An adaptive method for the recovery of missing samples from FHR time series

Vangelis P. Oikonomou*, Jiri Spilka†, Chrysostomos Stylios*, Lenka Lhostka†
**Laboratory of Knowledge and Intelligent Computing,*
Department of Informatics and Telecommunication Technology
TEI of Epirus, 47100, Arta, Greece, Email: viknmu@gmail.com, stylios@teiep.gr
†Department of Cybernetics, Faculty of Electrical Engineering,
Czech Technical University, Prague, Czech Republic, Email: spilka.jiri@fel.cvut.cz

Abstract

Missing data cause serious problem for automatic evaluation of the fetal heart rate(FHR) series. In this work we present an algorithm to surpress this problem. More specifically, an adaptive approach is proposed based on two steps. The first step concerns the reconstruction step where we obtain an estimate of the missing data using an empirical dictionary. The second step consists from the construction of the dictionary using the updated values from the first step. The above two steps are applied iteratively until convergence. The method adapts each time the dictionary and the reconstructed time series to the new information that we gain. Results on real and simulated experiments have shown the usefullness of our approach. More specifically, a comparison with cubic spline interpolation is performed and have shown that the proposed approach achieved 4 to 9dB better reconstruction ability.

1. Introduction

Oxygen inssuficiency for fetus during delivery could cause adverse sequels for the newborn. Hence, accurate evaluation of fetal status is crucial during pregnancy [1], [2]. A standard approach to monitor the fetal status is the Cardiotocography (CTG), which measures maternal uterine contractions (UC) and fetal heart rate (FHR). It is believed that these signals contains significant information about the underlying physiology of the fetus. The introduction of CTG in 1960 brought great expectations related to the outcome. However, after many years and studies, it has observed that it does not offer significant improvement in the delivery outcomes. Furthermore, it became the main suspect for increased rate of cesarean sections [3]. The major drawbacks of CTG are the poor standard of interpretation and the contribution of the human factor, demonstrated by high intra and inter observer variability [4]. To overcome these shortcomings we can either offer more training and education on the interpretation of standard or produce decision support systems to help the experts.

The general steps of a decision support system are: preprocessing of FHR time series, extraction of features and classification. More specifically, the preprocessing step includes segment selection, artefact removal and interpolation of missing samples [5]. The goal of preprocessing is to provides us a time series of high quality for further investigation. After preprocessing, the feature extraction step takes place. In the literature a vast amount of features have been applied for FHR analysis. Features from time and frequency domain, morphological features, as well as features based on non - linear analysis, are studied in the context of FHR analysis [5]–[7]. Also, significant effort has been consumed on the determination of usefull features for subsequent analysis of the time series [6], [8]. Finally, the classification of time series as normal or pathological is performed using a classifier, such as Support Vector Machines (SVM) [5], [7].

A subject that has been overlooked in the literature, related to FHR analysis, concerns the recovery of missing data. The FHR time series is a very noisy signal and a vast amount of data from it have been missing during the acquisition [9]. This happens due to the movement of the baby and the stress induced during labour. Because of missing segments in FHR time series researchers in this field are obliged to remove large parts of the signals from the subsequent analysis [5]–[7]. More specifically, to deal with this problem, linear interpolation is applied to small segments of the time series, while bigger segments have been removed, even the whole time series [6].

The recovery of missing data from observations is a problem that has attracted much attention in signal and image processing, and machine learning communities [10], [11]. In image processing community this problem is also called inpainting [12]. In our work, we borrow general ideas from these fields and we apply them for the recovery of missing samples from FHR time series. Furthermore, the recovery of missing data constitutes a linear inverse problem and it is an ill - posed problem since the degratation operator is ill behaved, as we will see in the next section. The concept of linear inverse problems is of the most active research field in signal and image processing. Methods based on deterministic approach or probabilistic inference have been extensively studied based on Markov Random Fields (MRF) [13] or on some functional spaces [14].

In many cases, it is usefull to represent the data under study into a domain which contains some usefull properties. Usually, this representation is achieved by applying a linear operator (or dictionary) on the data. An interesting scientific problem is how to choose this dictionary. In this field there are two general approches. In the first approach we can choose a pre-constructed dictionary such as wavelets, sines and cosines etc. Also, these dictionaries are accompanied by a strong theoretical basis that make them very attractive. However, these dictionaries are usefull when the signal under study possess's the anticipated properties. Alternativelly, we can resort to a tunable selection of a dictionary by adopting a learning point of view. In this option, we build a training database of signal instances and construct an empirically learned dictionary. The atoms of this dictionary come from the data and not from a theoretical model. One widely used approach for the construction of such dictionaries, which we adopt in our study, is the KSVD method [15].

In this work we proposed an adaptive algorithm for the recovery of missing samples from FHR time series. More specifically, our algorithm is of iterative nature and consists of two alternating steps, estimation of the dictionary and estimation of missing samples. The remainder of this paper is organized as follows: In section II we describe the proposed algorithm, giving the theoretical basis of what follows in experiments. To assess the performance of the proposed algorithm, we present in section III a comparison with cubic spline interpolation based on real and simulated experiments. Finally, in section IV we give some concluding remarks and future directions.

2. Methodology

Assume that the time series is arranged into a nx1 vector $\mathbf{f}$ such that $\mathbf{f} = \begin{bmatrix}\mathbf{y}\\ \mathbf{z}\end{bmatrix}$, where $\mathbf{y}$ is of size n_1x1 and contains the available (observed) samples and $\mathbf{z}$ is of size n_2x1 and denotes the missing samples, where $n_1 + n_2 = n$. Note here that the proposed algorithm works for arbitrary data missing patterns and the above notation is obtained after performing a reordering into the vectors. Let $\mathbf{M}$ denotes a n_1xn projection matrix such that:

$$\mathbf{y} = \mathbf{M}\mathbf{f}. \tag{1}$$

As we see the matrix $\mathbf{M}$ is a degradation operator that removes n_2 samples from the time series. It can be build by taking the nxn identity matrix and removing n_2 rows corresponding to the missing samples. Using the observation model described by Eq. 1 we can obtain the least square estimate such as:

$$\hat{\mathbf{f}}_{LS} = \mathbf{M}^T(\mathbf{M}\mathbf{M}^T)^{-1}\mathbf{y}. \tag{2}$$

However, this estimate does not offer any improvement into our problem since it places the zero value at the missing samples. To deal with this kind of problem a regularization term must be added into our model. Usually the regularization term is added into a model in the form of prior distribution, if we work with probabilistic modelling [16], or through a penalty term, if the deterninistic approach is adopted [17], as we performed in our study.

Using a nxn dictionary matrix $\mathbf{D}$ that contains n protype atoms for columns, we can represent a time series as a linear combination of these atoms, $\mathbf{f} = \mathbf{D}\mathbf{x}$, where $\mathbf{x}$ is the coefficient vector. So, our model takes the form:

$$\mathbf{y} = \mathbf{M}\mathbf{D}\mathbf{x}. \tag{3}$$

Obtaining an estimate for the coefficients $\hat{\mathbf{x}}$, we can reconstruct the original time series as:

$$\hat{\mathbf{f}} = \mathbf{D}\hat{\mathbf{x}}. \tag{4}$$

The classical approach to this type of inverse problems is to find the coefficients vector with the smallest ℓ_2 - norm:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_2 \text{ such that } \mathbf{y} = \mathbf{M}\mathbf{D}\mathbf{x}. \tag{5}$$

The solution to the above optimization problem is given in closed form as:

$$\hat{\mathbf{x}} = (\mathbf{M}\mathbf{D})^T(\mathbf{M}\mathbf{D}(\mathbf{M}\mathbf{D})^T)^{-1}\mathbf{y}. \tag{6}$$

Assuming that the coefficients vector $\mathbf{x}$ is of sparse nature then we can obtain an estimate by solving the following optimization problem:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_0 \text{ such that } \mathbf{y} = \mathbf{M}\mathbf{D}\mathbf{x}. \tag{7}$$

However, the above problem is NP-hard. One solution to the above difficulty is to tranform the problem into one which is more tractable. In that spirit we can replace the $\|\cdot\|_0$ with its convex approximation $\|\cdot\|_1$. More specifically, we consider the following problem:

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \|\mathbf{x}\|_1 \text{ such that } \mathbf{y} = \mathbf{MDx}. \quad (8)$$

The ℓ_1 minimization problem can be reformulated as a linear problem with equality constraints, and it can be solved by using interior - point methods [18].

The choice of matrix $\mathbf{D}$ is a difficult one. As we have mention before we can choose the dictionary from a set of predetermined dictionaries such as wavelet transforms or we can build one. In our study, we adopt the later approach. To construct the dictionary, the KSVD algorithm is embraced [15]. Next, key ideas and concepts related to KSVD algorithm is provided. More information about the properties of KSVD can be found in [15].

The KSVD algorithm is a generalization of the well - known K - means algorithm. Given a set of examples $\mathbf{Y} = \{\mathbf{y}_i\}_{i=1}^N$ the goal is to extract K atoms holding various properties. From these atoms we can construct the desired dictionary matrix $\mathbf{D}$. To achieve that

$$\min_{\mathbf{D},\mathbf{W}} \|\mathbf{Y} - \mathbf{DW}\|_F^2 \text{ subject to } \forall i \|\mathbf{w}_i\| \leq T_0 \quad (9)$$

To solve the above problem a two stage procedure is used. In the first stage the coefficients matrix $\mathbf{W}$ is estimated by using any pursuit method, while at the second stage each dictionary element of $\mathbf{D}$ is calculated along with its coefficients. More specifically, at the first stage we need to solve the following optimization problems:

$$\min_{\mathbf{w}_i} \|\mathbf{y}_i - \mathbf{D}\mathbf{w}_i\|_2^2 \text{ subject to } \|\mathbf{w}_i\| \leq T_0 \quad (10)$$

where T_0 is the number of non - zeros elements of coefficient vector $\mathbf{w}_i$.

As concerns the second stage, the penalty term can be written as $\|\mathbf{Y} - \mathbf{DW}\|_F^2 = \|\mathbf{E}_k - \mathbf{d}_k\mathbf{w}^k\|_F^2$ where $\mathbf{d}_k$ is the k column (or atom) of the dictionary and $\mathbf{w}^k$ is the k-th row of matrix $\mathbf{W}$. Using the SVD method we can find a rank - 1 approximation of $\mathbf{E}_k$ and hence alternative $\mathbf{d}_k$ and $\mathbf{x}^k$. However, this approach does not enforce the sparsity constraint on $\mathbf{w}^k$ and most likely the new vector $\mathbf{w}^k$ will have a dense structure. The solution to this is to transform our problem into new one where the sparsity structure remains. This can be achieved by using only the samples that use the atom $\mathbf{d}_k$. From another point of view we take only the non zeros values of vector $\mathbf{w}^k$. So, our new problem is:

$$\min_{\mathbf{d}_k,\mathbf{w}^k} \|\mathbf{E}_k^R - \mathbf{d}_k\mathbf{w}_R^k\|_F^2 \quad (11)$$

where $\mathbf{E}_k^R$ is the restricted matrix $\mathbf{E}_k$ and $\mathbf{w}_R^k$ is a vector contains the non zeros elements of $\mathbf{w}^k$. Finally, applying the SVD on $\mathbf{E}_k^R$, we obtain the updates for $\mathbf{d}_k$ and $\mathbf{w}_R^k$. As we can observed at the second step of the KSVD algorithm, we update not only the dictionary but also the coefficients. However, we keep the sparsity structure obtained from the first step. This observation is the key difference between the KSVD algorithm and various others algorithms with the same goal. Most algorithms on this field keep the coefficients fixed during the update of the atoms.

2.1. Proposed algorithm

Assuming that we have a dictionary that describes with acurracy our data then we can use Eq.(4) to obtain the reconstructed signal. The coefficients $\mathbf{x}$ can be computed by solving the problems described by Eq. (5) or Eq. (8). The issue is to find/choose a good dictionary for our missing values problem. As we have mention above, we resort to KSVD algorithm to solve this. However, a significant issue is how to determine the set of examples that must be feed to this algorithm. This issue is part of the overall strategy to attack the problem of missing values in a signal with considerable large size. The adopted strategy is to work on overlapping segments of our original signal.

The signal is split up into overlapping segments. More specifically, the signal is divided into L overlapped data segments of length M, overlapping by D points, i.e from the original signal $\mathbf{f}$ we get a set of segments $\{\mathbf{f}_i\}_{i=1}^L$ where each $\mathbf{f}_i$ is a Mx1 vector. Note here, that the segmentation above has been performed with respect to the true signal, not the observed. However, the same segmentation structure could be take if we placed the zero value at the positions of missing values.

A segment $\mathbf{f}_i$ possibly will have some missing values and to restore these values a reconstruction procedure, as those described above, is used. So, at this stage, we need to solve L minimization problems. Choosing the segment length M at appropriate level the computational cost could be encountered. Finally, window averaging all segments $\hat{\mathbf{f}}_i$ we obtain the reconstructed signal $\hat{\mathbf{f}}$, which will be use to create the training set for the KSVD algorithm. Using a similar segmentation procedure on the restored signal $\hat{\mathbf{f}}$ we obtain the desired training set for KSVD. To summarize, the proposed adaptive algorithm constists from two steps:

- reconstruct the original signal
- using the reconstructed signal finds a dictionary

The above two steps are applied iteratively until convergence. Note that in each iteration of the above algorithm the input in the KSVD is changing, and the proposed algorithm adapts each time the contents of dictionary with respect to the new reconstructed data.

3. Experiments

To evaluate our algorithm experiments have been performed on real data. Also, a comparison with spline interpolation is provided. More specifically, the experiments are divived into two cases. In the fist case, we use real data and remove randomly samples to obtain a missing values signal i.e. we emulate the real problem. The comparison is performed with respect to the output SNR defined as: $SNRout = 20\log_{10}\frac{\|\mathbf{s}\|^2}{\|\mathbf{s}-\hat{\mathbf{f}}\|^2}$ where $\mathbf{s}$ and $\hat{\mathbf{f}}$ are the true and the reconstructed signals, respectively. In the second case we applied our algorithm to real data with missing values and provided a qualitative analysis and comparison.

3.1. Data Description

The database of 552 records is a subset of 9164 intrapartum CTG recordings that were acquired between years 2009 and 2012 at the obstetrics ward of the University Hospital in Brno, Czech Republic. The CTG signals were carefully selected with clinical as well as technical considerations in mind. The database is described in depth in [19]. We present the main clinical parameters as mean (minimum, maximum). The parameters are maternal age 29.8 years (18, 46), parity 0.43 (0,7), gravidity 1.43 (1,11), gestational age 40 weeks (37,43), pH 7.23 (6.85,7.47), base deficit 4.6 mmol/l (-3.4,26.11), Apgar score at 5 minute 9.06 (4,10), and neonate's weight 3408 g (1970, 4750). The proportion of males and females were almost same 259 females and 293 males. We have decided to select recordings that ended as close as possible to the birth and that had in the last 90 minutes of labor at least 40 minutes of usable signal. Additionally since CTG signal at II. stage of labor is very difficult to assess [20], only those recordings which had II. stage at maximum 30 minutes-long were included. All data were coming from the STAN S21 machines using either direct FECG (102 records) scalp electrode or ultrasound probe (412 records) or combination of both (35 records). For the three records the information was not available. All recordings were sampled at 4Hz. The majority of babies were delivered vaginally (506) and rest using caesarean section (46).

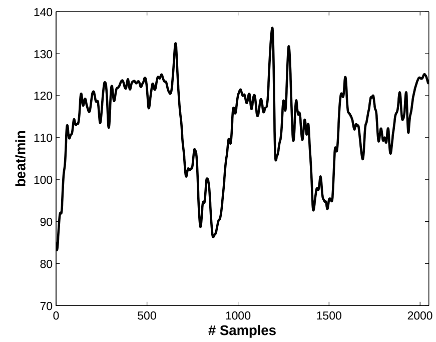


Figure 1. FHR signal used as ground truth in the simulated case.

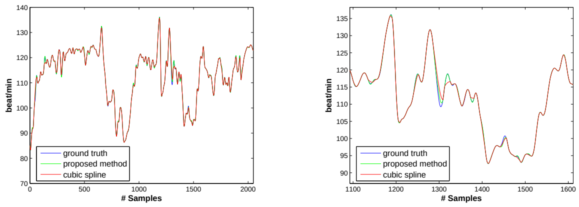
	$SNRout$(dB)	
% missing values	proposed (L2 norm)	spline
25 %	63.1071	56.7855
50 %	50.7095	41.9989
75 %	31.7955	27.9212

Table 1. Results

3.2. Simulated Experiments

A segment of 2048 samples FHR signal (see Fig. 1) is used to conduct our experiments. The above signal is obtained after performing a moving average procedure of ten points in the raw data as suggested in [21]. From this signal we remove samples randomly based on the uniform distribution. We vary the percentage of missing samples according to Table 1. The obtained results are shown in Table 1. These results are obtained after performing 20 Monte Carlo simulations, and hence represent mean values. It is clear that the proposed approach presents better results, in terms of $SNRout$, compared to the spline interpolation. More specifically, the difference between the two methods in the quality of the reconstructed signal, ranges from 4dB to 9dB.

In Fig. 2(a) we see the reconstructed signal using the proposed method and the cubic spline interpolation method. In this example the percentage of missing values was 75%. It is obvious that the proposed approach achieved better visual quality with respect to the reconstruction. When exists consecutive missing values the linear interpolation is not able to find the salient structure of the true signal. This effect can be observed in Fig. 2(b) which depticted a zoom in the reconstructed signals.



(a) Whole signal (b) Zoom

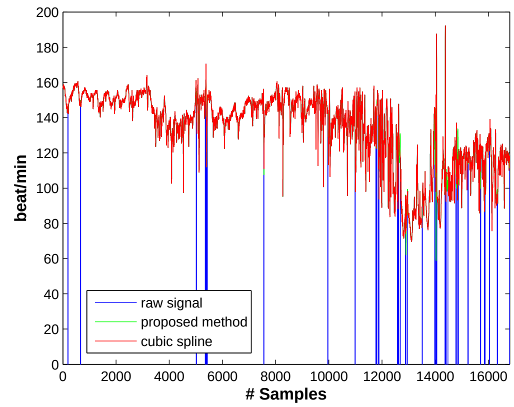
Figure 2. An Example of the reconstructed signal using the proposed method and the cubic spline interpolation.

3.3. Real data

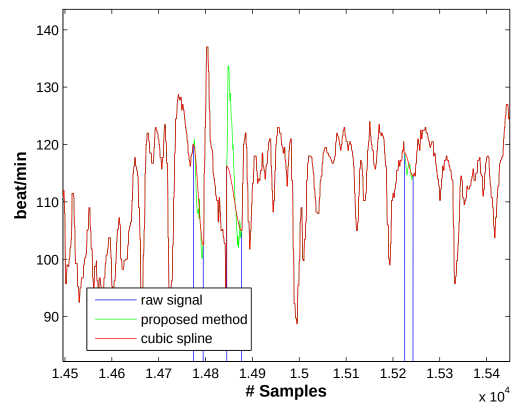
Finally, we applied the proposed algortihm in real situations and we provide a qualitative comparison with the cubic spline interpolation. A FHR time series was acquired from our database which presents missing values. The percentage of missing values was around 5% (see Fig. 3). In Fig. 3(a) we depict the raw signal and the reconstructed signals using the two approaches. While at the first look the two signal seems the same, a more carefull look reveals significant difference between the approaches. In Figs. 3(b) and 3(c) a zoom in the signals provides us with more information about the behaviour of the two approaches. More specifically, we can see that the proposed method reconstructs the missing values taking into account the neighborhood of the particular region. For example, in Fig. 3(b) the proposed method reveals a peak around position 14900, a finding that it is consistent with the structure of the signal in this region.

4. Conclusion

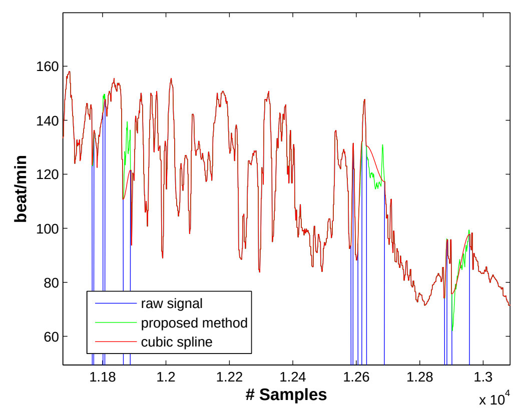
We have presented an algorithm to find the missing samples from FHR time series. The algorithm is of iterative nature and consists of two steps/stages. In the first step the reconstructed signal is computed using an adaptive dictionary. Then, using the reconstructed signal we calculated the new dictionary using the KSVD algorithm. The above two step are apllied iteratively until convergence is achieved. The initial results on small subset of database show the usefulness of our approach. Especially, when consecutive missing samples are present. In the future, we intend to study more intensively various aspects of the algorithms such as the length of the segments, the structure of missing samples and the learning of the dictionary. Also, it will be usefull to see how the proposed approach affects the subsequent analysis of FHR time series, i.e. feature extraction and classification.



(a) Whole signal



(b) Zoom - Region 1



(c) Zoom - Region 2

Figure 3. An Example of the reconstructed signal using the proposed method and the cubic spline interpolation on real situation.

Acknowledgment

This research work was supported by the joint research project "Intelligent System for Automatic

CardioTocoGraphic Data Analysis and Evaluation using State of the Art Computational Intelligence Techniques" by the programme "Greece-Czech Joint Research and Technology projects 2011-2013" of the General Secretariat of Research & Technology, Greek Ministry of Education and Religious Affairs, Culture and Sports, co-financed by Greece, National Strategic Reference Framework (NSRF) and the European Union, European Regional Development Fund. It is also support by the research program No. NT11124-6/2010 of the Ministry of Health Care of the Czech Republic.

References

[1] V. Doria, A. Papageorghiou, A. Gustafsson, A. Ugwumadu, K. Farrer, and S. Arulkumaran, "Review of the first 1502 cases of ecg-st waveform analysis during labour in a teaching hospital," *BJOG*, vol. 114, p. 1202 1207, 2007.

[2] M. Westerhuis, A. Kwee, A. van Ginkel, A. Drogtrop, W. Gyselaers, and G. Visser, "Limitations of st analysis in clinical practice: three cases of intrapartum metabolic acidosis," *BJOG*, vol. 114, p. 1194 1201, 2007.

[3] Z. Alfirevic, D. Devane, and G. M. L. Gyte, "Continuous cardiotocography (ctg) as a form of electronic fetal monitoring (efm) for fetal assessment during labour.," *Cochrane Database Syst Rev*, vol. 3, no. 3, p. CD006066, 2006.

[4] E. Blix, O. Sviggum, K. S. Koss, and P. Oian, "Inter-observer variation in assessment of 845 labour admission tests: comparison between midwives and obstetricians in the clinical setting and two experts.," *BJOG*, vol. 110, pp. 1–5, Jan 2003.

[5] J. Spilka, V. Chudacek, M. Koucky, L. Lhotska, M. Huptych, P. Janku, G. Georgoulas, and C. Stylios, "Using nonlinear features for fetal heart rate classification," *Biomedical Signal Processing and Control*, vol. 7, p. 350 357, 2012.

[6] B. Fulcher, A. Georgieva, C. Redman, and N. Jones, "Highly comparative fetal heart rate analysis," in *Engineering in Medicine and Biology Society (EMBC), 2012 Annual International Conference of the IEEE*, pp. 3135 –3138, 28 2012-sept. 1 2012.

[7] G. Georgoulas, C. Stylios, and P. Groumpos, "Feature extraction and classification of fetal heart rate using wavelet analysis and support vector machines," *International Journal on Artificial Intelligence Tools*, vol. 15, no. 03, pp. 411–432, 2006.

[8] V. Chudacek, J.Spilka, P. Janku, M. Koucky, L. Lhotska, and M. Huptych, "Automatic evaluation of intrapartum fetal heart rate recordings: a comprehensive analysis of useful features," *Physiological Measurement*, vol. 32, no. 8, pp. 1347–1360, 2011.

[9] J. Bernardes, C. Moura, J. M. de Sa, and L. Pereira-Leite, "The porto system for automated cardiotocographic signal analysis," *J. Perinat. Med.*, vol. 19, pp. 61–65, 1991.

[10] P. Stoica, J. Li, J. Ling, and Y. Cheng, "Missing data recovery via a nonparametric iterative adaptive approach," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pp. 3369 –3372, April 2009.

[11] A. Dempster, L. A., and R. D., "Maximum likelihood from incomplete data via the EM algorithm," *Journal of Royal Statistical Society, Series B*, vol. 39, pp. 1–38, 1977.

[12] M. Bertalmio, L., Vese, G. Sapiro, and S. Osher, "Simultaneous structure and texture image inpainting," *Image Processing, IEEE Transactions on*, vol. 12, pp. 882 – 889, Aug. 2003.

[13] S. Z. Li, *Markov Random Field Modeling in Image Analysis*. Springer Publishing Company, Incorporated, 2009.

[14] A. Tikhonov, "Solution of incorrectly formulated problems and the regularization method (english translation)," *Doklady Akademii Nauk SSSR*, vol. 151, pp. 501 –504, 1963.

[15] M. Aharon, M. Elad, and A. Bruckstein, "The k-svd: An algorithm for designing of overcomplete dictionaries for sparse representation," *IEEE Trans. On Signal Processing*, vol. 54, pp. 4311 – 4322, November 2006.

[16] D. Tzikas, A. Likas, and N. Galatsanos, "The variational approximation for bayesian inference," *Signal Processing Magazine, IEEE*, vol. 25, pp. 131 –146, November 2008.

[17] E. Candes and M. Wakin, "An introduction to compressive sampling," *Signal Processing Magazine, IEEE*, vol. 25, pp. 21 –30, March 2008.

[18] S. Boyd and L. Vanderberghe, *Convex Optimization*. Cambridge Univ. Press, Cambridge, England, 2004.

[19] V. Chudacek, J. Spilka, M. Bursa, P. Janku, L. Hruban, M. Huptych, and j. . P. L. Lhotska, title = Open access intrapartum CTG database: Stepping stone towards generalization of technical findings on CTG signals

[20] E. Sheiner, A. Hadar, M. Hallak, M. Katz, M. Mazor, and I. Shoham-Vardi, "Clinical significance of fetal heart rate tracings during the second stage of labor," *Obstet Gynecol*, vol. 97, pp. 747–752, May 2001.

[21] M. Signorini, G. Magenes, S. Cerutti, and D. Arduini, "Linear and nonlinear parameters for the analysis of fetal heart rate signal from cardiotocographic recordings," *Biomedical Engineering, IEEE Transactions on*, vol. 50, pp. 365 –374, March 2003.