Application of Adaptive Block Matching in the Extraction of Temporal Motor Activity Signals From Video Recordings of Neonatal Seizures

Nicolaos B. Karayiannis and Abdul Sami
Department of Electrical and Computer Engineering
N308 Engineering Building 1
University of Houston
Houston, TX 77204-4005
Karayiannis@uh.edu, Asami@mail.uh.edu

Abstract

This paper presents a procedure developed to extract quantitative information from video recordings of neonatal seizures in the form of temporal motor activity signals. The motor activity signals are extracted by tracking selected anatomical sites during the seizure using adaptive block matching. The motion of a block of pixels is quantified by searching for the most similar block of pixels in subsequent frames; this search is facilitated by employing various update strategies to account for the changing appearance of the block. The experiments indicate that the temporal motor activity signals extracted by the proposed procedure constitute an effective representation of videotaped clinical events and can be used for seizure recognition and characterization.

1. Introduction

Seizure occurrence represents the most frequent clinical sign of central nervous system disorders in the newborn [2], [10], [16]. These disturbances in cerebral function may result in significant long-term adverse sequelae such as neurological handicaps, mental retardation, and postnatal epilepsy [1], [10], [16]. Thus, the prompt recognition of seizures in the neonatal intensive care unit is very important with regard to diagnosis and management of underlying neurological problems.

The development of portable EEG/video/polygraphic monitoring techniques allows investigators to assess and characterize neonatal seizures at the bedside and permits retrospective review [1], [9], [10]. These techniques are relatively expensive, are generally used for only a few hours of monitoring, and may not be routinely available in many centers. Automated processing and analysis of video recordings of neonatal seizures can generate novel methods for extracting quantitative information that is relevant only to the seizure. The extraction of quantitative information from video recordings of neonatal seizures can be accomplished by two complementary procedures designed to extract temporal motion strength and motor activity signals from video [5], [6]. In principle, motor activity signals are obtained by projecting to the horizontal and vertical axes an anatomical site located at the body part affected by the seizure. The extraction of motor activity signals from video recordings of neonatal seizures relies on a procedure that can track the anatomical site of interest throughout the frame sequence. This paper presents the results of a study that relied on adaptive block matching to extract motor activity signals from the video recordings of neonatal seizures and other clinical events associated with high motor activity.

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

2. Extraction of temporal motor activity signals from video

The extraction of quantitative information from video recordings of neonatal seizures can be accomplished by projecting the location of selected anatomical sites to the horizontal and vertical axes. As the seizure progresses in time, these projections will produce temporal motor activity signals for the body parts affected by the seizure.

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. 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

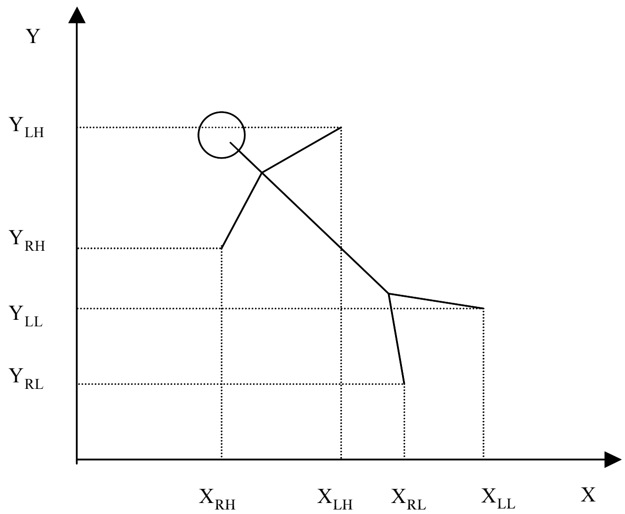

Figure 1: Generation of temporal motor activity signals by projecting four selected anatomical sites to the horizontal and vertical axes.

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 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.

The development of an automated procedure capable of tracking the site of interest in successive frames of the video recording was accomplished in [6] by employing a feature-tracking procedure often referred to as the KLT algorithm [8], [14]. The KLT algorithm automatically selects "good features" from the first frame of an image sequence. A good feature is one that can be tracked well throughout the entire image sequence [12], [14]. Although the KLT algorithm was generally successful, in some cases the algorithm lost some features that were located at moving body parts tracked throughout the image sequence. The susceptibility of the KLT algorithm to "lost features" motivated the tracking of a sufficiently large number of features within a predetermined radius from the selected anatomical site in the frame sequence. This paper presents an alternative procedure for the extraction of temporal motor activity signals, which relies on block matching.

3. Feature tracking based on block matching

Block matching is a popular correlation-based approach to motion estimation [4] and tracking [11], [13]. Block matching relies on the assumption that a block of pixels remains constant over time and motion [4]. This assumption is valid only if the frame rate is sufficiently high and the time period is short. The anatomical site to be tracked is typically defined as the center of a square block of pixels (e.g.15x15, 11x11), which is referred to as the *reference block*. The reference block is tracked by searching for the most similar block in subsequent frames according to some similarity measure. The search is typically constrained to a search window, which has to be chosen appropriately. A large search window allows to track any rapid motion that would have been lost if the search window were smaller. However, a very large window would increase the likelihood of mismatch. Moreover, increasing the size of the search window increases considerably the computational effort associated with block matching. The computation time can be greatly reduced by employing signature-based and suboptimal block matching techniques, such as the 2-D logarithmic search, three-step search, orthogonal search, one-at-a-time search, and cross search [3], [4], [7]. Although computationally effective, such block matching techniques may be inferior in terms of their reliability. The reference block, to which the block of pixels is matched, has to be updated in order to take into consideration the changes in the appearance of the target. Such an extension of block matching is often referred to as *adaptive block matching*. The update of the reference block can be implemented according to a variety of strategies, which include the single-frame, multiframe, and FIR update strategies tested and evaluated in this study.

4. Extraction of motor activity signals from video based on adaptive block matching

The application of adaptive block matching (ABM) in the extraction of temporal motor activity signals from video recordings of neonatal seizures involves several choices, which include the block size, the similarity measure, and the update strategy for the reference block.

4.1. Similarity measures

A *distortion function* is used to quantify the similarity between the target block and candidate blocks. Let $\{A(p,q)\}$ and $\{B(p,q)\}$ be the pixels of the $M \times N$ target and candidate blocks, respectively. The mean absolute difference (MAD) function is defined as

$$MAD = \frac{1}{MN}\sum_{p=1}^{M}\sum_{q=1}^{N}\left|A(p,q)-B(p,q)\right|. \quad (1)$$

The mean square difference (MSD) function is defined as

$$MSD = \frac{1}{MN}\sum_{p=1}^{M}\sum_{q=1}^{N}\left(A(p,q)-B(p,q)\right)^2. \quad (2)$$

The differences between the MAD and MSD criteria are often too subtle to be perceived in practice.

4.2. Size of reference block

The application of adaptive block matching in feature tracking requires the selection of the size of the reference block. The experiments tested reference blocks of sizes 5×5, 7×7, 11×11, and 15×15. The width of each reference block was deliberately chosen to be an odd number in order for the block to contain a center pixel that can be projected to the horizontal and vertical axis to produce motor activity signals. The experiments indicated that large reference blocks contain more information and are less susceptible to noise. Nevertheless, increasing the size of the reference block increases the computational effort associated with the extraction of motor activity signals. In addition, large reference blocks may include regions of the frames that represent background. In such a case, the reference block is likely to match the background of a certain set of frames instead of a site located on a body part affected by a seizure.

4.3. Update strategies

According to the adaptive block matching method, the reference block is tracked by searching for the most similar block in subsequent frames according to some similarity measure. Tracking of the reference block requires the adoption of a strategy to be used for updating the reference block throughout the frame sequence. The update strategies employed in this experimental study are described below [11]:

4.3.1. Single-frame update strategy. The simplest update strategy that can be employed for adaptive block matching is the single-frame strategy. According to this strategy, the reference block is replaced after every N_S frames by the block of pixels at the current tracking position. If $N_S = 1$, the reference block is updated after every frame. On the other hand, the reference block is never replaced if $N_S = \infty$. If the reference block is updated too often, then the feature tracked may be lost because of the accumulation of errors due to camera jitter and even finite precision. If the reference block is not updated often enough, the feature tracked may be lost again. This is due to the fact that the feature tracked can change with time to a degree that no good match exists between the feature and the reference block.

4.3.2. Multiframe update strategy. The multiframe update strategy searches for the best match in the current frame (say the n th frame) based on N_M reference blocks, with $N_M > 1$. The reference blocks employed by this update strategy are the best-matched blocks found in the N_M previous frames. Let B_n be a candidate block of pixels in the current frame and let M_k be the best match in the k th frame. According to the multiframe update strategy, the best match in the current frame is determined based on the following similarity measure

$$f_m(B_n, M_{n-1}, \ldots M_{n-N_M}) = \sum_{k=1}^{N_M} w_k \; f(B_n, M_{n-k}), \quad (3)$$

where $f(B_n, M_{n-k})$ measures the similarity measure between B_n and M_{n-k} and $\{w_k\}$ are real weights. The weights $\{w_k\}$ can be determined to ensure that the search for the best match in the current frame is influenced more intensely by the best matching blocks found in the most recent frames. Such a scheme can be realized by setting $w_k = \lambda\alpha^k$, with $\alpha < 1$, and by computing λ such that

$$\sum_{k=1}^{N_M} w_k = 1. \quad (4)$$

4.3.3. FIR update strategy. This update strategy searches for the best match in the current frame based on a reference block obtained as a linear combination of the best-matched blocks in the previous N_F frames. According to this update strategy, the reference frame block for the n th frame is obtained as

$$R_n = \sum_{k=1}^{N_F} a_k \; M_{n-k}, \quad (5)$$

where M_k is the best matching block in the k th frame and $\{a_k\}$ are real coefficients. The name of this update strategy underlines the resemblance of (5) to a linear finite impulse response (FIR) filter. If $N_S \le N_F$, $a_{N_S} = 1$, and $a_k = 0, \forall\; k \ne N_S$, then the update strategy based on (5) reduces to the single-frame strategy described above. In general, the coefficients $\{a_k\}$ can be selected to be decreasing functions of k in order to ensure that the reference block R_n resembles the best matching blocks in the most recent frames. The use of (5) ensures that the reference block adapts to the changing appearance of the feature tracked. This makes this update strategy capable of tracking the features throughout the given frame sequence. On the other hand, the weighted averaging involved in (5) tends to cancel the noise that might be present in the best-matched blocks. This makes this update strategy resistant to noise.

5. Experimental results

The three update strategies described above were employed to extract motor activity signals from two myoclonic and two focal clonic seizures selected from a database available at the Methodist Hospital in Houston. Temporal signals were also produced for two video recordings of normal infant behavior (random infant movements). This section presents some of the results of this study. Myoclonic and focal clonic seizures generally affect the infants' extremities. The anatomical sites to be tracked were located at the infants' extremities affected by the seizures. The experiments indicated that a reference block of size 15×15 balances the tradeoff between performance and computational effort. The search window was a square block of size 50×50. Feature tracking was attempted in the experiments by employing the single-frame update strategy with $N_S = 1, 5, 10, \infty$. It was found that the single-frame strategy with $N_S = 5$ combined resistance to noise with the capacity to track features throughout the frame sequence. The multiframe update strategy was used in the experiments with $N_M = 5$ and $\alpha = 0.9$. The FIR update strategy was tested using $N_F = 5$ and the same set of weights used in the multiframe strategy. Figures 2 and 3 show the temporal motor activity signals produced by the three update strategies tested in the experiments for a myoclonic and a focal clonic seizure, respectively. 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 Figures 2 and 3 can be used as a reference to verify the consistency of the temporal 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.

In the myoclonic seizure shown in Figure 2, the infant's left leg moves to the right of the frame between frames 10 and 16 (Figure 2 shows only frame 14). This movement was captured by the temporal signal obtained as the projection of the moving part to the horizontal axis. The temporal signal obtained as the projection of the moving part to the vertical axis indicates that the left leg also moves toward the top of the frame, which can be verified by comparing frames 0 and 14 of the sequence. The infant's left leg remains at an almost fixed position between frames 50 and 150. In this time interval, the temporal motor activity signals are almost flat. According to Figure 2, there is a noteworthy difference between the motor activity signals produced by the three update strategies. In this case, the main difference can be observed in the signals produced by projecting the anatomical site located on the infants left leg to the horizontal axis. The signal produced by the FIR update strategy reveals that the infant's leg moves to the right and then to the left just before frame 14. The same movement was also captured by the multiframe strategy. The multiframe update strategy revealed a weaker horizontal movement before frame 14. The amplitude of this movement was even lower in the motor activity signals produced by the single-frame update strategy. Finally, the motor activity signals produced for the myoclonic seizure shown in Figure 2 are consistent with the "jerky" movements that are the typical signatures of such events.

Figure 3 shows the temporal motor activity signals produced by adaptive block matching for a focal clonic seizure affecting the infant's right hand. Figure 3 indicates that the temporal signals produced by the two proposed procedures captured and quantified the rhythmicity that characterizes the movements of such clinical events. Frame-by-frame inspection of the video recording indicated that the temporal motor activity signal $Y_{RH}(t)$ produced by the single-frame update strategy does not constitute a satisfactory representation of this clinical event. On the other hand, the multiframe and FIR update strategies produced similar motor activity signals. In fact, the difference between the motor activity signals produced by these two update strategies for the focal clonic seizure shown in Figure 3 are too subtle to actually affect the quantification of motor activity.

6. Conclusions

This paper showed that adaptive block matching can be used to extract motor activity signals from video recordings of neonatal seizures. Adaptive block matching was tested in this application by employing three different update strategies for the reference block, namely the single-frame, multiframe, and FIR strategies. The outcome of the experiments indicated that the performance of adaptive block matching depends rather strongly on the update strategy employed for the reference block. More specifically, the FIR update strategy outperformed both the multiframe and single-frame update strategies. On the other hand, the multiframe update strategy performed better than the single-frame update strategy. Although the multiframe and FIR update strategies are more effective than the single-frame strategy, they incur higher

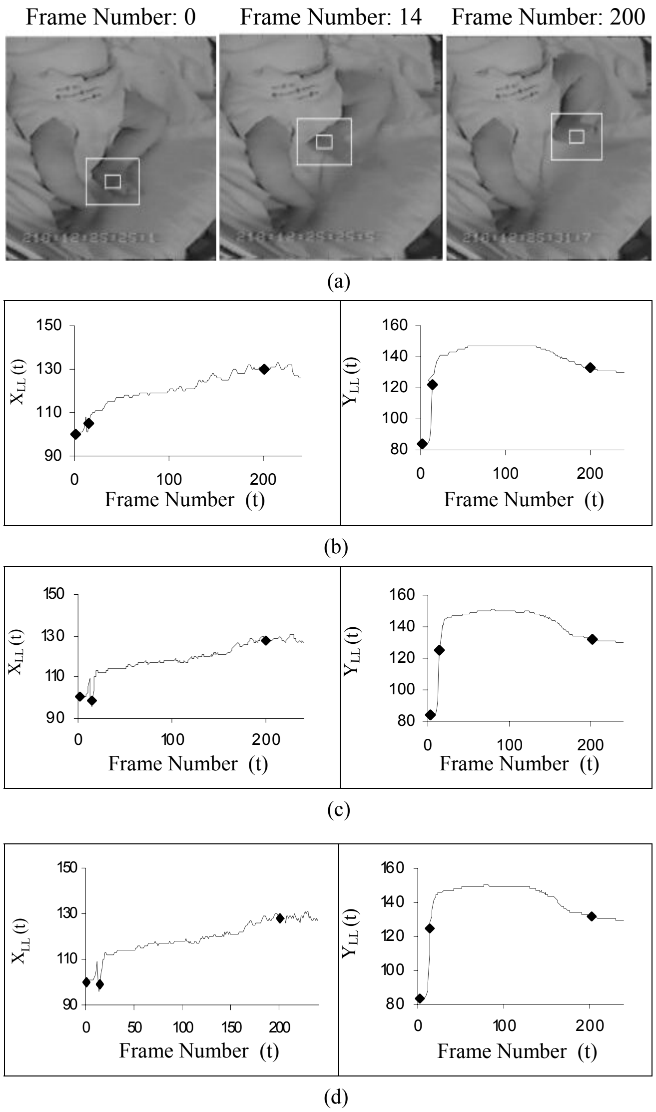

Figure 2: Temporal signals produced for a video recording of a myoclonic seizure affecting the infant's left leg: (a) selected frames of the sequence, temporal motor activity signals produced by adaptive block matching relying on: (b) the single-frame update strategy, (c) the multiframe update strategy, and (d) the FIR update strategy.

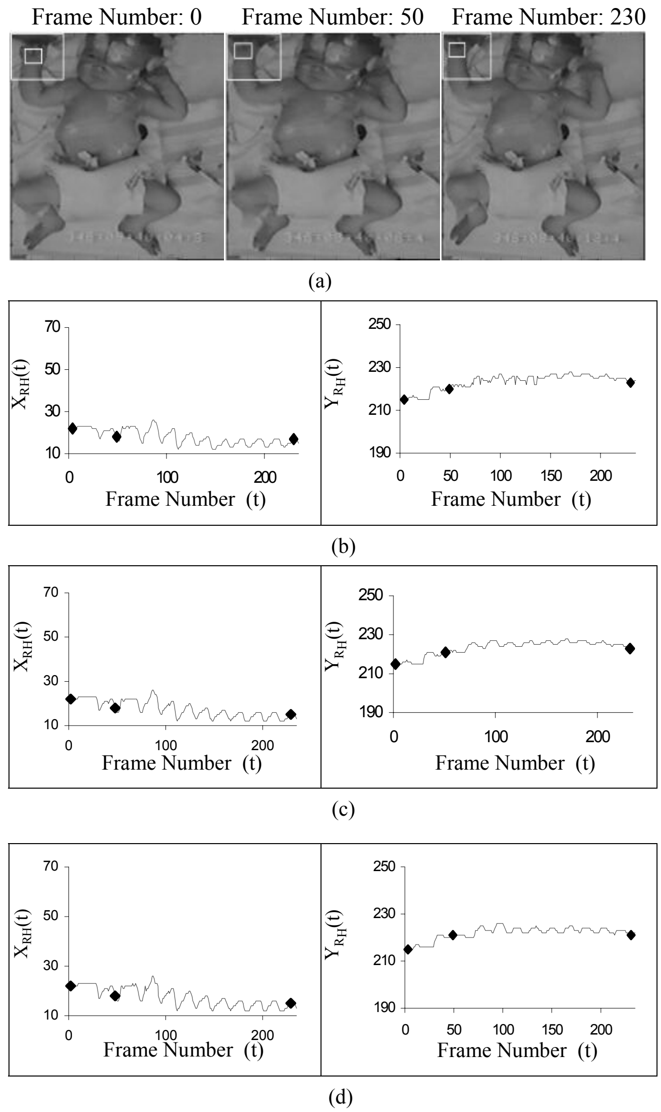

Figure 3: Temporal signals produced for a video recording of a focal clonic seizure affecting the infant's right hand: (a) selected frames of the sequence, temporal motor activity signals produced by adaptive block matching relying on: (b) the single-frame update strategy, (c) the multiframe update strategy, and (d) the FIR update strategy.

computational and memory cost. This study relied on exhaustive search to determine the best match for the reference block within the search window. The computational effort associated with adaptive block matching can be reduced considerably by employing signature-based and/or suboptimal search algorithms. Such search algorithms may have a negative impact on the reliability of the tracking procedure. Nevertheless, the tradeoff between tracking reliability and computational effort is worth investigating in order to enhance the practical value of the proposed procedure.

An interesting problem currently under investigation is the extraction of motor activity signals based on a *predictive block matching* method. Such an approach would attempt to predict the location of the block in the next frame according to some method similar with that employed by the KLT algorithm. The adoption of such an approach instead of a blind search is expected to improve the reliability of the block matching procedure while speeding up the search for the best match by allowing the reduction of the search window size.

It was found that adaptive block matching occasionally failed to deal effectively with the problem of occlusion, that is, the problem occurring when other body parts occlude the anatomical site tracked throughout the frame sequence. Occlusion over a short period of time does not have any substantial effect on the tracking because of the update strategies employed for the given reference block by adaptive block matching. However, partial occlusion over a long period of time can lead to mismatches that may result in lost features. A potential solution to this problem could be to divide the reference block into smaller blocks of the same size. Adaptive block matching would provide the motion vectors, which could be regularized using a vector median filter. The motion vectors of the unoccluded region would provide the basis for tracking the anatomical site of interest even if the site being tracked becomes partially occluded.

Further improvement and refinement of the procedure developed in this study can produce temporal motor activity signals that constitute a consistent and effective representation of focal clonic seizures, myoclonic seizures, and clinical events not due to seizures that are characterized by substantial motor activity of the infants' extremities. This can be accomplished by fine-tuning the proposed procedure on a large database of video recordings of neonatal seizures and clinical events not due to seizures, which is currently in progress. The fine-tuning of the proposed procedure is expected to enhance the statistical significance of the results and to verify the validity of the resulting motor activity signals.

The proposed approach was developed for myoclonic and focal clonic seizures and may not be suitable for other types of neonatal seizures involving subtle movements of body parts other than the extremities, such as ocular and oral-buccal-lingual seizures. Nevertheless, focal clonic and myoclonic seizures are characterized by movements of the extremities and constitute a large portion (up to 75%) of seizures observed in the neonate. Systems capable of identifying more subtle seizure types may be feasible with further development and refinement of the procedure outlined in this paper. The extension of the proposed procedure to more subtle neonatal seizures can certainly benefit from current technological developments in the recording and storage of high-resolution digital video.

7. References

[1] A. M. E. Bye, C. A. Cunningham, K.Y Chee, and D. Flanagan, "Outcome of neonates with electrographically identified seizures, or at risk of seizures," *Pediatr Neurol,* vol. 16, pp. 225-231, 1997.

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

[3] M. Ghanbari, "The cross-search algorithm for motion estimation," *IEEE Trans. on Communications*, vol. 38, no. 7, pp. 950-953, 1990.

[4] J. Jain and A. Jain, "Displacement measurement and its application in interframe image coding," *IEEE Trans. on Communications*, vol. 29, no. 12, pp. 1799-1808, 1981.

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

[6] 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 Trans. on Medical Imaging*, vol. 20, no. 9, pp. 965-980, 2001.

[7] T. Koga, K. Iinuma, A. Hirano, Y. Iijima, and T. Ishiguro, "Motion-compensated interframe coding for video conferencing," *Proc. of Nat. Telecom. Conf.*, November 1981, pp. 531-535.

[8] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," *Proc. International Joint Conf. on Artificial Intelligence*, Vancouver, Canada, August 1981, pp. 674-679.

[9] E. M. Mizrahi, "Neonatal electroencephalography: Clinical features of the newborn, techniques of recording, and characteristics of the normal EEG," *Amer. J. EEG Technol.,* vol. 26, pp. 81-103, 1986.

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

[11] A. M. Peacock, S. Matsunaga, D. Renshaw, J. Hannah, and A. Murray, "Reference block updating when tracking with block matching algorithm," *IEE Electronic Letters,* vol. 36, no. 4, pp. 309-310, 2000.

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

[13] H. Smith, C. Richards, S. Brandt, and N. Papanikolopoulos, "Visual tracking for intelligent vehicle-highway systems," *IEEE Trans. on Vehicular Technology*, vol. 45, no. 4, pp. 744-759, 1996.

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

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

[16] J. J. Volpe, *Neurology of the Newborn*, Philadelphia: W. B. Suanders, 1995.)