UDC 004.93'12

M. V. Davydov, PhD.,
Yu. V. Nikolski, ScD.,
S. M. Tykhanskyi

WEIGHTED OPTICAL FLOW MODEL FOR VISEMES RECOGNITION IN SIGN LANGUAGE TUTORING SOFTWARE

The problem of viseme recognition in video stream is considered. The problem arises as a part of sign language understanding and translation process. An approach that uses only dynamic viseme features is proposed. A weighted optical flow model is developed for modeling these features. Initialization, join, and distance operators are defined on weighted optical flow models for training and recognition purposes. The performance of the developed method was evaluated on a database of 119 visemas that belong to five classes. The recognition rate 69% was achieved for single user recognition.

Keywords: *weighted optical flow model, visual speech recognition, dynamic video features, recognition of visemas, machine learning, sign language, visible articulation, operations on weighted optical flow models, viseme recognition based on dynamic features*

М. В. Давыдов, канд.техн.наук,
Ю. В. Никольский, д-р техн.наук,
С. М. Тиханский

РАСПОЗНАВАНИЕ ВИЗЕМ УКРАИНСКОГО ЯЗЫКА ЖЕСТОВ ИЗ ВИДЕОЗАПИСИ С ИСПОЛЬЗОВАНИЕМ ВЗВЕШЕННОГО ОПТИЧЕСКОГО ПОТОКА

Рассматривается проблема распознавания визем в видеопотоке при решении проблемы распознавания и перевода жестового языка. Рассмотрено подход, который использует только динамические характеристики видео. Разработано взвешенную модель оптического потока для моделирования таких характеристик. Введен набор операций инициализации, объединения и вычисления расстояния над взвешенными моделями оптического потока для обучения моделей и распознавания визем. Разработанный метод протестировано на базе 119 визем, принадлежащих пяти классам. Процент правильного распознавания составил 69% при распознавании видеозаписей, отснятых с одного пользователя.

Ключевые слова: *взвешенная модель оптического потока, распознавание видимой артикуляции, динамические характеристики видео, машинное обучение, жестовый язык, видимая артикуляция, операции над взвешенными моделями оптического потока, распознавание визем на основе динамических характеристик*

М. В. Давидов, канд.техн. наук,
Ю. В. Нікольський, д-р техн.наук,
С. М. Тиханський

РОЗПІЗНАВАННЯ ВІЗЕМ УКРАЇНСЬКОЇ ЖЕСТОВОЇ МОВИ З ВІДЕОЗАПИСУ ІЗ ВИКОРИСТАННЯМ ЗВАЖЕНОГО ОПТИЧНОГО ПОТОКУ

Розглядається проблема розпізнавання візем у відеопотоці. Розглянута проблема постає при вирішенні проблеми розпізнавання та перекладу жестової мови. Розглядається підхід, який використовує лише динамічні характеристики відео для розпізнавання візем. Розроблено зважену модель оптичного потоку для моделювання цих динамічних характеристик. Введено операції ініціалізації, об’єднання та обчислення віддалі над зваженими моделями оптичного потоку для навчання та розпізнавання моделей. Розроблений метод був протестований на базі з 119 візем, які належать до п’яти класів. Відсоток правильного розпізнавання становить 69% при розпізнаванні записів, відзнятих з одного користувача.

Ключові слова: *зважена модель оптичного потоку, розпізнавання видимої артикуляції, динамічні характеристики відео, машинне навчання, жестова мова, видима артикуляція, операції над зваженими моделями оптичного потоку, розпізнавання візем на основі динамічних показників*

Introduction. Lip reading plays an important role in understanding sign language, because it helps to recognize signs and their meaning in case of ambiguity. Fast and robust viseme classification method is required for the development of an automatic sign language translation system. Viseme is a visible articulation unit that corresponds to one or more sounds.

The problem of viseme recognition lies in transforming a sequence of frames into a sequence of visemes. This task is challenging because input images quality depends a lot on the camera being used, lighting, and individual speech peculiarities (uniqueness of articulate manners). The problems that are mentioned above make it difficult to build a universal viseme recognition system.

In this paper, we introduce our work on automatic lip reading in sign language tutoring software. The obtained results can also be used to develop alternative interfaces and specialized devices such as TVs, game consoles, mobile devices, etc.

The main part of the paper is divided into three sections. Section 3 describes our approach to tackle face-tracking challenges in sign language tutoring software. Section 4 is focused on the proposed weighted optical flow model. The model itself is described and operations for joining models and calculating the distance between them are introduced. Viseme classifier that is based on the proposed model is described in Section 5. The latter part of the paper contains experimental results (Section 6) and conclusion (Section 7), where our work is summarized and directions for further research are outlined.

The main contribution of the paper is in introducing a weight parameter in discrete optical flow model and developing operations for joining two weighted discrete models for optical flow and estimating the distance between them. These operations were used to train and test the implemented viseme classifier.

In Ukrainian sign language, lip articulation has an important semantic meaning, because it allows understanding and interpreting a gesture properly in case of sign ambiguity. It is even more important than in a spoken language. The development of computer systems that would facilitate communication between people who speak a sign language and those who do not is not possible without a proper viseme recognition module.

Related work. Known models that are used in viseme recognition methods can be broadly divided into four groups: parametric, appearance-based, dynamic, and hybrid.

Parametric models are based on lip contour extraction and tracking. The lips can then be efficiently represented by “Snakes”, Active Shape Models (ASM), Active Appearance Models (AAM), and deformable templates [1, 2, 3]. The main drawback of parametric models is the difficulty in estimating them from video sequences where motion-blurred images are obtained.

Appearance models are based on extraction of image features from regions that contain lips. As a rule, these models are used with linear discriminant analysis or principal components analysis to reduce features space [4, 5]. The main disadvantage of appearance models is the lack of dynamic features.

Dynamic methods use the information about changes in facial movement in neighboring video frames [6, 7]. Hybrid methods combine the approaches mentioned above [8].

Viseme recognition rate for parametric and appearance models is up to 76 %, for dynamic models it is up to 66 %, and for hybrid models it reaches 78 % [1, 2, 6, 8]. However, the information about facial expressions is underrated, and optical flow recognition methods are not widely used. Nevertheless, the authors use optical flow to compute certain numerical characteristics of visible articulation [9].

Viseme recognition is a part of audio-visual speech recognition and visual speech recognition process. However, as it was already shown, visual speech recognition is ambiguous because one viseme can correspond to several spoken sounds. The development of cued language has greatly helped deaf people to understand articulated words properly.

The purpose of this article is to introduce a new dynamic viseme recognition method that is based on weighted optical flow model. The method does not use static lip shape characteristics that can be accurately estimated only at the stage of exposure. Instead, the classification decisions are made at the initial stage of viseme articulation. The advantage of this method is that the decisions can be made before using static methods. This is essential for real-time applications. In addition, the developed method can complement parametric or appearance-based methods to achieve better recognition rate.

Tracking Face and lips frame. Position of a gesturing person's face within a frame is not static; it constantly changes. The magnitude of the movements may be different, but they always take place. As any facial movement affects the subsequent computation of optical flow, it is very important to determine the position of face and lips in every frame as accurately as possible.

To detect a face in any image is to identify the smallest rectangle containing that face features. Furthermore, it is essential that the position of the rectangle in relation to the face be preserved all the time to make it possible to further compute optical flow. Already known methods for face detection in real-time video sequences [10, 11] fail to do that. For this reason, face tracking methods are used. They determine face position in each subsequent frame by comparing it with that in the preceding one [12, 13]. The detection method is used to determine the initial position of face, while the tracking method is utilized to track face in subsequent frames.

The problem of face detection is complicated because, in some cases, hand may cover face resulting in incorrect work of both algorithms. To solve this problem, a new *TrackFaceHybrid* algorithm for face detection and tracking is introduced. The algorithm involves the use of CAMSHIFT method to carry out face tracking and Viola-Jones method implemented by Rainer Lienhartom in the OpenCV library to initialize tracking and estimate whether or not any other object seizure took place.

TrackFaceHybrid algorithm consists of the following steps:

Step 1. $F := 0$. Set the counter of frames in which face positions are not verified to zero

Step 2. Take the next frame from video sequence. If there are no new frames than finish.

Step 3. Determine the position of face and its size in the frame by using Viola-Jones method.

Step 4. If face is detected, initialize CAMSHIFT algorithm to track the face, otherwise, proceed to step 2.

Step 5. Take the next frame and obtain the position of face p_1 and its size s_1 using CAMSHIFT method.

Step 6. Determine the position of face p_2 and its size s_2 by utilizing Viola-Jones method; seek only in the rectangle with its center at p_1 and its size 1,5 s_1.

Step 7. If face is not detected then $F := F + 1$.

Step 8. If $F > 5$ then proceed to step 1.

Step 9. If $|s_2 - s_1| > 0.5\, s_1$ then initialize CAMSHIFT algorithm with a new position of face p_2 and its size s_2.

Step 10. If video sequence still contains frames, go to step 5, otherwise finish.

The use of *TrackFaceHybrid* algorithm allowed eliminating the occasional unwanted switching of CAMSHIFT algorithm from face tracking to hand that covered the face. Step 9 allows handling the case when a person comes up to the camera or goes away from it.

Selection of face optical flow model for the construction of viseme classifier.

Efficient classifier construction involves reducing the input data redundancy and choosing only those characteristics that are important for viseme recognition. According to the authors, the characteristics may be the elements of face image optical flow.

Face optical flow is a vector field that denotes visible movements of face elements. Optical flow can be calculated by using dense and sparse methods [14]. As the methods for calculating dense optical flow are not fast enough to run in real time, Lucas-Canada sparse method is chosen [15]. The method allows obtaining optical flow field vectors in certain points, whose movements can be tracked with the smallest measure of inaccuracy.

FaceOpticalFlow algorithm for face optical flow computing consists of the following steps:

Step 1. Detect face in each frame of video sequence by using *TrackFaceHybrid* method.

Step 2. Scale the face to 100x100 pixels.

Step 3. In face image, select only the lower part of the face.

Step 4. Identify points good for tracking in each separate frame by utilizing GoodFeaturesToTrack method [16].

Step 5. Calculate optical flow vectors in these points by using Lucas-Canada method [15].

Let mathematical description of image elements movements in two consecutive frames of video sequence by means of replacement vectors of its characteristic points be a point model of optical flow. Optical flow point model is represented as a set of vectors $O = \{o_1, o_2, \ldots, o_n\}$, $o_i = (x_i, y_i, u_i, v_i)$, where x_i, y_i are the point coordinates in the first frame; u_i, v_i are the coordinates of the point movements between two consecutive frames. The obtained set of points can change from frame to frame, and that requires that additional transformations be carried out when comparing optical flows.

There are two types of elements of optical flow point model: those that are important for viseme recognition and those that are unnecessary and both increase the computation time and complicate the classifier performance. In order to consider the elements of optical flow that are important, a concept of their importance measure is introduced. It is defined by a weight coefficient.

To construct a viseme classifier, a training method is utilized. During the training, viseme optical flow characteristic points and their weights are calculated. Concepts of weighted optical flow and point model of weighted optical flow are presented.

Weighted optical flow is a vector field $(x, y) \to (u, v, w)$, where $(u, v)$is the replacement vector, w is the optic flow weight in some certain point – real number from interval [0;1] – where 0 is the minimum weight and 1 is the maximum weight. Weighted optical flow in a certain point is represented by vector $f = (x, y, u, v, w)$, where x, y are the coordinates of the point in the first frame, u, v are the coordinates of this vector movements between two consecutive frames, w is the weight. Point model of weighted optical flow $F = \{f_1, f_2, \ldots, f_n\}$is the set of vectors that show the weighted optical flow by means of the frame characteristic points. When calculating the distance between two optical flows, the point that is more important for viseme identification will be heavier.

In the course of training, a point model of weighted optical flow is created for each viseme. This model is revised with every use of a new training example. Such examples are pairs of consecutive frames from different viseme videos at the exposure stage.

The initial point model of weighted optical flow is obtained from the point model of face optical flow $O = \{o_1, o_2, ..., o_n\}$, $o_i = (x_i, y_i, u_i, v_i)$ by adding weight to its each point. The result is a weighted optical flow point model $F = \{f_1, f_2, ..., f_n\}$, $f_i = (x_i, y_i, u_i, v_i, w)$. Let the operation be called the *initialization operation* and be represented as follows:

$F = Init(O, w)$.

To provide the training function performance, an operation of *weighted combination* of two weighted point models is defined. This operation allows revising the model on the basis of a new training example. In different point models, the points for which the replacement vectors are defined do not coincide, so it requires recalculation.

$Join(F_1, F_2, \alpha)$ *weighted combination operation* is performed over two weighted optical flow models. The influence coefficient of the first of them is defined by α, $\alpha \in [0;1]$, the coefficient of the second of them is calculated by using formula $1-\alpha$. In particular, when $\alpha = 1$then only the first model is used in the combination, when $\alpha = 0$ then the second one is used.

$Join(F_1, F_2, \alpha)$ *weighted combination operation* is carried out by using the following algorithm:

Step 1. Multiply the weight of all elements $F_1$by α, and the weight of all elements F_2 by $1-\alpha$.

Step 2. Calculate the number of elements of the optical flows combination model by using formula $N = \max(|F_1|, |F_2|)$

Step 3. Combine models $F_R = F_1 \cup F_2$.

Step 4. While $|F_R| > N$ then perform steps 4.1-4.3.

Step 4.1. In set $F_R$select a pair of vector $\{f_1, f_2\} = \underset{\substack{f_1, f_2 \in F_R \\ f_1 \neq f_2}}{\arg\min}\ dist(f_1, f_2)$, where the distance between vectors $f_1=(x_1,y_1,u_1,v_1,w_1)$ and $f_2=(x_2,y_2,u_2,v_2,w_2)$ is calculated by using formula $dist(f_1,f_2)=\sqrt{(x_1-x_2)^2+(y_1-y_2)^2}$.

Step 4.2. Calculate vector f using formula $f=\beta f_1+(1-\beta)f_2$, $\beta=\dfrac{w_1}{w_1+w_2}$, where w_1 and w_2 are the weights of vectors $f_1=(x_1,y_1,u_1,v_1,w_1)$ and $f_2=(x_2,y_2,u_2,v_2,w_2)$ respectively.

Step 4.3 Recalculate model F_R by using formula $F_R:=(F_R\setminus\{f_1,f_2\})\cup\{f\}$.

Step 5. For every element $f\in F_R$, perform steps 5.1 – 5.6.

Step 5.1. Combine models $G:=F_1\cup F_2$.

Step 5.2 Recalculate the weight of each vector $g=(g_x,g_y,g_u,g_v,g_w)$, $g\in G$ by multiplying it by the coefficient inversely proportional to the distance from f to g,

$g'_w=g_w\cdot\min\{1,dist^{-1}(f,g)\}$.

Step 5.3. Compute the weighted average and standard deviations of replacements

$$W=\sum_{g\in G}g'_w\ ,\ u=\frac{1}{W}\sum_{g\in G}g_u g'_w\ ,\ v=\frac{1}{W}\sum_{g\in G}g_v g'_w\ ,$$

$$\sigma_u=\sqrt{\frac{1}{W}\sum_{g\in G}(g_u-u)^2 g'_w}\ ,$$

$$\sigma_v=\sqrt{\frac{1}{W}\sum_{g\in G}(g_v-v)^2 g'_w}\ .$$

Step 5.4. Calculate coefficients $k=\dfrac{1}{W}\max\limits_{g\in G}g'_w$ and $r=\sqrt{(\sigma_u{}^2+\sigma_v{}^2)/(u^2+v^2)}$.

Step 5.5. Compute a new weight by using formula

$$w=\begin{cases}\overline{w}+(1-\overline{w})(1-2r)(1-k), & if\ \ r<0,5\\ \overline{w}-\overline{w}\cdot(1-0,5\cdot r^{-1})(1-k), & if\ \ r\ge 0,5\end{cases},$$

where $\overline{w}=\dfrac{1}{W}\sum\limits_{g\in G}g_w g_w$.

Explanation. The coefficient of diversity r is taken into consideration when new weights are calculated. This coefficient shows how diverse are vectors of joined models in the neighborhood of point (g_x,g_y). The more vectors near this point are agreed, the smaller are dispersions σ_u and σ_v and the smaller is coefficient r. Thus, when $r<0{,}5$ (vectors are agreed in the neighborhood), the weight is increased by $(1-\overline{w})(1-2r)(1-k)$. Otherwise, when vectors are not agreed, the weight is decreased by $\overline{w}\cdot(1-0{,}5\cdot r^{-1})(1-k)$. Multipliers $(1-\overline{w})$ and $\overline{w}$ limit the speed of increase or decrease when the weight is close to 0 or 1. The multiplier $(1-2r)$ ensures weight increase when vectors are well agreed and the multiplier $(1-0{,}5\cdot r^{-1})$ forces weight decrease when vectors are not agreed well. The multiplier $(1-k)$ is used to prevent surplus weight change when neighborhood of point (g_x,g_y) contains few or only one vector with large weight. The larger is coefficient k, the larger is influence of vector with the largest weight in the neighborhood of (g_x,g_y). The value of $k=1$ means that only one vector was used.

Step 5.6. Set $f_u:=u$, $f_v:=v$, $f_w:=w$

Step 6. Return model F_R.

In the algorithm, steps 2-4 are responsible for the selection of model points and step 5 is responsible for the replacement recalculation in these points. At the same time, large standard deviation of the recalculation reduces the weight of the corresponding point.

To classify visemas, $Dist(F_1,F_2)$ *distance determination operation* is introduced. $Dist(F_1,F_2)$ Distance calculation is carried out using the following algorithm:

Step 1. Set $d:=0$, $W_d:=0$.

Step 2. For each element $f\in F_1$, perform steps 2.1 – 2.6.

Step 2.1. $G:=F_2$.

Step 2.2. Recalculate each vector weight $g=(g_x,g_y,g_u,g_v,g_w)$, $g\in G$ by multiplying it by the coefficient inversely proportional to the distance from f to g ,

$g'_w=g_w\cdot\min\{1,dist^{-1}(f,g)\}$.

Step 2.3. Compute the weighted average and standard deviations of replacements

$$W=\sum_{g\in G}g'_w\ ,\quad u=\frac{1}{W}\sum_{g\in G}g_u g'_w\ ,\quad v=\frac{1}{W}\sum_{g\in G}g_v g'_w\ ,$$

$$\sigma_u = \sqrt{\frac{1}{W}\sum_{g\in G}(g_u - u)^2 g'_w},$$
$$\sigma_v = \sqrt{\frac{1}{W}\sum_{g\in G}(g_v - v)^2 g'_w}$$

Step 2.4. Calculate coefficients

$k = \frac{1}{W}\max_{g\in G} g'_w$, and $r = \sqrt{(\sigma_u{}^2 + \sigma_v{}^2)/(u^2+v^2)}$.

Step 2.5. Calculate the weight by using formula

$$w = \begin{cases} \overline{w} + (1-\overline{w})(1-2r)(1-k), & if \;\; r < 0{,}5 \\ \overline{w} - \overline{w}\cdot(1-0{,}5\cdot r^{-1})(1-k), & if \;\; r \geq 0.5 \end{cases},$$

where $\overline{w} = \frac{1}{W}\sum_{g\in G} g_w g_w$.

Step 2.6. Calculate the coefficient of similarity $s = \frac{g_u u + g_v v}{\sqrt{u^2+v^2}\cdot\sqrt{g_u^2+g_v^2}}$. Coefficient s is in the range between -1 and 1. The vectors have the same direction, when $s=1$, and they have opposite direction when $s=-1$.

Step 2.7. Calculate $d := d + \frac{(1-s)^2}{4}\cdot g_w \cdot w$, $W_d := W_d + g_w \cdot w$.

Step 3. Return d/W_d .

Explanation. Steps 2.1 – 2.5 stands for calculation of the optical flow vector based on model F_2 at every point of F_1 model using interpolation. The similarity coefficient *s*, that is calculated in step 2.6, is a cosine of angle between optical flow vector from F_1 and F_2 models. The similarity coefficients are transformed into distance measures and are accumulated in step 2.7 with weight multipliers. The weighted average of these distances is calculated in step 3 and the result is returned. The returned value is a numerical estimate of dissimilarity between F_1 and F_2 models.

Viseme classifer constructing by using point model of weighted optical flow.

There are 15 visemes in Ukrainian sign language [17]. Some of them are very difficult to distinguish because of their visible articulation peculiarities — movements of parts of speech apparatus invisible from the outside. The authors studied the problem of recognizing of 5 visemas for letters A, O, У, I, E, the visible articulation of which differs much from one another.

To solve the problem, point models of weighted optical flows were built for each viseme. The construction was performed by calculating and combining the flows using a set of viseme video sequences.

For each viseme, n pairs of consecutive frames depicting face at the stage of exposure — preparation of speech apparatus to sound pronouncing — were selected [2]. The pairs were selected from different training videos.

The training was carried out by using *TrainVisemeModel (V)* algorithm.

TrainVisemeModel (FramePairs) algorithm

Input parameters. FramePairs list of pairs of consecutive frames depicting face at the stage of exposure.

Result. Point model of viseme weighted optical flow.

Step 1. For every pair from , calculate point models of face optical flows $O_1, O_2, ..., O_n$ by using *FaceOpticalFlow* algorithm.

Step 2. Complement all point models with weight $F_i = Init(O_i; 0{,}5)$, $i = 1,2,...,n$.

Step 3. Set $G := F_1$.

Step 4. For $i = 2,3,...,n$, carry out one by one $G := Join(G, F_i, i^{-1})$

Step 5. Return G – point model of weighted optical flow of viseme V.

The result of bulilding of prototype models for all visemes is a set of pairs $\langle V_i, F_{V_i}\rangle$, where V_i is the viseme indication and F_{v_i} is the prototype model of viseme optical flow V_i. For 5 visemes of Ukrainian language have $M = \{\langle 'A', F_A\rangle, \langle 'O', F_O\rangle, \langle 'У', F_У\rangle, \langle 'E', F_E\rangle, \langle 'I', F_I\rangle\}$. The example of vectors of point model of viseme “A” optical flow is depicted in Fig.1.

The built models are used to identify visemes in every frame of video sequence by calculating the distance from optical flow point model, calculated in frame, to weighted point models of visemes optical flows. The result of recognition is the viseme whose model distance is minimal.

The experimental results. In a series of experiments, the number of errors that occurred while applying the viseme identification method and the lip shape method was evaluated [1]. The experiments were conducted on the basis of video database of Ukrainian words comprising 50 videos with 119 visemas in total.

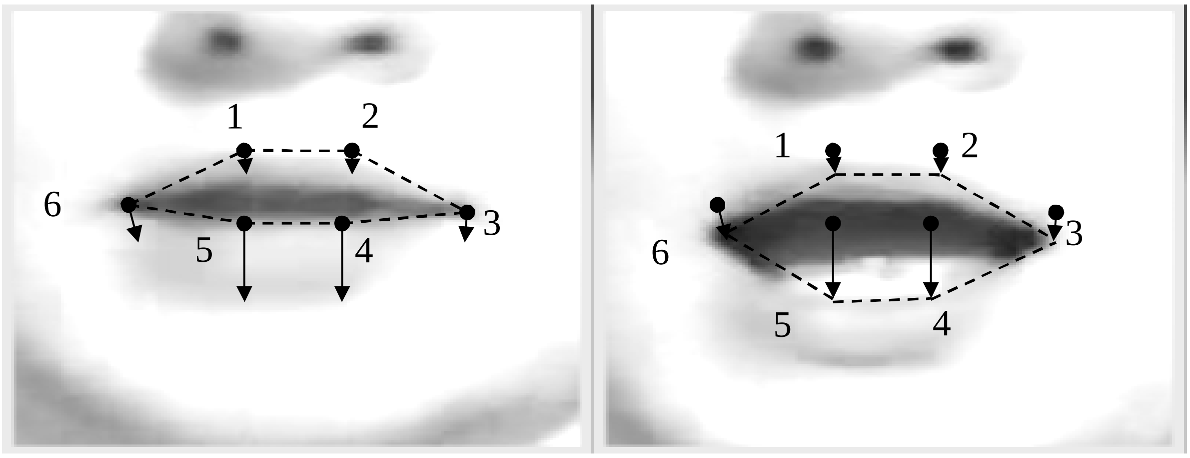

Fig. 1. Example of point model of viseme "A" optical flow. At the left, lips contour in the first frame of pair, at the right – in the second frame. In both images, vectors of point model of optical flow are put as arrows

Each experiment was conducted in the following way:

1. Selected a video from the database.
2. Applied the viseme identification algorithm for each pair of neighboring frames.
3. If at the stage of exposure (exposure frames were determined by the experimentalist manually from the moment lips started to move till the moment the movements stopped) viseme model appeared to be the closest to the computed model of optical flow of its neighboring frames, the viseme was considered to be correctly recognized.

After conducting a number of experiments, the percentage of correctly recognized visemas as compared to their general number was calculated for each viseme type. The obtained recognition results are depicted in Table. 1.

The experiments have shown that the results obtained after the usage of the optical flow method for "A" and "I, И" visemas recognition appeared to be better than those obtained after the lip shape method usage. Taking into account that the methods use fundamentally different approaches to viseme recognition, it is possible to combine their results to increase the percentage of recognition accuracy.

To evaluate the algorithm time characteristics, the average time needed to process one sequence video frame was determined by using a test database of videos.

The average time for processing one frame by using the method that uses point model of optical flow is 83 ms what is 28% lower than by using the method based on lip shape (121 ms). This allows processing video at the speed of 11.5 frames per second on a PC with Intel Pentium Dual-Core T3400 (2.16GHz) processor.

1. Comparison of visemas recognition accuracy index obtained by using different viseme recognition methods

Viseme	Number of visemas of the specific type in the database	Method based on lip shape (%)	Method based on optical flow (%)
A	14	71	78
O	19	63	57
У	30	86	80
E	44	75	63
I, И	12	66	75
Total	119	74	69

Conclusions. The method for viseme identification is proposed where the classification decision is made at the initial stage of viseme articulation. The method utilizes models based on dynamic features. Such dynamic features are optical flow models, which are calculated for neighbor video frames. The identification is done via comparison of optical flow models to a set of trained weighted optical flow models. A new algorithm is proposed for training weighted optical flow models as well as procedures for joining and comparison of such models. The recognition rate 69% was obtained for 5 visemas recognition of a single speaker. The obtained recognition rate for «A» and «I, И» visemas is higher than in static models. The use of dynamic approach for viseme recognition lets obtain recognition result earlier than in static methods. Further research will be focused on extension of viseme database and consideration of interpersonal dissimilarity of visible articulation.

References

1. Ipsic I. Speech and Language Technologies [Text] / I. Ipsic. – Rijeka, Croatia: InTech, 2011. – ISBN 978-953-307-322-4. – 344 p. [in English].

2. Давидов М. В. Метод ідентифікації візем за артикуляції у складі жесту української жестової мови [Текст] / М. В. Давидов, Ю. В. Нікольський, С. М. Тиханський // Інформаційні системи та мережі. Вісник НУ «Львівська політехніка». – Львів : 2011. – № 699. – С. 302 – 312 [in Ukrainian].

3. Patil, R. A. a. Features Classification Using Support Vector Machine for a Facial Expression Recognition System [Text] / R. A. a Patil, A. S. b, Mandal // Journal of Electronic Imaging. SSN: 10179909 – 2012. – 21 (4), art. no. 043003. DOI : 10.1117/1.JEI.21.4.043003 [in English].

4. Huang X. a. b. Towards a Dynamic Expression Recognition System under Facial Occlusion [Text] / [X. a. b. Huang, G. a Zhao, W. b Zheng and others] // Pattern Recognition Letters. – 2012, 33 (16). – P. 2181 – 2191. ISSN: 01678655. DOI : 10.1016/j.patrec.2012.07.015 [in English].

5. Belhumeur P. N. Eigenfaces vs. Fisherfaces: Recognition Using Class Specific Linear Projection [Text] / P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman // IEEE Transactions on Pattern Analysis and Machine Intelligence. ISSN : 01628828 –1997. – 19 (7). – P. 711 – 720. DOI : 10.1109/34.598228 [in English].

6. Glenn A. Martin. Lipreading by Optical Flow Correlation [Text] / Glenn A. Martin // Tech. report – Computer Science Department University of Central Florida. – 1995. – 7 p. [in English].

7. Liu J. a. b. Human Action Recognition Based on Latent Author-Action Models [Text] / J. a. b. Liu, T. a. Li, J. a. Yang // Journal of Computational Information Systems. ISSN : 15539105. – 2012. – 8 (24). – P. 10485 – 10492 [in English].

8. Alin G. Comparison Between Different Feature Extraction Techniques for Audio-Visual Speech Recognition [Text] / G. Alin, L. Rothkrantz, J. Wojdel, P. Wiggers // Journal on Multimodal User Interfaces, Springer. – 2007. – P. 7 – 20 [in English].

9. Karlsson S. M. Lip-Motion Events Analysis and Lip Segmentation Using Optical Flow [Text] / S. M. Karlsson, J. Bigun // Computer Vision and Pattern Recognition Workshops (CVPRW). – IEEE Com. Soc. Conference on. – 16-21 June 2012. – P. 138 – 145 [in English].

10. Robust Real-Time Object Detection [Text] / Paul Viola, Michael Jones // International Journal of Computer Vision – 2001. Citeseer. – Vol. 57. – Iss. 2. – P. 137 – 154 [in English].

11. Schneiderman, H. A Statistical Method for 3D Object Detection Applied to Faces and Cars [Text] / H. Schneiderman, T. Kanade // Computer Vision and Pattern Recognition, Proc. IEEE Conf. on. – 2000. – Vol. 1. – P. 746 – 751 [in English].

12. Cheng Y. Mean Shift, Mode Seeking, and Clustering [Text] / Y. Cheng // Pattern Analysis and Machine Intelligence, IEEE Trans. on. – 1995. – Vol. 17. – Iss. 8. – P. 790 – 799 [in English].

13. Shrinivasa Maika, C. L. An Intelligent Face Tracking System for Human-Robot Interaction Using Camshift Tracking Algorithm [Text] / [C. L. Shrinivasa Maika, Vibhor

Nikhra, Shashishekha R. Jha and others] // International Journal of Machine Intelligence. – Vol. 3. – Iss. 4. – 2011. – P. 263 – 267 [in English].

14. Liu H. Accurate Dense Optical Flow Estimation Using Adaptive Structure Tensors and a Parametric Model [Text] / H. Liu, R. Chellappa, A. Rosenfeld // Image Processing, IEEE Transactions on. – Oct. 2003. – Vol.12. – No. 10. – P. 1170 – 1180 [in English].

15. Lucas B. D. An Iterative Image Registration Technique With an Application to Stereo Vision [Text] / B. D. Lucas, T. Kanade // Proceedings of Imaging Understanding Workshop. – 1981. – P. 121 – 130 [in English].

16. Shi J. Good Features To Track [Text] / Jianbo Shi, Carlo Tomasi // IEEE Conf. on Computer Vision and Pattern Recognition Seattle. – June 1994. – P. 593 – 600 [in English].

17. Крак Ю. Синтез міміки та емоцій на обличчі людини для моделювання жестової мови [Текст] / Ю. Крак, А. Тернов // Вісник Київського національного університету ім. Тараса Шевченка. Кібернетика. – К. : Київський університет. – ISSN : 1728-2276. – – 2010. – № 10.– С. 52 – 55 [in Ukrainian].

Received 15.07.2013

References

1. Ipsic I. Speech and Language Technologies [Text] / I. Ipsic. – Rijeka, Croatia: InTech, 2011. – ISBN : 978-953-307-322-4. – 344 p. [in English].

2. Davydov M.V. The Method for Viseme Identification in Articulation of Ukrainian Sign Language Gestures [Text] / M. V. Davydov, I. V. Nikolski, S. M. Tykhanskyi // Information systems and networks. Proceedings of Lviv Polytechnic National University. – Lviv, 2011. – № 699. – P. 302 – 312 [in Ukrainian].

3. Patil R. A. a. Features Classification Using Support Vector Machine for a Facial Expression Recognition System [Text] / R. A. a Patil, A. S. b, Mandal // Journal of Electronic Imaging. SSN : 10179909 – 2012. – 21 (4), art. no. 043003. DOI : 10.1117/1.JEI.21.4.043003 [in English].

4. Huang X. a. b. Towards a Dynamic Expression Recognition System under Facial occlusion [Text] / [X. a. b. Huang, G. a. Zhao, W. b. Zheng and others] // Pattern Recognition Letters. – 2012. – 33 (16). – P. 2181 – 2191 [in English].
ISSN : 01678655.
DOI : 10.1016/j.patrec.2012.07.015.

4. Belhumeur P. N. Eigenfaces vs. Fisher-Faces: Recognition Using Class Specific Linear Projection [Text] / P. N. Belhumeur, J. P. Hespanha, D. J. Kriegman // IEEE Transactions on Pattern Analysis and Machine Intelligence.– 1997. – 19 (7). – P. 711 – 720.
ISSN : 01628828 [in English].
DOI : 10.1109/34.598228.

6. Glenn A. Martin. Lipreading by Optical Flow Correlation [Text] / Glenn A. Martin / Technical Report. – Compute Science Department University of Central Florida. – 1995. – 7 p. [in English].

7. Liu J. a. b. Human Action Recognition Based on Latent Author-Action Models [Text] / J. a. b. Liu, T. a. Li, J. a. Yang // Journal of Computational Information Systems. – P. 10485 –10492 [in English].
ISSN : 15539105 (2012) 8 (24).

8. Alin G. Comparison Between Different Feature Extraction Techniques for Audio-Visual Speech Recognition [Text] / [G. Alin, L. Rothkrantz, J. Wojdel and others] // Journal on Multimodal User Interfaces, Springer. – 2007. – P. 7 – 20 [in English].

9. Karlsson S. M. Lip-Motion Events Analysis and Lip Segmentation Using Optical Flow [Text] / S. M. Karlsson, J. Bigun // Computer Vision and Pattern Recognition Workshops (CVPRW). – 2012 IEEE Computer Society Conference on. – 16-21 June 2012. – P. 138 – 145 [in English].

10. Paul Viola. Robust Real-time Object Detection [Text] / Paul Viola, Michael Jones // International Journal of Computer Vision (2001), Citeseer. –Vol. 57. – Iss. 2. – P. 137 – 154 [in English].

11. Schneiderman H. A Statistical Method for 3D Object Detection Applied to Faces and Cars [Text] / H. Schneiderman, T. Kanade // Computer Vision and Pattern Recognition, 2000. – IEEE Conf. on. – Vol. 1. – P. 746 – 751 [in English].

12. Cheng Y. Mean Shift, Mode Seeking, and Clustering [Text] / Y. Cheng // Pattern

Analysis and Machine Intelligence, IEEE Transactions on. – Vol. 17. – Iss. 8. – 1995. – P. 790 – 799 [in English].

13. Shrinivasa Maika C. L. An Intelligent Face Tracking System for Human-Robot Interaction Using Camshift Tracking Algorithm [Text] / [Shrinivasa Maika C. L., Vibhor Nikhra, Shashishekha R Jha and others] // International Journal of Machine Intelligence. – Vol. 3. – Iss. 4. – 2011. – P. 263 – 267 [in English].
ISSN : 0975–2927 & E-ISSN: 0975–9166.

14. Liu H. Accurate Dense Optical Flow Estimation Using Adaptive Structure Tensors and a Parametric Model [Text] / H. Liu, R. Chellappa, A. Rosenfeld // Image Processing, IEEE Transactions on. – Vol. 12. – No. 10. – Oct. 2003. – P. 1170 – 1180 [in English].

15. Lucas B. D. An Iterative Image Registration Technique with an Application to Stereo Vision [Text] / B. D. Lucas, T. Kanade // Proceedings of Imaging Understanding Workshop. – 1981. – P. 121 – 130 [in English].

16. Shi J. Good Features To Track [Text] / Jianbo Shi, Carlo Tomasi // IEEE Conference on Computer Vision and Pattern Recognition Seattle. – June 1994. – P. 593 – 600 [in English].

17. Krak Yu. A Synthesis of Mimicry and Emotions is on Face of Man for the Design of Gesture Language [Text] / Yu. Krak, A. Ternov // Proceedings of Taras Shevchenko National University of Kyiv. Cybernetics. – Kyiv: Kyiv University.. – № 10. – 2010. – P.52 – 55 [in Ukrainian].
ISSN : 1728-2276.

Davydov, Maksym D.
Doktorant of Information Systems and Networks Department, Lviv Polytechnic National University,
Bandery str. 13,
Lviv,
Ukraine, 79000,
tel. (032)258-25-51

Nikolski, Iouri V.
Associate Professor of Information Systems and Networks Department, Lviv Polytechnic National University,
Bandery str. 13,
Lviv, Ukraine, 79000.

Tykhanskyi, Sergiy M.
Postgraduate student,
Lviv Polytechnic National University,
Bandery str. 13,
Lviv,
Ukraine, 79000.