Extraction of Temporal Motor Activity Signals From Video Recordings of Neonatal Seizures By Feature Tracking Methods Based on Deformable Motion Models

Nicolaos B. Karayiannis and Yaohua Xiong
Department of Electrical and Computer Engineering, University of Houston, Houston, TX 77204-4005

Abstract—**This paper presents a new feature tracking method for video. This method estimates the displacement of a feature between two successive frames by minimizing an error function defined in terms of the feature intensities at these frames. Feature tracking relies on an affine motion model, which can be used to track features that may be translated and deformed from one frame to the next. The proposed method is used to extract temporal motor activity signals from video recordings of neonatal seizures.**

Keywords—**Affine model, feature tracking, motor activity signal, deformable motion model, translation**

I. INTRODUCTION

Seizure occurrence represents the most frequent clinical sign of central nervous system disorders in the newborn [2], [6]. Video recording is typically used with synchronized EEG and other polygraphic measures to analyze the characteristics of a seizure after its recording [1], [6]-[8]. Computerized processing and analysis of video recordings of neonatal seizures can extract quantitative information that is relevant only to the seizure. This information can be used to: 1) refine the characterization of repetitive motor behaviors, and 2) facilitate the differentiation of certain clinical seizures from other abnormal paroxysmal behaviors not due to seizures.

Neonatal seizures can be quantified in terms of temporal motion strength and motor activity signals [3], [4]. Motor activity signals were extracted in a recent study by tracking certain features located at moving body parts affected by the seizure. Feature tracking relied on the KLT algorithm [5], [10]. Although the KLT algorithm was generally successful, in some cases the algorithm lost features that were located at moving body parts tracked throughout the frame sequence. The improvement of the feature tracking method employed by the KLT algorithm focused on various aspects of feature tracking, including the use of a motion model involving an affine transformation in addition to translation [9], [11]. Despite the computational overhead due to the use of an affine motion model, the tracker proposed in [9] did not perform well when used to extract motion-related information from video recordings of neonatal seizures. This experimental outcome provided the motivation for the study outlined in this paper, which focused on the development of a working feature tracker based on an affine motion model.

This work was supported by the National Institute of Biomedical Imaging and Bioengineering under Grant 1 R01 EB00183-01.

II. EXTRACTION OF TEMPORAL MOTOR ACTIVITY SIGNALS FROM VIDEO

Figure 1 illustrates the mechanism that can be used for generating temporal signals tracking the movements of different parts of the infant's body during focal clonic and myoclonic seizures [4]. Figure 1 depicts a single frame containing the sketch of an infant's body with four selected anatomical sites. In this particular configuration, X_{LL} and Y_{LL} represent the projections of the site located at the left leg to the horizontal and vertical axes, respectively. The projections of the sites located at the right leg, left hand, and right hand are denoted by X_{RL} and Y_{RL}, X_{LH} and Y_{LH}, and X_{RH} and Y_{RH}, respectively. As the infant moves its extremities, the locations of the sites in the frame will change, as will the projections of the sites to the horizontal and vertical axes. Recording the values of the projections from frame to frame of the videotaped seizure will generate four pairs of temporal signals, namely the signals $X_{LL}(t)$ and $Y_{LL}(t)$ for the left leg, the signals $X_{RL}(t)$ and $Y_{RL}(t)$ for the right leg, the signals $X_{LH}(t)$ and $Y_{LH}(t)$ for the left hand, and the signals $X_{RH}(t)$ and $Y_{RH}(t)$ for the right hand. For a given set of anatomical sites, each seizure will produce signature signals depending on its type and location.

III. DEFORMABLE MOTION MODELS

Consider a frame sequence $\{I(\mathbf{u},t)\}$, where $u=[x \;\; y]^{\mathrm{T}}$, $\mathbf{a}^{\mathrm{T}}$ denotes the transpose of a vector $\mathbf{a}$, x and y are the coordinates of a pixel in the frame, and $I(\mathbf{u},t)$ represents the intensity of the pixel from frame t located at (x,y). Let $I(\mathbf{v},t+\tau)$ be the intensity of a small region (i.e., a feature) at frame $t+\tau$, where $\mathbf{v}=f(\mathbf{u})$ represents the new coordinates of the pixels within this region at frame $t+\tau$. The function $f(.)$ determines the model of motion employed for feature tracking. If feature tracking is based on a pure translation model, the location of the feature at frame $t+\tau$ is given by

$$\mathbf{v} = \mathbf{u} + \mathbf{d}_{\mathbf{u}}, \tag{1}$$

where $\mathbf{d}_{\mathbf{u}} = \begin{bmatrix} d_x & d_y \end{bmatrix}^{\mathrm{T}}$ is the displacement vector.

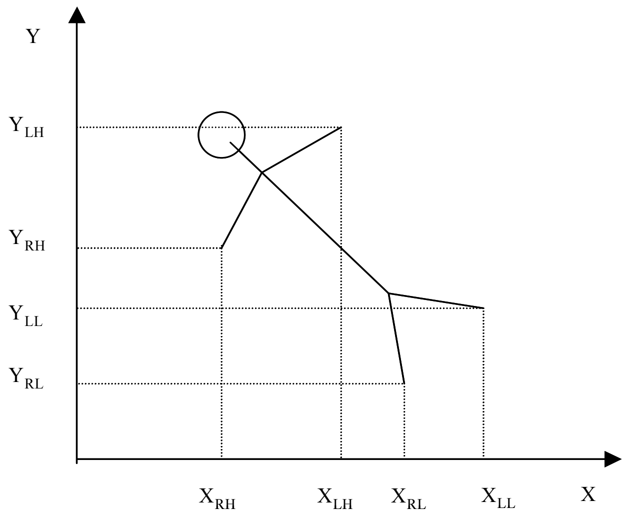

Figure 1: Extraction of temporal motor activity signals by projecting four selected anatomical sites to the horizontal and vertical axes.

Consider a linear motion model that implements an affine transformation, that is,

$$\mathbf{v} = \mathbf{u} + \mathbf{A}\mathbf{u} + \mathbf{d}_{\mathbf{u}}\,, \tag{2}$$

where

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix}, \tag{3}$$

and $\mathbf{d}_{\mathbf{u}} = \begin{bmatrix} d_x & d_y \end{bmatrix}^{\mathbf{T}}$. According to this motion model, the feature tracked may be displaced and deformed. The pure translation model in (1) is a special case of the linear motion model (2) that corresponds to $\mathbf{A} = \mathbf{0}$. If $\mathbf{A} \neq \mathbf{0}$, then the motion of the pixels within the feature tracked can be described by a vector $\mathbf{z} \in \mathbb{R}^{6\times1}$ that is formed in terms of the parameters of the motion model as $\mathbf{z} = \begin{bmatrix} a_{11} & a_{21} & a_{12} & a_{22} & d_x & d_y \end{bmatrix}^{\mathbf{T}}$. In such a case, the model (2) can also be written as $\mathbf{v} = \mathbf{u} + \mathbf{K}\,\mathbf{z}$, where

$$\mathbf{K} = \begin{bmatrix} x & 0 & y & 0 & 1 & 0 \\ 0 & x & 0 & y & 0 & 1 \end{bmatrix}. \tag{4}$$

It is assumed that, under the affine motion model $\mathbf{v} = \mathbf{u} + \mathbf{A}\mathbf{u} + \mathbf{d}_{\mathbf{u}}$, the intensities of the pixels within the feature tracked remain the same, that is,

$$I(\mathbf{z} + \delta\mathbf{z}, t + \tau) = I(\mathbf{z}, t). \tag{5}$$

The assumption is valid only for sufficiently high temporal sampling rates.

IV. FEATURE TRACKING

Tracking of a feature (i.e., a block of pixels) throughout a sequence of frames requires the development of a procedure for estimating the unknown vector $\delta\mathbf{z}$ between two successive frames in terms of the pixel intensities in these frames. This can be accomplished by minimizing the error [5], [10]

$$\varepsilon = \frac{1}{2}\sum_{W}\left[I(\mathbf{z} + \delta\mathbf{z}, t + \tau) - I(\mathbf{z}, t)\right]^2, \tag{6}$$

where W is a window located at the center pixel of the feature tracked. The minimization of ε can be made analytically tractable by approximating $I(\mathbf{z} + \delta\mathbf{z}, t + \tau)$ using a first-order Taylor expansion about $\mathbf{z}$ as

$$I(\mathbf{z} + \delta\mathbf{z}, t + \tau) = I(\mathbf{z}, t + \tau) + \mathbf{g}_{\mathbf{z}}^{\mathbf{T}}\,\delta\mathbf{z}, \tag{7}$$

where $\mathbf{g}_{\mathbf{z}} \doteq \nabla_{\mathbf{z}}\, I(\mathbf{z}, t+\tau)$ denotes the gradient of $I(.)$ with respect to $\mathbf{z}$, defined as

$$\mathbf{g}_{\mathbf{z}} = \begin{bmatrix} \dfrac{\partial I}{\partial z_1} & \dfrac{\partial I}{\partial z_2} & \cdots & \dfrac{\partial I}{\partial z_n} \end{bmatrix}^{\mathrm{T}}. \tag{8}$$

Using this approximation, the error defined in (6) becomes

$$\varepsilon = \frac{1}{2}\sum_{W}\left[I(\mathbf{z}, t + \tau) - I(\mathbf{z}, t) + \mathbf{g}_{\mathbf{z}}^{\mathrm{T}}\;\delta\mathbf{z}\right]^2. \tag{9}$$

The unknown vector $\delta\mathbf{z}$ can be obtained in terms of the gradient $\nabla_{\delta\mathbf{z}}\varepsilon$ of ε with respect to $\delta\mathbf{z}$ by solving the equation

$$\nabla_{\delta\mathbf{z}}\,\varepsilon = \sum_{W}\mathbf{g}_{\mathbf{z}}\left[I(\mathbf{z}, t+\tau) - I(\mathbf{z}, t) + \mathbf{g}_{\mathbf{z}}^{\mathrm{T}}\delta\mathbf{z}\right] = \mathbf{0}\,. \tag{10}$$

The equation (10) can also be written as [5], [10]

$$\mathbf{G}\,\delta\mathbf{z} = \mathbf{e}\,, \tag{11}$$

where

$$\mathbf{G} = \sum_{W}\mathbf{g}_{\mathbf{z}}\,\mathbf{g}_{\mathbf{z}}^{\mathrm{T}}\,, \tag{12}$$

and

$$\mathbf{e} = \sum_{W}\mathbf{g}_{\mathbf{z}}\left[I(\mathbf{z}, t) - I(\mathbf{z}, t+\tau)\right]. \tag{13}$$

The estimate $\delta\mathbf{z}$ obtained by solving (11) may not be particularly accurate. An alternative approach is to minimize the error in (9) by using an iterative optimization procedure, such as the Newton-Raphson method. In such a case, the new estimate $\delta\mathbf{z}^{new}$ of the unknown vector $\delta\mathbf{z}$ is obtained in terms of the current estimate $\delta\mathbf{z}^{old}$ as

$$\delta\mathbf{z}^{new} = \delta\mathbf{z}^{old} - \mathbf{H}^{-1}\,\nabla_{\delta\mathbf{z}}\,\varepsilon\,, \tag{14}$$

where $\nabla_{\delta\mathbf{z}}\varepsilon = -(\mathbf{e} - \mathbf{G}\,\delta\mathbf{z})$ is the gradient of ε with respect to $\delta\mathbf{z}$, and $\mathbf{H}$ is the Hessian matrix. The Hessian matrix for the error function in (9) can be obtained as $\mathbf{H} = \mathbf{G}$, where $\mathbf{G}$ is defined in (12). For $\nabla_{\delta\mathbf{z}}\varepsilon = -(\mathbf{e} - \mathbf{G}\,\delta\mathbf{z})$ and $\mathbf{H}^{-1} = \mathbf{G}^{-1}$, the update equation (14) becomes

$$\delta\mathbf{z}^{new} = \mathbf{G}^{-1}\mathbf{e}\,. \tag{15}$$

In this particular case, each iteration of the Newton-Raphson method is equivalent to solving the equation (11).

V. FEATURE TRACKING BASED ON DEFORMABLE MOTION MODELS

Consider the affine motion model defined in (2). Since $\mathbf{v} = \mathbf{u} + \mathbf{K}\,\mathbf{z}$,

$$\nabla_{\mathbf{z}} I(\mathbf{z}, t+\tau) = \nabla_{\mathbf{z}}(\mathbf{v})^{\mathbf{T}}\, \nabla_{\mathbf{v}} I(\mathbf{v}, t+\tau)\,, \tag{16}$$

where

$$\nabla_{\mathbf{z}}(\mathbf{v}) = \left[\frac{\partial \mathbf{v}}{\partial a_{11}} \;\; \frac{\partial \mathbf{v}}{\partial a_{21}} \;\; \frac{\partial \mathbf{v}}{\partial a_{12}} \;\; \frac{\partial \mathbf{v}}{\partial a_{22}} \;\; \frac{\partial \mathbf{v}}{\partial d_x} \;\; \frac{\partial \mathbf{v}}{\partial d_y} \right]. \tag{17}$$

Since $\mathbf{v} = \mathbf{u} + \mathbf{K}\,\mathbf{z}$, $\nabla_{\mathbf{z}}(\mathbf{v}) = \nabla_{\mathbf{z}}(\mathbf{u} + \mathbf{K}\,\mathbf{z}) = \mathbf{K}$. The gradient $\nabla_{\mathbf{v}} I(\mathbf{v}, t+\tau)$ can be computed in terms of the gradient $\mathbf{g}_{\mathbf{u}} \doteq \nabla_{\mathbf{u}}\, I(\mathbf{u}, t)$, defined as

$$\mathbf{g}_{\mathbf{u}} = \left[\frac{\partial I}{\partial x} \;\; \frac{\partial I}{\partial y} \right]^{\mathbf{T}}. \tag{18}$$

This can be accomplished by using (5), which can also be written as

$$I(\mathbf{v}, t+\tau) = I(\mathbf{u}, t)\,. \tag{19}$$

Since $\mathbf{v} = \mathbf{u} + \mathbf{K}\,\mathbf{z}$, taking the gradient with respect to $\mathbf{u}$ of both sides of (19) gives

$$\nabla_{\mathbf{u}}(\mathbf{v})^{\mathrm{T}}\, \nabla_{\mathbf{v}} I(\mathbf{v}, t+\tau) = \nabla_{\mathbf{u}} I(\mathbf{u}, t)\,, \tag{20}$$

where

$$\nabla_{\mathbf{u}}(\mathbf{v}) = \left[\frac{\partial \mathbf{v}}{\partial x} \;\; \frac{\partial \mathbf{v}}{\partial y} \right]. \tag{21}$$

Since $\mathbf{v} = \mathbf{u} + \mathbf{A}\mathbf{u} + \mathbf{d}_{\mathbf{u}}$, $\nabla_{\mathbf{u}}(\mathbf{v}) = \nabla_{\mathbf{u}}(\mathbf{u} + \mathbf{A}\mathbf{u} + \mathbf{d}_{\mathbf{u}}) = \mathbf{I} + \mathbf{A}$. Thus, (20) gives

$$\nabla_{\mathbf{v}} I(\mathbf{v}, t+\tau) = [(\mathbf{I} + \mathbf{A})^{-1}]^{\mathrm{T}}\; \mathbf{g}_{\mathbf{u}}\,. \tag{22}$$

The gradient $\mathbf{g}_{\mathbf{z}} = \nabla_{\mathbf{z}} I(\mathbf{z}, t+\tau)$ of $I(.)$ with respect to $\mathbf{z}$ can be obtained by combining (22) and (16) as

$$\mathbf{g}_{\mathbf{z}} = [(\mathbf{I} + \mathbf{A})^{-1} \mathbf{K}\,]^{\mathrm{T}}\; \mathbf{g}_{\mathbf{u}}\,. \tag{23}$$

For comparison, the gradient $\mathbf{g}_{\mathrm{z}}$ was obtained in [9] for the same model as

$$\mathbf{g}_{\mathrm{z}} = \mathbf{K}^{\mathbf{T}}\, \mathbf{g}_{\mathbf{u}}\,. \tag{24}$$

For $\mathbf{A} \neq \mathbf{0}$, (24) differs from the gradient formula (23) derived in this paper.

VI. EXPERIMENTAL RESULTS

Figures 2 and 3 show the motor activity signals extracted from the video recordings of neonatal seizures by utilizing the feature tracking methods based on pure translation and deformable motion models. The locations of the moving body parts during the clinical event are shown in representative frames of each video recording. The frames of the video recordings shown in the figures can be used as a reference to verify the consistency of the temporal motor activity signals with the corresponding clinical events. The values of the signals corresponding to the frames shown at the top of each figure are indicated by dots, while the moving body part in each video recording is shown within a box.

Figure 2 shows the temporal motor activity signals produced for a myoclonic seizure affecting the infant's right foot by the feature tracking methods relying on pure translation and deformable motion models. Both methods identified significant motor activity in the horizontal direction right after frame 160. The temporal signals shown in Figure 2(c) also reveal some motor activity along the vertical direction after frame 160, which was identified only by the feature tracking method relying on the deformable motion model. Frame-by-frame visual inspection of the entire frame sequence indicated that there is substantial motion in a short time internal around frame 160 but there is no substantial motion outside this interval (see frames 120, 162, and 200 in Figure 2). Note also that the temporal motor activity signals shown in Figure 2 are consistent with the rapid and "jerky" movements of the infants' extremities affected by myoclonic seizures.

Figure 3 shows the temporal motor activity signals produced by the feature tracking methods tested in the experiments for a focal clonic seizure afflicting the infant's right leg. Both feature tracking methods managed to track a feature located at the infant's right leg. In fact, both feature tracking methods tested on this focal clonic seizure captured and quantified the rhythmicity that is the signature characteristic of such clinical events; however, they produced slightly different motor activity signals.

VII. CONCLUSIONS

This paper introduced a new method for tracking features in video. The proposed method relies on an affine motion model that can track features that may be translated and deformed from one frame to the next. The proposed method was used to extract motor activity signals from the video recordings of neonatal seizures of the myoclonic and focal clonic type. The same task was also performed by a feature tracking method relying on a pure translation motion model. Unlike the feature tracking method developed in [9] using the same affine model, the feature tracking method proposed in this paper was successful in quantifying motor activity from video recordings of neonatal seizures. This experimental study did not reveal any significant differences between the feature tracking methods relying on the pure translation and deformable motion models. This can be attributed to the fact that the features tracked in these experiments occupied relatively small and uniform regions located at the body parts affected by the seizures. Thus, the pure translation motion model was probably sufficient for quantifying the motion of the features tracked in the two video recordings used in the experiments.

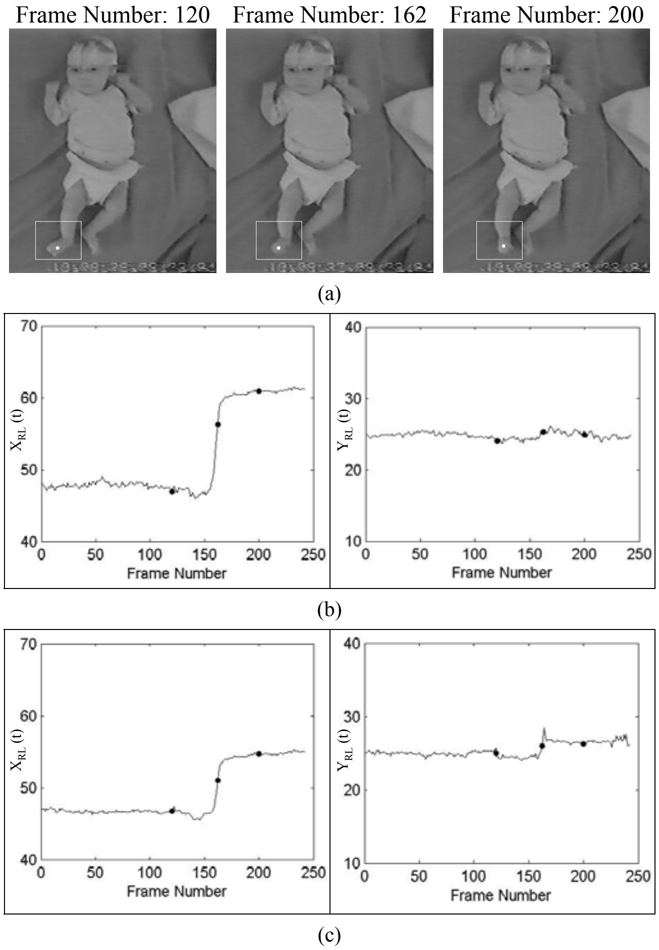

Figure 2: (a) Selected frames of a video recording of a myoclonic seizure affecting the infant's right foot, (b) motor activity signals produced by the feature tracking method based on a pure translation motion model, and (c) motor activity signals produced by a feature tracking method based on a deformable motion model.

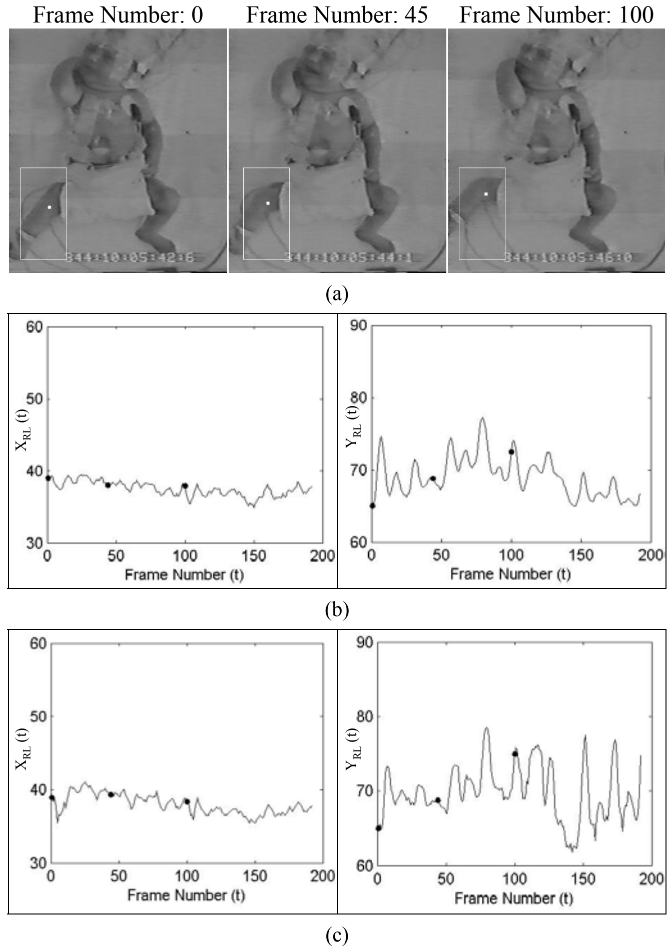

Figure 3: (a) Selected frames of a video recording of a focal clonic seizure affecting the infant's right leg, (b) motor activity signals produced by the feature tracking method based on a pure translation motion model, and (c) motor activity signals produced by a feature tracking method based on a deformable motion model.

REFERENCES

[1] C. D. Binnie, A. J Rowan, J. Overweg, H. Meinard, T. Wisman, A. Kamp, F. L. da Silva, "Telemetric EEG and video monitoring in epilepsy," *Neurology,* vol. 31, pp. 98-303, 1981.

[2] G. M. Fenichel, *Neonatal Neurology,* 3rd ed. New York, NY: Churchill-Livingstone, 1990.

[3] N. B. Karayiannis, "Advancing videometry through applications: Quantification of neonatal seizures from video recordings," *Proceedings of Fourteenth International Conference on Digital Signal Processing,* Santorini, Greece, July 1-3, 2002, pp. 11-21.

[4] N. B. Karayiannis, S. Srinivasan, R. Bhattacharya, M. S. Wise, J. D. Frost Jr., and E. M. Mizrahi, "Extraction of motion strength and motor activity signals from video recordings of neonatal seizures," *IEEE Transactions on Medical Imaging,* vol. 20, no. 9, pp. 965-980, 2001.

[5] D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereoscopic vision," *Proceedings of International Joint Conference on Artificial Intelligence*, Vancouver, Canada, August 1981, pp. 674-679.

[6] M. Mizrahi, "Neonatal seizures," in *Pediatric and Adolescent Medicine*, S. Shinnar, N. Amir, D. Branski (Eds.), Basel, Karger, vol. 6, pp. 18-31, 1995.

[7] M. Mizrahi and P. Kellaway,. "Characterization and classification of neonatal seizures," *Neurology,* vol. 37, pp. 1837-1844, 1987.

[8] F. Pierelli, G-E. Chatrian, W. W. Erdly, P. D. Swanson, "Long-term EEG-video-audio monitoring: Detection of partial epileptic seizures and psychogenic episodes by 24-hour EEG record review," *Epilepsia,* vol. 30, pp. 513-523, 1989.

[9] J. Shi and C. Tomasi, "Good features to track," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, WA, June 1994, pp. 593-600.

[10] C. Tomasi and T. Kanade, "Detection and tracking of point features," *Carnegie Mellon University Technical Report CMU-CS-91-132*, April 1991.

[11] T. Tommasini, A. Fusiello, E. Trucco, and V. Roberto, "Making good features track better," *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, Santa Barbara, CA, June 1998, pp. 178-183.