International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 01, January-2017

A New Approach for Lip Recognition

Bnar Azad* and Asaf Varol
Software Engineering Department,
College of Technology, Firat University
Elazig, Turkey

***Abstract*—There are many biometric recognition opportunities such as fingerprints, face, eye iris, vein, etc. that can be used for determining a person's identity. In the last decade, many papers have been published on biometric recognitions using lip location and its movements, but there are some difficulties and limitations initiated while using these biometric features for accurate and appropriate identifications. As a respond to these challenges, this theoretical and implemental research has been conducted. Innovations in lip-recognition usually incorporate lip localization and detection, highlight extraction, reading technique, and a combination of these strategies as well as the arrangement of the mouth. This research has investigated the features used for lip segmentation, techniques used for lip localization, lip recognition model, speech database, and mouth classification including the various applications of lip recognition technology. In addition, the research discusses the difficulties and limitations of lip recognition systems, and it also gives some recommendations and suggestions in terms of improving the accuracy of lip recognition using the findings of the practical work.**

Keywords*— *Face Detection; Lip Reading; Mouth Detection; Opencv; Skin Segmentation

I. INTRODUCTION

Lip recognition has recently gained ground in the community of computer visionaries. This growing interest comes from a broad range of applications that rely on the visual information or the visual information system may improve their overall system robustness and performance [1]. These applications include: identifying audio-visual speech, analysis of facial expressions and lip synchronization. Here, the challenge is mainly due to the high lip deformity which causes major changes such as, different lips' colors, lighting conditions, the appearance of the teeth and tongue in the presence of facial hair, and finding a strong lip and non-trivial tasks correctly. In recent years, many techniques have been suggested for detecting lip movement. One of the first methods used to achieve lip segmentation is edge information [2]. The broad techniques group is known as model-based where a model is built for lips and the configuration is to be described by a set of the model's parameters. These lip recognition techniques include deformable template [3], snakes [4], models based on active contour [5] etc. The advantage of these techniques is an important feature of small dimensions space parameter. In addition, they rotate, scale, and light constantly. However, these models are very difficult to build in general and they require a lot of training on their range of high variability cover lips. Moreover, the adjustment parameters are generally very difficult to achieve, and many need to select and reset manually. The color is very useful to recognize lips for the purpose of additional information and has been widely used [3, 4] [6, 7].

Apart from machine learning methods, there have been a number of supplementary ways to detect and read lips such as Haar algorithms. Those methods are injected into the application to observe the differences affecting the performance and accuracy of the project. These changes have positive and negative impacts on lip reading actions that are being explored in this article.

One of the aims of implementing this project is to produce a desktop lip reading application in order to perform machine learning algorithms. These algorithms are being tested with various attributes in this application to achieve the main objective of the research which is recognizing lip movements and convert it to speech. This has been the central goal of the research since the beginning.

Beyond this introductory section, the paper is organized as follows: the major development in lip recognition systems and processes are elaborated and discussed in section II. The technology behind the developments provided in section II is investigated in section III with a major focus on lip detection techniques and their classifications. The architectural design and practical implementation as well as the classification are presented and discussed in section IV, V, and VI respectively. Some useful information on the applications of lip detection technology is provided in section VII and the final remarks and conclusions are summarized in section VIII.

II. LIP RECOGNITION DEVELOPMENT

In 1954 Sumby proposed the idea of lip recognition. Here, the assumption made is based on lip movement as a source of the channel of visual information to understand the tone. Lip recognition aims to explain the movement of the lip which also means that individuals' facial expressions and emotional status can be detected by the movement of lip [8].

In 1984, the first lip-reading system from Illinois University was established by Petajan. Petajan collected movements of lips' information as a supplement to the ear's canal, and the combined reading of lip recognition and recognition of speech model. Reading labial recognition model receives input from the two best candidates per speech,

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 01, January-2017

in this method, the recognition rate or noise plus the speaker environment improvements [9]. Further development breakthrough was initiated in 1989 and 1991 when Artificial Neural Network (ANN) was introduced by Yuhas and Mase from MIT respectively [8].

In 1993, Hidden Markov Model (HMM) was applied in the field of lip recognition by Goldschen and in 1994 HMM and ANN were combined by Bregler and were used to read letters and words that are isolated [10].

In 2010, researchers at the Institute of Technology Karlsruhe in Germany developed a method of converting lip movements into a sound, which may allow phones to read lips. By using the mouth muscle measurements, the telephone will be able to know what is being said. Using this information, it can then transmit this data down the line [33].

In 2013, Microsoft launched Kinect2 which can read lips. Kinect2 accurately captures motion function, the depth information of the lip by the depth sensor to receive recognition of the lip to be identified. This is a combination of the history of lip gloss technological development, the disclosure of the eighties, the nineties are included on the accumulation of twenty-first-century breakthrough. This is realized in lip-reading as an important visual information which is growing hot area of research, not only in theory but also in practical applications. Lip reading software and hardware is growing rapidly, and moving gradually into the practical phase [8].

III. LIP RECOGNITION TECHNOLOGY

A collection of lip constraint strategies has been portrayed in the composition of lip recognition technology through the late 15 years. Conspicuous philosophies base on shading and power thresholding to partition the lips from whatever left from the face [11-13]. By and large, the lips are then arranged by fitting a shape model around the isolated mouth, where various procedures are investigated. Another surely understood procedure is the usage of snakes in the mix with mouth corner part identification [14]. Furthermore, shape designs have been used as a piece of solicitation to restrain the lip frames. Another procedure is to organize the zones in a photo according to the level and vertical power profiles, with unprecedented contemplations of the assorted tossing of shadows in the mouth zone [15]. There are a couple of conveyances that especially focus on consistent lip taking after. Every now and then they use the same strategies as indicated above, perhaps as unraveled or quicken varieties.

A. Features Used in Lip segmentation

The features utilized as a part of lip segmentation can be broadly separated into two primary classifications: appearance based components, and shape-based elements [16].

a) Appearance-based features

These can be basically gotten after face detection. A few strategies are utilized to discover facial components, for example, eyes and nostrils, in light of their relative position on the face. An illustration of the eyes confinement calculations can be found in [17]. Another sample of the appearance - based element incorporates pixel shading data or force esteem. Shading data gives advantages in either removing highlights, for example, lips, or smothering the undesirable ones. The appearance-based elements are ordinarily extricated from Region of Interest (ROI) utilizing picture changes, for example, changes to various shading space segments, where pixel estimations of the face/mouth ROI are utilized [16].

b) Shape-based features

They are by and large separated into geometric, parametric and measurable models of the lip shape, and are extricated utilizing systems, for example, snakes [18], template models [19, 20], appearance model, and active shape and model [21]. This component expects that a large portion of the data is contained fit as a fiddle of speaker's lips. Geometric elements, for example, tallness, width, and the border of mouth, can be promptly separated from the ROI. Then again, model-based components are acquired in conjunction with parametric or factual element extraction calculation.

B. Techniques used in Lip localization

The following are the fundamental techniques used in lip localization.

a) Region based technique

The locale based technique mainly uses the local measurement qualities to acknowledge lip tracking. Ordinary samples incorporate Deformable Template (DT), region based Active Contour Model (ACM), Active Shape Model (ASM), and Active Appearance Model (AAM) [22]. The DT figuring uses a local cost; a lip picture divided by the lip and non-lip ranges by a method for a parametric format, and it addresses the mouth area fittingly. Yuille introduced the initiative effort that demonstrates a lip format controlled by a game plan of parameters [23], and these parameters are changed by a method for a noteworthy minimizing taking care of so that the lip layout can encourage beyond what many would consider possible prices.

Region based technique can be divided into three main classes:

- The deterministic technique which depends on shading appropriation modeling and threading operations.
- The classification technique (managed or non-administered) which considers lip division as pixel class issue, for the most part between skin pixel and lip pixel.
- Lastly, the statistical technique which depends on mouth shapes and appearance. As every one of these techniques is area based, an exact lip division around the lip shapes is not generally accomplished.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 01, January-2017

a) *Deterministic technique*

In this strategy, no earlier information and no former models about the measurements of the lip and skin shading dissemination are used. The lip division is performed with a thresholding venture on the luminance or specific chromatic segment. Programmed calculation of the powerful edges in different lighting conditions is the main test and constraint of this technique [24]. One approach to determining the edge appears in [25]. With an example picture information set, the histogram limits esteem for every chrominance part shifts from the most reduced worth in the most astounding quality.

b) Classification technique

With face detection as a preprocessing step, the lip division can likewise be seen as a grouping issue with two classes: the lip class and the skin class. Using distinctive characteristics characterizing every class, the grouping technique used in face detection between the skin and non-skin class can likewise be connected to lip division [16]. The most commonly used strategy includes factual techniques (estimation hypothesis, Bayesian choice hypothesis, and Markov random field), neural systems, bolster vector machine and fluffy C-mean. They can be ordered into a regulated and unsupervised methodology. Administered strategies use earlier information about the distinctive classes. It involves the development of the training information set that covers an extensive variety of classes and ecological conditions.

c) Statistical technique

Another administered technique is the statistical shape models in which the training set is incorporated to depict the lip shape or appearance variety instead of the lip shading circulations. A shape model is found out from training a set of clarified pictures. The Principle Component Analysis (PCA) produces a little arrangement of parameters to drive the model. By minimizing a cost, the capacity of the component, the parameters are shifted to taking the limp form. This strategy is known as ASM [20]. AAM was introduced by Cootes and Taylor to include dim level appearance information in the training. Reference [20] is discussing a sample of ASM and AAM.

d) Contour based technique

It is a standout amongst the essential strategies for applications of human-machine, for instance, lip-acknowledgment and outward appearance examination [25]. Incidentally, it may be unnecessary to find a strong and exact extraction methodology because it can produce significant image errors. This can be because the speakers can have different skin tones which are light and dark, and the tongue and teeth may appear which changes the color of the mouth between the lips. There can be little difference in contrast between the facial skin and the lips, as well as the lip shapes may be different between different people. By the earlier decade, different procedures needed to be suggested to complete the lip structure, system. There divided into two classes: model-based methodology and edge-based strategy.

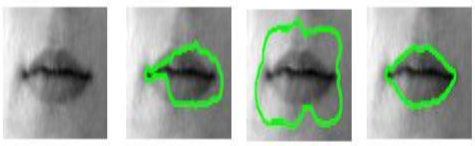
e) Edge-based strategy

It mainly uses the low-level spatial prompts, for example, edges, and shading, to fulfill lip confinement and eradication. The execution of such a system frequently rots when there is a poor many-sided quality amidst lip and encompassing skin regions [35].

f) Model-based methodology

This model forms a lip model and a little course of action of lip-parameters that mostly beats the prior one. Tests incorporate DT and ASM [25].

The conventional ACM grants an underlying shape to turn by minimizing a specific overall imperative ability to convey the pined for the division. Reference [25] is presenting the accomplishment of this methodology in its application area. This framework may delineate the exact parameters, uneven enlightenment, and teeth sway. Moreover, while things have heterogeneous estimations, it is found that the limited locale based ASM [26] can overall perform a satisfactory division result while the routine ACM misses the mark. From the LACM, the advancing bend parts in those close-by neighborhoods under the area inside the range and neighborhood outside the district independently. As need be, the limited essentialness for advancing and removing can be figured. In any case, less than ideal parameters, for instance, broad reach or distance away advancing twist in LACM can provoke the incorrect extricating results as shown in Fig. 1.



a b c d

Fig 1. (a) Mouth picture. (b) Extricated outcome by ACM. (c) Extracted outcome by LACM. (d) Separated outcome by the LACM with the best possible parameters [32].

IV. ARCHITECTURAL DESIGN

The lip recognition system using OpenCV described in this paper, as shown in Fig. 1, can be divided into the following three stages:

- The lower part of the face is extracted from the video sequences of the speaker using the Harr cascade.
- The mouth and lip are detected also by using Harr cascade.
- The classification is done using some the methods of machine learning such support vector machine and K-Nearest Neighbor (K-NN). The output of the system is speech detection. The architectural design showed in Fig. 2.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 01, January-2017

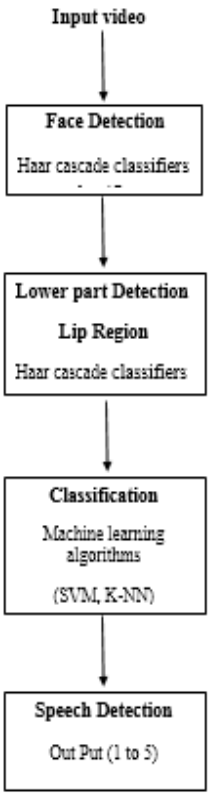


Fig. 2. System design steps

V. FACE AND LOWER PART DETECTION

The speaker images are extracted from the video. Face and mouth region are detected by using Haar cascade. Applying the face detection process then cropping the mouth region followed by mouth detection process. There are a lot of techniques for extracting the mouth region, some of which were discussed in previous sections. The machine learning (Viola Jones object recognizer) [31] is employed both to distinguish between the lower and upper parts of the face and to isolate the mouth area in the lower part of the face. The procedure requires two steps, the first is detecting the facial region of the person, and the second is using the fact that the mouth area of the face is in the lower half of the face as shown in Fig. 3.

VI. CLASSIFICATION

In this paper, K-NN and SVM are the two used algorithms for classification purpose. Support Vector Machine (SVM) is the algorithm used in computer vision. This method is working with higher dimensional space, by combining the features the new dimension are created. K-NN is also used for classification, and the class with the highest frequency using the K-most algorithm is selected as the output for each instance. Each instance is counted as a vote for their class, and the class with highest votes wins. At first, face and facial part are detected, then separated. This secures some information about the upper and lower parts. In further steps, the classification of a given sequence of images in speechless and speech images is conducted, as shown in the results presented in Fig. 4.

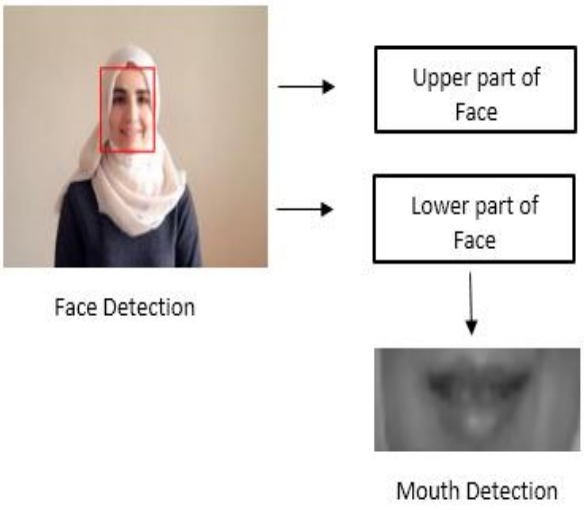


Fig. 3. Face and mouth detection.

Fig. 4. Sequence of images speechless and speech

VII. APPLICATIONS OF LIP RECOGNITION TECHNOLOGY

A. Aid Signals Acknowledgment

Motions and lip-acknowledgment are required. When they watch developments, they need a portable amplifier calms watch and moving shape and likewise facial appearance and lip moving. The conducted experiment on mouth shape movement acknowledgment can upgrade clarity of development acknowledgment. Its planned framework is the same as that of discourse acknowledgment.

B. Aid Discourse Acknowledgment

It is a principal usage of the lip-acknowledgment to join discourse acknowledgment. In the world of different situations and speakers, it can improve the rate of discourse acknowledgment to join the lip-acknowledgment and discourse acknowledgment.

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 01, January-2017

C. The Human-PC Interaction

Speech-recognition development ability to be related in the range of somatotype communication will get fresh method for human-computer interaction, for example, lip information methodology.

D. The Mouth-coding and Mouth Combination

As one might expect, the most important emotional indicator is the sentiment words, also called opinion words. These are words that are normally used to express positive or negative sentiments. For instance, great, awesome, and stunning are certain positive sentiment words, and awful, poor, and unpleasant are negative sentiment words. Aside from individual words, there are additional expressions and sayings, e.g., cost somebody dearly. Sentiment words and expressions are instrumental in sentiment analysis for evident reasons. A rundown of such words and expressions is known as a sentiment dictionary (or opinion vocabulary). Throughout the years, analysts have composed various calculations to assemble such vocabularies.

E. Authentication and Security

Biometric studies with the aid of the lip were investigated [27]. In [28], it demonstrates a part extraction technique that is based on the quadratic presentation framework for interior lip shapes and pictures were changed over into shading area to color area which is used to diminish influences of human's teeth. In [29], sound and apparent modalities are essential for discourse acknowledgment. It demonstrates another cross breed approach too oversees lip limitation and taking after. In [30], to beat the officially communicated lighting issues, different strategies with the iterative system, which basically depend on the conventional for structure, are proposed. They upgraded the dynamic structure model to see the structure.

F. Deaf-quiet Guide Instruction their Discourse Capacity

It is astoundingly chief for us to add to a mouth shape and the talk structure for the educators of needing a listening device calm school to demonstrate practically hard of hearing calm understudies.

VIII. CONCLUSION

In this paper, a new approach for lip recognition system was presented. This approach used the Haar cascade and two algorithms from machine learning: Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN). The Haar cascade was used to find the speaker's face and then the mouth region of the face. Both SVM and KNN were used to classify the data, however, for less than ten shapes, K-NN worked better on its own. The developed approach is able to detect and recognize lips effectively and accurately compared to the methods found in the literature. From the results analysis, it is clearly found that the method is superior because it allows the user to adapt the system for different shapes of faces. The lips were found automatically in the lower part of the mouth. The frame rate was adjustable according to the capabilities of the machine. A higher frame rate means much more data to process and thus will be too heavy for a slower machine. The new system has the ability to accept more training which means that additional face shapes different from those encountered may be used. This will help it to be more adaptive than other techniques. Although the proposed method has demonstrated a superior performance; yet, further validation and practical development are needed to ensure the feasibility of the proposed system. As further development is needed, the system should be re-evaluated and modified in order to obtain higher accuracy and better performance.

REFERENCES

[1] E. Skodras & N. Fakotakis, An unconstrained method for lip detection in color images, Acoustics, Speech and Signal Processing (ICASSP), 2011 IEEE International Conference, (2011) 1-4

[2] X. Zhang, and R.M. Mersereau, "Lip Feature Extraction towards an Automatic Speech Reading System", Int. Conf. on Image Processing (ICIP'00), Vancouver, 2000.

[3] A.W-C. Liew, S.H. Leung, and W.H. Lau, "Lip Contour Extraction from Color Images Using a Deformable Model", Pattern Recognition 35, Elsevier, pp. 2949-2962, 2002.

[4] P. Delmas, P.Y. Coulon, and V. Fristot, "Automatic Snakes for Robust Lip Boundaries Extraction", Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP'02), Phoenix, 2002.

[5] X. Liu, Y. Cheung, M. Li, and H. Liu, "A Lip Contour Extraction Method Using Localized Active Contour Model with Automatic Parameter Selection", Int. Conf. on Pattern Recognition (ICPR'10), Istanbul, 2010.

[6] N. Eveno, A. Caplier, and P.Y. Coulon, "A Parametric Model for Lip Segmentation", ICARCV'O2, Singapore, 2002.

[7] U. Meier, R. Stiefelhagen, J. Yang, and A. Waibel , "Towards Unrestricted Lip Reading", Int. Journal of Pattern Recognition and Artificial Intelligence (IJPRAI), 2000.

[8] G-L. Zheng, M-L. Zhu, & L-L. Feng, Review of lip-reading recognition, 2014 Seventh International Symposium on Computational Intelligence and Design (2014) 293-298.

[9] E. D. Petajan. Automatic lipreading to enhance speech recognition [D]. Ph. D. thesis, University of Illinois at Urbana_Champain,1984.Proceedings of the IEEE Communication Society Global Telecommunications Conference, November 26-29, Atlanta, Georgia

[10] C. Bregler, S .M. Omohundro. Surface Learning with Applications to Lipreading [M]. In J. D. Cowan, G. Tesauro and J. Alspector, editors, Advances in Neural Information Processing Systems, volume 6 .Morgan Kaufmann, 1994.

[11] X.J. Zhang, H.A. Montoya, and B. Crow. Finding Lips in Unconstrained Imagery for Improved Automatic Speech Recognition. Lecture Notes in Computer Science, 4781:185, 2007.

[12] LEL Moran and RE Pinto. Automatic extraction of the lips shape via statistical lips modeling and chromatic feature. In Electronics, Robotics and Automotive Mechanics Conference, 2007. CERMA 2007, pages 241–246, 2007.

[13] F. Schneiter. Lip Contour Localization using Statistical Shape Models. Master Thesis Supervised by Gabriele Fanelli Computer Vision Institute, Department of Electrical Engineering & Information Technology, ETH Zurich, Switzerland ¨ Summer 2009.

[14] W.C. Ooi, C. Jeon, K. Kim, D.K. Han, and H. KO. Effective Lip Localization and Tracking for Achieving Multimodal Speech Recognition. Pages 90–93, 2008.

[15] M. Li and Y. Cheung. Automatic lip localization under face illumination with shadow consideration. Signal Processing, 2009.

[16] B. N. Husain. FACE DETECTION AND LIP LOCALIZATION. A Thesis presented to the Faculty of California Polytechnic State University, San Luis Obispo, 2011. 1-121.

[17] Liew, Alan W., and Wang, Shilin. "Lip Modeling and Segmentation." Visual Speech Recognition: Lip Segmentation and Mapping. Hershey PA, USA: IGI Global, 2009. 70-127

(This work is licensed under a Creative Commons Attribution 4.0 International License.)

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 6 Issue 01, January-2017

[18] Bouvier, C.; Coulon, P.-Y.; Maldague, X.; "Unsupervised Lips Segmentation Based on ROI Optimization and Parametric Model," Image Processing, 2007. ICIP 2007. IEEE International Conference on , vol.4, no., pp.IV-301-IV-304, Sept. 16 2007-Oct. 19 2007
[19] Bouvier, C.; Coulon, P.-Y.; Maldague, X.; "Unsupervised Lips Segmentation Based on ROI Optimization and Parametric Model," Image Processing, 2007. ICIP 2007. IEEE International Conference on , vol.4, no., pp.IV-301-IV-304, Sept. 16 2007-Oct. 19 2007
[20] Gacon, P.; Coulon, P.-Y.; Bailly, G.; "Statistical active model for mouth components segmentation," Acoustics, Speech, and Signal Processing, 2005. Proceedings. (ICASSP '05). IEEE International Conference on, vol.2, no., pp. ii/1021- ii/1024 Vol. 2, 18-23 March 2005.
[21] Sadeghi, M.; Kittler, J.; Messer, K.; , "Modeling and segmentation of lip area in face images," Vision, Image and Signal Processing, IEE Proceedings - , vol.149, no.3, pp. 179- 184, Jun 2002
[22] Y-M. Cheung, X. Liu, X. You. A local region based approach to lip tracking. Pattern Recognition 45 (2012) 3336–3347
[23] A.L. Yuille, P.W. Hallinan, D.S. Cohen, Feature extraction from faces using deformable templates, International Journal of Computer Vision 8 (2) (1992) 99–111.
[24] Zhang, X.; Mersereau, R.M.; "Lip feature extraction towards an automatic speech reading system," Image Processing, 2000. Proceedings. 2000 International Conference on , vol.3, no., pp.226-229 vol.3, 2000
[25] Dargham, J.A.; Chekima, A.; "Lips Detection in the Normalised RGB Colour Scheme," Information and Communication Technologies, 2006. ICTTA '06. 2nd, vol.1, no., pp.1546-1551.
[26] Leung, Shu-Hung; Wang, Shi-Lin; Lau, Wing-Hong; , "Lip image segmentation using fuzzy clustering incorporating an elliptic shape function," Image Processing, IEEE Transactions on , vol.13, no.1, pp.51-62, Jan. 2004.
[27] B. S. Devi. LIP RECOGNITION FOR PERSON AUTHENTICATION. International Journal of Advanced Research in Electronics and Communication Engineering (IJARECE) Volume 4, Issue 4, April 2015, 1-4.
[28] Chen, Q.C. Deng, G.H. Wang, X.L and Huang, H.J. (2006) „An inner contour based lip moving feature extraction method for Chinese speech", Proceedings of International Conference on Machine Learning and Cybernatics , pp. 3859–3864.
[29] Wei, W.C. Jeon, C. Kim, K. Han, D.K and Ko, H. (2008) „Effective lip localization and tracking for achieving multimodal speech recognition", Proceedings of IEEE International Conference on Multisensor Fusion and Integration for Intelligent systems. pp. 90–93.
[30] H. Seyedarabi, W. Lee, and A. Aghagolzadeh, "Automatic lip tracking and action units classification using two-step active contour and probabilistic neural networks," in Proc. Can. Conf. Elect. Compute. Eng., 2007, pp. 2021–2024.
[31] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: IEEE Conference on Computer Vision and Pattern Recognition, vol. 1, 2001, pp. 511–518.
[32] L.Xin, M.Yiu, L.Meng and L.Hailin,"A Lip Contour Extraction Method Using Localized Active Contour Model with Automatic Parameter Selection", International Conference on pattern Recognition 2010.
[33] http://www.telecomspace.com/content/cebit-2010-silent-sound-technology-endless-possibilities

(This work is licensed under a Creative Commons Attribution 4.0 International License.)