

Intelligence System For Emotional Facial State Estimation During Inspection Control

Viktor Sineglazov

Aviation computer integrated
complexes department
National Aviation University
Kyiv, Ukraine
svm@nau.edu.ua

Roman Pantyeyev

Aviation computer integrated
complexes department
National Aviation University
Kyiv, Ukraine
romanpanteevmail@gmail.com

Ilya Boryndo

Aviation computer integrated
complexes department
National Aviation University
Kyiv, Ukraine
ib.mistlemagic@gmail.com

Abstract—The problem of customs control intelligence security system creations is considered. The necessity of the passenger emotions face changes analysis during control in airports is shown. The emotional changes of a human face correspond to internal reaction of the person on the posed control questions. As the solution of this problem it's proposed to apply the convolution neural network at the stage of micro emotion identification and the indistinct qualifier – at the stage of decision making for potential threats of passenger. As the indistinct qualifier it is offered to use the NEFCLASS neural network. The example of practical approach for micro emotion recognition by means of convolution neural network is given.

Keywords—combined convolution neural network, NEFCLASS, micro expressions, facial state recognition.

I. INTRODUCTION

Nowadays, the real importance is given to increasing the aircraft safety conditions, in particular during the passenger control. Commonly, the number of people for each security officer is too high to deal with them in restricted period of time. The employee of Aircraft Company is faced by a hard task, to ask the number of special questions to understand the emotional state of the passenger to successful admission of the flight. The main features that allow solving this problem is emotional changes of the passenger during the control conversation. The most successful application of this approach was done by El Al aircraft company (Israel). The challenge is to detect and recognize the facial micro changing. It's very difficult task especially with a large number of passengers (about 20-30 people on one security officer) and demand considerable preparation. In this work as the solution of given problem offered to use an intellectual analysis system which consists of the two-level neural networks of deep learning based on micro emotion recognition: at the first level the convolution neural network [1], on the second – the qualifier constructed with use of indistinct neural network, for example NEFCLASS [2] is applied.

II. REVIEW OF EXISTING SOLUTIONS

During the analysis of human emotional state recognition, it's necessary to consider “profiling” [3]. This is a set of psychological approaches for analyzing and predicting a persons behavior, based on particular features, appearance and verbal behavior. This technique can be successfully

applied in computer vision due to the classification of these features for potential danger prevention. Using the profiling method it is necessary to use a coding system of facial movements (SKLiD) [4]. By using the SKLiD, it is possible to create a facial model based on units of actions and a fixed period of time needed to act any emotion. Here, the units of action are the movements performed by individual muscles or a group of muscles.

The system also has a limited number of descriptors (unitary movements performed by a group of muscles: tightening the cheeks, stretching the eyelids, raising the wings of the nose, raising the upper lip, deepening the noselabial fold, raising the corners of the lips, dimpling the lips, lowering the corners of the mouth, lowering the lower lip, pulling off the lips) [5]. Each manifestation of facial emotions of a person can be described by a set of descriptors. As the apparent facial changes there also occur the micro emotions. They can be taken into account in more complicated recognition approaches. Table 1 describes the main facial changes relatively to the six standard types of emotions [6].

Motion units of the person can be divided into three groups conditionally:

- static – recognition using only the photo is possible;
- dynamic – it is necessary to continuous frame changing, key points initialization or obtaining the average value of distances between motion units;
- empty – actively participate in manifestation of emotions, however are not registered search algorithms (dimples on cheeks).

Now it is possible to review the following recognition methods of the human emotional state using a profiling approach. The most efficient ones are: holistic methods, local methods, methods of calculation of forms of objects, methods of calculation of dynamics of objects (Tab. 2) [7]. For the human face it is possible to initialize up to 80 facial landmarks. Commonly, it is borders of eyes, mouth and eyebrows. Molar muscles are not important feature for human expression recognition and its analysis. Each of emotions has the unique dynamics parameters.

TABLE I. RELATIONS OF EMOTIONAL FACIAL FEATURES CHANGING

Emotion	Eyebrow	Mouth
Surprise	Rise	Open
Fear	Rise and wrinkled	Open and stretch
Disgust	Decrease	Rise and ends will decrease
Anger	Decrease and wrinkled	Opens and ends will decrease
Happiness	Bends down	Ends will rise
Sadness	End part will decrease	Ends will decrease

TABLE II. METHODS FOR FACIAL EMOTIONAL STATE RECOGNITION OF HUMAN FACE

Methods	Holistic methods	Local methods
Methods for shapes calculations	Classificatory: Artificial Neural network, Random Forest, Adaboost, Gabor filters, 2D face models: AAM, ASM EBGMM.	Classificatory: Artificial Neural network, Bayes Classificatory, Adaboost, Geometric face models. Own vectors: PCA. Local histograms: HoG, LBP.
Methods for dynamics calculations	Optical flow, Dynamic models.	3D dynamic models Statistical models: HMM, DBN.

Therefore, profiling is the relevant method for emotional identifying and analyzing of an individual features, which makes it possible to significantly increase the level of security, for example, at airports.

These approaches were implemented in the following software for processing video images of a human face subject to emotions [7]:

1) Face Reader - developed by “Noldus Information Technology” (Netherlands), requires the facial video for proper identification. System capabilities: recognition of emotions, the definition of ethnicity, the use of the Active Template method, the creation of an artificial face model. Advantages: recognition with an accuracy of 89%; you can define emotions by frames or completely (video), full visualization (histograms, diagrams of emotions). Disadvantages: does not recognize children under five; inaccurate definition of emotions in a person with glasses; different skin color is perceived by the system in different ways; turned face is not detected.

2) Emotion Software and GladOrs application - developed by “Visual Recognition” (Netherlands). System capabilities: the system creates a 3D model of the face with the

identification of 12 key areas, such as the corners of the eye and the corners of the mouth. Pluses: recognizes anger, sadness, fear, surprise, disgust and happiness. The software is not demanding on the computer. Disadvantages: unknown details of the implementation algorithm (no defects identified).

3) Face Analysis System - developed by “MMER-Systems” (Germany). System features: Face overlay a certain deformable mask that allows you to calculate the necessary parameters in real time. Advantages: recognizes six basic emotions, determines gender, age, and ethnicity; identifies the person if the photo was previously uploaded to the database, has an additional module. Disadvantages: incomplete coverage of the loaded data, since you can work with a webcam; inaccurate results for uploading data.

The task of recognizing emotions is part of a comprehensive system for image analysis. For its proper operation, it is important to clearly identify the person’s face in the image and properly extract it for further processing, rely on a number of dynamic parameters.

Face detection algorithms can be divided into four categories [8]:

- empirical method;
- method of invariant signs;
- recognition on the template implemented by the developer;
- a method detection on external signs (the training systems).

The main stages of algorithms of empirical approach are:

- stay on the image of the person: eye, nose, mouth;
- detection: borders of the person, form, brightness, texture, color;
- combination of all found invariant signs and their verification.

Shortcoming is that this algorithm is very sensitive to degree of an inclination and turn of the head.

III. MATHEMATICAL PROBLEM STATEMENT FOR FACIAL STATE RECOGNITION

Mimic reactions of each person have a certain set of standard manifestation parameters and are divided into two categories: geometric and behavioral.

To describe the quantitative and qualitative parameters of the face (voluntary and involuntary) use a coding system of facial movements. In this case, the quantitative parameter is the intensity of movement from A to E.

The video data stream is a sequential set of frames. The goal of recognition is to merge faces on images into disjoint classes. The task of face recognition is formulated as follows:

it is required to build a recognition function (1) output that determines the image class w , presented by a vector of sign $(x_1(w), \dots, x_n(w))$.

$$F(w) = (F_1(w), F_2(w), \dots, F_k(w)) \quad (1)$$

In this case a class is the one of six basic emotions of the person.

$$F_k(w) = \begin{cases} 1, & \text{if } w \in \Omega_k \\ 0, & \text{if } w \notin \Omega_k \end{cases} \quad (2)$$

Search of a solution is carried out by using of artificial neural networks.

The invariant is the property of some class (set) of mathematical objects remaining invariable at conversions of a certain type [8].

The invariant moments represent characteristic signs which can meet in each picture. Most often persons on video frames are exposed to the different deformations inherent to mimicry of the person. In such conditions it is necessary to tell about "pseudo-invariants" [9].

It is reasonable to convert the color images in a gray-scale

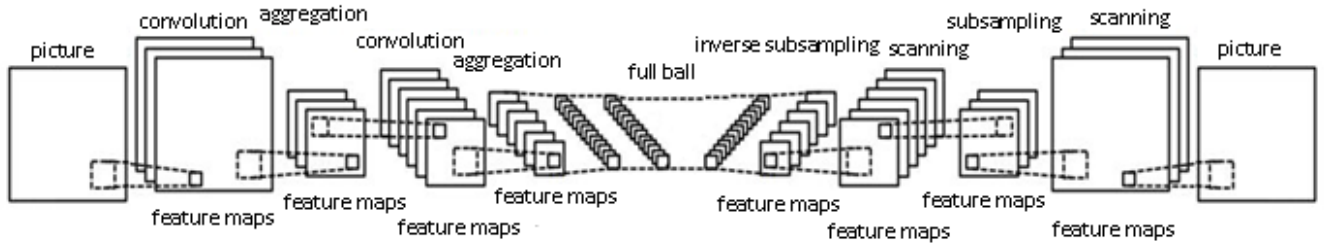


Fig. 1. Structural scheme of combined convolutional neural network

style. After preprocessing and normalization the sample represents a matrix of pixels, each of which matters brightness in the range $[0 \dots 1]$ (Fig. 1).

IV. PROBLEM SOLUTION

The solution of a problem of recognition of emotions belongs to a problem of classification, i.e. the neural network should carry the received data set to emotions, answering to the set of parameters. Let's consider the mathematical description of a problem of recognition:

Let the set of M of images of persons (emotion, for example, surprise) is given as $\{w_1, \dots, w_n\}$, each of which has a vector of values of signs (mimic signs) [9]:

$$x_i = (x_{i1}, \dots, x_{im}) \quad (3)$$

Vector of signs are carried by experts to some classes:

$$\begin{aligned} \Omega_1 &= \{w_1, \dots, w_{m_1}\}, \Omega_2 = \{w_{m_1+1}, \dots, w_{m_1+m_2}\}, \dots \\ \Omega_k &= \{w_{m_{k-1}+1}, \dots, w_{m_k}\}, \end{aligned} \quad (4)$$

where $m_1 + m_2 + \dots + m_k = m$, $m = |M|$.

For a solution of this task in work the combined convolution neural network which consists of convolution neural network, the qualifier and scanning neural network is used (Fig. 2). Use of this network allows selecting emotional reaction of each of above-mentioned features. Data of exits of this network are inputs of the indistinct qualifier which allows making on the basis of the analysis of reaction of separate elements of the person decisions concerning threats which this passenger can represent.

The combined convolution neural network solves such problems as:

- 1) Recognition and description of features of the person.

This task demands definition of position of bodies (eyes, a nose, a mouth, etc.) on a face and also forms of these bodies should be defined.

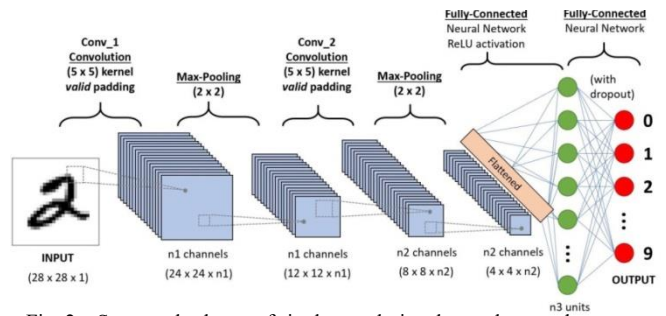


Fig. 2. Structural scheme of single convolutional neural network

- 2) Classification.

According to information on features of the person it is necessary to define what type of emotion is present at the image. Then it is necessary to define information on mood for further development of the intelligent interface.

The combined network consists of convolution neural network, some qualifier and re-convolution neural network. The conceptual structure of such network presented on a Figure 1. Such architecture allows not only to distinguish image elements, but also to notice on it recognition elements.

Re-convolution neural network is specula reflection of convolution neural network.

The ability of the multilayer neural networks trained by method of gradient descent to creation of difficult multidimensional areas on the basis of a large number of the training examples allows applying them as the qualifier to image identification.

Despite it, in traditional full-meshed neural network there is a number of the shortcomings lowering efficiency of their work. First of all, it is the big size of images (the image is understood as the graphical representation of a recognizable image presented in the form of set of pixels), which can reach several hundred. For correct training in such data it is required to increase number of the hidden neurons that leads to increase in number of parameters, and, as a result, lowers training speed, demands the big training selection. But the biggest restriction of such networks is that they do not differ in invariance to different deformations, for example, to transfer or insignificant distortion of an input signal.

The convolution neural network systems (Conv Nets or CNNs) are the logical instrument receives an input parameters as image in the set of pixels view, finds some features on it and due to it sets the parameters (weighted coefficients) to wide data objects in the images and be able to highlight any special things among all over objects. Conv Nets requires the less clean processing power relatively to other processing algorithms. Unlike to standard filter methods that working as hard-engineered unit, the convolution neural network can achieve it through the training processes.

The structure of a Conv Nets is same as connectivity image of human brain biological neurons and was based on group of the “Visual Cortex”. Each one neuron responds to stimuli only in a specially-restricted area of the visual field known as the receptive area. A set of such areas overlap to cover the entire visual area. A Conv Nets are able to clearly capture the spatial and temporal dependencies in any input image data through the application based on relevant filters. CNN structure performs a great setting to the image dataset with help of reduction in the number of parameters involved and reusability of weights. Therefore, the neural network can be trained to understand the sophistication of the image better. On the Fig. 2 presented the conceptual structure of convolution neural network.

As the input commonly we will get the RNG image which means R – red color, G – greed and B – blue. Also, image can be in different color types: gray-scale, HSV, RGB, CMYK, etc.

For the while, take into the mind which the computational power needed for process the image of 8K (7680×4320) sizes. The main purpose of Conv Nets is to compress it into much more lighter state without losing important parts (features). It’s necessary to take care when we’re designing structure with low features learning possibilities and huge data sets (Fig. 3).

On the basis of results of check of the combined

convolution neural network on test selection it is possible to determine a number of significant parameters [10]:

1. Number of convolution layers.
2. Number of aggregation layers.
3. Mutual placement of convolution layers and aggregation layers.
4. Convolution layers (for each layer separately):
 - Convolution core size (for each layer separately);
 - Number of features maps (for each layer separately);
 - Padding size (for each layer separately);
 - Parameter of ending effect.
5. Aggregation layers (for each layer separately):
 - Aggregation core size;
 - Aggregation core function.
6. Fully-connected layers (for each layer separately):
 - Number of fully-connected layers;
 - Sizes of each layer;
 - Classificatory type: auto encoder.
7. Existence of extracting operation for each layer: – extracting percent and random function.

For optimization of structure and parameters of convolution neural network the genetic algorithm is used [10].

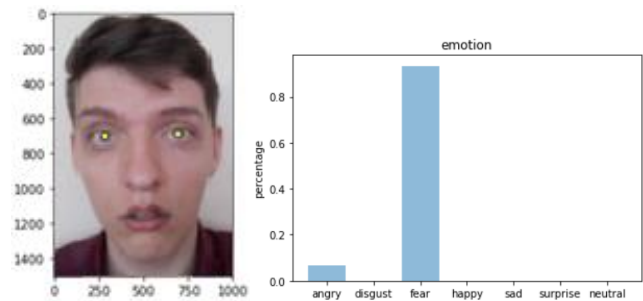


Fig. 3. Facial expression recognition example obtained using complex CNN

V. CONCLUSION

In this work the effective approach for emotional state recognition of human face using digital images analysis is proposed. It is developed the ways of application the combined convolution neural network for assigned task and algorithms of digital image processing was applied. For the solution of the classification task it is used fuzzy neural network classifier such as “NEFCLASS”, inputs of which are outputs of CNN. Given approach has the acceptable recognition level and good enough accuracy. This system can be successfully applied to perform the security purposes in the airports and able to increase the security level. However the system has some limitations as the sensitivity to lightning

level, dependence on head rotation degree. These limitations will be eliminated at continuation of this work in the following directions: gain of compensation of light difference using the complication of an algorithm of adaptive filtering; expansion the range of emotions. In figure 3 presented the result of combined convolution neural network for expression and analysis of human emotional state.

REFERENCES

- [1] A. Krizhevsky, I. Sutskeyever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Network" 2012. (*references*)
- [2] D. Nauck, R. Cruse, "NEFCLASS – a neuro-fuzzy approach for classification of data", January 1995.
- [3] Y. M. Volinsky-Bosmanov "Profayling. Technology for preventing crime acts" in Magnetism, vol. III, G. T. Rado and H. Suhl, Eds. New York: Academic, 2015, pp. 220.
- [4] R. Li, S. Ma, "Impact of depression on response to comedy: A dynamic facial coding analysis" *Journal of abnormal psychology* 116 (4): 804–9, 2007.
- [5] B. A. Knyazev, Y. E. Gapanyuk "Recognition of aberrant behavior of the person" *Journal of engineering*, 2013, p 512.
- [6] P. Ekman and W. Friesen, "Facial Action Coding System: A Technique for the Measurement of Facial Movement", consulting Psychologists Press, Palo Alto, 1978.
- [7] D. Stutz, "Introduction to Neural Networks. Seminar on Selected Topics in Human Language Technology and Pattern Recognition", 2014.
- [8] D. A. Tatarenkov, "Analysis of face recognition methods on images", 2015, p. 270.
- [9] G. Deco and D. Obradovic, "An Information-theoretic Approach to Neural Computing, Springer", New-York, 1996.
- [10] G. K. Voronosky, S. N. Petrashev, S. A. Sergeev, "Genetic algorithms. Artificial intelligence systems and problems of virtual reality", 1997.