A technique for detecting diagnostic events in video channel of synchronous video and electroencephalographic monitoring data

D Murashov[1], Yu Obukhov[2], I Kershner[2] and M Sinkin[3]

[1]Federal Research Center "Computer Science and Control" of Russian Academy of Sciences, Moscow, Russia, 119333
[2]Kotel'nikov Institute of Radio Engineering and Electronics of RAS, Mokhovaya str., 11-7, Moscow, Russia, 125009
[3]N.V. Sklifosovsky Research Institute for Emergency Medicine of Moscow Healthcare Department, Bolshaya Sukharevskaya square, 3, Moscow, Russia, 129090

e-mail: d_murashov@mail.ru

Abstract. In this paper, a technique for automated detecting diagnostic events in the video channel of video and electroencephalographic monitoring data is presented. The technique is based on the analysis of the quantitative features of facial expressions in images of video data. The analysis of video sequences is aimed at detecting a group of frames characterized by high activity of frame regions. For detecting the frames, a criterion computed from the optical flow is proposed. The preliminary results of the analysis of real clinical data are presented. The intervals of synchronous muscle and brain activity, which may correspond to an epileptic seizure, are detected. These intervals can be used for diagnosing epileptic seizures and distinguishing them from non-epileptic events. Requirements for video shooting conditions are formulated.

1. Introduction

This paper is aimed at the solution of the problem for automated detecting diagnostic events in the video channel of synchronous video and electroencephalographic monitoring data. Video-electroencephalographic (VEEG) monitoring is a method for long-term synchronous registration of electroencephalography (EEG) and video image. Simultaneous video recording the clinical condition of the patient and the bioelectric activity of the brain (i.e. EEG) allows one to diagnose epileptic seizures reliably and distinguish them from events of non-epileptic nature [1, 2].

The duration of EEG monitoring is usually 24 hours or more, and if used in intensive care units it can last for weeks. Visual analysis of large amounts of data obtained during the long-term VEEG monitoring requires huge labor costs and special training of clinical neurophysiologists. This determines the urgency of developing new methods for detection, quantitative analysis, and classification of diagnostic objects and the exception of artifacts in long-term VEEG monitoring of patients.

The methodology of EEG analysis is traditionally based on the visual analysis of curves. Experts identify non-artifact fragments of the record and analyze its background structure, single (epileptiform) graph elements and their special patterns, which are specific for different clinical conditions [3]. In most cases, the algorithmic capabilities of the software for video EEG instruments are limited to preprocessing multi-channel EEG signals, indicating the likelihood of record artifacts, calculating inter-channel coherence and sources of electrical activity. To simplify the assessment of large volumes of visual information, a mathematical analysis of the oscillations with a graphical presentation of the results, notably a quantitative EEG (qEEG), is used [4]. However, the method of visual presentation of quantitative EEG in the form of trends and histograms does not take into account many artifacts, in particular, chewing and movement of the patient's head. When detecting high-amplitude plots on the qEEG histogram, the physician needs to revise the in video record a fragment of interest for visual assessment and differentiation of an epileptic and artifact event. For this purpose, they use not only the electrographic pattern but also the analysis of the video. In this case, trained EEG technical and medical staff visually scanning all VEEG data and marking specific events [2].

Recently, a number of methods for automatic detecting seizures according to EEG data have been proposed [3, 5-7]. However, all of them use only the native EEG without taking into account the video image, and their accuracy is insufficient for widespread clinical practice. The analysis of publications in periodicals and monographs in the subject domain showed the lack of publications on methods for automatic recognition of epileptic seizures in video sequences obtained during the video-EEG monitoring. Therefore, it is necessary to develop methods and algorithms for automated detection of diagnostic events in long-term video-EEG recordings, which will improve the reliability of their classification and significantly reduce the time for analyzing large amounts of video EEG data and increase their diagnostic significance.

In this paper, we propose a technique for detecting diagnostic events in the video channel of the video-EEG monitoring data of patients in coma.

2. Detecting events in VEEG data

When developing a technique for automated detection of diagnostic events in VEEG data, it is assumed that the decision on event detection is made when the specific features are detected simultaneously in EEG and video channels. This will make it possible to avoid false alarms caused by activity in only one of the data channels. For example, if the camera fixed the movement of a patient, which is not associated with convulsions, or the appearance of medical staff in the frame.

In this case, it is necessary to analyze informative areas in patient images with visible particular muscle contractions. Informative areas are usually associated with the details of a person’s face (eyes, nose, and mouth). Analysis of video sequences with recorded seizures showed a variety of appearance of these seizures. For example, in the case of a non-convulsive attack, only rather weak muscular contractions are observed in the region of the patient’s mouth. At the same time, the rest of the facial muscles remain motionless. In another case, more intense contractions of the muscles of the mouth and periodic movements of the head with immobile muscles in the eye area can be observed. In a number of cases, intense contractions of the muscles are visible all over the face, and contractions of the neck muscles and head movements are also possible. In the absence of seizures, the frames of the video data are relatively static for the studied group of patients.

One of the possible approaches for detecting events in a video channel can be associated with an analysis of the dynamics of the details of a person’s face (eyes, nose, mouth). The literature presents a wide range of methods for localizing these details [8-10]. It should be noted that the images of video sequences obtained during VEEG monitoring have the following features. First, an arbitrary angle of video recording the patient's face (see Figure 1). This feature eliminates methods based on the property of facial symmetry, and methods that require a full frontal image of the face. Second, medical equipment, partially covering the details of the face (see Figure 1(a)). This circumstance also complicates the task of localizing diagnostically important regions. Third, informative regions may not be associated with characteristic points of the face (eyes, corners of the mouth, etc.). Such regions, for example, may be neck areas. Therefore, conventional methods for localization of characteristic points and details of the face may not be applicable. In paper [11], the authors proposed to use displacement vectors of scene objects, which are calculated as projections of the optical flow vectors onto the floor plane, for abnormal behavior detection in the video surveillance systems. This method has demonstrated efficiency in a wide range of operating conditions and scenes.

In this paper, we propose to detect diagnostic events using the magnitude of the criterion characterizing the degree of activity of the region of interest. The region of interest will be the part of the frame that includes the patient's face, head, and neck areas (see Figure 1). As a criterion of the activity of the region of interest, the total optical flow calculated for each frame of the video sequence is used:

$$J(i) = \sum_{x=0}^{W-1}\sum_{y=0}^{H-1}\sqrt{V_x^2(x,y,i)+V_y^2(x,y,i)} + \delta(i),\ i=1,...,N, \tag{1}$$

where $J(i)$ is a criterion value calculated in the frame number i; W, H are the frame width and height; $V_x(x,y,i)$ and $V_y(x,y,i)$ are the optical flow values in axial directions X and Y in the frame number i at a pixel with coordinates (x,y); $\delta(i)$ is a noise.

a) b)

Figure 1. Frames from long-term VEEG records.

Since the noise component is present in the model (1), the smoothed value of the activity criterion $\hat{J}(i)$ should be used to detect events. The smoothed $\hat{J}(i)$ value is obtained using a discrete version of the Kalman-Bucy filtering algorithm [12]. We apply the Kalman-Bucy algorithm since it provides the optimal estimate in the sense of minimum error variance. The decision to fix a diagnostic event is made according to the threshold rule. To avoid false alarms of the detector due to short-term spikes, the decision about the occurrence of an event is made if the value of $\hat{J}(i)$ exceeds a predetermined threshold in a sequence of frames not shorter than M. Thus, the decision rule is formulated as follows:

$$Event = \begin{cases} 1,\ if\ \hat{J}(i) \geq T\ u\ i-i_0 \geq M; \\ 0,\ if\ \hat{J}(i) < T\ or\ i-i_0 < M, \end{cases} \tag{2}$$

where $Event$ is an event indicator, T is a threshold value, i_0 is a frame number from which the inequality $\hat{J}(i) \geq T$ is taking place, M is the length of the sequence of frames required to make a decision about the appearance of a diagnostic event. The threshold value is defined as follows:

$$T = \hat{J}_0 + k\sigma, \tag{3}$$

where $\hat{J}_0$ is computed as a mean value of $\hat{J}(i)$ in a fragment of video sequence with low dynamics of the scene, σ is a standard deviation of $\hat{J}(i)$, k is a coefficient.

Thus, the algorithm for detecting diagnostic events in the video channel of VEEG monitoring data consists of the following steps.

1. Reading frame number 1 of a video sequence.

2. Computing the total optical flow in the region of interest in the video sequence frame according to the formula (1).
3. Computing the smoothed value of the activity index $\hat{J}(i)$.
4. Checking conditions (2) and (3). If the condition $\hat{J}(i) \geq T$ is satisfied, the current frame number $i_0 = i$ is stored. If the condition is not satisfied, go to step 1.
5. Repeating steps 1–3. If the conditions $\hat{J}(i) \geq T$ and $i - i_0 \geq M$ are satisfied, an event is detected. If not, go to step 1.

In the next section, an experiment aimed at testing the proposed technique is described.

3. Experiment

The developed technique is implemented in the MatLab software environment. For computing $V_x(x, y, i)$ and $V_y(x, y, i)$ in (1), Lucas–Kanade algorithm [13] is applied. This algorithm for computing the optical flow is chosen from the condition of the highest performance in comparison with other methods. The magnitude of the smoothed activity index $\hat{J}(i)$ is estimated using a discrete version of the Kalman-Bucy filtering algorithm [12]. The values of the filtering algorithm parameters are selected when processing test video sequences, and based on the best ratio of the error and speed values.

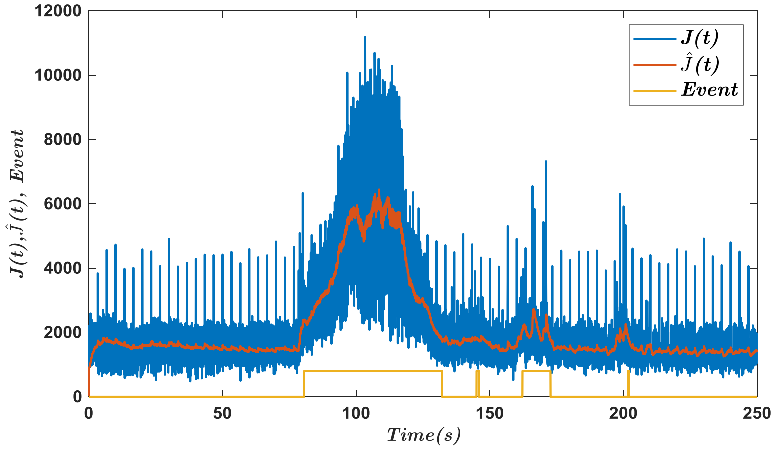
The developed technique was applied to five videos of patients in a coma. In three records epileptic seizures are detected, including non-convulsive ones. For each of the five video sequences, the parameters of the decision rule (2), (3) were determined from fragments with low scene dynamics. The number of frames corresponding to the shortest event duration in condition (2) was chosen equal to 75, which corresponds to a time interval of 2.5 seconds. The value of the coefficient in (3) was selected as $k = 1.1$. Examples of graphs of the criterion $J(t)$ and its smoothed value $\hat{J}(t)$, as well as the event indicator $Event$ for two fragments of the VEEG monitoring of the patient shown in Figure 1(a) are given in Figures 2 and 3. Here and below, instead of the variable i designating the frame number, we use the variable $t = i / FrameRate$, where t is the time, and $FrameRate$ is the frame rate of the video. In the experiment, we processed videos with a frame rate equal to 30 frames per second. Figure 2 (a) demonstrates the detection of diagnostic events. For the videos, corresponding to the graphs shown in Figures 2 (a) and 3, the parameters in expression (3) for determining the threshold value T were found to be equal to $\hat{J}_0 = 1604$ and $\sigma = 127.9$.

Figures 2 (b, c) show the graphs of the processed EEG signals from the synchronous recording of VEEG monitoring. Figure 2 (b) shows the projection of the ridge of the wavelet spectrogram on the power spectral density (PSD) and time axes. Using an adaptive threshold, the ridge points (maximum values of the power spectral density at each time point), which lie above this threshold, were calculated. Close points of the ridge, lying above the threshold, were combined into clusters - fragments of the ridge, marked in black on the graph. These ridge fragments are interpreted as episodes of suspicious activity, similar to an epileptic seizure [7].

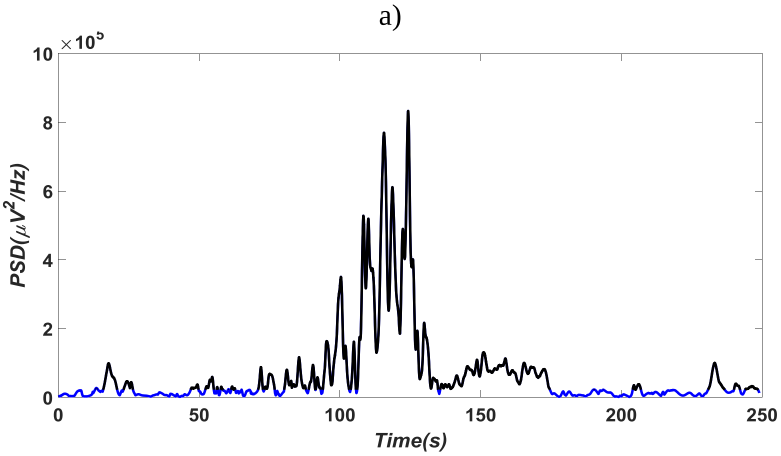
Figure 2 (c) shows the EEG signal in one of the channels, filtered by a Butterworth filter of the 8th order with a passband from 5 to 22 Hz and notch filters at frequencies multiple of 50 Hz. The sampling frequency of the signal is equal to 500 Hz. The black color indicates the suspicious intervals obtained by analyzing the ridges of the wavelet spectrograms. From Figures 2 (a-c) one can see the intervals of synchronous muscle and brain activity, which may correspond to epileptic seizures. These intervals are found between 75 and 130 seconds, 144 and 147 seconds, 165 and 175 seconds, and between 202 and 207 seconds.

The graphs in Figure 3 correspond to video record without seizures.

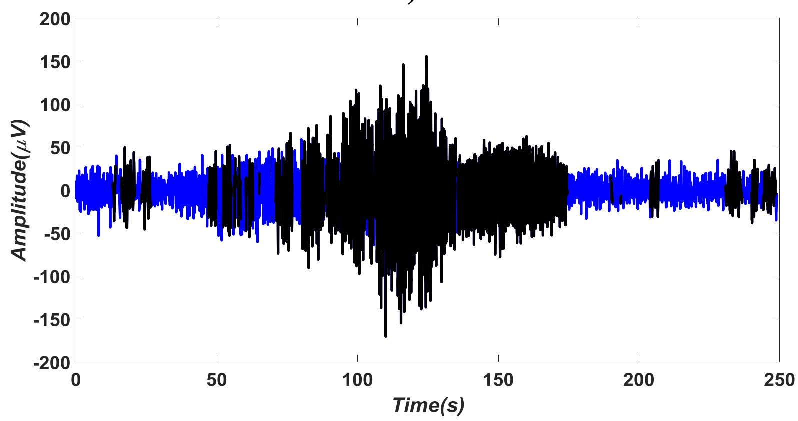
A fragment of VEEG data from another patient recorded a non-convulsive epileptic seizure. Figure 4 shows the results of processing synchronous video and EEG channels of this recording. Figure 4 (a) demonstrates the detection of diagnostic events in the video channel of the VEEG record. This figure shows manifestations of seizure, which is poorly expressed, but distinguishable in $\hat{J}(t)$ graph.



a)



b)



c)

Figure 2. Graphs obtained from VEEG monitoring data illustrating detection of diagnostic events: (a) graphs of the criterion $J(t)$, estimate $\hat{J}(t)$, and event indicator $Event$; (b) projection of the ridge of the wavelet spectrogram on the power spectral density (PSD) and time axes; (c) EEG signal in one of the channels, filtered by a Butterworth filter of the 8th order and notch filters.

Figure 4 (b) shows the projection of the ridge of the wavelet spectrogram on the power spectral density (PSD) and time axes. Figure 4 (c) illustrates the EEG signal in one of the channels, filtered by a Butterworth filter of the 8th order and notch filters.

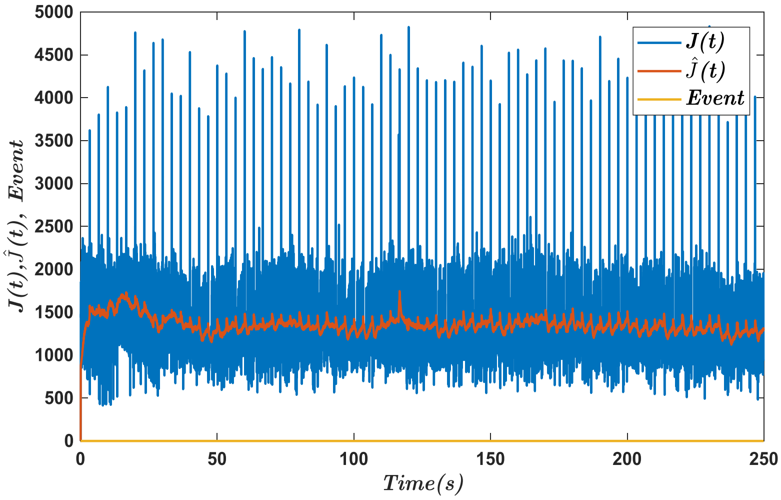


Figure 3. Graphs of the criterion $J(t)$, smoothed estimate $\hat{J}(t)$, and event indicator $Event$. The analyzed video does not contain seizures.

In Figures 4 (a-c) the intervals of synchronous muscle and brain activity, which may correspond to an epileptic seizure, can be seen. These intervals are visible between 7 and 45 seconds, 50 and 53 seconds, and between 95 and 98 seconds.

For the video, corresponding to the graphs shown in Figure 4 (a), the parameters in expression (3) for the threshold value T were computed equal to $\hat{J}_0 = 670.44$ and $\sigma = 12.29$.

It follows from Figures 2, 3, and 4 that diagnostic events can be detected using the criterion (1) and the rule (2)-(3) at different angles of shooting and with partial occlusion of the patient's face with medical equipment.

4. Requirements for recording video data of VEEG monitoring

Requirements for recording video data of VEEG monitoring are derived from the need for reliable detection of diagnostic events. For this, the field of view must be selected so as to provide the necessary dynamic range of values of the criteria used for event detecting. In the field of view of the camera should not get the details of the scene, generating a strong noise background. Essential for the video channel is the immobility of the camera. In the video sequence captured by the unfixed camera, the event cannot be detected due to the high level of noise in the optical flow caused by camera movement. Camera resolution should allow fixing facial expressions and muscle contractions of small amplitude. Based on the analysis of the video channel of the video EEG monitoring data, the following requirements for the video shooting parameters are formulated. First, the field of view of the video camera should cover the patient's head and neck. Secondly, the camera should be fixed. Thirdly, the resolution of the camera matrix should not be lower than HD.

5. Conclusions

A technique for automatic detection of diagnostic events based on the analysis of the quantitative characteristics of the patient's activity in video records is proposed. Analysis of video sequences is aimed at detecting a group of frames with high scene dynamics. A criterion computed from the optical flow magnitude is applied. The preliminary results of the analysis of real clinical data for patients in a coma are presented. The results of the analysis showed the efficiency of the proposed algorithm at different angles of shooting and partial occlusion of the patient's face with the details of medical equipment. The comparison of the results of diagnostic event detection from the video record with data obtained from the synchronous EEG showed the possibility of reliable diagnosing epileptic seizures and distinguishing them from non-epileptic events. Future research will be aimed at applying pattern classifiers for detecting epileptic seizures based on a joint analysis of synchronous EEG and video channels.

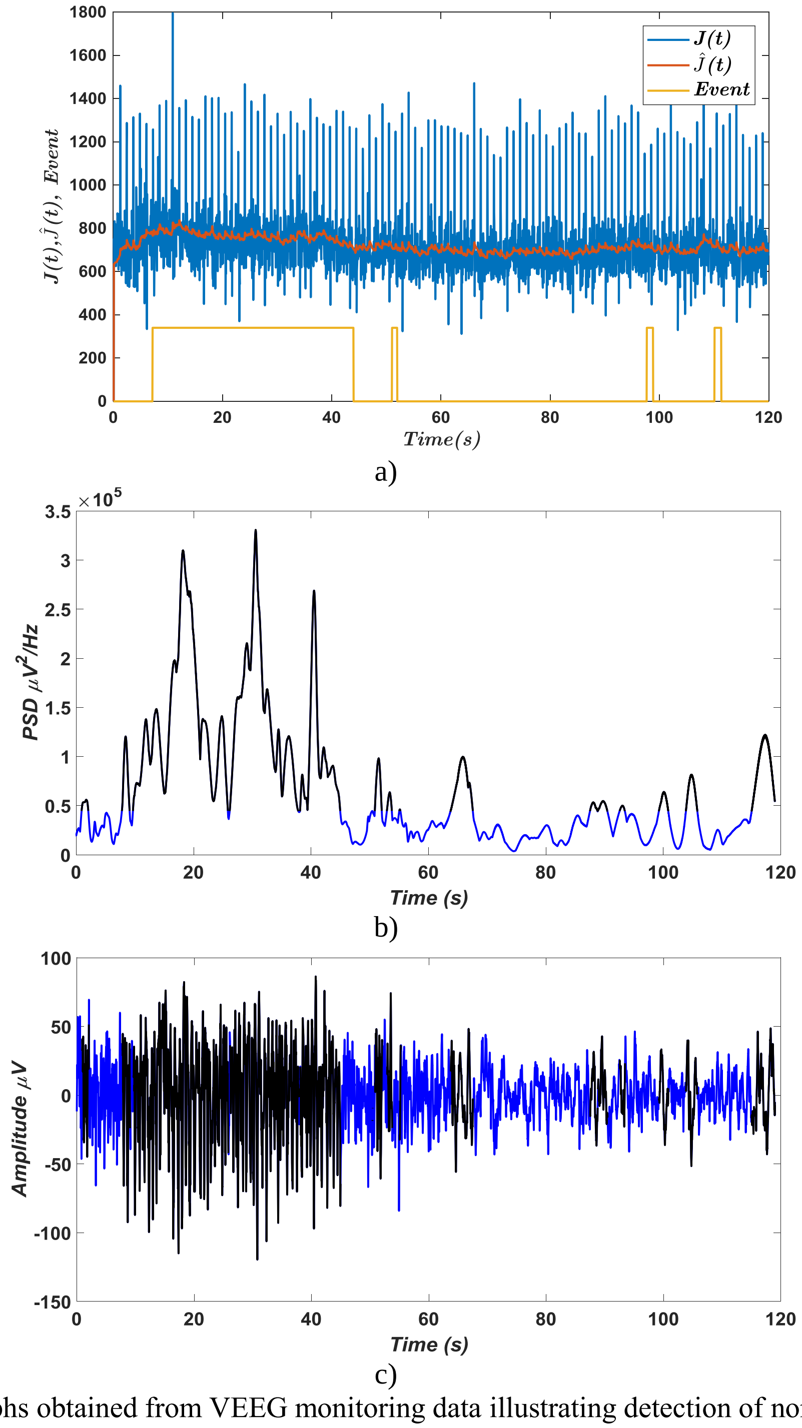


Figure 4. Graphs obtained from VEEG monitoring data illustrating detection of non-convulsive seizure: (a) graphs of the criterion $J(t)$, estimate $\hat{J}(t)$, and event indicator *Event*; (b) projection of the ridge of the wavelet spectrogram on the PSD and time axes; (c) EEG signal in one of the channels, filtered by a Butterworth filter of the 8th order and notch filters.

6. References

[1] Gravino G, Galea B, Soler D, Vella N and Aquilina J 2016 Video-EEG Long Term Monitoring as a new service at Mater Dei Hospital *Malta Medical Journal* **28** 46-54

[2] Lee Y Y, Lee M Y, Chen I A, Tsai Y T, Sung C Y, Hsieh H Y and Wu T 2009 *Chang Gung Med J* **32** 305-312
[3] Hirsch L, Brenner R 2011 *Atlas of EEG in critical care* (John Wiley & Sons)
[4] Duffy F H, Hughes J R, Miranda F, Bernad P and Cook P 1994 *Clinical Electroencephalography* **25** VI-XXII
[5] Tzallas A T, Tsipouras M G and Fotiadis D I 2007 *Computational Intelligence and Neuroscience* **2007** 80510
[6] Antsiperov V E, Obukhov Y V, Komol'tsev I G and Gulyaeva N V 2017 *Pattern Recognition and Image Analysis* **27** 789-803
[7] Obukhov K, Kershner I, Komol'tsev I and Obukhov Y 2018 *Pattern Recognition and Image Analysis* **28** 346-353
[8] Yow K C and Cipolla R 1997 *Image and vision computing* **15** 713-735
[9] Viola P and Jones M J 2004 *International journal of computer vision* **57** 137-154
[10] Singh A, Patil D, Reddy M and Omkar S N 2017 Disguised face identification (DFI) with facial keypoints using spatial fusion convolutional network *IEEE International Conference on Computer Vision* 1648-1655
[11] Shatalin R A, Fidelman V R, Ovchinnikov P E 2017 Abnormal behavior detection method for video surveillance applications *Computer Optics* **41(1)** 37-45 DOI: 10.18287/2412-6179-2017-41-1-37-45
[12] Kalman R E, Bucy R S 1961 *Journal of basic engineering* **83** 95-108
[13] Lucas B D, Kanade T 1981 *IJCAI81* 674-679

Acknowledgments
This research is partially funded by RFBR, grant № 18-29-02035.