INVESTIGATION OF RECOGNITION METHODS OF A PERSON'S FATIGUE USING A 3D CAMERA

Bakhmet A.G., student of IPZ-16-2 group,
e-mail: andrii.bakhmet@nure.ua.
Supervisor: Ph.D., Assoc. Prof.Nazarov O.S.
Kharkiv National University of Radio Electronics

The object of research in the presentation is methods of analysis of image recognition. Algorithms and methods of face recognition are considered. The result of the work is the software for the analysis of human drowsiness with the help of Intel® RealSense 3D-camera.

The analysis of facial expressions originates from the XIX century. Even Darwin in 1872 demonstrated the universality of expressions of man and animal.

In 1971, Ekman and Frice identified six primary emotions, each of which contains a certain content and is characterized by a special expression of the person. They are called: joy, sorrow, fear, disgust, surprise and rage. These emotions appeared to be universal for people of different nationalities and cultures.

Traditionally, the expression of a person's face was studied using either two-dimensional static images or two-dimensional video sequences. However, the analysis based on 2d technology is not capable of working with a large amount of emotion variations. The experiment conducted by Beijing Envada Electric Power Engineering Technology Co., Ltd has shown that the use of three-dimensional technologies for emotion generation yields a much more accurate result than two-dimensional technology.

Consequently, 2D technologies are characterized by the inability to process large volumes of variations of emotions and inadequate precision.

At the same time, it should be noted that the number of studies on the use of a three-dimensional model in face recognition is insufficient. The major obstacle to such research was the lack of a publicly accessible database of three-dimensional facial expressions.

Currently, virtually all existing methods for the analysis of facial expressions and recognition systems are based on static images or dynamic video of two-dimensional databases. In this case, the factors that contribute to the improvement of the decision of the face recognition problem are found. For example, it has been determined that performance in modeling is reduced with head-turn angle or a change in lighting.

A common theme in the study of two-dimensional recognition of facial expressions is that the person is a flat pattern, similar to a two-dimensional geometric form, perceived with certain textures. Such representation leads to the fact that changes in emotions are characterized by measurements performed on a flat image. However, in real life, the main feature of faces is a three-dimensional surface, rather than a two-dimensional picture. The perception of the face as a mobile relief surface instead of a flat pattern in three-dimensional analysis methods has theoretical value and practical application.

Students from the State University of New York at Binghamton developed a three-dimensional face-to-face DB, which included:

– initial three-dimensional forms of face expression;
– two-dimensional structural features of the face.

In total, about 2500 models for 100 items were presented. The developed database was a valuable resource for evaluating the algorithm of facial expressions transcription using the three-dimensional analysis method.

In terms of psychology, it has been discovered that the human visual system can perceive and understand the built-in functions contained on a three-dimensional face, even if such features do not appear in the corresponding two-dimensional flat images. This is due to the fact that the viewer actually represents the possible shape of the face of the person when constructing representations for recognition, that is, human recognition of two-dimensional facial expressions is now much better than machine recognizing.

An important factor in recognizing facial expressions is the movement of the skin. Slight movements are demonstrated by skin in the cheeks, forehead, in the space between the nose and the corners of the mouth, eyes, chin, or mouth.

These areas contain a large number of valuable properties of the surface (convex, concave, or other three-dimensional primitive objects) and can play an important role in determining the differences between subtle expressions of the face. Based on 2D approaches, it is unlikely that three-dimensional surface features (for example, wrinkles) and deep movements will be detected.

In addition, 2D technologies show very high contrast characteristics in protruding areas (e.g. eyes, nose, mouth). Consequently, combined use of 2D and 3D technologies is needed. The basic approach in this case should be based on a three-dimensional analysis. This may allow you to study the subtle structure change for universal and complex expressions.

One of the problems arises in the analysis of the face, taking into account the movement of the head and posture of a person, because people rarely express emotions without the motion of a head or spontaneous posture. With:

– using only the frontal representation puts a precise analysis of expressions under the threat, since the head is an important factor that, in combination with facial expression, reflects the real emotion of the person;
– a significant change in the posture of the head will result in a change in the lighting on the face, which may result in the face being unavailable for analysis. All this increases the complexity to trace the features of the face and posture in the 2D plane.

With regard to the recognition of drowsiness, these technologies are actively engaged in Toyota Europe, which since 2006 has been developing the 3D Face Tracker, designed to analyze the state or mood of the driver, identifying expressions on their faces. The approach is to focus on 3D, not on the 2D model, which is used by most surveillance systems as a more reliable one. Using the images of real people to create a 3D model, the system controls 238 points on the driver's face to analyze its expression.

One can conclude that the most promising is the use of the 3D model to determine the emotional state of the person and identification of drowsiness.

There are several algorithms for recognizing emotions and drowsiness:

1. Flexible comparison method on graphs.

The method is an elastic comparison of graphs with weighted vertices and edges applied to the surface of a three-dimensional model of face with a reference model.

At the stage of determining the emotion the reference graph remains unchanged, and the input graph is deformed to fit the first. This algorithm is used both for two-dimensional input values, in which the graph represents a rectangular grid (Fig. 1), and for three-dimensional, in which the graph represents the structure formed by the anthropometric characteristics of the person (Fig. 2).

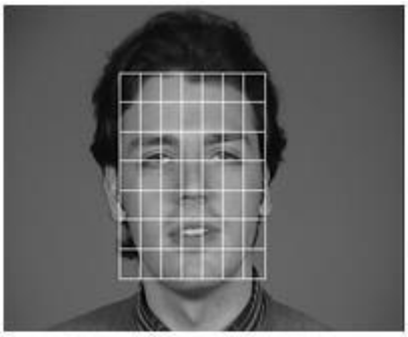

Figure 1.

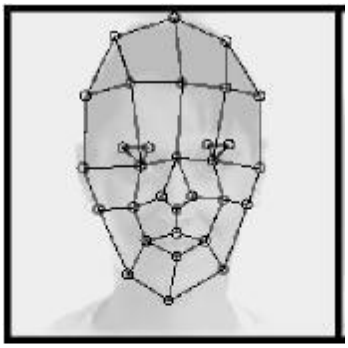

Figure 2.

At the next stage, each vertex of the graph calculates the values of the characteristics. For this purpose, the complex values of the Gabor filter line or their ordered sets-the Gabor wavelets that are calculated in a certain local area of the vertex of the graph locally by summing the value of the brightness of the pixels with the Gabor filters (Fig. 3)

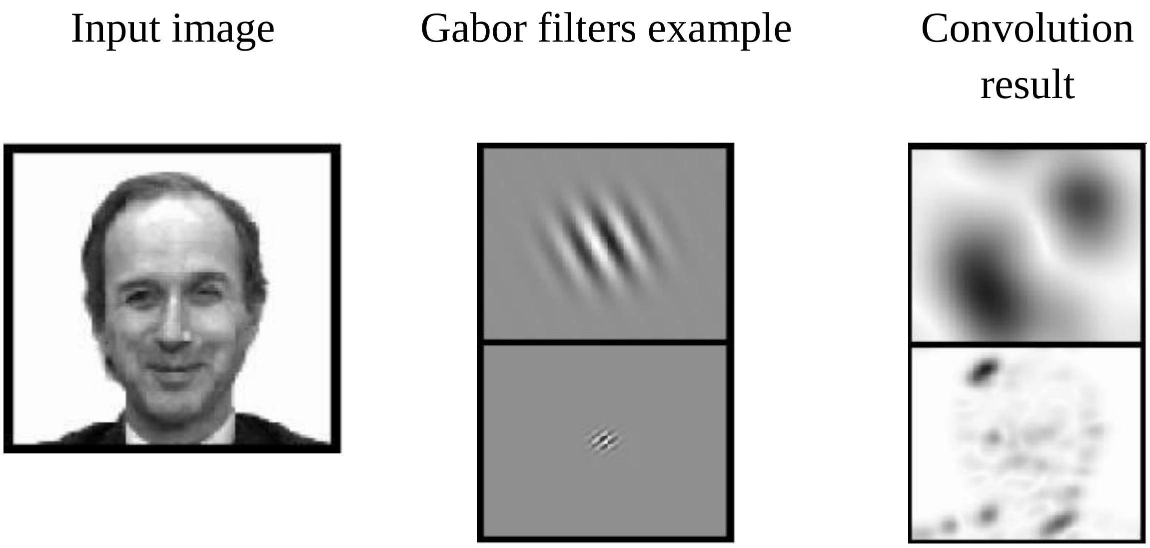

Figure 3.

At the third stage, the edges of the graph are weighted by the distances between adjacent vertices. The difference (distance or discriminatory characteristic) between two graphs is calculated using the deformation function, which takes into account: the difference between the values of the signs calculated at the vertices of the graph and the degree of deformation of the edges of the graph.

The stage of the deformation of the graph is ensured by the displacement of each vertex at a certain distance in a certain direction relative to the initial position of the graph and the choice of a position in which the difference between the values of the signs (Gabor filter responses) at the vertex of the deformed graph and the vertex of the reference graph corresponding to it is minimal. This operation is performed sequentially for each vertex of the graph until the smallest total difference between the two graphs is achieved. This deformation procedure should be performed for all reference emotions embedded in the system database. The result of the algorithm is the standard with the best value of the deformation function.

Different publications say about the accuracy of the definition is close to 95-97%, even in the presence of a deformation of the person's model or changes in the angle of the person up to 15 degrees relative to the device capture 3D model of the face.

Disadvantages:

A. The high computing value of this approach.

B. Low productivity with the addition of new standards and the linear dependence of the algorithm's running time on the size of the base of these standards.

Since the driver's face will be directly ahead of the camera and the algorithm requires high computational power, this algorithm is not recommended. It is recommended to use the algorithm in cases when it is important to record the change in the angle of the person.

2. Neural Networks (NN).

One of the most commonly used variants is NN, which is built on a multilayer perceptron, which allows to classify the input data according to the previous training of the network.

Training NN happens on a set of instructive examples. The essence of training is to adjust the weight of connections between neurons in the process of solving an optimization problem by the method of gradient descent. In the process of training there is: automatic extraction of signs; determination of their importance; construction of interconnections between them.

It is anticipated that a trained NN will be able to apply the experience gained in the training process to unknown patterns due to the ability to summarize the decision. The best results in the field of person recognition were the burial neural network (BNN), which is the development of the ideas of such an architecture of the NN, such as: a cognitron; neocognitron

The success is due to the possibility of taking into account the two-dimensional topology of the image in contrast to the multilayer perceptron. Distinctive features of BNN are:

- local receptor fields that provide local two-dimensional connectivity of the neurons;
- general scales that provide detection of some features in any place of the image;
- hierarchical organization with spatial sampling.

Thanks to this innovation, BNN provides partial resilience to scale changes. This applies to:

- displacement;
- turning;
- change of the angle;
- other distortions and deformations.

The testing of BNN on a database containing reference patterns of emotions on faces with small changes in lighting, scale, rotations in space and position showed a 96% recognition accuracy.

The development of the BNN was achieved in the development of the DeepFace system, which Facebook Company took to recognize the identity of its social network user. All the peculiarities of the architecture of this system are private.

Disadvantages:

A. The need for a complete retraining of the network with the addition of a new standard, which is a rather time and resource-consuming process.

B. Training NN has a number of problems of a mathematical nature, which can include the hit in a local optimum, the choice of an optimal optimization step and etc.

B. Problems in choosing the network architecture (number of neurons, layers, the nature of links).

Due to the prolonged training and retraining of the NN when adding a new standard, this algorithm is not recommended. It can be used in software development with pre-defined set of users.

3. Hidden Markov Models (HMM).

HMM use the statistical properties of the signals and take into account directly their spatial characteristics. Elements of the model are:

- set of hidden states;
- the set of observed states;
- matrix of transition states;
- matrix of transitive probability;
- initial probability of states.

When recognizing an object:

- Markovski models are generated for a given database are checked;
- it is necessary to find the maximum probability that the sequence of observations for a given object is generated by the corresponding model.

Disadvantages:

A. For the algorithm it is necessary to select the model parameters for each database with the standards.

B. The training algorithm only maximizes the feedback of each image on its model, but does not minimize the feedback on other models.

C. Long comparison with reference models.

There are currently no examples of commercial use of HMM for recognizing emotions or drowsiness.

4. The method of the main components.

Once received in the tutorial sample of images, the set of eigenvectors is used to encode all other faces of images represented by a weighted combination of these eigenvectors. Using a limited number of eigenvectors, one can obtain a brief approximation of the input image of the face, which can then be saved in the database in the form of a vector of coefficients, which is at the same time the search key in the base of these emotions.

Initially, the whole instructive set of faces will be transformed into one common data matrix, where each line is a single copy of the image of a face that is arranged in a string. All people of the instructive set should be reduced to one size and with normalized histograms.

Then the data rationing and bringing the lines to the mean value with the calculation of the dispersion is done, the covariance matrix is calculated. For the obtained covariance matrix, the problem of determining proper eigenvalues and eigenvectors corresponding to them is solved. Next, the sort of eigenvectors in order of decreasing eigenvalues and leaving only the first k vectors.

Disadvantages:

A. In the cases where significant changes in the lighting or change of face are present, the efficiency falls significantly.

The algorithm is recommended for the development of software for stationary computers and laptops. In this case, the light from the monitor provides lighting required for the algorithm.

5. Active models of external appearance.

AAM (Active Appearance Models) are statistical models of images that can be adjusted to a real image by means of different types of deformations. This type of model in a two-dimensional version was proposed by Tim Couste and Chris Taylor in 1998.

The advantages of this algorithm are:

– high recognition accuracy;
– stability in the analysis of deformed models;
– no need for long-term adjustment or program training;
– high number of research and available software based on this approach.

The only drawback of the algorithm is the volume of required mathematical calculations that a graphics processor built in RealSense ™ technology cameras handles.

Based on the comparative characteristics, it was decided to use active models of appearance.

The analysis of the possibilities of recognition algorithms in the work, allowed to make a decision on the use of active models of external appearance for the development of face recognition.

Intel® RealSense was chosen as hardware for the program. Camera with RealSense ™ technology consists of:

– webcams with Full HD video support;
– infrared depth sensor, which is used together with infrared laser projector;
– graphics processor, i.e. a specialized microchip for the primary processing of the data stream in such algorithms as depth calculation, barrier filter, etc. The built-in graphical processor essentially unloads the central processor of the computer and suppresses the work of the camera as a whole;
– stereo microphone for high-quality speech recognition.

The essence of the three-dimensional analysis lies in the fact that the depth sensor, which is part of the camera, sends an IR beam and measures the return time. Knowing the speed of light, you can easily calculate the distance from the camera to the object. Thus, we can solve the problem of low-light model.

The disadvantage of this technology can be:

- Demand for computing resources, since RealSense ™ cameras run on Intel® processors of the fourth generation and Windows 8 and further;
- high cost for mass implementation.

Java was used to develop the software.

The application highlights the area of the face (Detection) and the eye areas (Landmark) during operation.

The program responds to:

- eye closure, including incomplete one;
- For a long-term (more than 2 seconds) turn of the head forward or backward.

The system generates an audible signal as a respond.

The program has been tested on 5 different individuals. Stable work results were demonstrated, head and eye closure were detected in 95% of cases.

References.

1. P. Ekman and W. Friesen. Facial Action Coding System. New York: Consulting Psychologists Press, 1977.

2. Y. Yacoob and L. Davis. Recognizing human facial expressions from long image sequences using optical flow. IEEE Trans. PAMI, 18 (6), 1996.

3. G. Donato, M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski. Classifying facial actions. IEEE Trans. PAMI, 21(10):974-989, 1999.

4. M. Lyons, et al. Automatic classification of single facial images. IEEE Trans. PAMI, 21(12):1357-1362, 1999.

5. Y Chang, C. Hu, and M. Turk. Probabilistic expression analysis on manifolds. CVPR’04, Washington DC, 2004.

6. F. Bettinger and T.F.Cootes. A Model of Facial Behavior. In FGR’04.

7. Y Tian, T. Kanade, and J. Cohn. Recognizing action units for facial expression analysis. IEEE Trans. on PAMI, 23(2), 2001.

8. M. Bartlett, J. Hager, P. Ekman, and T. Sejnowski. Measuring facial expressions by computer image analysis. Psychophysiology, 36, 1999.

9. G. Littlewort, M. Bartlett, I. Fasel, J. Susskind, and J. Movellan. Dynamics of facial expression extracted automatically from video. In CVPR Workshop on FPIV'04, 2004.

10. Cohen, F. Cozman, N. Sebe, M. Cirelo, and T. Huang. Semisupervised learning of classifiers: Theory, algorithms for Bayesian network classifiers and application to human-computer interaction. IEEE Trans. PAMI, 26(12), 2004.

11. Cohen, N. Sebe, A. Garg, L. Chen, and T. Huang. Facial expression recognition from video sequences: temporal and static modeling. CVIU, 91(1), 2003.

12. Y Zhang and Q. Ji. Active and dynamic information fusion for facial expression understanding from image sequences. IEEE Trans. PAMI, 27 (5):699-714, May 2005.

13. Barr A. S. The use of curvelets in face recognition / A. S. Barr. – URL: http://www.andybarr.com/mypapers/final2.pdf.

14. Анализ существующих подходов к распознаванию лиц / В. Коломиец – Электронный ресурс: https://habrahabr.ru/company/synesis/blog/238129/.

15. L Wiskott, N Krüger, N Kuiger, C Von Der Malsburg. Face recognition by elastic bunch graph matching. IEEE Transactions on pattern analysis and machine intelligence 19 (7), 775-779.