A Comprehensive Method for Assessing the Blepharospasm Cases Severity

Vitoantonio Bevilacqua[1]([envelope]), Antonio Emmanuele Uva[2], Michele Fiorentino[2],
Gianpaolo Francesco Trotta[2], Maurizio Dimatteo[1], Enrico Nasca[1],
Attilio Nicola Nocera[1], Giacomo Donato Cascarano[1], Antonio Brunetti[1],
Nicholas Caporusso[1], Roberta Pellicciari[3,4], and Giovanni Defazio[3]

[1] Department of Electrical and Information Engineering, Polytechnic of Bari,
Via Orabona 4, 70125 Bari, Italy
`vitoantonio.bevilacqua@poliba.it`
[2] Department of Mechanical and Management Engineering, Polytechnic of Bari,
Via Orabona 4, 70125 Bari, Italy
`antonio.uva@poliba.it`
[3] Department of Basic Medical Sciences, Neuroscience and Sensory Organs,
University of Bari Aldo Moro, Piazza Giulio Cesare, 11, 70120 Bari, Italy
`giovanni.defazio@uniba.it`
[4] Department of Neurology and Psychiatry, Sapienza University of Rome,
Rome, Italy
`ro.pellicciari@gmail.com`

Abstract. Blepharospasm is characterized by bilateral, synchronous, and symmetric involuntary orbicularis oculi muscle spasms leading to partial/total eyelid closure. We proposed a comprehensive method for assessing the severity of blepharospasm cases, based on expert observation of fixed-length video recordings and natural feature detection algorithms. The developed software detects involuntary spasms and blinks in order to evaluate BSP severity. In this work we have considered a new BSP severity scale to realize an objective evaluation of BSP severity.

1 Introduction

Blepharospasm (BSP) is characterized by bilateral, synchronous, and symmetric involuntary orbicularis oculi (OO) muscle spasms leading to partial/total eyelid closure [1–3]. Spasms may be brief or prolonged. Usually, they are aggravated by bright light, stress, or voluntary muscle contractions; conversely, they are reduced by attention-demanding tasks, such as writing. Additional symptoms include: increased blinking rate, sensory tricks, apraxia of eyelid opening (AEO) and dystonia in other body parts [4,5]. As clinical management is symptomatic and often incomplete, research focuses on new treatment strategies. Periodic botulinum neurotoxin (BoNT) injections, into the affected muscles, provides symptomatic treatment with reasonable efficacy and safety. Yet, the development of new more effective therapies [6] requires sensitive and objective methods to rate symptoms severity [7].

Determining the efficacy of treatments requires a system for measuring the severity of symptoms. Existing clinical rating scales are inherently subjective. Objective methods, such as kinematic measures, are expensive and not easily adoptable for widespread clinical use. This critical need for development of an objective, sensitive BSP rating scale has not moved forward. Current rating scales, such as the Burke-Fahn-Marsden Dystonia Rating Scale (BFM) [8], Global Dystonia Rating Scale (GDRS) [9], Jankovic Rating Scale (JRS) [10] and the recently developed rating scale for BSP remain based on the subjective evaluation of a clinician. To this end, video analysis might offer a better alternative. Videos, recorded during clinical examinations, enable more authorized expert to review a consistent set of data, without requiring them to attend recordings or examinations. Moreover, video recordings do not require additional resources and patient set-up time, as would be the case with equipment for kinematic measurements, so that the clinical setting time is not affected. The wide availability of inexpensive digital video cameras also enables video analysis to be performed with a rapidly growing suite of more recent and advanced artificial intelligence algorithms.

2   Related Works

Potential solutions include the Computer Expression Recognition Toolbox (CERT), previously available from the University of California, San Diego (UCSD) for academic use, now available as the Facet commercial software from iMotions.com. CERT combines artificial intelligence algorithms taken from the domains of computer vision, pattern recognition and machine learning. CERT is able to evaluate blepharospasm severity according to the following scales:

- Burke-Fahn-Marsden (BFM):
  - no dystonia present
  - slight: occasional blinking
  - mild: frequent blinking without prolonged spasms of eye closure
  - moderate: prolonged spasms of eye closure, but eyes open most of the time
  - severe: prolonged spasms of eye closure, with eyes closed at least 30% of the time
- Global Dystonia Rating Scale (GDS):
  - no dystonia present
  - minimal dystonia
  - moderate dystonia
  - severe dystonia
- Jankovic Rating Scale (JRS):
  - none
  - minimal, increased blinking present only with external stimuli
  - mild, but spontaneous eyelid fluttering (without actual spasm), definitely noticeable, possibly embarrassing, but not functionality disabling
  - moderate, very noticeable spasm of eyelids only, mildly incapacitating
  - severe, incapacitating spasm of eyelids and possibly other facial muscles

Previous BSP severity scales do not include factors, such as duration, because of the relatively short observation periods. Indeed video-recordings have been performed according to the following standardized video protocol:

– at rest, eyes open for 10 s
– at rest, eyes gently close for 10 s
– at rest, after opening eyes, for an additional 10 s
– forced eyelid closure for 3 times, then observation effect for 5 s after each closure

However, with the new BSP severity scale, video-recordings were performed according to a standardized video protocol lasting approximately 5 min:

– Patient at rest, eyes open (10 s)
– Patient voluntarily performs a forceful eye closure followed by eye re-opening (repeated 5 times, one cycle per second)
– Patient at rest, eyes open (10 s)
– Voluntary gentle eye closure followed by eye reopening (repeated 5 times, one cycle per second)
– Patient at rest, eyes open (10 s)
– The doctor asks the patient the following questions: are you able to suppress eye closure? How? Are you able to voluntary suppress it? Or do you need to touch your eyes, face or neck?
– The patient answers the questions (at least 50 s)
– The doctor asks the patient to write a standard sentence 3 times (today it is a nice sunny day)
– Patient at rest, eyes open (at least 150 s). In the last 120 s, we counted the number of tonic OO spasms (and measured their duration), the number of blinks and clonic OO spasms. Patients were instructed to avoid antagonistic gestures

In this work we used the aforementioned video-recordings standard protocol to analyse the last 120 s of video and to evaluate BSP severity, according to the BSP severity scale described in Fig. 1 [11].

3   Methods

The process of analysing BSP symptoms consists in two steps:

– Signal extraction: frame-by-frame video analysis to measure the geometry of specific facial features [12]. Data are collected as digital signals which will exhibit local minima at frames where phenomena occur
– Signal processing: signal trend analysis and threshold evaluation to distinguish BSP-related phenomena from natural movements of the face [13]

3.1   Signal Extraction

In the signal extraction step, the dlib frontal face detector outputs a region of interest, which is scanned by the dlib shape predictor [14] to estimate the

Intensity rating

ITEM A1) Type of eyelid spasm
▯ Brief ($<$ 3 sec duration) eyelid spasms with complete rim closure = score 1
▯ Prolonged ($\geq$ 3 sec duration) eyelid spasms with partial rim closure = score 2
▯ Prolonged ($\geq$3 sec duration) eyelid spasms with complete rim closure = score 3

ITEM A2) Apraxia of eyelid opening
▯ Yes = score 2
▯ No = score 0

ITEM A3) OO spasms occur during writing:
▯ Yes = score 1
▯ No = score 0

ITEM A4) Average duration of prolonged eyelid spasm with complete rim closure recorded while patient at rest, eyes open, for two minutes. Calculate the correspondent tertile as follows:
▯ I tertile = 3 to 4 sec = score 1
▯ II tertile = 4.1 to 5 sec = score 2
▯ III tertile = $>$ 5 sec = score 3

Frequency rating

ITEM B1) Count "number of blinks + brief eyelid spasm / min" (patient at rest, eyes open, for two minutes) and calculate the corresponding tertile as follows:
▯ I tertile = 1 – 18 blinks + brief spasm / min = score 1
▯ II tertile = 19 – 32 blinks + brief spasm /min = score 2
▯ III tertile = $>$ 32 blinks + brief spasm / min = score 3
ITEM B2) Count number of prolonged eyelid spasm with complete rim closure / min" (patient at rest, eyes open, for two minutes) and calculate the corresponding quartile as follows:
▯ I tertile = 1 – 3/ min = score 1
▯ II tertile = 3.1 – 7/min = score 2
▯ III tertile = $>$ 7/ min = score 3
Total score = Intensity + Frequency = (A1 + A2 + A3 + A4) + (B1 + B2)

Fig. 1. BSP severity scale

coordinates of 68 facial landmark points. A subset of such points is then used to perform simple geometric calculations.

For each frame i, the function $Y(i)$ is computed as shown in Eq. 1.

$$Y(i) = \frac{T(i)}{L(i)} \tag{1}$$

where:

– T(i): height of the triangle shown in Fig. 2(b)
– L(i): distance between the nose tip and the bridge as shown in Fig. 2(c)

The tip-bridge distance is a normalising divisor that prevents sudden signal variation due to head rotation. The output of this step is a signal that will exhibit local minima in correspondence of spasm events.
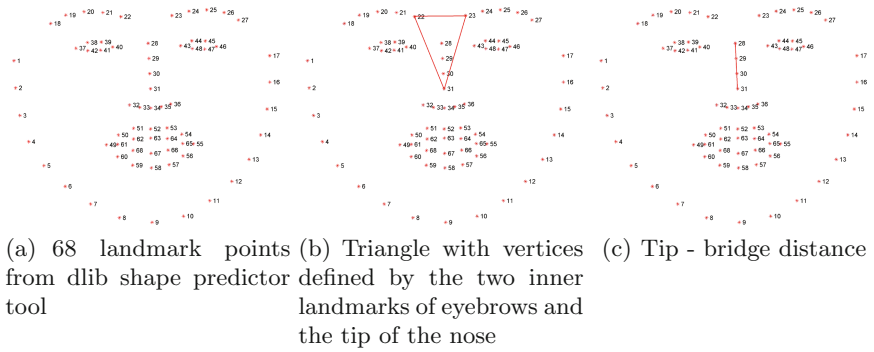
(a) 68 landmark points from dlib shape predictor tool

(b) Triangle with vertices defined by the two inner landmarks of eyebrows and the tip of the nose

(c) Tip - bridge distance

Fig. 2. Using of dlib shape predictor tool

Figure 3 shows a simulation of spasm event.

(a) Triangle in case of opened eyes

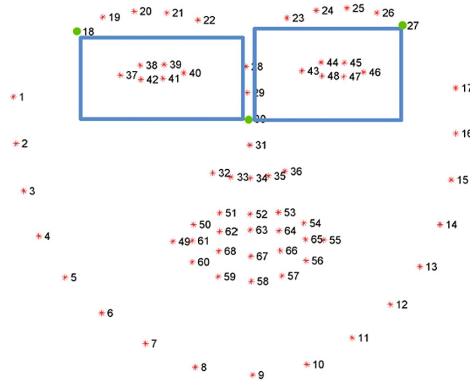(b) Triangle during spasm event

Fig. 3. Spasm simulation

Fig. 4. Estimated regions of interest.

An involuntary OO muscle spasm occurs when both eyes are closed. Additionally, we developed a blink detector to evaluate the severity of the symptoms. We employ a template matching technique using SIFT descriptors to mark closed eyes. First, a domain expert selects one or many sample frames (templates) in which the patient has his eyes open. For each template, eye regions are estimated using selected landmarks, as shown in Fig. 4; then, key points are extracted from these regions using the FAST algorithm [15]. Every key point is uniquely described by the SIFT algorithm. Each descriptor is obtained considering a region of 16×16 pixels around a key point. Then, the region is divided in 16 subregions of 4×4 pixels and an 8-bin orientation histogram is computed for each subregion. Each 8 bin sequence forms the key point descriptor (128 values). Key points and their relative descriptors are extracted in each frame during video processing. Using a K-nearest neighbour algorithm for each key point descriptor of the i_{th} frame, we use a classification technique to find similar key points, based on Euclidean distance, between template and the i_{th} frame. In order to remove outliers, we consider the distance between the i_{th} and the closest-to-template key point descriptor and the distance between the i_{th} and the second closest-to-template key point descriptor [16]. If the ratio of the distances exceeds a 0.7 threshold the match shall not be valid. As a result, the output of the algorithm is the product between similar key points in right and left eye.

Figure 5 shows a simulation of blink event and regions from which extract key points.

The face region, which the dlib shape predictor requires as an input parameter, is computed by the dlib frontal face detector, which is based on the Histogram of Oriented Gradients (HOG). For the purposes of this work, the dlib face detector provides an advantage in terms of accuracy and speed over the OpenCV face detector, which is based on Haar Cascade [17]. Still, video processing is affected by some delay caused by the face detector, which is executed on every frame. However, the video recording protocol adopted in our set-up requires the patient to sit in front of a fixed camera: hence we can assume that the position of the face does not
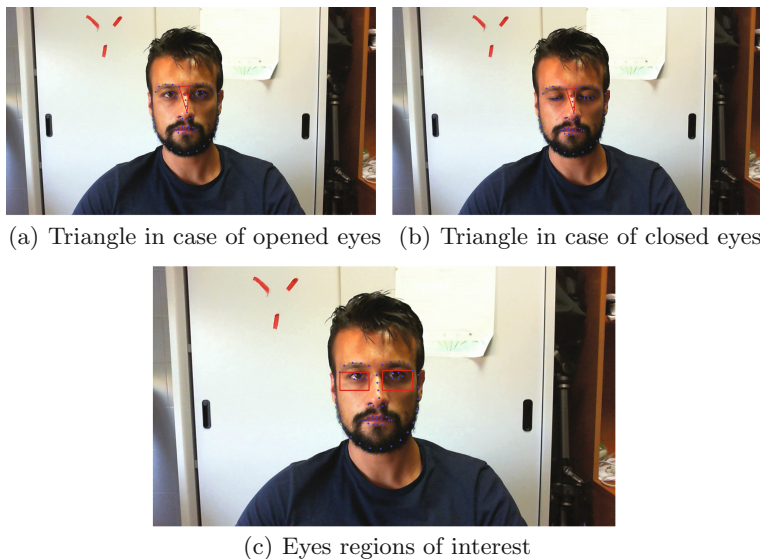
(a) Triangle in case of opened eyes   (b) Triangle in case of closed eyes

(c) Eyes regions of interest

Fig. 5. Blink simulation and eyes regions of interest

significantly change between consecutive frames. A naive solution to decrease processing time consists in running the face detector every k frames. A more elegant approach for detecting changes in the shape of the face is based on the Frobenius distance, shown in Eq. 2.

$$F = \sqrt{\sum_{i=0}^{67}((x_i - \bar{x})^2 + (y_i - \bar{y})^2)} \qquad (2)$$

where:

- (x_i, y_i): the i_{th} point coordinates
- $(\bar{x}, \bar{y})$: the centroid coordinates

If the dlib shape predictor tool is executed on a region that does not contain a full face, the Frobenius distance higher. If it exceeds a certain threshold, it is necessary to execute the face detector. In this work, we used the following algorithm to reduce processing time:

- read the i_{th} frame and run both the face detector and the dlib shape predictor tool on it
- read the $(i_{th} + 1)$ frame and use the previously detected landmark points to evaluate face region position on the frame
- run dlib shape predictor tool on the $(i_{th} + 1)$ frame
- evaluate Frobenius distance for the $(i_{th} + 1)$ and i_{th} frame: if the difference between the $(i_{th} + 1)$ and i_{th} exceeds a certain threshold read the $(i_{th} + 2)$ frame and run face detector alternatively use the previously detected landmark points to evaluate face region position

Table 1. Difference between detector runs number using Frobenius distance or sub samples on 4 videos

| Total number of frame	Detector runs (with Frobenius distance)	Detector runs (every 3 frames)
3881	56	1293
3861	77	1287
3915	266	1305
3829	26	1276

Table 1 compares the number of detector runs using Frobenius distance and a constant sub samples (every 3 frames) on 4 different videos. The improvement is considerable and the use of Frobenius distance does not result in any deterioration in the quality of the extracted signals.

3.2   Signal Processing

The beginning of a spasm is detected by signal analysis on $Y(i)$. The main goal is to find $Y(i)$'s local minima and to determine a threshold value to distinguish BSP phenomena from natural movements. During a spasm the patient's eyes are closed. The ending of a spasm is detected by the analysis of the signal generated by the number of similar key points found during video processing. The main goal is to find local minima in which the frame is characterized by open eyes. This information, in turn, enables to evaluate the BSP severity using the BSP severity scale in Fig. 1.

The $Y(i)$ signal, an example is shown in Fig. 6, shows a noise component due to two main factors:

– Head movements causing shifts in the signal. To cancel this noise source, we apply a detrending filter to the signal.
– Small variations of landmark points position happen also when the patient is steady. To delete this noise source, we apply a smoothing filter to the signal.
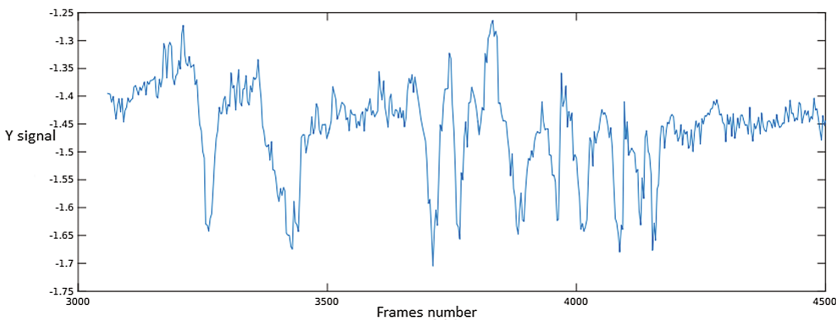
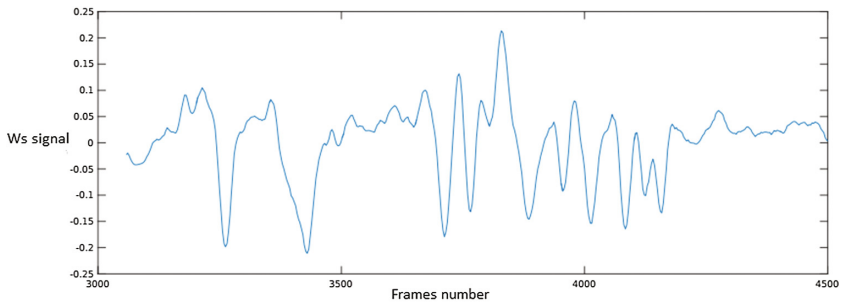**Fig. 6.** The raw signal example relative to spasms detection

Fig. 7. The output signal example relative to spasms detection after pre-processing operations

To mitigate the noise component caused by head movements, the $Y(i)$ signal is estimated by a low order polynomial function $p(i)$. The output of this process is a new signal $W(i) = Y(i) - p(i)$. To mitigate the noise component caused by small variations of landmark points position, we apply a Savitzky-Golay filter to $W(i)$ signal. An example of $W_s(i)$ signal is shown in Fig. 7.

In order to determine a threshold value to distinguish BSP phenomena from natural face movements, we apply the find-peaks MATLAB function to find all $-W_s(i)$'s peaks that are $W_s(i)$'s local minima. Using the find-peaks MATLAB function, we obtain an array of peak prominences, which is then used to create an histogram where the binning is realized to set the ranges of values in different bins to an equal amplitude. By doing this, less prominent peaks due to signal noise are isolated in the bottom half of the histogram. To determine a threshold value we consider P_k for $k = 1, 2, ..., n$ the values set of most important prominences and B_k for $k = 1, 2, ..., n$ the number of elements in bin k. The threshold value T, shown in Eq. 3, is a halved weighted average where $B_1, B_2, ..., B_n$ are the weights.

$$T = \frac{\sum_k P_k B_k}{2 * \sum_k B_k} \tag{3}$$

This approach is useful to isolate prominences that are repeated frequently in the signal, indicating possible spasms. The threshold value evaluation to distinguish BSP phenomena from natural face movements is done using the first part of the video in which the patient performs voluntary spasms and blinks. The same pre-processing tasks are applied to the signal extracted from open eye detection. In this case, finding a constant threshold value is useful to detect when eyes are open or closed. Figure 8 shows an output signal related to the detection of eyes open.

This signal does not enable to choose a constant threshold value. As a result, we manually solved the issue. During video processing the application runs a software frame grabber that enables the user to choose an indefinite number of frames with eyes open. The higher the number of selected templates the better the performance.
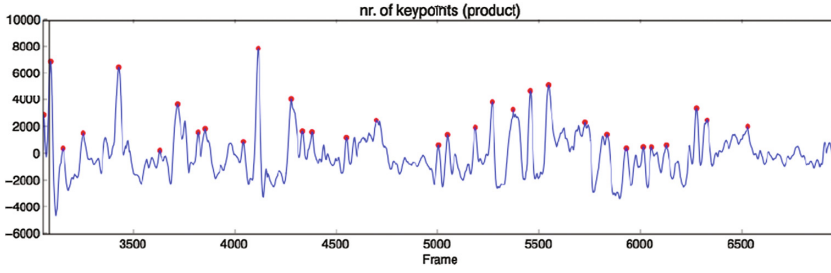
Fig. 8. The output signal example relative to eyes open detection step after pre-processing operations

4   Results

Two videos were considered. The first one shows a subject exhibiting both short and long spasms along with apraxia. Conversely the patient in the second video shows both enduring and multiple very short spasms. Figure 9 shows the results obtained with data from the first patient.

The BSP phenomena are detected accurately. However to evaluate intensity and frequency rating according to the BSP severity scale in Fig. 1 it is useful to detect one additional phenomenon, in order. Apraxia of lid opening (ALO) is a nonparalytic motor abnormality characterized by difficulty initiating the act of lid elevation after lid closure [18]. The algorithm that detects apraxia consists of the following steps:

– Run the algorithm to detect spasm
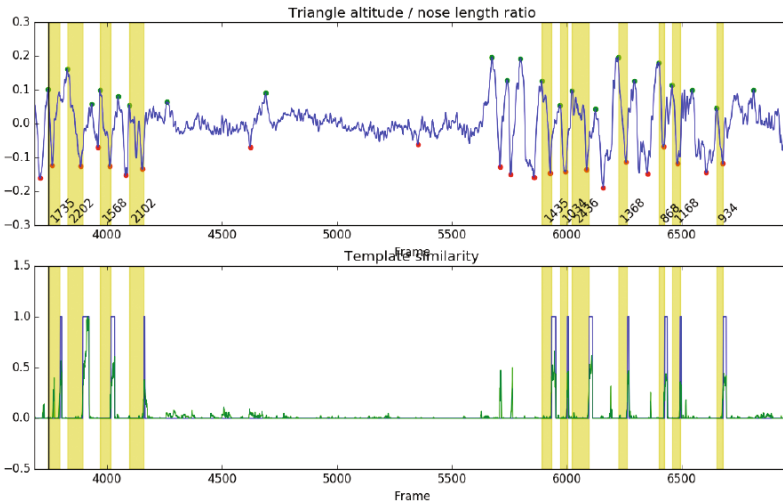– Run the algorithm to detect opened eyes

Fig. 9. The results relative to the first patient

- Consider the highest peaks of signal related to spasm detection after pre-processing that are not in any spasm time interval
- For each highest peak, find a maximum in the open eye detection signal after pre-processing in the next 2 s
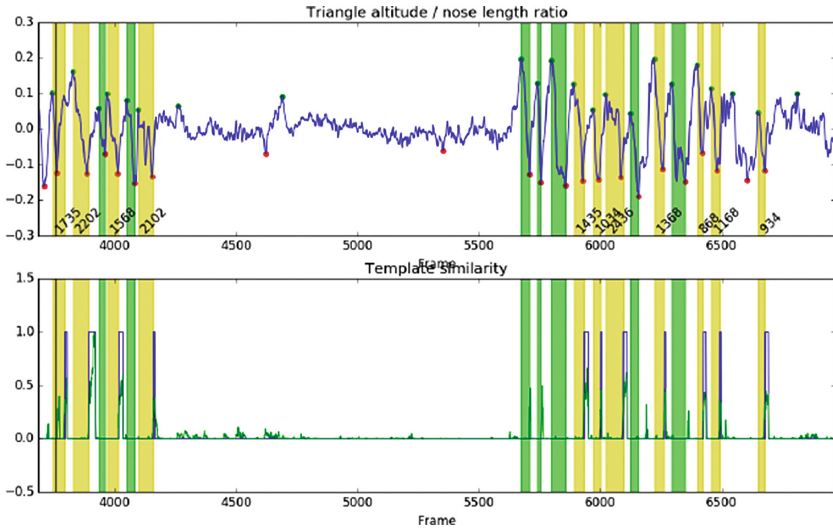- If there is a maximum then apraxia is detected.

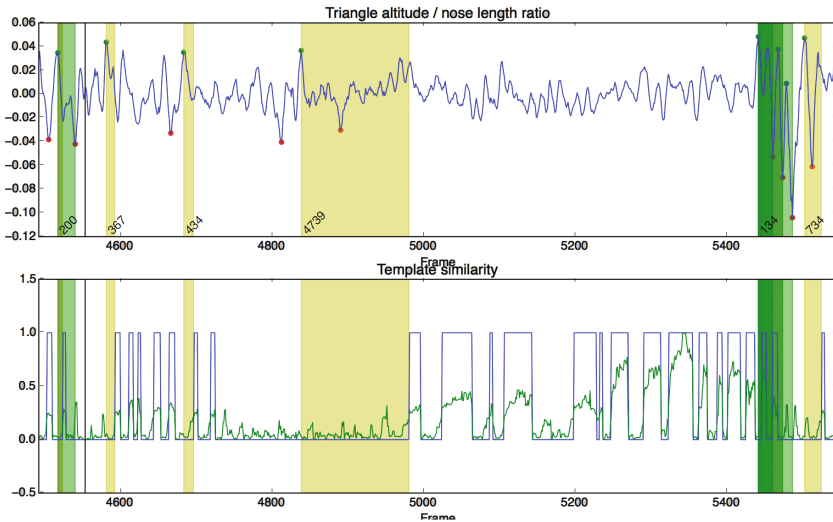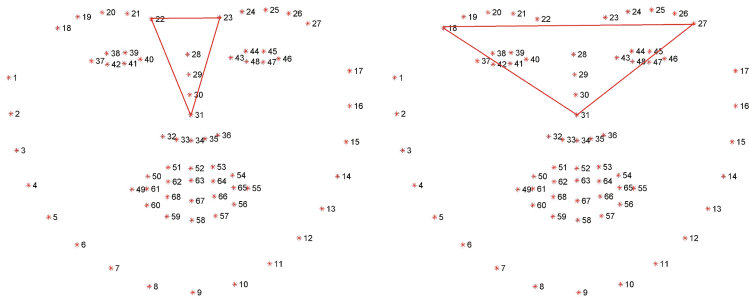**Fig. 10.** The results relative to the first patient with apraxia detection (Color figure online)

Fig. 11. The results relative to the second patient

(a) Triangle with vertices defined by the two outer landmarks of eyebrows and the tip of the nose.

(b) Triangle where the endpoints defined by the left and right outer eyebrows form the base, whereas the tip of the nose is the remaining vertex

Fig. 12. Two different triangles to resolve the issue of persistent spasm

Data resulting from the apraxia detection algorithm are shown in Fig. 10. Green bars are relative to apraxia phenomena.

Figure 11 shows the results from the second patient. Similarly, all phenomena are detected accurately. However, data show a problem caused by a phenomenon knowns as *persistent spasm*. Indeed a spasm occurs when the triangle in Fig. 12(b) is subject to height reduction. A persistent spasm is characterized by a fixed height of the triangle in Fig. 12(a). So, in this case, all eyes open detected by software should be brief spasms.

5 Conclusion

In this work we conducted a qualitative analysis to evaluate blepharospasm (BSP) severity. Based on both video observation and comparison of results from algorithms, we were able to detect involuntary spasms and blinks. The software developed has been used to extract data that will be used to implement an algorithm that evaluates BSP intensity and frequency rating using a new BSP severity scale. In the future, our results will help define a gold standard on a relevant number of patients, which, in turn, will enable to conduct a quantitative analysis and to improve algorithms in order to resolve issues such as, persistent spasm.

References

1. Grandas, F., Elston, J., Quinn, N., Marsden, C.D.: Blepharospasm: a review of 264 patients. J. Neurol. Neurosurg. Psychiatry **51**(6), 767–772 (1988)
2. Hallett, M., Evinger, C., Jankovic, J., Stacy, M.: Update on blepharospasm: report from the BEBRF international workshop. Neurology **71**(16), 1275–1282 (2008)

3. Jinnah, H.A., Berardelli, A., Comella, C., De Fazio, G., DeLong, M.R., Factor, S., Galpern, W.R., Hallett, M., Ludlow, C.L., Perlmutter, J.S.: The focal dystonias: current views and challenges for future research. Mov. Disord. **28**(7), 926–943 (2013)
4. Defazio, G., Berardelli, A., Hallett, M.: Do primary adult-onset focal dystonias share aetiological factors? Brain **130**(5), 1183–1193 (2007)
5. Abbruzzese, G., Berardelli, A., Girlanda, P., Marchese, R., Martino, D., Morgante, F., Avanzino, L., Colosimo, C., Defazio, G.: Long-term assessment of the risk of spread in primary late-onset focal dystonia. J. Neurol. Neurosurg. Psychiatry **79**(4), 392–396 (2008)
6. Price, K.M., Ramey, N.A., Richard, M.J., Woodward, D.J., Woodward, J.A.: Can methylphenidate objectively provide relief in patients with uncontrolled blepharospasm? A pilot study using surface electromyography. Ophthalmic Plast. Reconstr. Surg. **26**(5), 353–356 (2010)
7. Wabbels, B., Jost, W.H., Roggenkämper, P.: Difficulties with differentiating botulinum toxin treatment effects in essential blepharospasm. J. Neural Transm. **118**(6), 925–943 (2011)
8. Burke, R.E., Fahn, S., Marsden, C.D., Bressman, S.B., Moskowitz, C., Friedman, J.: Validity and reliability of a rating scale for the primary torsion dystonias. Neurology **35**(1), 73–73 (1985)
9. Comella, C.L., Leurgans, S., Wuu, J., Stebbins, G.T., Chmura, T.: Rating scales for dystonia: a multicenter assessment. Mov. Disord. **18**(3), 303–312 (2003)
10. Jankovic, J., Orman, J.: Botulinum a toxin for cranial-cervical dystonia a double-blind, placebo-controlled study. Neurology **37**(4), 616–616 (1987)
11. Defazio, G., Hallett, M., Jinnah, H.A., Stebbins, G.T., Gigante, A.F., Ferrazzano, G., Conte, A., Fabbrini, G., Berardelli, A.: Development and validation of a clinical scale for rating the severity of blepharospasm. Mov. Disord. **30**(4), 525–530 (2015)
12. Loconsole, C., Chiaradia, D., Bevilacqua, V., Frisoli, A.: Real-time emotion recognition: an improved hybrid approach for classification performance. In: Huang, D.-S., Bevilacqua, V., Premaratne, P. (eds.) ICIC 2014. LNCS, vol. 8588, pp. 320–331. Springer, Cham (2014). doi:10.1007/978-3-319-09333-8_35
13. Bevilacqua, V., D'Ambruoso, D., Mandolino, G., Suma, M.: A new tool to support diagnosis of neurological disorders by means of facial expressions. In: 2011 IEEE International Workshop on Medical Measurements and Applications Proceedings (MeMeA), pp. 544–549, May 2011
14. Kazemi, V., Sullivan, J.: One millisecond face alignment with an ensemble of regression trees. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1867–1874 (2014)
15. Rosten, E., Drummond, T.: Machine learning for high-speed corner detection. In: Leonardis, A., Bischof, H., Pinz, A. (eds.) ECCV 2006. LNCS, vol. 3951, pp. 430–443. Springer, Heidelberg (2006). doi:10.1007/11744023_34
16. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. Int. J. Comput. Vis. **60**(2), 91–110 (2004)
17. Viola, P., Jones, M.: Rapid object detection using a boosted cascade of simple features. In: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, vol. 1, p. I-511. IEEE (2001)
18. Defazio, G., Livrea, P., Lamberti, P., De Salvia, R., Laddomada, G., Giorelli, M., Ferrari, E.: Isolated so-called apraxia of eyelid opening: report of 10 cases and a review of the literature. Eur. Neurol. **39**(4), 204–210 (1998)