{x}_{02},\mathrm{y}_{02},\mathrm{z}_{02})$,… (x_{0M},y_{0M},z_{0M}). This is a quadratic big problem which can be solved using the interior point method by[26].They can classify the pattern into one of the face classes by computing the distance to all of the identity surfaces as the Euclidean distance between z_0 and the corresponding point on the identity surface called (z) .

$$\mathrm{d}=\|z_0 - z\| \quad (3)$$

An object trajectory is obtained by projecting the face patterns into the KDA feature space. In same time , according to the pose information about the face patterns. They can build the model trajectory on the identity surface of each subject using the same pose information and temporal order of the object trajectory. Those two kinds of trajectories, i.e. object and model trajectories, encode the spatio-temporal information on the tracked face. Hence , the recognition problem can be solved by matching the object trajectory to register for the set of model trajectories. The primitive achievement of trajectory matching face is applying by computing the trajectory distances, it reaches to the time of the frame called (t).

$$\mathrm{d_m} = \sum_{\mathrm{i=1}}^{\mathrm{t}} \mathrm{w_i}\, \mathrm{d_{mi}} \tag{4}$$

where d, the pattern distance between the face pattern catches in the frame and the identity surface of the subject, is computed from (3), and $(\mathrm{w_i})$ term is the weight on this distance.

3.3 Trajectory for sequence position lips

The novel trajectory Lead to the lips sample selection approach is proposed. In training, the image samples are sequences(S) encoded in low-dimensional visual feature vector. The feature vector is used to train HMM trajectory λ model that is a statistical model. The trained model gives the best feature trajectory by using a maximum likelihood (MLL) that is sensitive. The last status is to reconstruct the optimal feature trajectory drawback by $\bar{S}$term in the original high-dimensional sample space. The low-dimensional visual parameter trajectory to samples in the sample space. In implementing used the HMM lead to predicted trajectory $\bar{\bar{V}}$, a smooth image sample sequence$\bar{\bar{s}}$is sought more best from the sample library and the mouth sequence is then returned back to a background primitive recorder for video. The lips images, has a large number of the Eigen lips contained of the accumulated variance. The visual feature of each lips image is formed by its PCA vector, $V^T = s^T$w where w is the projection matrix made by number Eigen lips. We use the specially algorithm to specify to the best visual parameter vector sequence.

$$\mathrm{V} = [V_1^T, V_2^T, \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots V_T^T]^T \tag{5}$$

By giving maximization for the maximum likelihood (LM) algorithm . The HMM predicted visual parameter trajectory had detailed to move a compact description , in the lower level eigen-lips space. However, the lips image sequence shown at the top of is blurred due to dimensionality reduction in PCA and MLL-based model parameter estimation and trajectory is obtained . To solve this blurring, suggest the trajectory is leading to real sample sequence approach to constructing from. Hence , the detailed movement in the visual trajectory is reconstructed and image real sample rendering is truth. This was propose [30].The unit obtained in concatenative speech synthesis , the cost count for a sequence of trajectory called *T* choice samples are the weighted sum of the target and concatenation costs:

$$\mathrm{C}(\hat{\mathrm{v}}_1^T, \hat{\mathrm{S}}_1^T) = \sum_{l=1}^{T} \omega^t\, c^t(\hat{\mathrm{v}}_i, \hat{\mathrm{s}}_i) + \sum_{i=2}^{T} \omega^c\, c^c(\hat{\mathrm{v}}_{i-1}, \hat{\mathrm{s}}_i) \tag{6}$$

The target cost of an image sample (s) is dependent over the measured of the Euclidean distance between their PCA vectors.

The concatenation cost is measured based on the normalized 2-D cross correlation (NCC) between two image samples $\hat{s}_i$ and $\hat{s}_J$. Since the correlation coefficient ranges in value from -1.0 to 1.0, NCC is in nature a normalized similarity score, proposed by [1].

$$c^t(\hat{v}_i, \hat{s}_i) = \|\hat{v}_i - \hat{s}_i^{\boldsymbol{T}}\| \quad (7)$$

4. EXPERIMENT AND RESULTS

4.1 Expire Vector Feature

They created separate shape and appearance model to encode any view independently, hence the shape feature (parameter, p), and app feature (parameter λ) ,this introduced by [10].They need for two phased for associated the shape and appearance parameters. First, they used primitive way by combining the feature vectors (CFV) , which called a cat and the second is concatenating the features and reduce the dimensionality using PCA, proposed by [14] which used (csam*)*feature. CFV improved by using an LDA over window for the set of frames. It is proposed by [15] which went to represented (hilda*)* features in the frontal lip-reading, It has applied for two features are the discriminating , it is introduced by [9] For all features, a z-score normalization is used, which has been visible to develop the separability between the features of the classes by [31] . The best viewing angle for the primitive features, i.e., those that aren't relative to a third PCA or an LDA i.e (shape, app and cat) seems to be more one angle of a view**.**

4.2. Expire for across multiple camera

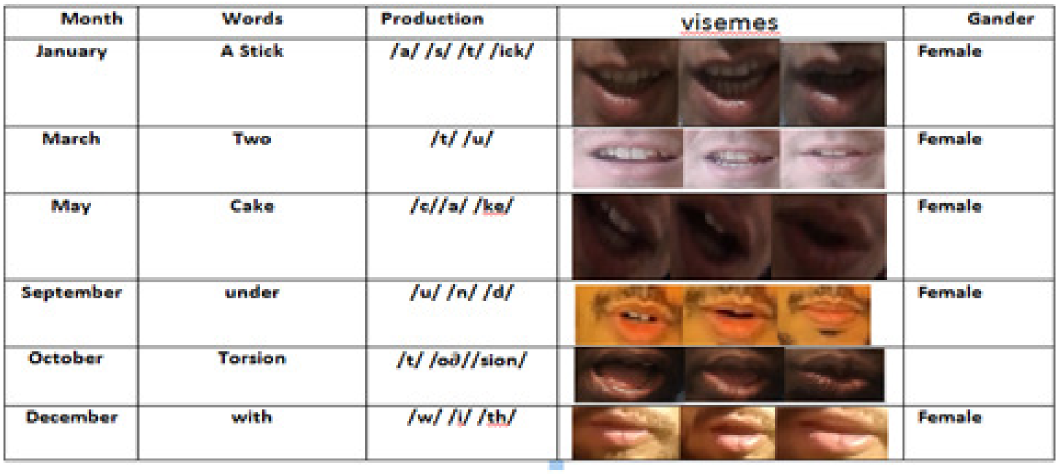
They used audio-visual speech dataset base on called LiLIR. The dataset contains multi-camera, multi-angle recordings of a speaker recite a lesson200 sentences from the supplier arrangement Corpus. The structure and size of the LiLIR dataset has enable to train the hidden Markov models (HMMs) onset of the word for visual speech units, such visemes, hence the number words as representative for the vocabulary of the database and it is approximately 1000 words, datasets used to satisfy automatic lip-reading for the each view camera . The dataset contains multiple type of camera such as two HD cameras recording .As well, there are three SD camera sand 60 viewpoints. All cameras were synchronization locked during recording, proposed by[13][32].

4.3. Expire Result

In[36],used the mouth cavity region implemented in a banalization process, As well , a time change of the two features is expressed as a two-dimensional a trajectory of the lip motion of the target word . Observing these lip images, in the table .1, it can show us the visual speechless person's male concept that 30th frame lip image expressed the shape of ” Stick”, 40st, 30th, and 50th frames expressed ”cake”, ”torsion”, "two" ,"with", and ”under”, respectively. in the table .2, it can show us the visual speechless person's female concept, that 30th frame lip image expressed the shape of ” Stick”, 40st, and 50th frames expressed ”cake”, ” Torsion”, "two" ,"with", and ” Under”, respectively.

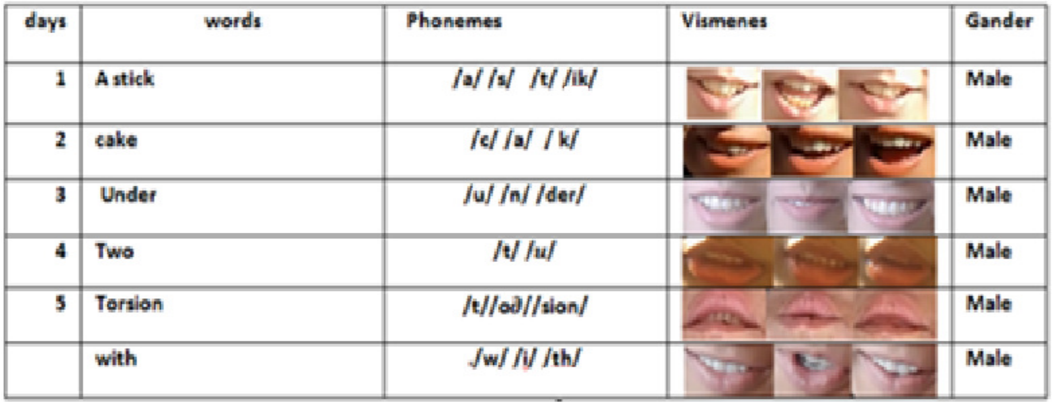
Proposed by [35] ,recorded 50 times for each word , and recorded 500 image sequence one subject. The image size is 320×240 pixels and the frame rate is 30 frame/sec. The lip detection was applying for 500 image sequences. Time passes along the direction of the arrow base on the automatic of the lip in an utterance. It can show us that trajectory in all words is being drawn by the visual observation. The recognition process was applying with feature sets. Figure.1 shows a trajectory which the word is ”a stick ” torsion", "cake" ,"under", "with" and "two", respectively .The horizontal axis is area **s** , and the vertical axis is an aspect ratio area(*A*). Plotted circle marks in this figure are the position of vector features of all the frame, and these marks are connected with a line.

Table 1. Illustrating Visual Speech Person's Female Concept



Month	Words	Production	visemes	Gander
January	A Stick	/a/ /s/ /t/ /ick/		Female
March	Two	/t/ /u/		Female
May	Cake	/c//a/ /ke/		Female
September	under	/u/ /n/ /d/		Female
October	Torsion	/t/ /oə//sion/		
December	with	/w/ /i/ /th/		Female

Table .2.Illustrating Visual Speech Person's Male Concept



days	words	Phonemes	Vismenes	Gander
1	A stick	/a/ /s/ /t/ /ik/		Male
2	cake	/c/ /a/ / k/		Male
3	Under	/u/ /n/ /der/		Male
4	Two	/t/ /u/		Male
5	Torsion	/t//oə//sion/		Male
	with	./w/ /i/ /th/		Male

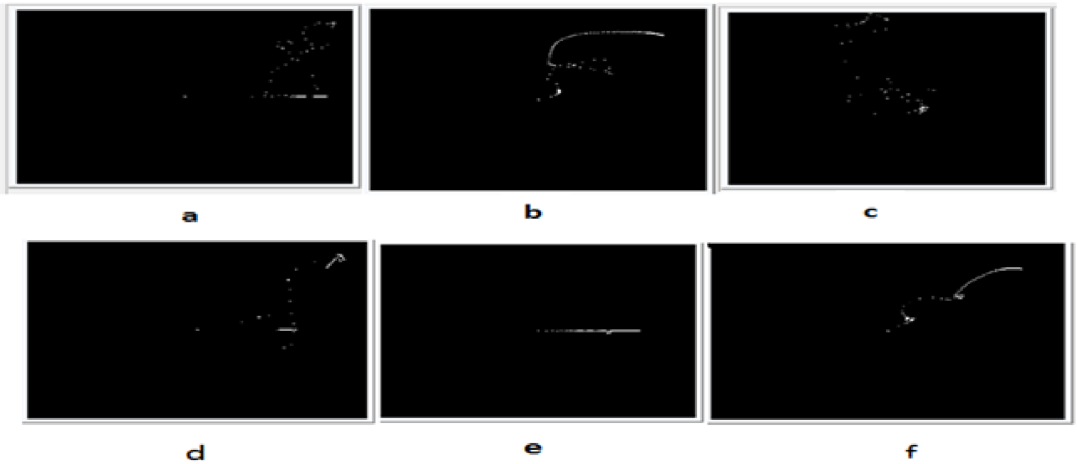


Figure.1. Image of computed trajectory and correspond Lip

5. CONCLUSION

We propose a trajectory-guided, real sample concatenating approach for synthesizing high-quality automatic image -real articulator. Objectively, we evaluated the performance of our system in terms of speaker's face identification by using automatic lip reading represented in the visual domain. The system framework using the signature of the visemes approaches by track trajectory for lip contour extraction   as represented   the whole word. The target word is recognized based on the word's English included two types of gender female and males were shown that recognition using the trajectory vector feature is obtained the vocabulary of the database is approximately more than 100 words, data sets used to satisfy the automatic lip-reading across multi-view camera.

REFERENCES

[1] A. Hunt and A. Black, "Unit selection in a concatenative speech synthesis system using a largespeech database," Proc. ICASSP 1996, pp. 373-376.

[2] D. Sweet's, J. Weng, Using discriminant eigen features for image retrieval, IEEE Transactions on Pattern Analysis and Machine Intelligence 18 (8) (1996) 831–836.

[3] B. Scholkopf, A. Smola, K.-R. Muller, Kernel principal component analysis, in: W. Gerstner, A. Germond, M. Hasler, J.-D. Nicoud (Eds.),Artificial Neural Networks—ICANN'97, Lecture Notes in Computer Science, Springer, Berlin, 1997, pp. 583–588.

[4] V. Roth, V. S. Solla, T. Leen, K.-R. Mu¨ller,"Steinhage, Nonlinear discriminant analysis using kernel functions", in: (Eds.), Advances in Neural Information Processing Systems 12, MIT Press, Cambridge,MA, 1999, pp. 568–574.

[5] S. Sako, K. Tokuda, T. Masuko, T. Kobayashi, and T. Kitamura, "HMM-based Text-To-Audio-Visual Speech Synthesis," ICSLP 2000.

[6] L. Xie, Z.Q. Liu, "Speech Animation Using Coupled Hidden Markov Models," Pro. ICPR'06, August 2006, pp. 1128-1131

[7] W. C. Yau, D. K. Kumar, and S. P. Arjunan. "Visual speechrecognition using dynamic features and support vectormachines, International Journal of Image and Graphics, vol.8,pp. 419-437, 2008.

[8] T. J. Hazen, "Visual model structures and synchrony constraints for audio-visual speech recognition," IEEE Transactions on Speech and Audio Processing, 14(3), (2006),1082-1089.

[9] E. Ong, Y. Lan, B. Theobald, H. R., and R. Bowden, "Robust facial feature tracking using selected multi-resolution linear predictors," in Proc. of ICCV, 2009.

[10] C. G. Fisher, “Confusions among visually perceived consonantsnants,”Journal of Speech and Hearing Research, vol. 11, pp. 796–804, 1968.

[11] T. Cootes, G. Edwards, and C. Taylor, “Active appearance models,” IEEE Trans. Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pp. 681–685, June 2001.

[12] Y. Lan, R. Harvey, B. Theobald, E.-J. Ong, and R. Bowden, “Comparing visual features for lip-reading,” in Proceedings of Proceedings of International Conference on Auditory-Visual Speech Processing, 2009, pp. 102–106.

[13] Y. Lan, B. Theobald, R. Harvey, E.-J. Ong, and R. Bowden, “Improving visual features for lip-reading,” in Proceedings of Proceedings of International Conference on Auditory-Visual Speech Processing, 2010.

[14] T. Cootes and C. Taylor, “Statistical models of appearance for computer vision,” ImagingScience and Biomedical Engineering, University of Manchester, Tech. Rep., 2004.

[15] G. Potamianos, C. Neti, J. Luettin, and I. Matthews, “Audiovisual automatic speech recognition: An overview,” in Issues in Visual and Audio-visual Speech Processing. MIT Press, 2004.

[16] Cetingul, H.E. Yemez, Y. Erzin, E. Tekalp, A.M. ,“Discriminative lip-motion features for biometric speakeridentification”, in IEEE ICIP, 2004, vol.3, pp.2023-2026.

[17] L. Sirovich and M. Kirby, “Low-dimensional procedure for the characterization of human faces,” J. Opt. Soc. Amer., A, vol. 4, pp. 519–524,1987.

[18] M. Turk and A. Pentland, “Eigenfaces for recognition,” J. Cogn.Neurosci., vol. 3, pp.71–86, 1991.

[19] S. Lankton, A. Tannenbaum, Localizing region-based active contours, IEEE Transactions on Image Processing 17 (11) (2008) 2029–2039.

[20] S. Wang, W. Lau, S. Leung, Automatic lip contour extraction from color images, Pattern Recognition 37 7 (12) (2004) 2375–2387.

[21] Y.M.Cheung , X.Liu , X. You .A local region based approach to lip tracking,Int. JournalPattern Recognition,. Vol. 45 Iss. 9 pages 3336–3347, Sep,2012 .

[22] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: active contour models. In Proc. 1st Int. Conf. on Computer Vision, pages 259–268, 1987.

[23] J.J. Koenderink and A.J. Van Doorn. Affine structure from motion. J. Optical Soc. of America A., 8(2):337–385, 1991.

[24] A. Blake, R. Curwen, and A. Zisserman. A framework for spatio-temporal control in the tracking of visual contours. Int. Journal of Computer Vision,11(2):127–145, 1993.

[25] S. Ullman and R. Basri. Recognition by linear combinations of models. IEEE Trans. Pattern Analysis and Machine Intelligence, 13(10):992–1006, 1991.

[26] R. Vanderbei. Loqo: An interior point code for quadratic programming. Technical report ,Princeton University, 1994. Technical Report SOR 94-15.

[27] Ezzat, T. and Poggio, T. 1996. Facial analysis and synthesis using image-based methods. In IEEE International Conference on Automatic Face & Gesture Recognition, Vermont,US, pp. 116–121.

[28] T. Cootes, G. Edwards, and C. Taylor. Active appearance models. In EuropeanConferenceon Computer Vision, volume 2, pages 484–498, Freiburg, Germany, 1998.

[29] T.Vetter,. and V.Blanz, . 1998. Generalization to novel views froma single face image. In FaceRecognition: From Theory to Applications,(Eds.), Springer-Verlag, pp. 310–326.

[30] M. A. Fischler and R. A. Elschlager. The representation and matching of pictorial structures. IEEE. Trans. Computers, C-22(1), 1973.

[31] Y. Lan, B. Theobald, R. Harvey, E.-J. Ong, and R. Bowden, “Improving visual features for Lip-reading,” in Proceedings of Proceedings of International Conference on Auditory-Visual Speech Processing, 2010

[32] P. Price, W. Fisher, J. Bernstein, and D. Pallett, “Resource management RM2 2.0,” Linguistic Data Consortium,Philadelphia, 1993.

[33] C.Bregler., Y,Konig., (1994) “Eigenlips For Robust Speech Recognition”, Proc.ofICASSP'94,Vol. II, Adelaide, Australia, p669-672.

[34] C.Neti ., G. Potamianos., J.Luettin., (2000) “Audio-visual speech recognition”, Final Workshop 2000 Report, Center for Language and Speech Processing, The Johns Hopkins University, Baltimore, MD.

[35] T. Saitoh and R. Konishi, “Lip reading based on sampled active contour model,” LNCS3656, pp. 507-515, September 2005.

[36] M. J. Lyons, C.-H. Chan, and N. Tetsutani, “Mouth Type: text entry by hand and mouth,”Proc. of Conference on Human Factors in Computing Systems, pp. 1383-1386, 2004.

[37] A. Khan, W. Christmas, and J. Kittler. Lip contour segmentation using kernel methods and level sets. In ICVS, volume 4842:II, pages 86–95, 2007.

[38] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham. Active shape models -their training and application. CVIU, 61(8):36–59, 1995.

[39] C. Xu and J.L. Prince. Gradient vector flow: A new external force for snakes. InCVPR, pages 66 – 71, 1997.

[40] K.F. Lai, C.M. Ngo, and S. Chan. Tracking of deformable contours by synthesis and match. In ICIP, volume 1, pages 657 – 661, 1996.

[41] T. Wakasugi, M. Nishiura, and K. Fukui. Robust lip contour extraction using separability of multi-dimensional distributions. In FGR, pages 415 – 420, 2004.

[42] P. Delmas, N. Eveno, and M. Li´evin. Towards robust lip tracking. In ICPR, 2002.

[43] N. Eveno, A. Caplier, and P.Y. Coulon. Jumping snakes and parametric model for lip segmentation. In ICIP, volume 2, pages 867 – 870, 2003.

[44] S. Stillittano and A. Caplier. Inner lip segmentation by combining active contours and parametric models. In VISAPP, pages 297 – 304, 2008.

[45] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. Internation Journal of Computer Vision, 1(4):321 – 331, 1988.

[46] M. S´anchez, J. Matas, and J. Kittler. Statistical chromaticity models for lip tracking with b-splines. In AVBPA, pages 69–76, 1997.

[47] L. Zhang. Estimation of the mouth feaures using deformable templates. In ICIP, volume 3, pages 328 – 331, 1997.

[48] S. Werda, W. Mahdi, and A. BenHamadou. Colour and geometric based model for lip segmentation. In ICIP, pages 9 – 14, 2007.

[49] A.E. Salazar, J.E. Hernandez, and F. Prieto. Automatic quantitative mouth shape analysis. Lecture Notes in Computer Science, 4673:416–423, 2007.

[50] J.A. Dargham and A. Chekima. Lips detection in the normalised rgb colour scheme. In ICTTA, volume 1, pages 1546 – 1551, 2006.

[51] E. Gomez, C. M. Travieso, J. C. Briceno, and M. A. Ferrer. Biometric identification system by lip shape. In ICCST, pages 39 – 42, 2002.

[52] H. Bunke and T. Caelli. Hidden Markov Models: Applications in Computer Vision. World Scientific Publishing Co., 2001.

[53] F. Luthon, A. Caplier, and M. Li´evin. Spatiotemporal mrf approach to video segmentation Application to motion detection and lip segmentation. Signal Processing, 76(1):61 – 80, 1999.

[54] X. Zhang and R.M. Mersereau. Lip feature extraction towards an automatic speechreading system. In ICIP, volume 3, pages 226 – 229, 2000.

[55] Y. Nakata and M. Ando. Lipreading methods using color extraction method andeigenspace technique. Systems and Computers in Japan, 35(3):1813 – 1822, 2004.

[56] K. Fukunaga. Inrtroduction to statistical pattern recognition. Academic Press, 1990.

[57] R. O. Duda, P. E. Hart, and D. G. Stork. Pattern Classification. Wiley, 2001.

[58] J.M. Zhang, D.J. Wang, L.M. Niu, and Y.Z. Zhan. Research and implementation of real time approach to lip detection in video sequences. In ICMLC, pages 2795 – 2799, 2003.

[59] R. Kaucic and A. Blake. Accurate, real-time, unadorned lip tracking. In ICCV, pages 370–375, 1998.

[60] W. Rongben, G. Lie, T. Bingliang, and J. Linsheng. Monitoring mouth movement For driver fatigue or distraction with one camera. In ITSS, pages 314–319, 2004.

[61] A.C. Hulbert and T.A. Poggio. Synthesizing a color algorithm from examples.Science, 239(4839):482 – 485, 1988.

[62] N. Eveno, A. Caplier, and P.Y. Coulon. New color transformation for lips segmentation.IEEE Fourth Workshop on Multimedia Signal Processing, pages 3 – 8,2001.

[63] Y. Wu, R. Ma, W. Hu, T. Wang, Y. Zhang, J. Cheng, and H. Lu. Robust lip localization using multi-view faces in video. In ICIP, pages IV 481 – 484, 2007.

[64] L.E. Mor´an and R. Pinto. Automatic extraction of the lips shape via statistical lips modelling and chromatic feature. In CERMA, pages 241 – 246, 2007

AUTHOR

Adil A. Aboshana received the B.S. mathematical from university salah-adeen , Science Iraq in 1985 and 1989 respectively ,and he received M.S. degrees from department of the information Technology, Science, University of Utara –Malaysia ,in 2007 and 2009, respectively. He is currently studying PhD in the department of information system. His research interests include, pattern recognition and image computer vision, processing with applications to biometrics. He is one of the participants who received paper award at the conference IEE in 2015