Recognition of human emotional state using the SVM and ANN algorithms

Astakhov Dmitriy
Volgograd State Technical University
Volgogorad, Russia
astakhov.d.a@gmail.com

Kataev Alexander
Volgograd State Technical University
Volgogorad, Russia
alexander.kataev@gmail.com

***Abstract*—The work compares the results of the work of the machine learning algorithm SVM, ANN on a set of data from the normalized key points and the distances between them. The process of preparing the training base and normalization algorithms is described.**

***Keywords*—system of recognition of emotions, key points, machine learning, face recognition, ann, svm, emotions**

I. Introduction

Long before speech arose, interaction between people was realized through non-verbal communication. Change in facial expression has become one of the most significant means of conveying the emotions and intentions of a person.

In recent years, the task of automatically analyzing the emotional state of a person has attracted a large number of researchers.

In particular, in the work led by Zachelova-Zotova A.V. the questions of definition of emotional reactions of the person on mimicry, nonverbal movements and a voice are considered [1]. The task of recognizing dynamic gestures of a person is considered in the works of Devyatkov V.V. and Alfimtseva A.N. [2]. In his doctoral dissertation Dementienko V.V. the system of automatic prediction of the driver's falling asleep due to his blinking and movement of the eyes is considered [3]. In the work of Zeifeng Shan, an algorithm for detecting emotions based on local binary patterns is considered [4].

Recently, machine learning methods such as SVM and ANN have often been used to define emotions, as can be seen from such works as [5]. The purpose of this study is to compare the work of these methods with different sets of data.

II. Determination of the emotional state of a person

It is known that the expression of emotions can be very diverse and vary depending on individual characteristics, as well as on the situational context. The cultural context plays an important role when it comes to emotions that do not belong to the category of basic ones, since the methods of expressing complex emotional states adopted in a certain community are different. If you observe a person, you can see that most emotions are rapidly changing each other and making it difficult to recognize [6]. In addition, manifestations that correspond to "mixed" emotions are difficult to recognize, because behind them lies the whole complex of feelings experienced by a person. Another factor that affects the effectiveness of the recognition of emotions is the possibility of faking emotional behavior. Cultural requirements, personal perceptions of admissible or any considerations arising from a person's perception of the actual situation (the desire to hide something or demonstrate a feeling that is not really there) - these and other factors affect the expression of a person's emotions. Automatic identification of images (text, sound, face, person, objects, etc.) with the help of a computer is one of the most important directions in the development of artificial intelligence technologies, which makes it possible to give the key to understanding the features of the work of the human intellect. Research methods of automatic recognition of emotions allows you to give the computer the ability to assess the mood of a person, for this purpose, the algorithm uses the recognition of emotions, presented in "Fig. 1".

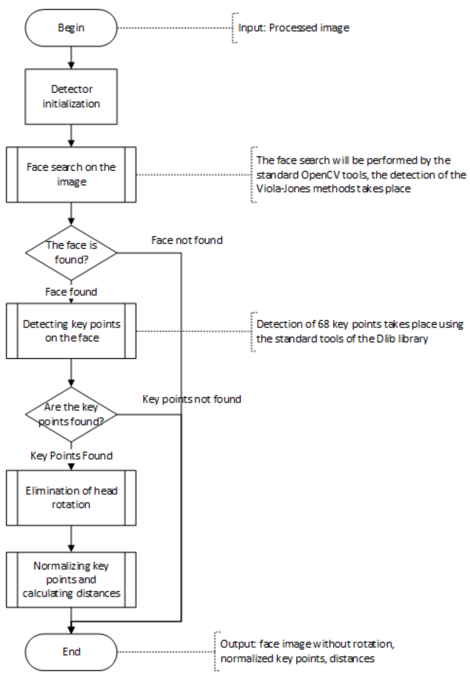


Figure 1. The main algorithm for recognizing emotions.

III. PREPARING A SAMPLE FOR TRAINING

For learning machine learning algorithms, the Extended Cohn-Kanade Database (CK +) sample [10] was used, consisting of 11,061 photographs, with a resolution of 640 × 490 pixels, in * .png format. In the CK + database 623 structures with emotions are marked out. The sample is divided into 8 classes of emotions: 1 - anger, 2 - contempt, 3 - disgust, 4 - fear, 5 - happiness, 6 - sadness, 7 - surprise, 8 - normal. In the future, for brevity, the classes of emotions will be denoted by the corresponding numbers.

The original database was reworked in such a way as to minimize recognition errors, for this purpose, an image with the most "expressed" emotion was selected for each emotion and a person and a mirror image was made for each image to increase the sample size. The resulting volume of the processed database produced 2016 photographs. “Fig. 2” shows the distribution of the volume of images for each class.

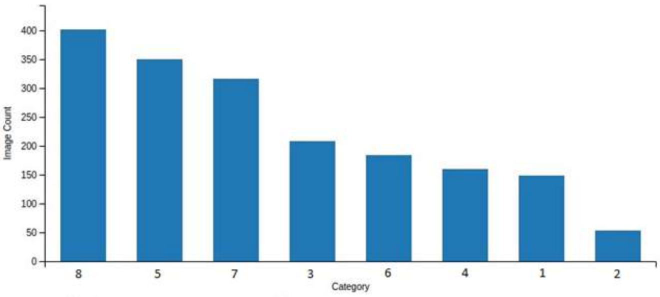


Figure 2. The diagram of the distribution of the data volume by classes.

Uneven distribution of images by classes is associated with the limited sample size for each class in the original CK + database, classes 5, 7, 8 have more data in the source database than classes 1, 2, 3, 4, 6. Example of images used in the database is shown in “Fig. 3”.

Figure 3. An example of images from the database, the classes are located on the left to the right (the leftmost image corresponds to class 1, the rightmost to the 8th class) (The Extended Cohn-Kanade Database).

For further work, the sample was normalized. By normalization is meant - an image in which the person is located without turns, inclinations and 95% of the image is the person itself. An example of normalized images for each emotion class is shown in “Fig. 4”. The image size after normalization is not fixed, it ranges from 201 x 199 pixels to 306 x 275 pixels, type of images * .png.

A. Facial coding system

For the encoding of facial movements, the systems of FACS and EmFACS are used. Facial Action Coding System (FACS) - is a system for the taxonomy of human facial expressions [7]. This standard is generally accepted for the systematic classification of the physical expression of emotions [8].

Figure 4. An example of images from a normalized base, classes are located on the left to the right (the leftmost image corresponds to class 1, the rightmost to the 8th class).

Emotional Facial Action Coding System (EmFACS) - this system is considered only the coding of facial movements associated with emotions [9].

These systems were developed by Paul Ekman and Wallace Friesen in 1978.

With the use of FACS, it is possible to manually code any practical, anatomically possible facial expression, constructing it from the actions of specific units of action and the time required by them to reproduce a particular facial expression. The FACS determines the units of action that cut or relax one, or more muscles.

This system allows you to align which key points are involved in facial movements. An example of the encoding of motion movements is shown in “Fig. 5”.

Figure 5. Encoding facial movements in the FACS system.

B. Key Points

For the process of highlighting the key points on the face, a third-party Dlib library was used, which allows 68 face key points to be found on the image “Fig. 6”.

In accordance with the CLRD, for the training from 68 key points, 44 points describing the facial movements were selected, for this purpose points describing the contour of the face (points 0 - 16) were excluded, which do not affect the description of facial movements in the emFACS system. The next step for each image from the sample is two sets of data: normalized key points (“Fig. **??**”) and distances between the key points (“Fig. **??**”).

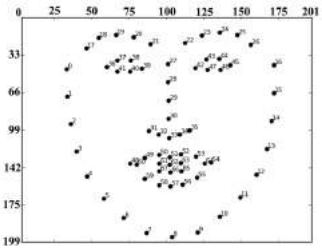
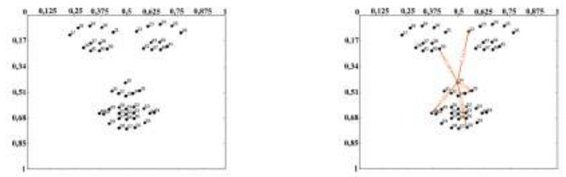


Figure 6. Example of detecting key points by the dlib library.



a) b)

Figure 7. Example of data used, figure (a) - normalized key points, figure (b) - distance between key points

To the normalized face image, methods for detecting key points are applied, resulting in the output of 68 key points in the coordinate system of the normalized image. For normalization of data, all coordinates are given to the range [0, 1].

Calculation of normalized key points was carried out according to the formula:

$$x'_i = \frac{x_i}{\text{width}}; y'_i = \frac{y_i}{\text{height}} \quad (1)$$

Calculation of distances between the key points was carried out according to the formula:

$$d_i = \sqrt{(x_c - x'_i) + (y_c - y'_i)} \quad (2)$$

C. System Training

The training was done using the built-in machine learning in the OpenCV library.

To train the algorithms, the prepared database was divided into two parts: training and test. For testing, 15% of the total sample size was selected, which is 272 images, the training base consists of 1744 images.

For the training of SVM, the type C_SVC was used, which makes it possible to classify into 8 classes using a multiplier C with a value of 0.5. The type of the kernel was chosen to be polynomial with parameters: a level equal to 0.001, coef0 equal to 0.1, gamma equal to 0.00015.

The network architecture for learning the ANN method on normalized key points consists of 4 layers: an input layer of 44 neurons, two intermediate layers of 88 neurons each, and an output layer of 8 neurons. The sigmoidal activation function is used, the network is trained by the method of back propagation of the error with the scale parameter equal to 0.1. The maximum number of iterations was set to 1000. The selection of parameters for both methods of machine learning is empirical.

IV. Preparing a sample for training

The work of the machine learning methods SVM and ANN was carried out on two sets of data: normalized key points and distances between the key points. The results of the experiments carried out for each method of machine learning with different data sets.

Table I
THE RESULT OF TESTING THE SVM METHOD ON NORMALIZED KEY POINTS

Expected emotion	**Recognized emotion**							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	**30.62**	0.00	22.50	0.00	0.00	31.25	0.00	15.62
2	3.70	**0.00**	3.70	0.00	16.66	11.11	0.00	64.81
3	0.00	0.00	**31.77**	0.00	17.77	14.77	0.00	35.69
4	0.00	0.00	0.00	**37.07**	17.53	0.00	12.69	32.69
5	0.00	0.00	3.33	0.00	**93.33**	0.00	0.00	3.33
6	0.00	0.00	9.37	6.13	9.37	**48.88**	0.00	26.25
7	0.00	0.00	0.00	4.16	0.00	4.17	**91.67**	0.00
8	0.00	0.00	0.56	1.11	4.33	1.67	0.00	**90.94**

Analyzing the data obtained during the testing ("Tab. I"), we can conclude that the SVM method tested on a set of key points showed low accuracy for this classification problem. This is due to the fact that the training sample does not have a uniform distribution of the number of images by classes, it can be observed that in most classes the percentage of recognized emotions refers to grades 5, 7 and 8, and the parameters chosen in training could not be found to be the most optimal.

Table II
THE RESULT OF TESTING THE SVM METHOD ON DISTANCE DATA

Expected emotion	**Recognized emotion**							
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	**28.13**	0.00	15.63	0.00	3.13	31.25	0.0	21.88
2	1.85	**0.00**	12.96	1.85	1.85	1.11	0.00	70.37
3	0.00	0.00	**50.00**	0.00	50.00	0.00	0.00	100
4	0.00	0.00	0.00	**69.23**	7.69	19.23	3.85	0.00
5	3.33	0.00	6.67	3.33	**76.67**	3.33	0.00	6.67
6	3.13	0.00	15.63	6.25	3.13	**56.25**	0.00	15.63
7	0.00	0.00	0.00	12.50	12.50	4.17	**70.83**	0.00
8	1.72	0.00	8.33	1.39	2.94	1.33	0.00	**86.28**

The SVM method tested on a set of distances between key points ("Tab. II"), showed itself the same way as when testing at key points, low accuracy for the task of classifying emotions. This is due to the fact that the training sample does not have a uniform distribution of the number of images by classes, it can be observed that in most classes the percentage of recognized emotions refers to 4, 5, 7 and 8 classes. But, nevertheless, the result of training at these distances proved to be more effective than at these key points by 2.17

The ANN method tested on a set of key points ("Tab. III") showed the accuracy of the recognition of emotions equal to 53.5%, therefore, this approach may be applicable when working with real data. The inaccuracy of the classification

Table III
THE RESULT OF ANN TESTING ON NORMALIZED COORDINATES

Expected emotion	Recognized emotion							
	1	2	3	4	5	6	7	8
1	**37.50**	0.00	3.12	0.00	9.37	21.87	28.12	0.00
2	0.00	**27.78**	0.00	1.85	7.40	33.30	29.62	0.00
3	0.00	0.00	**21.30**	0.00	0.00	34.80	21.70	22.20
4	0.00	3.84	0.00	**54.10**	3.84	5.380	23.80	9.04
5	0.00	0.00	0.00	0.00	**50.00**	16.40	14.50	19.10
6	0.00	0.00	0.00	0.00	0.00	**71.87**	8.12	20.00
7	0.00	0.00	0.00	3.33	0.00	4.20	**77.91**	14.36
8	0.00	1.38	0.00	4.17	2.70	0.00	3.61	**88.14**

is due to the fact that the training sample does not have a uniform distribution of the number of images by classes, and the parameters chosen in training could not be found to be the most optimal.

Table IV
THE RESULT OF TESTING THE ANN METHOD ON DISTANCE DATA

Expected emotion	Recognized emotion							
	1	2	3	4	5	6	7	8
1	**41.87**	0.00	0.00	0.00	6.25	12.5	18.75	20.62
2	0.00	**38.51**	0.00	1.85	1.85	14.81	11.11	21.85
3	0.00	0.00	**50.00**	0.00	0.00	0.00	0.00	50.00
4	0.00	0.00	0.00	**58.46**	0.00	11.53	16.92	13.07
5	0.00	0.00	0.00	0.00	**60.00**	0.00	13.33	26.67
6	0.00	0.00	0.00	0.00	3.12	**60.62**	9.37	26.87
7	0.00	0.00	0.00	0.00	0.00	0.00	**87.50**	12.50
8	0.00	1.38	0.00	0.00	0.00	4.16	16.66	**77.77**

The ANN method tested on a sample from the distances between the key points (“Tab. IV”) showed a classification accuracy of 53.5%, which is 6.5% better than on data with key points. Therefore, this approach may be applicable when working with real data. The inaccuracy of the classification is due to the fact that the training sample does not have a uniform distribution of the number of images by classes, and the parameters chosen during training could not be found to be the most optimal.

Table V presents a comparative analysis of the results obtained.

Table V
SUMMARY TABLE OF TEST RESULTS

Method	Dataset	Accuracy of recognition,%
SVM	Key Points	52.50
	Distance	54.67
ANN	Key Points	53.50
	Distance	60.00

Thus, analyzing the data obtained for each test example, it can be concluded that SVM and ANN methods showed very close results in the accuracy of the classification of emotions. But the ANN method showed greater recognition accuracy than the SVM method at a distance of 5.33%, and at these key points by 1%.

As you can see, the accuracy of the classification increases, if you use the data based on the distance between the key points for learning. For the SVM method, the accuracy increased by 2.67%, and for the ANN method the accuracy increased by 6.50%.

CONCLUSION

In the paper, the machine learning methods ANN and SVM were tested on a set of data from normalized key points and the distances between them. Both methods have an acceptable detection level. However, during the research it was revealed that the ANN method on the set of distances showed the best result.

Both of these methods can be used in the interface of an intelligent system to create conditions for natural and intuitive human-machine interaction.

REFERENCES

[1] The task of creating a system of automated recognition of emotions / A. V. Zachelova-Zotova [and others] // Open semantic technologies for the design of intelligent systems: materials of the International. scientific-techn. Conf. OSTIS-2012 (Minsk, February 16-18, 2012) / BSUIR. - Minsk, 2012. - P. 347-350.

[2] Devyatkov, V. V. Recognition of manipulative gestures / V. V. Devyutkin, A. N. Alfimtsev // Bulletin of the Bauman Moscow State Technical University. N. E. Bauman. - 2007. - vol. 68, i. 3. - P. 56-75.

[3] Dementienko, V. V. Physical principles of constructing systems for the safe monitoring of the human operator status: the author’s abstract. dis. ... Dr. techn. Sciences / V. V. Dementienko. - Moscow, 2010. - 38 p.

[4] Caifeng Shan, Shaogang Gong, Peter W. McOwan .: Facial expression recognition based on Local Binary Patterns: A comprehensive study. Image and Vision Computing (2008).

[5] Enikilopov S. N., Kusnetsova Yu. M. .: The Task of Recognition of Violent Situations Using the Automatic Systems and Methods of Artificial Intelligence. Journal of Psychology and Law, vol. 2, pp. 1- 16. Moscow psychological and pedagogical University, Moscow (2011).

[6] Rozaliev, V. L. Modeling the emotional state of a person on the basis of hybrid methods / V. L. Rozaliev // Software products and systems. - 2010. - No. 2. - P. 141-146.

[7] LeCun, Y. Convolutional Networks for Images, Speech, and Time Series / Y. LeCun, Y. Bengio // The Handbook of Brain Theory and Neural Networks. - USA: MIT Press, 2002. - P. 276-279.

[8] Jizheng Yi, Xia Mao, Lijiang Chen, Yuli Xue, Angelo Compare facial expression by identifying individual differences in facial structure and texture. IET Computer Vision 2014, vol. 8, pp. 429-440. The Institution of Engineering and Technology (2014).

[9] Panagiotis Perakis, Georgios Passalis, Theoharis Theoharis, Ioannis A. Kakadiaris .: 3D Facial Landmark Detection & Face Registration. Department of Informatics and Telecommunications University of Athens, Greece (2010).

[10] The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression / P. Lucey [et. al.] // Proceedings of IEEE CVPR Workshop on Biometrics, 13-18 Jun 2010, San Francisco, CA, USA / IEEE Computer Society. - San Francisco, 2010. - P. 94-101.

РАСПОЗНАВАНИЕ ЭМОЦИОНАЛЬНОГО СОСТОЯНИЯ ЧЕЛОВЕКА С ИСПОЛЬЗОВАНИЕМ АЛГОРИТМОВ SVM И ANN

Астахов Д.А., Катаев А.В.
Волгоградский Государственный Технический Университет

В работе сравниваются результаты работы алгоритмом машинного обучения SVM, ANN на наборе данных из нормализованных ключевых точек и дистанций между ними. Проводится описание процесса подготовки базы для обучения и алгоритмов нормализации.