34th Annual International Conference of the IEEE EMBS
San Diego, California USA, 28 August - 1 September, 2012

Using Spatio-Temporal Interest Points (STIP) for myoclonic jerk detection in nocturnal video

Kris Cuppens[1], Chih-Wei Chen[2], Kevin Bing-Yung Wong[2], Anouk Van de Vel[3], Lieven Lagae[4], Berten Ceulemans[3], Tinne Tuytelaars[5], Sabine Van Huffel[6], Bart Vanrumste[1] and Hamid Aghajan[7]

Abstract— **In this study we introduce a method for detecting myoclonic jerks during the night with video. Using video instead of the traditional method of using EEG-electrodes, permits patients to sleep without any attached sensors. This improves the comfort during sleep and it makes long term home monitoring possible. The algorithm for the detection of the seizures is based on spatio-temporal interest points (STIPs), proposed by Ivan Laptev, which is the state-of-the-art in action recognition [8]. We applied this algorithm on a group of patients suffering from myoclonic jerks. With an optimal parameter setting this resulted in a sensitivity of over 75% and a PPV of over 85%, on the patients' combined data.**

I. INTRODUCTION

About 25% of the patients suffering of epileptic seizures, which is almost 1% of the world's population, cannot be controlled by either medication or surgery. The gold standard in epilepsy monitoring uses EEG-electrodes attached to the scalp. However, these electrodes are difficult to attach, hamper the patient's sleep during the night, and therefore are prohibiting long term home monitoring.

During the last decade researchers started to investigate ways of detecting seizures with a motor component in a less intrusive way by means of accelerometers or video. Last year, the first detectors that are built in in a wristwatch were presented by BioLert (the EpiLert watch) and Smart Monitor Company (the SmartWatch). Both systems have recently been validated in clinical studies [11] [7], mainly on generalized tonic-clonic patients. Jallon et al. [5], Cuppens et al. [3] and Nijsen et al. [12] used multiple accelerometers attached to the extremities for seizure detection. A general conclusion from these papers is that large and intense seizures (tonic-clonic, hypermotor) can be detected with a high sensitivity and a low number of false positives, and that smaller seizures (myoclonic jerks) can be detected but with a higher number of false positives. The multi-modal approach used by Conradsen et al. [2] gives better results on patients with smaller seizures. Video based detection mainly focuses on the usage of markers [10] [1] or other ways to track limbs, like using colored pyjamas [4]. Karayiannis et al. [6] didn't use any markers, but the moving limbs of the patients were clearly visible as they were monitored in the Neonatal Intensive Care Unit (NICU) of the hospital. The best obtained result had a sensitivity above 90% and a specificity above 85%, in patients with myoclonic and focal seizures.

Vision based human action recognition has been widely studied, and various methods have been proposed [13]. Based on representations, these approaches can be broadly categorized into global representations and local representations. In the former, the entire body or its articulated poses are encoded in the model. In this work, because the body is usually occluded and not fully observed, local representations are used, where the observation is described as a collection of local descriptors. In particular, we use space-time interest point detectors and descriptors proposed by Ivan Laptev in [9], which achieves state-of-the-art performance in action recognition in video on real life actions. In that study, different realistic actions from movies (such as kissing, answering the phone and getting out of a car) are learned. This method outperforms the other algorithms on the KTH actions dataset [14] and reaches an accuracy of 91.8%.

The purpose of our study is the following, we want to keep track of the number of seizures during the night in order to have an objective measure for the neurologist. We want to obtain this without interfering too much in the patient's environment, so without removing the blanket or attaching markers.

II. METHOD

The general approach for classifying the nocturnal movements in our database is based on the state-of-the-art method proposed in [9]. In a first step, interest points in the video are found. Afterwards, spatio-temporal features from these interest points are extracted. Using a bag-of-features approach, features are fed into a support vector machine (SVM) to generate a classification model. Each part of this procedure is explained in the following sections.

*Research supported by Research Council KUL:GOA-MANET, IWT: TBM070713-Accelero, Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, 'Dynamical systems, control and optimization, 2007-2011); EU: Neuromath (COST-BM0601). Kris Cuppens is supported by a PhD grant of the Agency for Innovation by Science and Technology in Flanders (IWT).

[1]K. Cuppens and B. Vanrumste are with MOBILAB of the K.H.Kempen University College, Geel, Belgium and with KU Leuven, Department of Electrical Engineering (ESAT) SCD-SISTA and IBBT Future Health Department, Leuven, Belgium kris.cuppens at khk.be

[2]K. Wong and C.-W. Chen are with Ambient Intelligence Research (AIR) Lab, Department of Electrical Engineering, Stanford University, Stanford, California USA.

[3]A. Van de Vel and Berten Ceulemans are with the University Hospital of Antwerp, Antwerp, Belgium

[4]L. Lagae is with the University Hospital Leuven, Belgium

[5]T. Tuytelaars is with KU Leuven, ESAT-PSI and IBBT Future Health Department, Leuven, Belgium

[6]S. Van Huffel is with KU Leuven, Department of Electrical Engineering (ESAT) SCD-SISTA and IBBT Future Health Department, Leuven, Belgium

[7]H. Aghajan is with with Ambient Intelligence Research (AIR) Lab, Department of Electrical Engineering, Stanford University, Stanford, California USA and with TELIN/IBBT, Ghent University, Belgium

Fig. 1. STIPs detected within one frame, represented by the circles.

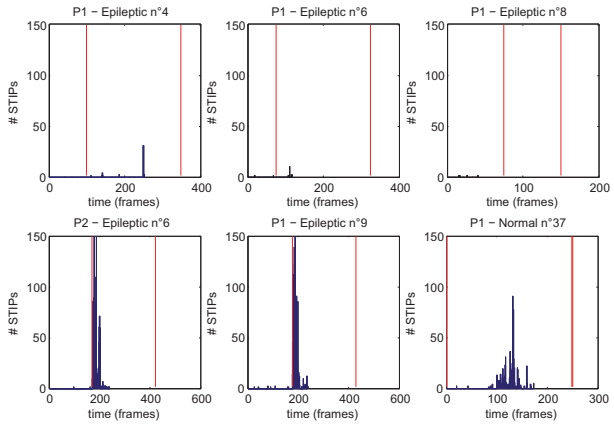

Fig. 2. STIPs of myoclonic and normal movements. The x-axis indicates the frame number of the video sequence. The y-axis represents the number of STIPs found in each frame. The boundaries indicate the beginning and ending of each seizure (or normal movement). The parameter combination generating the most STIPs is used here ($T = 10^{-12}$, $k = 5.10^{-5}$).

A. Description of dataset

The data was recorded at the Pulderbos Rehabilitation Center for Children and Youth in Zandhoven, Belgium. The patients included in the dataset were between the age of 3 and 7 years, suffering from myoclonic jerks. These seizures manifest themselves as short and jerky movement in one or more of the limbs. The videos were recorded during the night with a near infrared camera, with a resolution of 720 by 576 pixels at 25 frames per second. We processed data from 3 patients for a total of 6 nights.

B. Spatio-temporal interest points (STIP) and feature extraction

The Spatio-Temporal Interest Points detection, proposed by Laptev [8] searches for corners in the 3 dimensional video space (two spatial and one temporal dimension) based on the Harris corner detection. We apply this method on our data to find interest points. Figure 1 shows one frame of a video with the detected interest points marked by circles. The size of the circles represents the scale at which they are found.

The interest points are found using multiple spatio-temporal scales. The appearance and the motion around each interest point is then represented by the histograms of oriented gradients (HoG) and the histograms of optical flow (HoF). These are calculated from the neighborhood around the interest points. This is done by dividing this neighborhood in multiple cuboids in the spatial and temporal space. From every cuboid a four bin HoG and five bin HoF is calculated. However, in our study we only use the HoF as these represent the motion. Indeed, the appearance of the body part which is involved in the epileptic motion can differ and it does not contribute to the distinction between normal and epileptic movement. Furthermore, the appearance is also influenced by the blanket, whereas the motion is not as affected. As the neighborhood is divided into 18 cuboids, this results in 90 HoF features.

The motions that occur during the myoclonic seizures are often small. In the algorithm we can change the value of some parameters to increase the number of detected STIPs. A threshold T is used to eliminate interest points with a low probability. When this threshold is lowered, more noisier interest points are detected. A value k is used in the Harris corner detection to specify the maximum sharpness of the detected corners. If this value is lowered, sharper corners (but also more line-shaped rather than corner-shaped points) will be detected. A third value we altered is the window length, as the movement data is segmented in non-overlapping windows. Each window is considered as an example to be used in the training or test phase.

C. Classification

The HoF features derived from each interest point are used to classify the windows into myoclonic or normal movement. To do this, a bag-of-features approach is applied, as in [9], on the features derived from the STIPs.

A subset of the data is used for the creation of the bag-of-features vocabulary. We use 50 clusters in this study; the cluster centers are determined by a k-means clustering on a training set. All the STIPs within one window will be assigned to their according cluster center, i.e. to the cluster center at the smallest distance in the HoF feature space. After this step, a histogram is built with each bin representing one cluster center. This histogram is fed to a support vector machine (SVM) for classification. The use of this approach reduces the n 90-dimensional data points (STIPs) in a window, to one 50-dimensional point for each window, with n the number of STIPs in this window.

In the classification, two third of the data is used as training set and one third as test set. The SVM used for the classification has a radial basis function (RBF) kernel. We optimized the SVM model using the total sum of the misclassifications as cost function. All the tests are performed in a 15 fold randomization, randomizing the samples used in the training and test set.

III. RESULTS

A. Results on global dataset

The first test is done on all the patients combined, and on every patient individually. In this test, we used one night of data for all patients.

When we look at the results, we observe that the patient specific approach outperforms the approach combining the data of all patients. The results over all patients seem to improve with an increasing window length. The mean sensitivity over all settings of T and k increases from 11% to 37% when increasing the window length from 1 to 10 seconds. The best result is obtained using a 10 second window with $T = 10^{-10}$ and $k = 5.10^{-5}$, reaching a sensitivity of 56% and a PPV of 72%. The performance improves when we consider the patients individually. For patient 2 (P2), an increased window length from 1 to 10 seconds raises the average sensitivity from 22% to 72% and the PPV from 45% to 77%. The parameter combination $T = 10^{-12}$ and $k = 5.10^{-4}$ gives the best performance for most window lengths. For the ten second window a sensitivity of 84% and a PPV of 93% is obtained. However, for patients 3 (P3) and 1 (P1) no usable results could be obtained for half of the parameter combinations, as there were not enough seizures containing STIPs to train and test the classification model and obtain the performance results. Even when using the combination with the lowest threshold, some seizures had only a few or even no STIPs, which is also visible in figure 2. For example, seizure 8 of P1 has no STIPs during the seizure. Of course it is hard or even impossible for the algorithm to correctly classify such small motions. Inspecting the video data, we could see that there were several seizures with a low intensity. Also, for some patients, myoclonic seizures manifested themselves during other, non-epileptic movement.

B. Results on subset

To overcome the problems stated in the previous section, we validated the algorithm again on a subset of the data. Indeed, for the approach we use it is necessary that there is enough similarity within the seizure class. Furthermore some normal movement events contain jerk-like movement. To have an idea if this algorithm could work, we selected a subset of epileptic and non-epileptic data according to the following inclusion criteria:

- the jerks should be well visible in the video (thus resulting in a sufficient amount of STIPs);
- the jerks should be isolated in time (no influence of normal movement or other seizures);
- the normal sequences should not contain jerks.

With these criteria we composed a subset of data from 3 patients, containing in total 14 myoclonic jerks and 26 normal movements. The average length of each sequence is 40 frames. For P2, we obtained 7 seizures, and for P1, we obtained 6 seizures. This permits us to do training and testing across both patients, by training the model on one patient and evaluating it on the other, to verify the models genericity. For P3, we obtained one seizure. This one seizure is only used when combining the data from all patients.

In figure 3, the results are shown for the subset of the patients with myoclonic seizures. The results shown are averaged out over the different values for T and k, and show the influence of the window length. For all patients individually and for the patients combined, the ten second

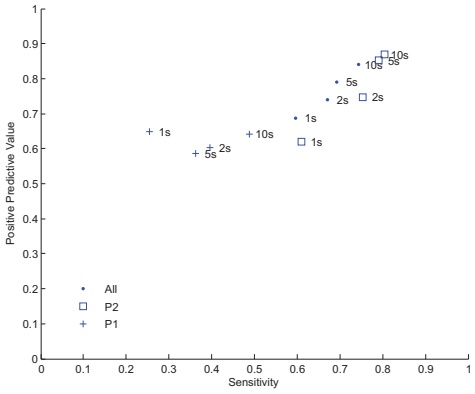

Fig. 3. The performance expressed in sensitivity and Positive Predictive Value on the subset of the data. The given performances are averaged out over the values of T and k, and show the influence of the window length. The window length is shown next to each data point, the marker shape indicates the corresponding dataset.

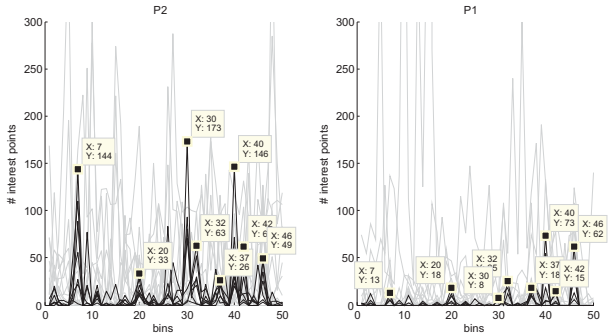

Fig. 4. The histograms for WiL (left) and VHJ (right). The black and the gray curves are the histograms of the epileptic and normal movements, respectively. The bins that have an increased value for both patients are highlighted and labeled. It is also visible that in general the number of STIPs for the movements of VHJ is lower than for WiL.

window gives the best results. For the patients combined, the performance increases from an average sensitivity of 60% to 74% and a PPV of 69% to 84%. The best performance reaches a sensitivity of 77% and a PPV of 87%. For P2 this is an increase from 61% to 80% and from 62% to 87%, respectively. For P1 the sensitivity increases from 25% to 49% but the PPV stays the same (a little decrease) from 65% to 64%. But the results are more spread out with respect to the different threshold combinations. The most optimal combination for P2 is a window length of 5 or 10 seconds with $T = 10^{-12}$ and $k = 5.10^{-5}$, so the combination with the highest number of STIPs. This resulted in a sensitivity of 97% and a PPV of 100%. The best result for P1 is a sensitivity of 70% and a PPV of 75% obtained for a window length of 10 seconds with $T = 10^{-9}$ and $k = 5.10^{-4}$. We see in this patient that the performance increases for this 10 second window with a decreasing value for T.

Figure 4 shows the 50 bins for the movements in patient P2 and P1, as explained in section II-C. The histograms shown in black represent the myoclonic jerks, the ones in gray are normal movement. The histograms here are not normalized, so the y-axis shows the absolute number of STIPs that are

assigned to the different clusters (bins). We can observe that there is a similarity in the bins (e.g. bin 20, 40 and 46) with the highest number of STIPs assigned to, for the epileptic movement in both patients. The histograms shown here are generated using a window length of 10 seconds, a threshold of 10^{-10} and a k-parameter of 5.10^{-4}. Because there is similarity in the histograms, it means that it should be possible to train the classification model on one patient, and test it on the other one. We have tested this for all the different parameter combinations, using all the data from both patients (and thus without a randomization).

When inspecting the performance we notice that there is a large spread in the results regarding the parameter combination. When averaging out the results for each window length, we also can see here that a larger window length increases the performance. When we train the classification model on P2 and test on P1, the average sensitivity increases from 31% to 56% from a 1 second window to a 10 second window, whereas the PPV stays the same at 78%. The other way around we see an increase of the sensitivity and PPV from 50% to 64% and from 58% to 80%, respectively. In the second test (training on P1 and testing on P2) we observe that using a threshold of 10^{-12} and a k-parameter of 5.10^{-4} gives the best performance, and for a 10 second window it even reaches a sensitivity and PPV of 100%.

IV. DISCUSSION

This study shows that we are able to detect the jerky movement of myoclonic seizures. The performance increases with the window length. Also, the patient specific approach gives better results than the group specific approach.

The influence of increasing or decreasing T or k is not always consistent. The results often differ when using a different parameter combination. Adding more STIPs (lowering T or k) can increase the number of characteristic features. But lowering the thresholds too much increases also the noisy features. Depending on how intense the motion is, the ideal value differs. However, we can conclude that the window of 10 seconds gives the best results in any case. And when we apply this approach on a larger scale, we assume that we get more conclusive results for an optimal T and k, which will allow us to choose a fixed value for both.

The results we get from the cross training and testing on two patients are encouraging, although the result depends strongly on which combination of parameters is used.

The calculation of the STIPs and the derived features is computationally intensive. Real-time processing is not feasible for the moment. However, if we could downscale the algorithm and tune it for our specific application, an increase in speed may be obtained. Furthermore, the use of the windowing limits the real-time detection to the length of the chosen window. If we e.g. use a window of 5 seconds, since all the features can only be calculated after completion of the recording of the corresponding epoch.

A downside of video detection in general in this setup is that the video cannot record subtle movement that occur under the blankets. For these type of movements, other modalities have to be used which may include accelerometers, thermal infrared cameras (if the blankets don't mask the heat too much), or depth cameras.

As future work we want to investigate the combination of accelerometer and video data to optimize the results, or even use other modalities. However attaching more sensors will reduce the benefit of the non-contacting video sensor. Another possible idea to reconstruct movement under the blankets is to model the patient's posture (e.g. if we know the orientation of the head, we can estimate how the patient is positioned under the blankets).

V. CONCLUSIONS

The application of the STIP method on the myoclonic data gives promising initial results. The best obtained result over all the patients combined reaches a sensitivity of 77% and a PPV of 87%. We can conclude that a longer window gives better results in detecting seizures. However if the seizures are too subtle, the method seems not able to detect them.

REFERENCES

[1] Chen L, Yang X, Liu Y, Zeng D, Tang Y, Yan B, Lin X, Liu L, Xu H, Zhou D (2009) Quantitative and trajectory analysis of movement trajectories in supplementary motor area seizures of frontal lobe epilepsy. Epilepsy & Behavior 14(2):344–353

[2] Conradsen I, Beniczky S, Wolf P, Henriksen J, Sams T, Sorensen HBD (2010) Seizure onset detection based on a uni- or multi-modal intelligent seizure acquisition (uisa/misa) system. 2010 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2010) pp 3269–72

[3] Cuppens K, Lagae L, Ceulemans B, Van Huffel S, Vanrumste B (2009) Detection of nocturnal frontal lobe seizures in pediatric patients by means of accelerometers: a first study. EMBC: 2009 Annual International Conference of the Ieee Engineering in Medicine and Biology Society, Vols 1-20 pp 6608–6611

[4] Haiping L, How-Lung E, Mandal B, Chan DWS, Yen-Ling N (2011) Markerless video analysis for movement quantification in pediatric epilepsy monitoring. 2011 33rd Annual International Conference of the IEEE Engineering in Medicine and Biology Society

[5] Jallon P (2010) A bayesian approach for epileptic seizures detection with 3d accelerometers sensors. 2010 32nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC 2010) pp 6325–8

[6] Karayiannis NB, Tao GZ, Frost JD, Wise MS, Hrachovy RA, Mizrahi EM (2006) Automated detection of videotaped neonatal seizures based on motion segmentation methods. Clinical Neurophysiology 117(7):1585–1594

[7] Kramer U, Kipervasser S, Shlitner A, Kuzniecky R (2011) A novel portable seizure detection alarm system: Preliminary results. Journal of Clinical Neurophysiology 28(1):36–38

[8] Laptev I (2005) On space-time interest points. International Journal of Computer Vision 64(2-3):107–123

[9] Laptev I, Marszalek M, Schmid C, Rozenfeld B (2008) Learning realistic human actions from movies. 2008 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)

[10] Li ZJ, da Silva AM, Cunha JPS (2002) Movement quantification in epileptic seizures: A new approach to video-eeg analysis. IEEE Transactions on Biomedical Engineering 49(6):565–573

[11] Lockman J, Fisher RS, Olson DM (2011) Detection of seizure-like movements using a wrist accelerometer. Epilepsy & Behavior 20(4):638–641

[12] Nijsen TME, Aarts RM, Cluitmans PJM, Griep PAM (2010) Time-frequency analysis of accelerometry data for detection of myoclonic seizures. IEEE Transactions on Information Technology in Biomedicine 14(5):1197–1203

[13] Poppe R (2010) A survey on vision-based human action recognition. Image and Vision Computing 28(6):976–990

[14] Schuldt C, Laptev I, Caputo B (2004) Recognizing human actions: A local svm approach. In: 17th International Conference on Pattern Recognition (ICPR), pp 32–36