

## PARAMETRIC REPRESENTATION OF THE SPEAKER'S LIPS FOR MULTIMODAL SIGN LANGUAGE AND SPEECH RECOGNITION

D. Ryumin <sup>a, b</sup>, A. A. Karpov <sup>a, b</sup>

<sup>a</sup> St. Petersburg Institute for Informatics and Automation of the Russian Academy of Sciences (SPIIRAS), Saint-Petersburg, Russian Federation - dl\_03.03.1991@mail.ru, karpov@iias.spb.su

<sup>b</sup> ITMO University, Saint-Petersburg, Russian Federation - daryumin@corp.ifmo.ru

Commission II, WG II/5

**KEY WORDS:** Sign language, Gestures, Speech recognition, Computer Vision, Principal Component Analysis, Machine learning, Face detection, Linear contrasting

### ABSTRACT:

In this article, we propose a new method for parametric representation of human's lips region. The functional diagram of the method is described and implementation details with the explanation of its key stages and features are given. The results of automatic detection of the regions of interest are illustrated. A speed of the method work using several computers with different performances is reported. This universal method allows applying parametrical representation of the speaker's lips for the tasks of biometrics, computer vision, machine learning, and automatic recognition of face, elements of sign languages, and audio-visual speech, including lip-reading.

### 1. INTRODUCTION

According to the World Health Organization's (WHO) statistics, in 2015 there were about 360 million people (5%) in the world (328 million adults and 32 million children) suffering from hearing loss (WHO, 2015). Currently, governments of many countries in collaboration with international research centers and companies are paying great attention to the development of smart technologies and systems based on voice, gesture and multimodal interfaces that can provide communication between deaf people and those with normal hearing (Karpov and Zelezny, 2014).

Deaf or "hard of hearing" people communicate by using a sign language (SL). All gestures can be divided into static or dynamic (Plouffe and Cretu, 2016). Sign languages recognition is characterized by many parameters, such as the characteristics of the gestural information communication channel, the size of the recognition vocabulary, variability of gestures, etc. (Zhao et al., 2016). The communication of information is carried out using a variety of visual-kinetic means of natural interpersonal communication: hand gestures, facial expressions, lip articulation (Karpov et al., 2013).

In addition, future speech interfaces for intelligent control systems embrace such areas as social services, medicine, education, robotics, military sphere. However, at the moment, even with the availability of various filtering and noise reduction techniques, automatic speech recognition systems are not able to function with the necessary accuracy and robustness of recognition (Karpov et al., 2014) under difficult acoustic conditions. In recent years, audiovisual speech recognition is used to improve the robustness of such systems (Karpov, 2014). This speech recognition method combines the analysis of audio and visual information about the speech using machine vision technology (so-called automatic lip reading).

### 2. DESCRIPTION OF THE PARAMETRIC REPRESENTATION METHOD

Parametric video representation methods belong to the methods of low and medium processing levels, as informative features, extracted from the input video stream, are calculated at the output.

A functional diagram of the method parametric lips representation is presented in Figure 1. The data from recorded video files or directly from the video camera are delivered at the output of the developed method. Stream processing of video frames is carried out. The image of each frame is subjected to linear contrasting (Al-amri et al., 2010). Then a graphical face area is being determined in the full image, using the Viola-Jones algorithm (Viola and Jones, 2004 and Castrillyn et al., 2011). The complexity of finding faces in the frames is due to the presence of different positions and facial expressions. Next, the eye region in the found face area is being determined using a classifier. The found region is being colored in monochrome black. Nose and mouth regions are being found using classifiers trained on the images. The remaining steps are identical to that for the eyes search, except that the mouth region is not colored but is converted from color image into a grayscale image, and also the color histogram is being equalized (Yoon et al., 2009). The final stages of the method perform the following:

- Change the image size that contains the mouth region;
- Analyze principal components of the features (Candes et al., 2011) (Principal Component Analysis, PCA);
- Saving the counting results in a special text file.

Stepping through of the method looks as follows. Input parameter of the method is the full path to the video file in the format Audio Video Interleave (AVI). Resolution of video frames should be at least 640x480 pixels in the 24 bits format and with a frequency of 30 Hz.

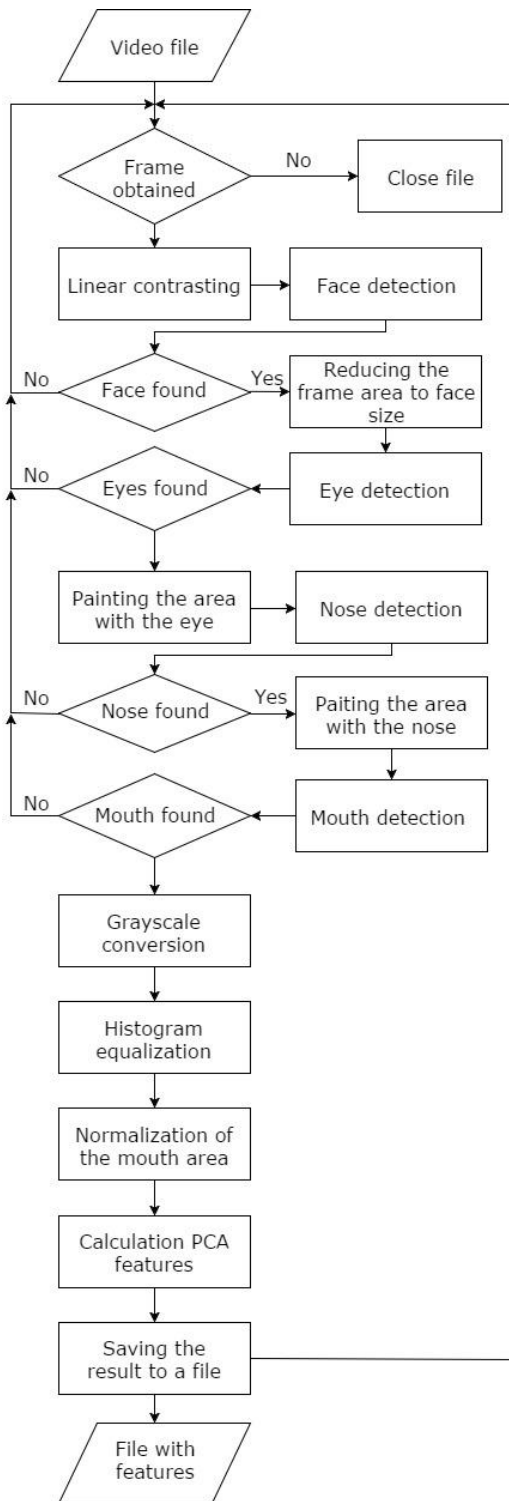


Figure 1. Functional diagram of parametric video representation method

The memory of the storage device in the size of 1 byte (8 bits) is given for digital representation of individual pixel  $x$  of any of the channels (red, green, blue). This means that brightness values are in the range  $[0, 255]$ . At the same time the black color corresponds to the value 0, and the white - value 255. If to take  $x_{\min}$  for minimal brightness, and  $x_{\max}$  - for maximum brightness, can determine how they obtained values differ from the brightness range. The image looks unsaturated and

unacceptable for further digital processing in the event of significant differences of the compared ranges. In this regard, each iteration begins with the improvement of the dynamic range of the produced colour image (linear contrasting) using element wise linear transformation, illustrated by the following formula:

$$y = a \times x + b \quad (1)$$

where  $a$  = minimum output brightness  $y_{\min}$   
 $b$  = maximum output brightness  $y_{\max}$

In the end the system of equations looks like:

$$\begin{cases} y_{\min} = a \times x_{\min} + b \\ y_{\max} = a \times x_{\max} + b \end{cases} \quad (2)$$

which are given to the type regarding the values  $a$  and  $b$  :

$$y = \frac{x - x_{\min}}{x_{\max} - x_{\min}} (y_{\max} - y_{\min}) + y_{\min} \quad (3)$$

Examples of linear contrasting at  $y_{\min} = 0$  and  $y_{\max} = 255$  are presented in Figure 2b. Comparing the original images (Figure 2a), with the processed images. (Figure 2b) one can see the significant improvement in the visual quality of the latter.

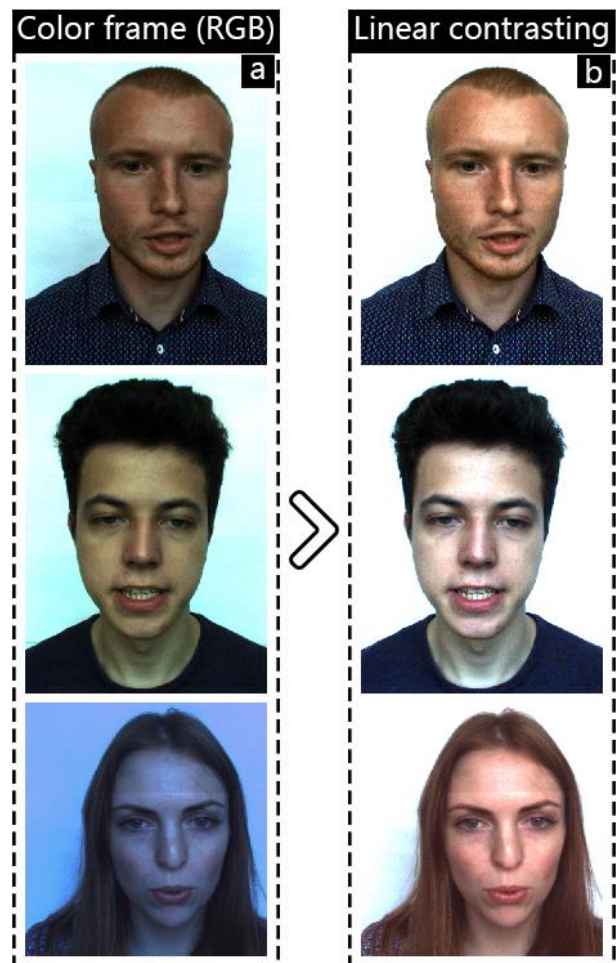


Figure 2. Examples of linear contrasting

Next, search of a graphic speaker's face carries out using the Viola-Jones algorithm, which has been offered Michael Jones and Paul Viola (Viola and Jones, 2004 and Castrillyn et al., 2011). This algorithm allows finding a variety objects in the image in real time. Its main task is detection of persons, however can be used also in different kind of object recognition. This algorithm is implemented in the Open Source Computer Vision Library (OpenCV, 2016) and has the following characteristics: the image has the integral representation (Gonzalez and Woods, 2007), which allows finding the necessary objects at a high speed:

- The necessary object is being searched using Haar's 19 signs (Haar Cascades, 2008) and (Padilla et al., 2012);
- Boosting is used, which is the selection of the most suitable characteristics for the desired object in the selected part of the image (Viola and Jones, 2004);
- A classifier previously trained on faces, (Bruce et al., 2016), which are scalable up to the 20x20 pixels size, accepts features at the input, and outputs the binary result "true" or "false";
- Cascades of signs, consisting of several classifiers, are used to quickly discard windows in which the object is not found (Felzenszwalb et al., 2010).

Educating of classifier takes long time, but the search for objects (faces) is fast. In addition, it has a quite low probability of false positive as the search is carried out on the principle of a scanning window the general view of which is as follows. Let  $R$  us all spatial area occupied by the image which includes the desired object (face) and is represented in an integral form. The integral representation of the image in the Viola-Jones algorithm is aimed a quick calculation of the total brightness of an arbitrary rectangular area on the input image. The calculation time remains unchanged in relation to the space of the chosen area.

The integral representation of the image is a matrix  $I$ , whose dimensions coincide with the dimensions of the original image. Each element stores the sum of intensity of all pixels located to the left and above this element. Elements of the matrix  $I$  are calculated by the following formula:

$$I(x, y) = \sum_{i=0, j=0}^{i \leq x, j \leq y} \times R(i, j) \quad (4)$$

where  $R(i, j)$  = the value of the brightness of the pixel of the original image

Any element of the matrix  $I(x, y)$  stores the sum of pixels in the rectangular from  $(0,0)$  to  $(x, y)$ , in other words value of the each pixel  $(x, y)$  is equal to the sum value of all pixels located to the left and above this pixel  $(x, y)$ . Calculation of a matrix happens on the following formula:

$$I(x, y) = R(x, y) - I(x-1, y-1) + I(x, y-1) + I(x-1, y) \quad (5)$$

Processing takes a linear time which is proportional to number of pixels in the image, due to this the integral image is calculated in one pass. In addition, the sum of the pixels of the rectangle having an arbitrary space is possible to calculate using such integral matrix in a short period of time.

$$rec\ tan\ gle_i = \{x, y, w, h, \alpha\} \quad (6)$$

where  $x, y$  = the origin of the rectangular area regarding  $I$   
 $w$  = width  
 $h$  = height  
 $\alpha$  = the angle of the rectangular area

Receiving signs is made using the Haar's filter. It compares the brightness in two rectangular areas of the image. The standard Viola-Jones algorithm uses Haar's primitives presented in Figure 3.

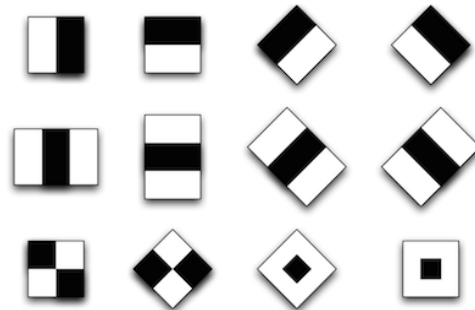


Figure 3. Haar's primitives in the standard Viola-Jones algorithm

The library of computer vision of OpenCV uses an expanded Viola-Jones algorithm with additional signs which are presented in Figure 4.

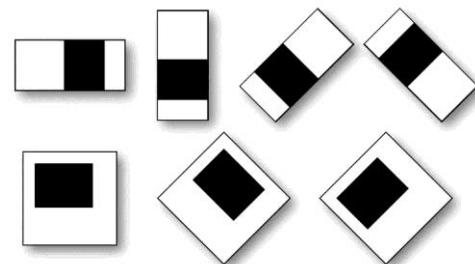


Figure 4. Additional Haar's primitives

The value of any of the given signs is calculated on the basis of the expression:

$$F = X - Y \quad (7)$$

where  $X$  = the sum of values of the brightness of points located on white part of sign

$Y$  = the sum of values of the brightness of points located on black part of sign

These calculations are performed using the integral representation of the image presented earlier and allow to determine the exact value of differences of the brightness on axes  $X$  and  $Y$ .

In the end the principle of the scanning window is based on the following steps:

- The scanning window with a step to 1 pixel and the size which is equal to the predetermined values is sequentially moved along the original image;
- The location of the signs using the scaling and rotation of the signs in the scanning window is calculated in each window;
- The scanning is sequentially happened at different scales;

- The image remains unchanged, the only scanning window is scaled (the size of the window is increased);
- The found signs as input parameter are passed on the cascade classifier, which gives the result.

In the process of search is not prudent to calculate all the signs, because time of treatment will increase. Therefore, the classifier is trained to react only to a certain subset of signs. This problem of classification in the Viola-Jones algorithm is solved with the help of machine learning and the Adaptive Boosting method (AdaBoost, adaptive amplification of classifiers) (Wu and Nagahashi, 2014). Some examples of determination of the speaker's face are presented in Figure 5.

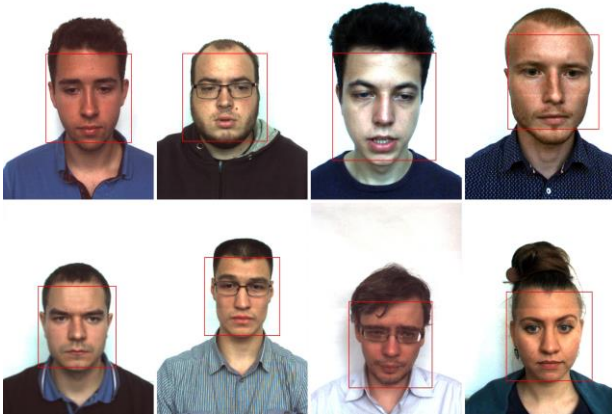


Figure 5. Examples of determination of the speaker's face

A distinctive feature of the method of parametric lip representation, presented in Figure 1, is the obligatory face area detection only in the first frame. Further, in case of receiving a negative result from the classifier, the coordinates will be drawn from the previous successful operations.

The next operation is aimed at reducing of the frame to area with the face (Figure 6). Cutting off the excessive area of the image allows reducing the further processing time.

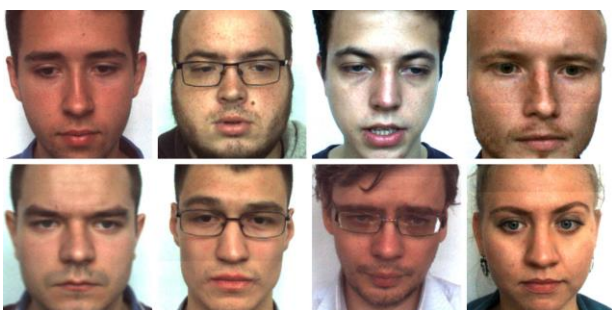


Figure 6. Examples of reduction of the frames to area with the speaker's face

The next step of the method submitted is aimed searching for the speaker's eye using the Viola-Jones algorithm. Instead of the detector of face is used the detector of eye, which provides models a graphical eye region trained on the images of the 11x45 pixels size. This search is performed only if the face area is successfully found.

As is the case with face area detection, eyes needed to find only in the first frame. Subsequently, any mistake is will lead to borrowing of coordinates of the eye region from the previous

successful operations. Some examples of determination of the speaker's eye are presented in Figure 7.



Figure 7. Examples of determination of the speaker's eye

Next, the eye region is colored in monochrome black (Figure 8). Such procedure allows excluding cases of definition of the nose region in the eye region.

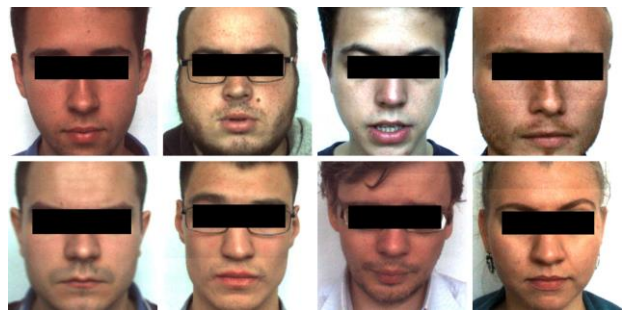


Figure 8. Examples of the monophonic painting of the eye region

Further the detector of eye is replaced by the classification detector of the nose region which is trained on the images of the 15x18 pixels size. Current search is performed only if the face area, and then eyes is successfully found.

As is the case with the previous area detection, the nose region had to be found only in the first frame. If there is an error extraction of coordinates for necessary area is made the previous successful operations. Some examples of determination of the different speaker's nose are presented in Figure 9.

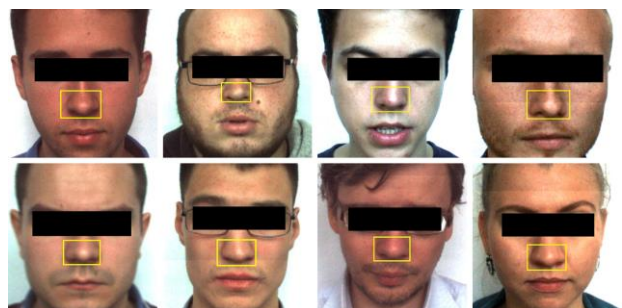


Figure 9. Examples of determination of the different speaker's nose

The nose region is colored by monochrome black (Figure 10) to exclude the possibility of further incorrect classification of lips, as was done with the eye region earlier.

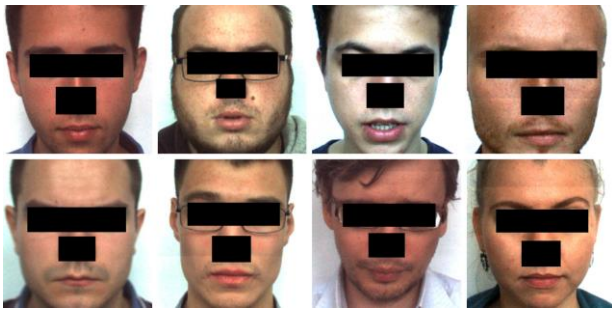


Figure 10. Examples of the monophonic painting of eye and nose regions

The Viola-Jones algorithm classifies the lip region using a mouth detector (trained on the images of the 15x25 pixels) at the end of recognition. This search will be implemented only in case of successful determination of all previous areas (face, eye, and nose). In analogy to previous activities, the lip region had to be correctly found only in the first frame. If there is an error extraction of coordinates for necessary area is made the previous successful operations. Some examples of determination of the different speaker's lips are presented in Figure 11.



Figure 11. Examples of determination of the different speaker's lips

Before calculating Principal Component Analysis (PCA) of the features of the lips region a grayscale transformation must be fulfilled (Figure 12a), which is based on transformation of the making pixels of red, green and blue channels of the color image in the corresponding brightness values. Then, leveling the histogram the resulting type of which must respond to a uniform distribution to realize for the levels of brightness of image to the same frequency (Figure 12b).

The process looks as follows in detail. The image has dimension  $M$  of pixels down and  $N$  pixels across and also the level of quantization of brightness  $J$ . The general dimension of a matrix of the original image is equal  $M * N$ . It followed that on average one level of brightness contains the number of pixels calculated on a formula:

$$n_0 = \frac{M}{J} \times N \quad (8)$$

The distance between discrete brightness levels from  $f_i$  to  $f_{i+1}$  in the histogram of the original image is equally in the histogram of the original image  $\Delta f$ , but each level of the histogram contains different number of pixels. After the alignment procedure, the distance  $\Delta g_i$  between  $g_i$  and  $g_{i+1}$  is distinguishable, and the number of pixels at each level, on average, one and is equal  $n_0$ . For this purpose the sum of all  $f_i$  is starting from 0, in increments of 1, until the value  $n_0$  had been reached. Then the level of initial  $f_i$  are assigned to all

summed pixels, after further processing is carried out starting with level  $f_i$  at which it was stopped. In case taken singly level exceeds the values  $n_0$  in  $q$  time, the connection  $q$  levels remain unfilled.

Finally, corresponding rectangular regions are normalized to the size 32x32 pixels (Figure 12c).

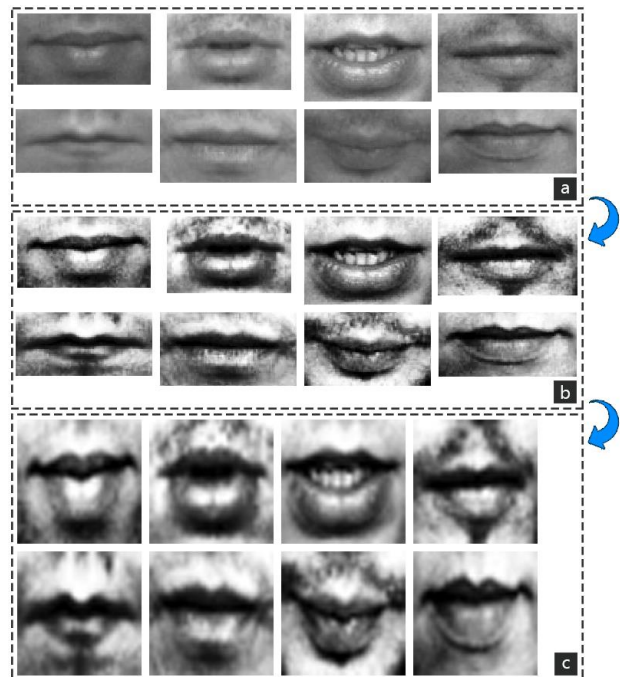


Figure 12. a) grayscale transformation, b) leveling of histogram, c) normalization of mouth region

The pixel visual signs using PCA of visual of the lips region are applied as the visual signs describing a shape of lips of the person (sound). This method is implemented in the library of computer vision of and is applied to compression of information without essential losses of informational content. It's to a linear orthogonal transformation of the input vector  $X$  dimension  $N$  of the output vector  $Y$  smaller dimension  $M$ . And components of the vector  $Y$  are uncorrelated and the general dispersion after transformation remains. The matrix  $X$  consists of all examples of images training set. The choice of the first  $M$  principal component breaks the vector space on the principal (own) space containing of principal components and its orthogonal addition. In this case, the input vectors represent the images centered and provided to a single scale in the task of images analysis. The input image degraded into a set of linear coefficients called by principal components using pre-processing display matrices. The sum of principal components multiplied by the corresponding own vectors is reconstruction of the image. The principal components (several tens) are calculated for each graphical interest area, all other components code only small differences between the standard and noise, and its can be not considered.

PCA projection is precomputed from a set of training images containing the mouth region of a man (Kaehler and Bradsky, 2017). To calculate the PCA projection are used the following data:  $U = (u^1, u^2, \dots, u^N)^T$  -  $N$  - dimension vector containing the sequence of pixels of some image with sizes

$W \times H$ . There is also a set of  $M$  input vectors, vectors  $\{U_1, U_2, \dots, U_m\}$ , vector of expected value  $\mu$  and covariance matrix  $C$ , which are defined as follows:

$$\mu = \frac{1}{M} \sum_{k=1}^m U^k \quad (9)$$

$$C = \frac{1}{M} \sum_{k=1}^M (U^k - \mu)(U^k - \mu)^T = \frac{1}{M} \sum_{k=1}^M U^k \times U_k^T - \mu\mu^T \quad (10)$$

The vector of the sums  $\mu$  and private covariance matrix  $\tilde{C}$  are defined according to formulas:

$$\tilde{\mu} = \sum_{k=1}^M U^k \quad (11)$$

$$\tilde{C} = \sum_{k=1}^m U_k U_k^T \quad (12)$$

The first  $p$  largest own (characteristic) numbers of the vector  $\lambda = \{\lambda_1, \lambda_2, \dots, \lambda_p\}$  and the corresponding own vectors  $V = (V_1, V_2, \dots, V_p)^T$  are used, so the PCA projection of the input vector  $U$  in the subspace  $Y = (y_1, y_2, \dots, y_p)^T$  is calculated by the formula:

$$Y = V(U - \mu) \quad (13)$$

The resultant vector is normalized using own numbers:

$$y_i^n = \frac{y_i}{\sqrt{\lambda_i}} \quad (14)$$

After principal component analysis of the mouth region, this vector will have the dimension of 32 components:

$$Y^n = (y_1^n, \dots, y_p^n)^T \quad (15)$$

The main advantage of applying the principal component analysis is storage and retrieval of image in large database, reconstruction of images. The possible disadvantage of the method is high standards for conditions to shooting of images. Images should be available in close conditions of illumination, the same foreshortening. Qualitative pre-processing leading the images to standard conditions (scale, turn, centering, and alignment of brightness, cutting off of a background) should be carried out.

Processor	Random access memory, GB	Type of hard disk	Processing speed of frame, ms
Intel Core i7-3770K 3,5 GHz	32	HDD	≈192
Intel Core i5-4210H 2,9 GHz	4	HDD	≈419

Table 1. The average processing speed of video frames using different computer systems

Calculation of the average speed of work of the proposed method of parametric video signals representation was made on computers with different performances, whose parameters are presented in Table 1.

Phases of work of the method of parametric lip are presented in Figure 13.

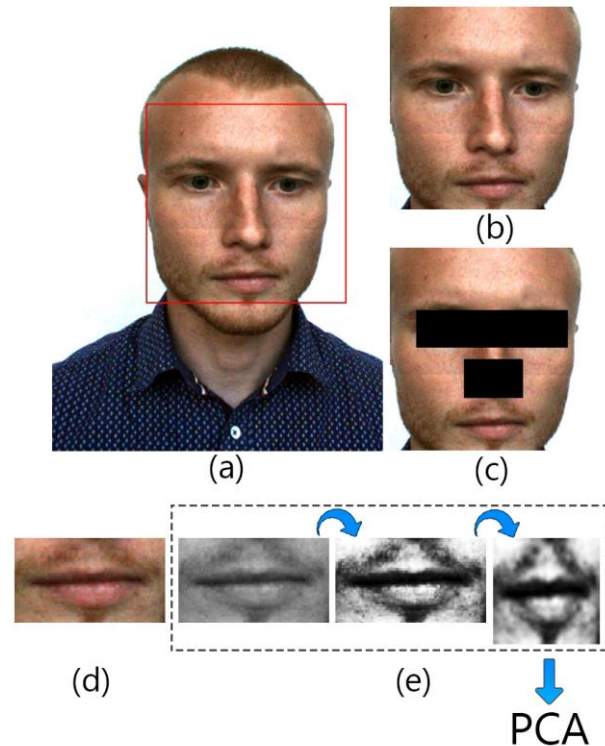


Figure 13. Parametric lip representation method: a) face area detection, b) reducing of the frame to area with the speaker's face, c) monophonic painting of eyes and nose regions, d) lips region detection, e) normalization of the mouth region

The testing has proved that in average it is possible to process video sequences 5 frames in 1 second.

### 3. CONCLUSIONS AND FUTURE WORK

The proposed method allows determining speaker's lip region in the video stream during articulation and it represents the region of interest as a feature vector with the dimensionality of 32 components. Due to its versatility, this parametric representation of video signals can be used for different tasks of biometrics, computer vision, machine learning, automatic systems for recognition of face (Lei et al., 2012), speech and elements of sign languages (Karpov et al., 2016). In our current research the proposed parametric representation is used for multimodal recognition of elements of the Russian sign language and audio-visual Russian speech.

In further research, we are going to use statistical approaches for recognition of gestures and audio-visual speech including lip-reading, e.g. based on some types of Hidden Markov Models, as well as deep learning approaches with deep neural networks (Kipyatkova and Karpov, 2016). Also we plan to apply a high-speed video camera with  $> 100$  fps in order to keep dynamics of gestures and lips (Ivanko and Karpov, 2016).

## ACKNOWLEDGEMENTS

This research is partially supported by the Council for Grants of the President of Russia (project MD-254.2017.8), by the Russian Foundation for Basic Research (projects 16-37-60100 and 15-07-04415), as well as by the Government of the Russian Federation (grant No. 074-U01).

## REFERENCES

- Al-amri et al., 2010. S. S. Al-amri, N. V. Kalyankar, S. D. Khamitkar, Linear and non-linear contrast enhancement image. *International Journal of Computer Science and Network Security*, Vol. 10, Iss. 2, pp. 139-143.
- Bruce et al., 2016. B. R. Bruce, J. M. Aitken, J. Petke, Deep Parameter Optimisation for Face Detection Using the Viola-Jones Algorithm in OpenCV. *International Symposium on Search Based Software Engineering (SSBSE)*, North Carolina, USA, pp. 238-243.
- Candes et al., 2011. Emmanuel J. Candes, Xiaodong Li, Yi Ma, J. Wright, Robust principal component analysis?. *Journal of the ACM (JACM)*, Vol. 58, No. 3. pp. 1-37.
- Castrillyn et al., 2011. M. Castrillyn, O. Deniz, D. Hernandez, J. Lorenzo, A Comparison of Face and Facial Feature Detectors based on the Viola-Jones General Object Detection Framework. *Machine Vision and Applications*, Vol. 22, Iss. 3. pp. 481-494.
- Felzenszwalb et al., 2010. Pedro F. Felzenszwalb, Ross B. Girshick, D. McAllester, Cascade object detection with deformable part models. In Proc. 23th IEEE Conference on Computer Vision and Pattern Recognition (CVPR), San Francisco, USA, pp 2241-2248.
- Gonzalez and Woods, 2007. Rafael C. Gonzalez, Richard E. Woods, *Digital Image Processing* (3rd Edition). Pearson, 976 p.
- Haar Cascades, 2008. OpenCV. Haar Cascades, <http://alereimondo.no-ip.org/OpenCV/34.version?id=63> (15 Mar. 2017).
- Ivanko and Karpov, 2016. D. Ivanko, Alexey A. Karpov, An Analysis of Perspectives for Using High-Speed Cameras in Processing Dynamic Video Information. *SPIIRAS Proceedings*, Iss. 44(1), pp. 98-113.
- Kaehler and Bradsky, 2017. A. Kaehler, G. Bradsky, *Learning OpenCV 3*, California. O'Reilly Media, 1024 p.
- Karpov et al., 2013. Alexey A. Karpov, Z. Krnoul, M. Zelezny, A. Ronzhin, Multimodal Synthesizer for Russian and Czech Sign Languages and Audio-Visual Speech. In Proc. 15th International Conference on Human-Computer Interaction HCI International-2013, Springer LNCS 8008, Las Vegas, USA, pp. 520-529.
- Karpov, 2014. Alexey Karpov, An Automatic Multimodal Speech Recognition System with Audio and Video Information. *Automation and Remote Control*, Vol. 75, Iss. 12, pp. 2190-2200.
- Karpov et al., 2014. A. Karpov, K. Markov, I. Kipyatkova, D. Vazhenina, A. Ronzhin, Large vocabulary Russian speech recognition using syntactico-statistical language modeling. *Speech Communication*, Vol. 56, pp. 213-228.
- Karpov and Zelezny, 2014. Alexey A. Karpov, M. Zelezny, Bilingual multimodal system for text-to-audiovisual speech and sign language synthesis. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*, Vol. 5, pp. 92-98 (In Rus.).
- Karpov et al., 2016. Alexey Karpov, I. Kipyatkova, M. Zelezny, *Automatic Technologies for Processing Spoken Sign Languages. 5th Workshop on Spoken Language Technologies for Under-resourced languages (SLTU)*, Vol. 81, pp. 201-207.
- Kipyatkova and Karpov, 2016. I. Kipyatkova, Alexey A. Karpov, Variants of Deep Artificial Neural Networks for Speech Recognition Systems. *SPIIRAS Proceedings*, Iss. 49(6), pp. 80-103.
- Lei et al., 2012. Z. Lei, S. Liao, A. K. Jain, S. Z. Li, Coupled discriminant analysis for heterogeneous face recognition. *IEEE Trans. Inf. Forensics Security*, Vol. 7, No. 6, pp. 1707-1716.
- OpenCV, 2016. <http://opencv.org> (14 Mar. 2017).
- Padilla et al., 2012. R. Padilla, C. F. F. Costa Filho, M. G. F. Costa, Evaluation of Haar Cascade Classifiers Designed for Face Detection. *International Journal of Computer, Electrical, Automation, Control and Information Engineering*, Vol. 6, No. 4. pp. 466-469.
- Plouffe and Cretu, 2016. P. Guillaume, Ana-Maria Cretu, Static and Dynamic Hand Gesture Recognition in Depth Data Using Dynamic Time Warping. *Aims & Scope*, Vol. 65, Iss. 2. pp. 305-316.
- Viola and Jones, 2004. P. Viola, M. Jones, Robust real-time face detection. *Int. Journal of Computer Vision*, Vol. 57, No. 2. pp. 137-154.
- WHO, 2015. World Health Organization, Deafness and hearing loss, <http://www.who.int/mediacentre/factsheets/fs300/en/> (13 Mar. 2017).
- Wu and Nagahashi, 2014. S. Wu, H. Nagahashi, Parameterized AdaBoost: Introducing a Parameter to Speed Up the Training of Real AdaBoost. *Aims & Scope*, Vol. 21, Iss. 6. pp. 687-691.
- Yoon et al., 2009. H. Yoon, Y. Han, H. Hahn, Image contrast enhancement based sub-histogram equalization technique without over-equalization noise. *World Academy of Science, Engineering and Technology International Journal of Electrical, Computer, Energetic, Electronic and Communication Engineering*, Vol. 3, Iss. 2. pp. 189-195.
- Zhao et al., 2016. Y. Zhao, Y. Liu, M. Dong, S. Bi, Multi-feature gesture recognition based on Kinect. *IEEE International Conference on Cyber Technology in Automation, Control, and Intelligent Systems (CYBER)*, Chengdu, China, pp. 392-396.