# Markerless Video Analysis for Movement Quantification in Pediatric Epilepsy Monitoring

Haiping Lu, How-Lung Eng, Bappaditya Mandal, Derrick W. S. Chan and Yen-Ling Ng

*Abstract*— This paper proposes a markerless video analytic system for quantifying body part movements in pediatric epilepsy monitoring. The system utilizes colored pajamas worn by a patient in bed to extract body part movement trajectories, from which various features can be obtained for seizure detection and analysis. Hence, it is non-intrusive and it requires no sensor/marker to be attached to the patient's body. It takes raw video sequences as input and a simple user-initialization indicates the body parts to be examined. In background/foreground modeling, Gaussian mixture models are employed in conjunction with HSV-based modeling. Body part detection follows a coarse-to-fine paradigm with graph-cut-based segmentation. Finally, body part parameters are estimated with domain knowledge guidance. Experimental studies are reported on sequences captured in an Epilepsy Monitoring Unit at a local hospital. The results demonstrate the feasibility of the proposed system in pediatric epilepsy monitoring and seizure detection.

## I. INTRODUCTION

Epilepsy is a common neurological condition in pediatrics and there have been many efforts devoted to automate epileptic seizure detection [1]. Detection can be used to trigger an alarm during severe seizures for medical assistance, and to help in evaluation of treatment effects [2]. Many existing seizure detection systems rely on electroencephalographic patterns. For example, the work in [3] identifies features important for seizure recognition from multi-channel electroencephalogram (EEG) signals [4], [5] rearranged into a third-order tensor [6]. This approach requires patients to remain attached to EEG equipment. Another approach is to detect seizures based on the visual analysis of signals from accelerometers worn on the patient's body [2]. Such attached sensors may lead to discomfort and are subject to dislocation.

Since epileptic seizures often induce uncoordinated movement in a patient's body such as jerks or stiffening, such movement is a relevant clinical factor in seizure identification and it can be extracted from video recordings [7]. The third approach utilizes markers attached to a patient's body for the quantification of movement patterns in patients during epileptic seizures [7], [8]. These markers often need manual identification with mouse clicking and pose difficulties in defining correspondence so only short sequences can be analyzed. We can see that for a pediatric population, these solutions are problematic. Thus, this paper focuses on non-

Haiping Lu, How-Lung Eng and Bappaditya Mandal are with the Institute for Infocomm Research, Agency for Science, Technology and Research, Singapore {hlu, hleng, bmandal}@i2r.a-star.edu.sg

Derrick W. S. Chan and Yen-Ling Ng are with KK Women's and Children's Hospital, Singapore {Derrick.Chan.WS, Ng.Yen.Ling}@kkh.com.sg

intrusive markerless video-based seizure detection in pediatric epilepsy monitoring.

To the best of the authors' knowledge, the only existing non-intrusive markerless video-based seizure detection system is the one developed by the University of Houston for neonates [9], [10]. Anatomic sites on the moving body part are selected by thresholding the magnitudes of the motion vectors and then the selected sites are tracked. The motion of body parts is quantified by temporal motion-strength signals extracted from video segments. This approach relies purely on motion information so the level of details in quantifications is limited and coarse. As a motion-based method, tracking tends to become unreliable over time so the processed video segments are only up to 20 seconds long in [9], [10]. In clinical studies, this system is only applied to manually selected sequences to differentiate neonatal seizures from random movements [9], [10].

In this paper, we investigate the development of a non-intrusive markerless video analytic system for pediatric epilepsy monitoring. The objective is to quantify body part movements in greater details than the motion-based system in [9], [10], and comparable to the marker-based or accelerometer-based systems. We propose such a system that only requires a patient to wear a pajama with specific colors, without attaching any sensor or marker to his/her body. This design enables us to model the patient's body parts. Video of patient activities in a clinical epilepsy monitoring unit (EMU) are captured using a camera mounted on the ceiling. Epochs of activities are excerpted and subjected to automated video analysis following simple manual initialization. The position and angle of patient's limbs are automatically extracted and displacement, velocity and frequency of limb movements are estimated for identifying distinct characteristics between seizure and non-seizure activities. We have analyzed recorded video data and observed sustained displacement from baseline and presence of strong oscillation during focal motor seizures, which are not present during interictal activity.

## II. MARKERLESS VIDEO ANALYSIS FOR MOVEMENT QUANTIFICATION

This section describes the proposed video analytic system. In our design, we assume: 1) only the patient is in the view; 2) the background is fixed in the duration of the monitoring sequence to be analyzed; 3) the patient is not covered by blanket or any other cover so that the pajama worn by the patient is visible from video cameras mounted on the ceiling; 4) the patient is wearing a pajama of customized colors.
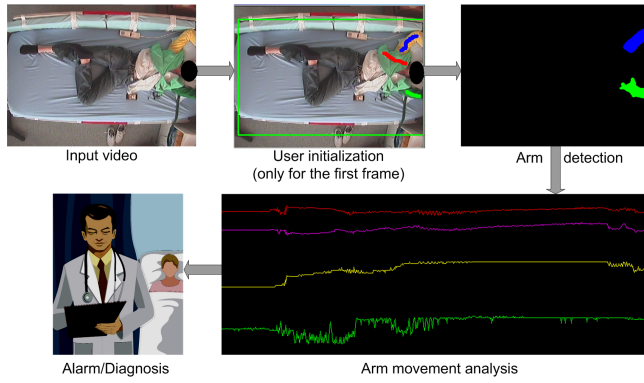
Fig. 1. Process flow of the proposed system for pediatric epilepsy monitoring. The patient's face is masked in this paper for anonymity and it is not part of the user initialization.

Figure 1 illustrates the proposed system. In this example, only the arm movements are quantified. The monitored patient wears specially designed color pajama, with no sensor/marker attached. Video cameras are mounted on the ceiling observing the patient in the EMU. Video sequences captured by the video cameras are fed into a computer system. The user provides initialization by indicating a region of interest, the two foreground colors, and one background color using mouse, which correspond to the green rectangle, the blue line and the green line, and the red line shown in the top middle of Fig. 1, respectively. In the figure, the patient's face is masked for anonymity and this is not part of the user initialization. This simple user-initialization is needed only for the first frame of a monitoring sequence to be analyzed.

After user initialization, several fully-automatic modules follow, including construction and update of background and foreground models, detection of patient body parts, and estimation of body part parameters.

### A. Foreground and Background Modeling

The hue, saturation, value (HSV) color space is employed for modeling the foreground and background in color video frames. The hue components play the most significant role in the modeling. The reason is that due to our specially designed colored pajamas, the color information will be very distinct for the limbs and this enables us to detect body part reliably.

Two foreground models $F_1$, $F_2$ and one background model $B_1$ are initialized based on the median values $\{H_{f1}, S_{f1}, V_{f1}\}$, $\{H_{f2}, S_{f2}, V_{f2}\}$, $\{H_{fb}, S_{fb}, V_{fb}\}$ of the HSV components of the user-selected foreground pixels and background pixels, respectively. Pixels with HSV values within a range $\{H_T, S_T, V_T\}$ from $\{H_{f1}, S_{f1}, V_{f1}\}$, $\{H_{f2}, S_{f2}, V_{f2}\}$, and $\{H_{fb}, S_{fb}, V_{fb}\}$ set the masks $F_1$, $F_2$, and $B_1$, respectively. The thresholds are set to $\{H_T = 0.03, S_T = 0.15, V_T = 0.15\}$ for HSV range $[0, 1]$ in our experiments, where the hue components play a major role. In addition, a gray mask $B_2$ is generated from pixels with saturation value $S$ less than a threshold $G_T$ ($G_T = 0.2$ for range of 0 to 1 in our simulations).

Following the simple HSV-based modeling, Gaussian mixture models (GMMs) are built as in [11]. For a color video

frame $j$, the array $\mathbf{z}_j = (z_1, ..., z_n, ..., z_N)$ of $N$ pixels where $z_n = (H_n, S_n, V_n)$, $n \in [1, N]$ in HSV space. Let an array $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)$, $\alpha_n \in \{0, 1\}$ denote the labeling of each pixel as background ($\alpha_n = 0$) or foreground ($\alpha_n = 1$). Two GMMs with $K(= 5)$ components each are defined for background and foreground pixels, parameterized as [11]

$$\boldsymbol{\theta} = \{\pi(a, k), \mu(a, k), \Sigma(a, k), \alpha = 0, 1, k = 1, ..., K\}, \quad (1)$$

where $\pi$ is the weight, $\mu$ is the mean, and $\Sigma$ is the covariance matrix. The vector $\mathbf{k} = \{k_1, ..., k_n, ..., k_N\}$ with $k_n \in \{1, ..., K\}$ indicates the component of the background or foreground GMM each pixel belongs to, assigning according as $\alpha_n = 0$ or 1.

### B. Coarse-to-Fine Detection of Body Parts

The detection of patient's body parts follows a coarse-to-fine approach adapted from [12], where the body parts are roughly located using simple methods and then more computationally expensive methods are utilized to refine the detection results.

The coarse detection simply produces the masks $B_1 \cup B_2$ and $F_1 \cup F_2$. Figure 2 illustrates the coarse detection using the HSV-based model. Figures 2(a) and 2(b) depict two foreground segments detected ($F_1$ & $F_2$), with one clean and one noisy, respectively. Figures 2(c) and 2(d) show the background segment detected $B_1$ and the gray mask generated $B_2$, respectively. The coarse detection phase also makes use of the detection in previous frame to make sure that the shifting of the bounding box of the detected body part is bounded by a maximum amount.

After the coarse detection, the fine detection employs the popular GrabCut algorithm in [11] to produce foreground with large connected region, as shown in Fig. 3(a). In GrabCut, a trimap $T$ is consisted of three regions: $T_B$, $T_F$ and $T_U$, corresponding to the initial background, foreground, and uncertain pixels, respectively. Pixels belonging to $T_U$ are to be classified as either background or foreground. We detect two color segments separately. Based on the HSV-based modeling, we initialize the trimap as $T_B = \{B_1 \cup B_2 \cup F_2\}$,
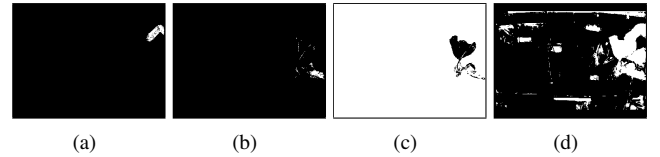


Fig. 2. Illustration of the HSV-based model for body part detection: (a) a clean foreground segment, (b) a noisy foreground segment, (c) a background segment, and (d) a gray mask.
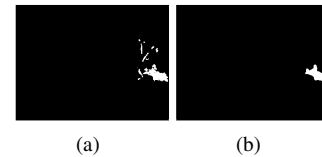


Fig. 3. Illustration of body part detection refinement: (a) the foreground segment in Fig. 2(b) after Graph-Cut-based segmentation, and (b) the final foreground segment after refinement of Fig. 3(a).

$T_F = F_1$, and $T_U = \{z_n \notin T_B \cup T_F\}$ for color 1 and $T_B = \{B_1 \cup B_2 \cup F_1\}$, $T_F = F_2$, and $T_U = \{z_n \notin T_B \cup T_F\}$ for color 2. For the detection or classification of pixels in $T_U$, $\boldsymbol{\alpha}$ is initialized as $\alpha_n = 0$ for $n \in T_B$ and $\alpha_n = 1$ for $n \in T_U \cup T_F$. The background and foreground GMMs are then initialized from $\boldsymbol{\alpha}$ with $k$-means clustering. Each pixel is then assigned a GMM component $k_n$ and GMM parameters are learned. The segmentation is then estimated through Graph Cut [13] by maximizing the following Gibbs energy [11]

$$\mathbf{E}(\boldsymbol{\alpha}, \mathbf{k}, \boldsymbol{\theta}, \mathbf{z}_j) = U(\boldsymbol{\alpha}, \mathbf{k}, \boldsymbol{\theta}, \mathbf{z}_j) + V(\boldsymbol{\alpha}, \mathbf{z}_j). \quad (2)$$

The data term $U$ is defined as [11]

$$U(\boldsymbol{\alpha}, \mathbf{k}, \boldsymbol{\theta}, \mathbf{z}_j) = \sum_n - \log p(z_n | \alpha_n, k_n, \boldsymbol{\theta}) - \log \pi(\alpha_n, k_n), \quad (3)$$

where $p(\cdot)$ is a Gaussian probability distribution. The smoothness term $V$ is computed as [11]

$$V(\boldsymbol{\alpha}, \mathbf{z}) = \gamma \sum_{(m,n) \in \mathbf{C}} [\alpha_n \neq \alpha_m] e^{-\beta \|z_m - z_n\|^2}, \quad (4)$$

where $\mathbf{C}$ is the set of neighboring pixel pairs, $\gamma$ is usually set to 50 and

$$\beta = \frac{1}{2 < \| z_m - z_n \|^2 >}, \quad (5)$$

where $< \cdot >$ is the expectation over a frame [14].

Next, a refinement procedure is introduced by keeping only one connected region with the largest normalized area for each foreground color, followed by median filtering. The foreground segment in Fig. 3(a) after refinement is shown in Fig. 3(b).

### C. Domain-Knowledge-Based Body Part Parameter Estimation

For quantitative clinical analysis, we estimate the parameters for a patient's body part movements. From the detected body parts as described in the previous subsection, we estimate their parameters by adapting the approach in [15], [16].

As in [15], [16], dynamic parameters are most significant for the analysis of body part movements. There are two basic categories of dynamic parameters. One is the position or location of body parts and the other is the orientation (or angle) of body parts. The estimation of body part position parameters is through: 1) determine which boundary point is the body part of interest based on domain knowledge; and 2) calculate an average position for the body part position for robust estimation. The estimation of body part orientation parameters is through: 1) detection of edge in the detected body part segment; 2) determine the most reliable edge to use for estimation of the body part orientation; and 3) estimate the orientation of the most reliable edge as the body part orientation based on the median values of the orientations for the pixels on that edge. In addition, several constraints are enforced to encode domain knowledge, e.g., the change of parameters in consecutive frames are capped to a maximum value.

## III. EXPERIMENTAL STUDIES

This section presents experimental studies of the proposed system. The experimental video data are collected from epileptic patients in the EMU at the KK Women's and Children's Hospital, Singapore. The video sequences are captured at 12 frames per second and the resolution is $384 \times 288$. The parameter settings in each module are set as described in Section II. To mimic EEG tracings for the clinician's easy adaptation in reading, we follow [7] to display the quantified movement information as movement trajectories. In this paper, we present the results for the arm movements of a patient for demonstration.

Figures 4 and 5 show the simulation results for a typical sequence of 767 frames with seizure events, and a typical sequence of 765 frames with normal events, respectively. Both sequences have a duration of 64 seconds. Four sample frames from each sequence are shown in Figs. 4(a)-4(d), and Figs. 5(a)-5(d). The patient's face is masked for anonymity. Four body part parameters are plotted in Figs. 4(e) and 5(e), where the horizontal axis is the time axis. The top two curves are the horizontal and vertical positions of the right shoulder of the patient, while the bottom two curves are the orientations for the right and left upper-arms.
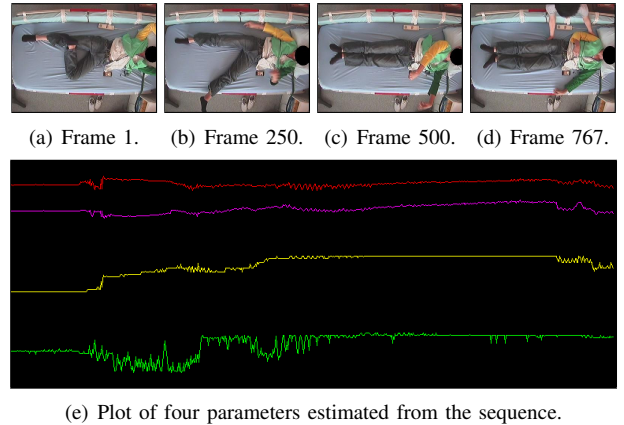


(a) Frame 1.  (b) Frame 250.  (c) Frame 500.  (d) Frame 767.



(e) Plot of four parameters estimated from the sequence.

Fig. 4. Illustration of movement quantification for a sequence with seizure events.



(a) Frame 1.  (b) Frame 250.  (c) Frame 500.  (d) Frame 765.



(e) Plot of four parameters estimated from the sequence.
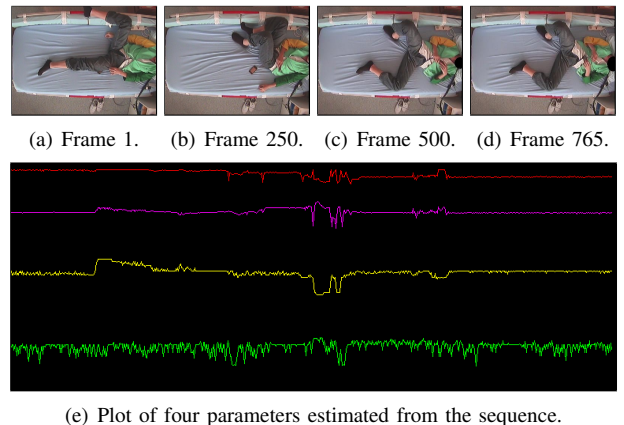
Fig. 5. Illustration of movement quantification for a sequence with normal (non-seizure) events.

(a) The displacement feature.
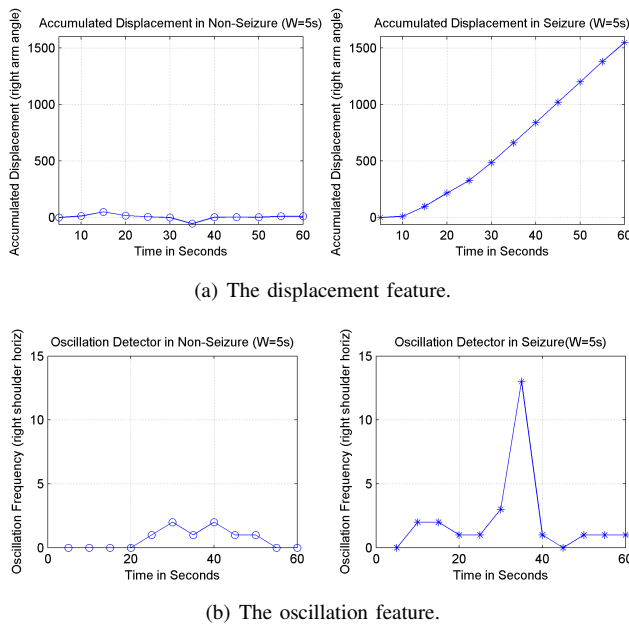


(b) The oscillation feature.

Fig. 6. Illustration of the distinct features between a seizure event and a non-seizure event (normal activity).

From the trajectories of the four arm parameters, we can observe some distinct motion patterns for a seizure event and a normal (non-seizure) event. The most distinctive characteristics of the seizure event from the normal event include the sustained displacement of the right arm angle from the baseline, and the presence of strong oscillation in the positions of the right shoulder.

Based on the observations above, we have designed two simple seizure detectors based on the displacement and oscillation features. The duration of each epoch $W$ is fixed at 60 frames or 5 seconds. The displacement feature is calculated as the accumulated absolute difference from the baseline (the first epoch), while the oscillation feature is the frequency of the strongest oscillation.

Figure 6(a) depicts seizure detection based on sustained displacement of right arm angle from baseline, the third trajectory in Figs. 4(e) and 5(e). Based on the detector response, seizure can be detected at 20 seconds using a threshold of 250. Figure 6(b) illustrates seizure detection based on the horizontal oscillation of the right shoulder, the first trajectory in Figs. 4(e) and 5(e). From the detector response, seizure can be detected around 30 seconds using a threshold of 5.

## IV. CONCLUSIONS

This paper presents a non-intrusive markerless video analytic system for quantification of body part movement for pediatric epilepsy patients. The proposed system makes use of colored pajamas and needs a simple user-initialization. It involves HSV-based and GMM-based modeling, Graph-Cut-based segmentation, coarse-to-fine detection, and domain-knowledge-based body part parameter estimation. The experimental studies show the feasibility of quantitative movement analysis and seizure detection, with distinct characteristics observed between ictal and interictal body movements. Furthermore, the proposed system is promising in a wider application field of fine semiology investigation.

## V. ACKNOWLEDGMENTS

## REFERENCES

[1] S. C. Schachter, J. Guttag, S. J. Schiff, D. L. Schomer, and S. Contributors, "Advances in the application of technology to epilepsy: The cimit/nio epilepsy innovation summit," *Epilepsy & Behavior*, vol. 16, no. 1, pp. 3–46, Sept. 2009.

[2] T. M. E. Nijsen, J. B. A. M. Arends, P. A. M. Griep, and P. J. M. Cluitmans, "The potential value of three-dimensional accelerometry for detection of motor seizures in severe epilepsy," *Epilepsy & Behavior*, vol. 7, no. 1, pp. 74–84, Aug. 2005.

[3] E. Acar, C. A. Bingol, H. Bingol, R. Bro, and B. Yener, "Seizure recognition on epilepsy feature tensor," in *Proc. 29th Annual Int. Conf. of the IEEE Engineering in Medicine and Biology Society*, 2007, pp. 4273–4276.

[4] B. Blankertz, R. Tomioka, S. Lemm, M. Kawanabe, and K.-R. Müller, "Optimizing spatial filters for robust EEG single-trial analysis," *IEEE Signal Processing Mag.*, vol. 25, no. 1, pp. 41–56, Jan. 2008.

[5] H. Lu, H.-L. Eng, C. Guan, K. N. Plataniotis, and A. N. Venetsanopoulos, "Regularized common spatial pattern with aggregation for eeg classification in small-sample setting," *IEEE Trans. Biomed. Eng.*, vol. 57, no. 12, pp. 2936–2946, Dec. 2010.

[6] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A survey of multilinear subspace learning for tensor data," *Pattern Recognition*, vol. 44, no. 7, pp. 1540–1551, July 2011.

[7] Z. Li, A. M. da Silva, and J. P. S. Cunha, "Movement quantification in epileptic seizures: A new approach to video-eeg analysis," *IEEE Trans. Biomed. Eng.*, vol. 49, no. 6, pp. 565–573, June 2002.

[8] L. Chen, X. Yang, Y. Liu, D. Zeng, Y. Tang, B. Yan, X. Lin, L. Liu, H. Xu, and D. Zhou, "Quantitative and trajectory analysis of movement trajectories in supplementary motor area seizures of frontal lobe epilepsy," *Epilepsy & Behavior*, vol. 14, no. 2, pp. 344–353, Feb. 2009.

[9] N. B. Karayiannis, G. Tao, J. D. F. Jr., M. S. Wise, R. A. Hrachovy, and E. M. Mizrahi, "Automated detection of videotaped neonatal seizures based on motion segmentation methods," *Clinical Neurophysiology*, vol. 117, no. 7, pp. 1585–1594, July 2006.

[10] N. B. Karayiannis, Y. Xiong, G. Tao, J. D. F. Jr., M. S. Wise, R. A. Hrachovy, and E. M. Mizrahi, "Automated detection of videotaped neonatal seizures of epileptic origin," *Epilepsia*, vol. 47, no. 6, pp. 1585–1594, July 2006.

[11] C. Rother, V. Kolmogorov, and A. Blake, ""GrabCut": interactive foreground extraction using iterated graph cuts," *ACM Transactions on Graphics (Proceedings of ACM SIGGRAPH 2004)*, vol. 23, no. 3, pp. 309–314, Aug. 2004.

[12] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "Coarse-to-fine pedestrian localization and silhouette extraction for the gait challenge data sets," in *Proc. IEEE Conference on Multimedia and Expo*, July 2006, pp. 1009–1012.

[13] Y. Boykov and V. Kolmogorov, "An experimental comparison of min-cut/max-flow algorithms for energy minimization in vision," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 26, no. 9, pp. 1124–1137, Sept. 2004.

[14] Y. Boykov and M.-P. Jolly, "Interactive graph cuts for optimal boundary & region segmentation of objects in n-d images," in *Proc. IEEE Conference on Computer Vision*, vol. 1, 2001, pp. 105–112.

[15] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A layered deformable model for gait analysis," in *Proc. IEEE International Conference on Automatic Face and Gesture Recognition*, Apr. 2006, pp. 249 – 254.

[16] H. Lu, K. N. Plataniotis, and A. N. Venetsanopoulos, "A full-body layered deformable model for automatic model-based gait recognition," *EURASIP Journal on Advances in Signal Processing*, vol. 2008, 2008, article ID 261317, 13 pages, doi:10.1155/2008/261317.