A video-based eye pupil detection system for diagnosing bipolar disorder

GÖKAY AKINCI

EDİZ POLAT

ORHAN MURAT KOÇAK

A video-based eye pupil detection system for diagnosing bipolar disorder

Gökay AKINCI,[1] **Ediz POLAT,**[1,*] **Orhan Murat KOÇAK**[2]
[1]Electrical and Electronics Engineering Department, Faculty of Engineering, Kırıkkale University,
Kırıkkale, Turkey
[2]Department of Psychiatry, Faculty of Medicine, Kırıkkale University, Kırıkkale, Turkey

Abstract: Eye pupil detection systems have become increasingly popular in image processing and computer vision applications in medical systems. In this study, a video-based eye pupil detection system is developed for diagnosing bipolar disorder. Bipolar disorder is a condition in which people experience changes in cognitive processes and abilities, including reduced attentional and executive capabilities and impaired memory. In order to detect these abnormal behaviors, a number of neuropsychological tests are also designed to measure attentional and executive abilities. The system acquires the position and radius information of eye pupils in video sequences using an active contour snake model with an ellipse-fitting algorithm. The system also determines the time duration of the eye pupils looking at certain regions and the duration of making decisions during the neuropsychological tests. The tests are applied to 2 different groups consisting of people with bipolar disorder (bipolar group) and people without bipolar disorder (control group) in order to mathematically model the people with bipolar disorder. The mathematical modeling is performed by using the support vector machines method. It is a supervised learning method that analyzes data and recognizes patterns for classification. The developed system acquires data from the being tested and it classifies the person as bipolar or nonbipolar based on the learned mathematical model.

Key words: Bipolar disorder, Eye pupil detection and tracking, support vector machines

1. Introduction

Eye pupil motion detection and tracking systems are used in several fields such as human computer interaction and neuroscience, and especially for diagnosing neuropsychiatric disorders. Pupil detection and tracking systems have been used for neuropsychological investigation of neuronal activity [1–3]. Pupil motion detection and tracking have been proven to be useful in studying human visual attention [4–7]. Eye detecting and tracking techniques/systems can be divided into 3 categories [8]. The first category includes the systems that employ electrooculography, which records the electric potential differences of the skin. The second category includes the systems that employ a scleral contact lens or search coil, which uses a contact lens in the eye. The final category includes the systems that employ video/image-based techniques that use image processing and computer vision algorithms. Unfortunately, the common problem of the electrooculography and scleral contact lens or search coil techniques is the use of intrusive and expensive sensors [9]. These invasive techniques can quickly become tiresome or uncomfortable for the user. Video/image-based techniques have minimized this invasiveness to some degree. Hence, these techniques are considered the least invasive of these modalities and are the most popular

*Correspondence: edizpolat@gmail.com

methods for eye/pupil detection and tracking. Near-infrared illumination and image-based techniques can be used together to minimize the illumination effects. The near-infrared light wavelength is between 700 and 900 nm. It is invisible to the human eye and it does not become tiresome or uncomfortable for the user. Near-infrared illumination also helps reveal the detailed structure of dark parts such as the eye's pupil. The pupil of the eye absorbs much light, but reflects less light. The iris reflects strongly in the infrared light. The strong infrared light reflectance yields high contrast images that are particularly beneficial to eye pupil detection, and in this way the eye's pupil can be seen more clearly for near-infrared illumination. Many image processing techniques have been developed for eye pupil and iris segmentation. Most of the iris and eye pupil segmentation methods use an Integro differential operator and Hough transform in order to localize the iris and eye pupil boundary accurately. Daugman proposed an Integro differential operator in order to detect the eye pupil and iris boundary [10]. Wildes used the Hough transform method and a voting procedure in order to locate the eye pupil and iris boundary [11].

In this study, a video-based robust and accurate eye pupil detection system and neuropsychological tests are designed and implemented for diagnosing bipolar disorder. The system employs an active contour snakes method [12] to detect and acquire data including eye pupil position in video sequences. The time durations of the eye pupils looking in certain directions are also determined for later usage in the tests. The tests are designed such that the cognitive process and ability-related problems including concentrating, remembering, and making decisions in people with bipolar disorder can be detected using the system. The system is able to learn and mathematically model the patterns of the bipolar group using the support vector machine (SVM) method [13], which is a supervised learning method that analyzes data and recognizes patterns for classification.

The rest of the paper is organized as follows. The details of the developed system, active contour snake model, test interface, feature extraction, and statistical analysis of data are described in Section 2. Subsequently, support vector machines for classification processes are also described in Section 2. Finally, Section 3 concludes the paper.
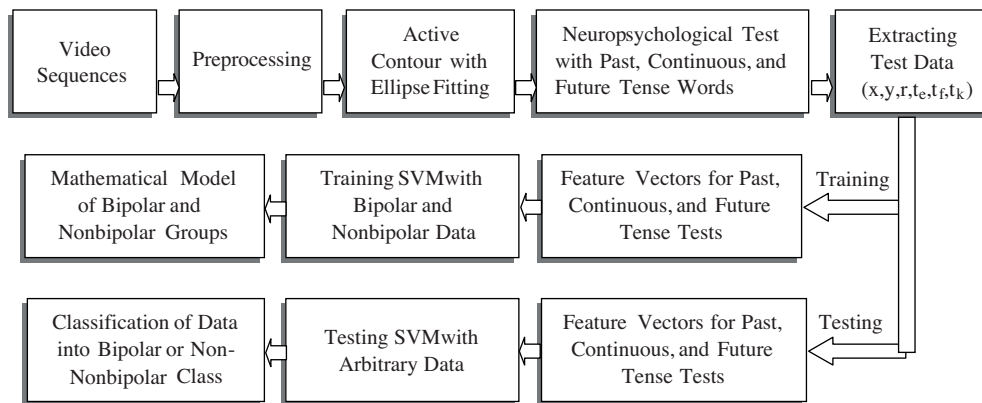
Figure 1. Overview of the developed system.

2. The developed system

The main purpose of this study is to develop a robust and accurate video-based eye pupil detection system that can be used to diagnose bipolar disorder by employing neuropsychological tests. People with bipolar disorder experience changes in cognitive process and abilities including reduced attentional and executive capabilities and impaired memory. In order to detect these abnormal behaviors, a number of neuropsychological tests are

designed to measure attentional and executive abilities. The tests are designed such that the cognitive process and ability-related problems including concentrating, remembering, and making decisions in people with bipolar disorder can be detected using the system. The Integro differential operator, Hough transformation, and active contour snake model with ellipse-fitting algorithms are frequently used in this field of study for detection of eye pupils in images. These 3 methods were implemented and quantitatively compared and the best method for the system was determined in our previous study [14]. The Integro differential operator and Hough transform only work when the iris and eye pupils have a circular boundary [15]. The snake model with an ellipse-fitting algorithm localizes perfectly, while the other 2 methods fail. Experimental results with quantitative analysis show that the snake model with ellipse-fitting algorithm is more robust and reliable. Thus, the snake model with ellipse-fitting algorithm is used in the proposed system. The system is able to learn and mathematically model the patterns of the bipolar group using the SVM method, which is a supervised learning method that analyzes data and recognizes patterns for classification. The overview of the developed system is given in Figure 1. The details of the system are described in the following sections.

2.1. Snake model with ellipse-fitting algorithm

Active contour snake models were first created by Kass, Witkin, and Terzopoulos as a means of locating shapes in images. Snake models have been successfully applied in a variety of problems in computer vision and image analysis. An active contour, also called a snake, is an energy-minimizing spline. It is basically a dynamic closed curve that moves within an image and tries to lock onto other curves. The snake has 2 types of energy, as given in Eq. (1). The internal energy is defined by the shape of the snake and is made up of continuity energy and curvature energy. Continuity energy defines how evenly spaced the contour's points are from each other. Curvature energy defines how smooth the contour is. The external energy is the image energy.

$$E(\vec{v}) = E_{internal}(\vec{v}) + E_{external}(\vec{v}) \tag{1}$$

In the preprocessing step in our system, the video sequences are first converted into gray scale to reduce the processing time and space for each frame. A Gaussian filter is then applied to reduce the general noise and illumination effects. Resulting images become clearer and smoother, which provides more robust pupil boundary detection. Pupil boundaries are detected by applying the active contour snake model. The ellipse-fitting algorithm is then applied to these detected boundaries. Finally, the eye pupils are detected with the image coordinates (x, y) and ellipse parameters (width and height). Examples of detected eye pupils in video sequences are shown in Figure 2. The snake algorithm is applied to the whole image without giving any special starting point. The weight of continuity energy, the weight of curvature energy, and the weight of image energy are experimentally chosen to be 0.20 for the best detection of the pupil. The minimal number of contour points that must be moved during any iteration to keep the iteration process running is chosen to be 10. If, at some iteration, the number of points is less than this criterion, the algorithm terminates.

In the literature, the Integro differential operator and Hough transformation are also used for pupil detection in images. Active contour snake model, Integro differential operator, and Hough transformation algorithms were all implemented and quantitatively compared in our previous study and the best method is used in this study. The detection results for these algorithms are shown in Figure 3.

In order to quantitatively compare the algorithms, 2 video sequences are used. First, eye pupil coordinates and radius values in all frames are manually extracted to obtain the true (ground truth) values. Second, all 3 algorithms are applied to the video sequences to determine eye pupil coordinates and radius values. Finally,

the mean-squared error (MSE) is used to quantify the difference between values obtained by the algorithms and the true values in all frames. The results are shown in Tables 1 and 2.
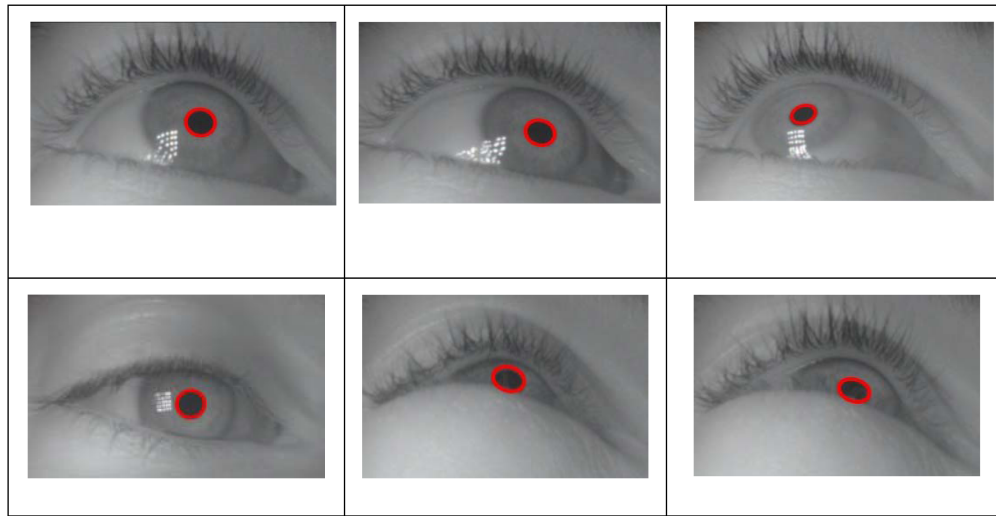
Figure 2. Eye pupil detection results using active contour snake model with ellipse-fitting algorithm.
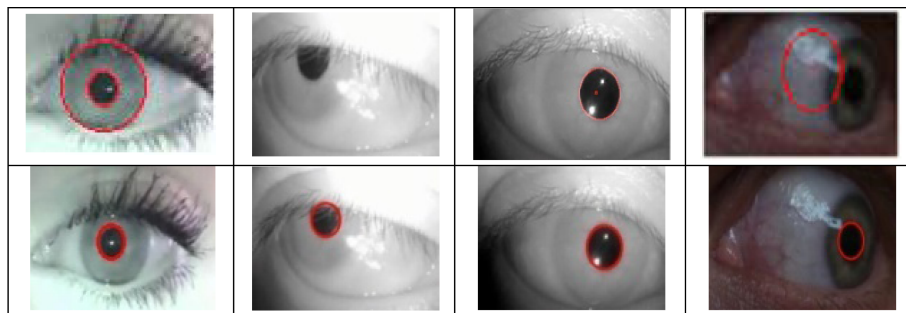
Figure 3. Eye pupil detection results using Integro differential operator algorithm (first row) and Hough transformation algorithm (second row).

Table 1. Quantitative comparison using MSE values for the 3 algorithms for Video-1.

Video-1 MSE	Integro differential	Hough transformation	Active contour snake
Eye pupil coordinates (x, y)	3.0348	3.0015	0.8944
Eye pupil radius (r)	0.4000	0.6633	0.3464

Table 2. Quantitative comparison using MSE values for the 3 algorithms for Video-2.

Video-2 MSE	Integro differential	Hough transformation	Active contour snake
Eye pupil coordinates (x, y)	2.0494	2.3874	0.3162
Eye pupil radius (r)	0.4472	0.7071	0.3162

As shown in the tables, the eye pupil localization and radius errors are much smaller for active contour snake algorithm used in this study. Smaller MSE errors provide more robust, accurate, and reliable detection of

the pupils. In addition, the eye pupil detection rate (DR) (**??**) and false alarm rate (FAR) (**??**) are also defined to quantify the performance of the system as follows:

- TP (True positive): Number of correctly detected eye pupils in the video sequence (i.e. Figure 3, first row, first image).

- FP (False positive): Number of incorrectly detected eye pupils in the video sequence (i.e. Figure 3, first row, last image).

- FN (False negative): Number of not detected eye pupils in the video sequence (i.e. Figure 3, first row, second image).

$$\text{DR} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad \text{FAR} = \frac{\text{FP}}{\text{TP} + \text{FP}} \tag{2}$$

The results of the detection rate and false alarm rate for 2 video sequences are given in Table 3. As is seen, the active contour snake algorithm has the highest detection rate and the lowest false alarm rate among the others.

Table 3. Detection and false alarm rates for the 3 algorithms for Video-1 and Video-2.

Video-1	Integro differential	Hough transform	Active contour snake
Detection rate (DR)	0.625 (62.5%)	0.829 (82.9%)	0.953 (95.3%)
False alarm rate (FAR)	0.286 (28.6%)	0.341 (34.1%)	0.146 (14.6%)

Video-2	Integro differential	Hough transform	Active contour snake
Detection rate (DR)	0.750 (75.0%)	0.864 (86.4%)	0.959 (95.9%)
False alarm rate (FAR)	0.250 (25.0%)	0.136 (13.6%)	0.042 (4.2%)

2.2. Test interface and feature extraction for classification

The algorithms are implemented using MS Visual Studio C++ and OpenCV [16] to prepare the neuropsychological test interface to observe the differences between the bipolar group and the control (normal) group. The test is designed to measure the time duration of reading and confirming (making a decision) and the time duration of eye pupils looking at certain regions while reading test words during the test. Since the bipolar group has difficulties in cognitive process and abilities including reduced attentional and executive capabilities, it is expected to have a different test result patterns in the bipolar group. In particular, the perception of tense suffixes in the bipolar disorder group is thought to be different than perception of the tense suffixes in the control (normal) group. In Turkish, tense suffixes are "-di" for past tense, "-yor" for continuous tense, and "-cek" for future tense. These suffixes are appended to the end of the verb and they indicate the status of the verb of sentences. Neural mechanisms that make up mental representation of status of the verb are still an active research area. However, the mental model can provide the shaping of linguistic expressions and functional imaging studies are performed within this scope. A problem with creating a mental model of the status of the verb in the bipolar disorder group is thought to be determined by using tense suffixes to indicate the status of the verb of the sentence [17]. Hence, these suffixes are employed in the neuropsychological test used in this study. Sample images of the test interface and the developed eye pupil detection system are given in Figure 4.
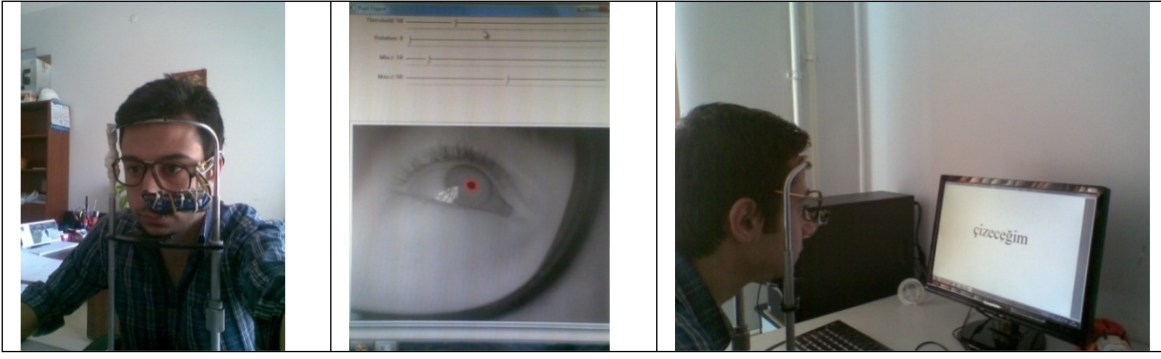
Figure 4. The developed system (left). Detected eye pupil (middle). The test interface (right).

The test is applied to bipolar and control (normal) groups, whose members' ages range from 20 to 50 years old, to collect and later mathematically model the data. A number of selected test words are displayed on a computer screen one at a time and the person looks at the word and clicks the mouse to continue with the next word. The selected words consist of Turkish meaningful words such as "okudum" (I have read), "okuyorum" (I am reading), and "okuyacağım" (I will read) and Turkish meaningless words such as "kalpardım", "kalparıyorum", and "kalparacağım". The eye pupil detection system detects the eye pupil in video frames acquired by the camera attached to the glasses the person is wearing. While the person reads the test word, the system determines the following parameters:

- The (x, y) image coordinates of the eye pupil,

- The width and height (w, h) of the ellipse that fits on the eye pupil,

- The time in milliseconds pupil is looking at these coordinates constantly,

- The part of the word the pupil looking at (1- Root, 2- Suffix, 3- Other place on the screen),

- The tense of the word (1- Past, 2- Continuous, 3- Future),

- Whether the word is meaningful or meaningless (1- Meaningful, 2- Meaningless),

- The number of times the pupil looked at the coordinates,

- The time in milliseconds to read the word.

Sample data acquired from the test are given in Table 4.

In order to mathematically model the acquired data using the SVM, a number of features that identify the bipolar and control (normal) groups are needed. For that purpose, a ratio of reading time duration of the tense suffix (t_e) to reading time duration of the whole word (t_f) can be defined as an indication of calling the word from the mental lexicon, as shown in Eq. (3). This ratio can be used to analyze the effect of reading time duration of tense suffix.

$$\text{Suffix Ratio} = \frac{\text{Reading time duration of suffix } (t_e)}{\text{Reading time duration of word } (t_f)} \tag{3}$$

The second feature is the ratio of last location reading time duration (t_k) to reading time duration of the word (t_f), as shown in Eq. (4). The last location reading time can be defined as the time duration for the last location where the eye pupil looked just before deciding.

Table 4. Sample data acquired from a person in the control group for a future tense meaningful word.

Time Elapsed (ms)	Pupil Coordinates (x, y)	Ellipse width, height	Time duration at (x, y)	Part of word at (x, y) 1- Root 2- Suffix 3- Other
33	(763, 312)	18, 19	33	2
66	(763, 312)	18, 19	66	2
99	(668, 326)	19, 19	33	2
132	(588, 312)	18, 19	33	1
165	(538, 312)	19, 19	33	1
198	(538, 312)	19, 19	66	1
231	(588, 312)	18, 19	33	1
264	(650, 326)	19, 19	33	2
297	(763, 312)	18, 19	33	2
330	(853, 312)	18, 19	33	2
363	(853, 312)	18, 19	66	2

$$\text{Last Location Ratio} = \frac{\text{Last location reading time duration} \quad (t_k)}{\text{Reading time duration of word} \quad (t_f)} \tag{4}$$

Other features that help distinguishing bipolar and normal people are whether the test word is meaningful or meaningless (C), the number of times of strolling over the word before deciding (T), the pupil looking at the part of the word (root or suffix) just before deciding (L), and the reading time duration of the test word (t_f). The feature vector in Eq. (5) is used in the SVM to model and classify the bipolar and normal people:

$$F = \left[\left(\frac{t_e}{t_f} \right), (L), \left(\frac{t_k}{t_f} \right), (C), (T), (t_f) \right]. \tag{5}$$

The details of the SVM method are given in the following sections.

2.3. Statistical analysis of acquired data

The data acquired from the bipolar disorder group and the control (normal) group are statistically compared in this section. The data are collected from 20 bipolar and 20 normal people whose ages are in the range of 20 and 50 years old.

Table 5 shows the locations (red dots) where the eye pupil looked at while reading the test word "vurdum" (word with past tense), "vuruyorum" (word with continuous tense), and "vuracağım" (word with future tense) for a normal and a bipolar person. It can be seen from the table that the bipolar person spends more time (more red dots) to complete the test due to reduced attentional and executive capabilities.

During the test, people are asked whether the test word on the screen is meaningful or meaningless (C). The total number of correct and incorrect answers obtained from the bipolar disorder group and the control group are given in Table 6. It is observed that the bipolar disorder group has made more mistakes than the control (normal) group.

The average values for the suffix ratio t_e/t_f are calculated for each test word having continuous, past, and future tense suffixes in order to observe the effect of tense suffix on calling the word from the mental lexicon. The average values for the suffix ratio t_e/t_f are given in Table 7.

Table 5. The locations (red dots) where the eye pupil looked at while reading the test word.

	Normal	Bipolar
Testword with past tense	vurdum	vurdum
Testword with continuous tense	vuruyorum	vuruyorum
Testword with future tense	vuracağım	vuracağım

Table 6. Total number of correct and incorrect answers given to meaningful and meaningless words.

Meaningful/meaningless (C)	Total number of correct answers	Total number of incorrect answers
Bipolar group	8910	90
Control group	8990	10

Table 7. Average values for the t_e/t_f for each tense suffix.

Suffix ratio (t_e/t_f)	Past tense Suffix ("-di")	Continuous tense suffix ("-yor")	Future tense Suffix ("-cek")
Bipolar group	0.529941	0.658065	0.638980
Control group	0.575861	0.641582	0.684621

It is observed that the average value for t_e/t_f for the past tense suffix is smaller than the t_e/t_f values for continuous and future tense suffixes. In addition, the average value for the t_k/t_f is calculated for each tense suffix in order to observe the effect of the tense suffix on calling the word from the mental lexicon. The average values for t_k/t_f are given in Table 8.

Table 8. Average values for the t_k/t_f for each tense suffix.

Last location ratio (t_k/t_f)	Past tense suffix ("-di")	Continuous tense suffix ("-yor")	Future tense suffix ("-cek")
Bipolar group	0.185501	0.161086	0.157723
Control group	0.192377	0.169792	0.188087

The average values for number of times strolling over the word before deciding (T) are calculated for each tense suffix. The average values are given in Table 9.

Table 9. Average values for number of times strolling over the word for each tense suffix.

Number of strollings of the word (T)	Past tense suffix ("-di")	Continuous tense suffix ("-yor")	Future tense suffix ("-cek")
Bipolar group	2.493333	2.326667	2.506667
Control group	1.953333	2.046667	1.886667

It is also observed that the total reading time for the bipolar disorder group is longer than that for the control (normal) group. Moreover, the bipolar people generally look at the roots and suffixes of words more than the control group people do. Average values for reading time duration of the test word (t_f) are calculated for each tense suffix. The average values are given in Table 10.

Table 10. Average values for reading time duration of the test word (t_f) for each tense suffix.

Reading time of test word (t_f) (ms)	Past tense suffix ("-di")	Continuous tense suffix ("-yor")	Future tense suffix ("-cek")
Bipolar Group	403.04	418.88	454.52
Control Group	295.46	315.26	335.5

## 2.4. SVM for modeling and classification

The SVM is a supervised learning method that analyzes data and recognizes patterns for classification. The SVM takes the feature vector determined from the eye pupil detection system and predicts membership in 1 of 2 classes: bipolar or nonbipolar. An SVM training algorithm builds a mathematical model from a set of training examples (the data collected from people with bipolar disorder using the eye pupil detection system), each marked as belonging to 1 of 2 categories. Given (x, y)-labeled data, the purpose of the SVM algorithm is to learn a classifier function that maps the data into 2 classes. The output of the classifier depends on the sign of a linear function, as shown in Eq. (6):

$$f(x) = sign\left(\sum_{i=1}^{N} a_i y_i k(x_i, x) + b\right), \tag{6}$$

where x is the input feature vector, b is the bias, and a and y are the weights of the support vectors. A hyperplane is defined such that it divides the multidimensional feature space into 2 half-spaces representing 2 classes. If the training set is linearly separable (the bipolar and nonbipolar data are each on their own side of the hyperplane), the SVM algorithm finds the hyperplane with maximum margin between classes. An overview of the classification system is given in Figure 5.

$$F = \left[\left(\frac{t_e}{t_f}\right), (L), \left(\frac{t_k}{t_f}\right), (C), (T), (t_f)\right]$$

$$F = \left[\left(\frac{t_e}{t_f}\right), (L), \left(\frac{t_k}{t_f}\right), (C), (T), (t_f)\right]$$
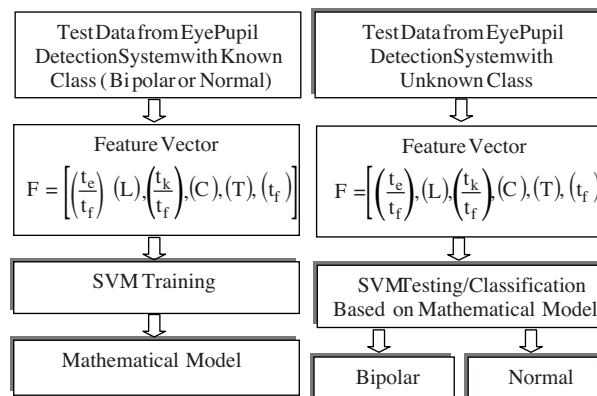
Figure 5. Overview of the classification system.

The kernel used in the implementation of the SVM is the radial basis function (RBF) with penalty parameter for outliers with values $C = 1000$ and kernel parameter $\gamma = 0.1$. The RBF kernel nonlinearly

maps features into a higher dimensional space and can handle the case when the relation between class labels and features is nonlinear. The parameter p (epsilon), which is the maximum allowable distance between feature vectors from the training set and the fitting hyperplane, is chosen as $= 10^{-6}$. In the classification system, a total of 165 feature vectors are used to evaluate the classification performance. These vector data are collected from 35 normal and 20 bipolar people. The system has classified 159 of them correctly and 6 of them incorrectly. The success rate can be given as:

$$\text{Classification success rate} = \frac{\text{Number of correctly classified vectors}}{\text{Total number of vectors}}$$

The resulting success rate of the system is $\frac{159}{165} = 0.9636\,(96.36\%)$.

3. Conclusions

In this study, a video-based eye pupil detection system was developed for diagnosing bipolar disorder. The performances of 3 different eye pupil detection algorithms were compared quantitatively. Localization precision, detection rate, and false alarm rate were all determined and it was found that the active contour snake model is more robust and accurate for the proposed system. In order to detect abnormal behaviors in bipolar people, a number of neuropsychological tests were also designed and implemented to measure attentional and executive abilities. The tests were designed such that the cognitive process and ability-related problems including concentrating, remembering, and making decisions in people with bipolar disorder can be detected using the system. The system is able to learn and mathematically model the patterns of bipolar people using the SVM method. The developed system acquires data from the person being tested and it classifies the person as bipolar or nonbipolar based on a learned mathematical model. A total of 165 data (feature vectors) from 55 people (normal and bipolar) were used to evaluate the classification performance of the system. The system was able to deliver a classification success rate of 96.36%.

References

[1] J.R. Leigh, D.S. Zee, The Neurology of Eye Movements, Oxford, Oxford University Press, 1983.

[2] D.A. Leopold, N.K. Logothetis, "Activity changes in early visual cortex reflect monkeys' percepts during binocular rivalry", Nature, Vol. 379, pp. 549–553, 1996.

[3] I. Kagan, M. Gur, D. M. Snodderly, "Spatial organization of receptive fields of V1 neurons of alert monkeys: comparison with responses to gratings", Journal of Neurophysiology, Vol. 88, pp. 2557–2574, 2002.

[4] K. Rayner, "Eye movements in reading and information processing: 20 years of research", Psychological Bulletin, Vol. 124, pp. 372–422, 1998.

[5] K. Rayner, T.J. Smith, G.L Malcolm, J.M. Henderson, "Eye movements and visual encoding during scene perception", Psychological Science, Vol. 20, pp. 6–10, 2008.

[6] J.M. Henderson, A. Hollingworth, "Eye movements, visual memory, and scene representation", in M.A. Peterson and G. Rhodes (Eds.), Analytic and Holistic Processes in the Perception of Faces, Objects, and Scenes", New York, Oxford University Press, pp. 356–383, 2003.

[7] J. Treuting, "Eye tracking and the cinema: a study of film theory and visual perception", SMPTE Motion Imaging Journal, Vol. 115, pp. 31–40, 2006.

[8] A.T. Duchowski, Eye Tracking Methodology: Theory and Practice, Berlin, Springer, 2007.

[9] R. Bates, H. Istance, L. Oosthuizen, P. Majaranta, "Survey of de-facto standards in eye tracking", COGAIN Conference on Communication by Gaze Interaction, 2005.

[10] J.G. Daugman, "How iris recognition works", IEEE Transactions on Circuits and Systems for Video Technology, Vol. 14, pp. 21–30, 2004.

[11] R Wildes, "Iris recognition: an emerging biometric technology", Proceedings of the IEEE, Vol. 85, pp. 1348–1363, 1997.

[12] M. Kass, A. Witkin, D. Terzopoulos, "Snakes: active contours models", International Journal of Computer Vision, Vol 1988, pp. 321–331, 1987.

[13] V.N. Vapnik, The Nature of Statistical Learning Theory, Berlin, Springer 1995.

[14] G. Akıncı, E. Polat, O.M. Koçak, "Bipolar bozukluk teşhisi için video tabanlı göz izleme sistemi", 20. IEEE Sinyal İşleme ve İletişim Uygulamaları Kurultayı, 2012 (in Turkish).

[15] J. Daugman, "New methods in iris recognition", IEEE Transactions on Systems, Man, and Cybernetics Part B: Cybernetics, Vol. 37, pp. 1167–1175, 2007.

[16] G. Bradski, "The OpenCV Library", Dr. Dobb's Journal of Software Tools, Vol. 25, pp. 122–125, 2000.

[17] G. Akıncı, E. Polat, O.M. Koçak, "Neuropsychiatric disorders classification using a video based pupil detection system", International Symposium on Innovations in Intelligent Systems and Applications, 2012.