Optic Cup Characterization through Sparse Representation and Dictionary Learning

Valery Naranjo*†, Carlos J. Saez*, Sandra Morales*†, Kjersti Engan‡ and Soledad Gómez§

*Instituto de Investigación e Innovación en Bioingeniería, I3B,
Universitat Politècnica de València, Camino de Vera s/n, 46022 Valencia, Spain
Email: {vnarajo,sanmomar}@i3b.upv.es

†Grupo Tecnologías de Informática Aplicadas a la Oftalmología, Unidad Conjunta UPV- FISABIO, Spain
‡Department of Electrical Engineering and Computer Science, University of Stavanger, N-4036 Stavanger, Norway
§Universidad Católica de Valencia, Valencia, Spain

Abstract—This paper describes how to construct a probability map using sparse representation and dictionary learning to indicate the probability of each optic disk pixel of belonging to the optic cup. This probability map will be used in the future as input to a method for automatically detecting glaucoma from color fundus images. The probability map was obtained constructing a model (using the Bayes classifier) which takes into account texture information, by means of sparse representation and RLS-DLA dictionary learning technique, and intensity information. Several experiments on a private database are presented in this work. The results are compared with the segmentation made by specialists, highlighting the promising performance of this technique in difficult cases where the optic cup is barely visible.

I. INTRODUCTION

Glaucoma is a progressive disease of the optic nerve caused by high intraocular pressure due to a bad drainage of the ocular fluid. Clinically, it causes a progressive and irreversible loss of the visual field that progresses to the complete loss of vision. It is currently the second leading cause of blindness worldwide and affects one in every hundred people under 50 and one in ten over 80 years [1]. In most cases, glaucoma is belatedly detected when visual field loss is irreversible. Currently, there is no cure for glaucoma damage and, therefore, early detection and prevention is the only way to prevent progression to total vision loss.

The optic nerve damage is manifested in a change in the appearance of the optic cup, so that the ratio between the optic cup and optic disk area, also known as cup-to-disk ratio (CDR), is related to the presence or absence of glaucoma. For healthy patients, CDR falls in the range of 0.3 to 0.5 and for glaucoma it is higher than 0.5. The sight of vision is completely lost at the CDR value of 0.8 [2]. An example of the difference between a fundus with and without glaucoma is illustrated in Fig. 1.

The state-of-the-art methods for cup detection are usually based only on intensity features using different segmentation methods such as level-sets methods [3], superpixel classification [4], cup segmentation using r-bends information or vessel geometry and Hough transforms [5]. These methods are based on intensity and does not take texture information into account. When the optic cup is barely visible, these methods usually produce segmentation errors. Moreover, this kind of
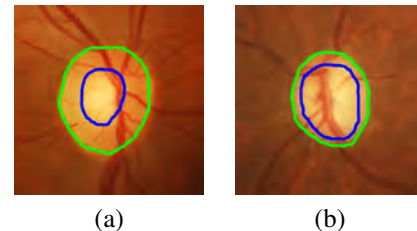

Fig. 1. Glaucoma effects on the cup-to-disk ratio: (a) Healthy fundus and (b) Glaucomatous fundus. Optic disk is marked in green and optic cup in blue.

algorithms produces a binary image which indicates if a pixel belongs to the cup or not. Our aim in this work is double: to include texture in the cup identification process and to define a probability for each pixel of being cup or no cup that can be used as a feature in a glaucoma detection algorithm. So, this paper focuses on optic cup characterization by using dictionary learning and sparse representation techniques. In particular, the image database was divided into two subsets: training and test. Training images were used to generate two dictionaries within the optic disk (cup and no-cup) through dictionary learning techniques. From these dictionaries, intensity and texture information were extracted. In the proposed approach, texture information was obtained by reconstructing the image through its sparse representation with the created dictionaries. Making use of texture and intensity information, a classification model was built. This model is used on test images to obtain a probability map that indicates the probability of belonging to the optic cup for each pixel in the optic disk.

The rest of the paper is organized as follows: Section II describes the material and methods used. Section III addresses the method proposed for optic cup characterization. Section IV shows the results obtained by the method. Finally, Section V closes the paper with the conclusions and future work lines.

II. MATERIAL AND METHODS

A. Material

The material of this work are images with the optic disk (OD) and the optic cup (OC) previously segmented by specialists composing the ground truth. The dataset was composed of 53 fundus images belonging to Hospital 12 de Octubre from

Madrid (Spain) [6]. 23 of them were normal (without known pathology) and the remaining 30 images were classified as glaucomatous.

This dataset was randomly divided into two subsets: training and test. The training set contains 70% of the images and the test set the remaining 30%, i.e. 37 images for training and 16 for testing.

B. Principal Component Analysis

The central idea of principal component analysis (PCA) is to reduce the dimensionality of a data set consisting of a number of interrelated variables, while retaining as much as possible of the variation present in the data set. This is achieved by transforming to a new set of variables, the principal components (PCs), which are uncorrelated, and ordered so that the first few retain most of the variations present in all of the original variables [7].

A three-channel image transformed to a principal component space creates three new channels in which the first (the most significant) contains the most structural contrast and information. The rank for each axis in the principal set represents the significance of that axis as defined by the variance in the data along that axis. Thus, the first principal axis is the one with the highest amount of scatter in the data and consequently the greatest amount of contrast and information, while the last principal axis represents the least amount of information such as noise and image artefacts [8]. In this case, the kth PC, z_k, is given by

$$z_k = \boldsymbol{\alpha}_k{}'\mathbf{f} = \alpha_{k_R}f_R + \alpha_{k_G}f_G + \alpha_{k_B}f_B, \quad (1)$$

where $\mathbf{f}(\mathbf{x}) = (f_R(\mathbf{x}), f_G(\mathbf{x}), f_B(\mathbf{x}))$ represents a RGB image, $\boldsymbol{\alpha}_k$ is a vector of constants, $'$ denotes transpose and $k \in \{1, 2, 3\}$. Specifically, principal-component axes ($\boldsymbol{\alpha}_k$) will be the eigenvectors of the covariance matrix.

C. Sparse Representation

Sparse representation is a signal model that suggests that a signal can be represented sparsely in a domain, usually represented by atoms collected as columns in a dictionary matrix. The dictionary D is a matrix $N \times K$, which contains K prototype signals of length N, also referred to as atoms. The model assumes that for any signal x, there exists a sparse linear combination well approximated of atoms from D. The approximation of x can be written as

$$x \approx Dw, \qquad \|w\|_0 \ll N \quad (2)$$

where w is a vector containing the coefficients. Most of the entries in w are zero and the operator $\|.\|_0$ counts the number of non-zero elements in the vector. Typically it is assumed that the dictionary is redundant in describing x.

So, given the dictionary D, the approximation $\hat{x}$ of signal x can be written as $\hat{x} = Dw$, and the representation error or residue can be written as $r = x - \hat{x} = x - Dw$. Most of the entries of w are zero, s is the number of non-zero coefficients, and s/N is the sparseness factor.

A common way to find w, i.e the sparse approximation problem, is solving the following equation:

$$w_{opt} = \underset{w}{\operatorname{argmin}}\{\|w\|_p + \gamma\|x - Dw\|_2^2\}, \quad p \in \{0, 1\}. \quad (3)$$

The problem with p = 0 (to minimize the number of non-zero coefficients) is NP-hard, but an approximate solution can be found by greedy methods. Alternatively the problem can be relaxed by setting p = 1 (to minimize the sum of absolute values) providing a convex problem that can be solved (LASSO) [9]. Both problem resolutions start with an all zero vector w, which is the solution when γ is close to zero. As γ factor increases the solution is getting more dense. In this work, SPAMS library, SPArse Modeling Software [10], was used to solve this problem.

D. Dictionary Learning

Dictionary Learning is often formulated as the problem of finding a dictionary such that the approximations of many vectors, the training set, are as good as possible given a sparseness criterion on the coefficients, i.e. allowing only a small number of non-zero coefficients for each approximation. Let X be a matrix containing K signals to be represented.

A common set up for the dictionary learning problem starts by obtaining a training set, a collection of training vectors, each of length N. The training vectors are usually K vectors collected from matrix X, resulting a training set of size $N \times K$. The aim of dictionary learning algorithms is to find both a dictionary D_{opt} of size $N \times L$ and a corresponding set of coefficients W_{opt} of size $K \times L$ such that the representation error is minimized and W_{opt} fulfils the imposed sparseness criterion. This optimized dictionary D_{opt} is composed by the L most representative atoms, being $L = 2N$.

The dictionary learning problem can be formulated as an optimization problem with respect to the coefficient matrix W and the dictionary D as:

$$\{D_{opt}, W_{opt}\} = \underset{D,W}{\operatorname{argmin}} \sum_{i=1}^{K} \|w_i\|_p + \gamma\|X - DW\|^2. \quad (4)$$

Different methods have been proposed to solve the previous optimization problem [11]–[14]. In this work, the Recursive Least Squares Dictionary Learning Algorithm (RLS-DLA) [15] was used. In this method, a single training vector x_i or a mini-batch (subset) of X is processed in each iteration solving the following equation:

$$D_i = B_i A_i^{-1}, \qquad being \quad \begin{cases} A_i = \lambda_i A_{i-1} + W_i W_i^T \\ B_i = \lambda_i B_{i-1} + X_i W_i^T \end{cases} \quad (5)$$

where $A_1 = W_1 W_1^T$ and $B_1 = X_1 W_1^T$. The current dictionary D_{i-1} is used to find the corresponding coefficients W_i.

III. PROPOSED METHOD

A. Pre-processing

Since we are only interested in the optic disk area, the rest of the image is removed remaining only that area of interest, using the manual segmentation of the optic disk provided by the specialists. The goal of this step is to reduce the computational cost.

After image cropping, PCA is applied to transform the input image to grey scale. This technique combines the most significant information of the three components RGB in a single image so that the structures of the retina are better appreciated. As seen in Subsection II-B, the first PC is the most significant so, z_1, defined as equation (1), is chosen as the input image of the method presented in this paper.

B. Dictionary generation

First, two dictionaries D_c and D_{nc} are generated, where D_c is a dictionary trained for representing the optic cup area and D_{nc} is a dictionary trained for representing the optic disk area without considering the optic cup. Both dictionaries are generated from the information provided by the training images on which these structures were previously segmented by a specialist (Fig. 2).
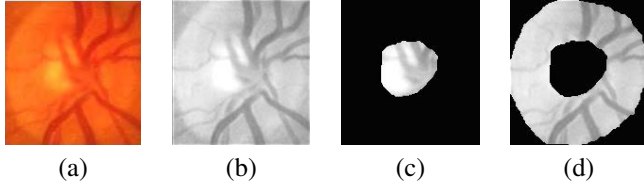

Fig. 2. Significant image areas: (a) Region of interest of the original image, (b) Region of interest of the first PC (z_1), (c) Optic cup area and (d) Optic disk area (no cup). (c) and (d) images were obtained using the cup segmentation ground truth.

Image blocks of $\sqrt{N} \times \sqrt{N}$ pixels are selected in both areas and reshaped into a column of length N. Image blocks are selected through a sliding window using maximum overlapping by sliding only one pixel position at the time. This process is shown in Fig. 3(a).
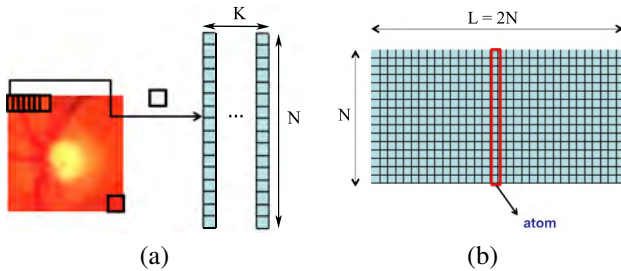

Fig. 3. Dictionary generation process: (a) Original training set and (b) Learned dictionary D_{opt}.

All the information extracted from the training images is stored in a training set X. Note that the boundary pixels between cup and no-cup areas are excluded from the training set. The dictionaries are initialized by using L=2N normalized example vectors from the training set (randomly chosen). All training vectors are used in the learning process. Dictionary learning technique is applied to use only relevant and necessary information. In particular, RLS-DLA is used (see Subsection II-D). λ is initialized to 0.995 and slowly increased to 1. The result of this algorithm is a learned dictionary as it is represented in Fig. 3(b).

C. Feature extraction

After dictionary generation, the following features are extracted:

1) Texture extraction: Texture extraction is performed by computing the residue of the image after sparse representation using the learned dictionaries. For each pixel (i, j) in the input test image I_m, a test vector, $x_{i,j}$ is formed by the column stacking of the $\sqrt{(N)} \times \sqrt{(n)}$ neighborhood around the pixel (i, j). A sparse representation of $x_{i,j}$ is found using D_c in equation 3, giving $\hat{x}_{i,j}^c$ and the corresponding residual $r_c = ||x_{i,k} - \hat{x}_{i,j}^c||$ placed in the right position in a residual image: $R_c(i, j) = r_c$. The same is done using D_{nc}, giving a different residual image, R_{nc}.

We expect the residual value at a specific pixel position inside the optic cup to be lower in the R_c image, and for a pixel position outside the optic cup we expect the residual value to be lower in the R_{nc} image. By nature, texture is a regional property that should not change pixel by pixel. This suggests a smoothing of the residue images. So, previously to the texture feature image computation, each residue image is filtered with a low-pass Gaussian filter of size 9×9 and variance 5 [16]. Finally, R_c and R_{nc} are combined to calculate the texture feature image, F_t, as:

$$F_t = \frac{R_{nc}}{R_{nc} + R_c}. \tag{6}$$

F_t is an image with the same size as the input image, providing a textural feature for each pixel. High values of F_t would suggest that the corresponding pixel belongs to the optic cup and low values suggest that the pixel is outside the optic cup.

2) Intensity extraction: The intensity feature corresponds with the intensity level of each image pixel calculated within a neighborhood. Thus, the original grey-scale image is filtered with a low-pass filter of size 3×3 giving an intensity feature value for each pixel of the input image. These values are normalized by subtracting the mean and dividing by the standard deviation giving place to the final intensity feature F_i.

D. Classification model

The two feature images extracted form the training images (F_t and F_i) are used to generate a model from Bayes classifier. As Bayes classifier uses supervised learning, optic cup masks are also needed to indicate the cup area.

Once the model is created, F_t and F_i feature images must be computed for each test image giving a feature vector for each pixel. These feature vectors are fed through the classifier, and

the output is interpreted as a *a posteriori probability* rather than a binary class label. The output probabilities from the classifier are reorganized as an image, providing a probability map for each image [17]. This map shows the probability of each pixel of belonging to the optic cup.

IV. RESULTS

The performance of the proposed procedure for optic cup characterization was evaluated on the private image database described in Subsection II-A. Different experiments were conducted to find out the best configuration of the proposed method. The first experiment tests different block sizes (3×3, 5×5 and 8×8) for the generation of D_c and D_{nc} dictionaries. Fig. 4 shows the obtained results. Red color corresponds to a high probability for that pixel to belong to the optic cup, and dark blue to a low probability. Size 5×5 obtained the best results determined by visual inspection and this size is a trade-off between the cup area and the definition of its border. The second experiment tests two classification models, the first of them considers only intensity information (F_i) and the second one combines intensity and texture information ($F_i + F_t$). In Fig. 5 probability maps obtained for both classification models are depicted. The probabilities for pixels belonging to the optic cup are higher in the probability maps obtained using the multivariate model.
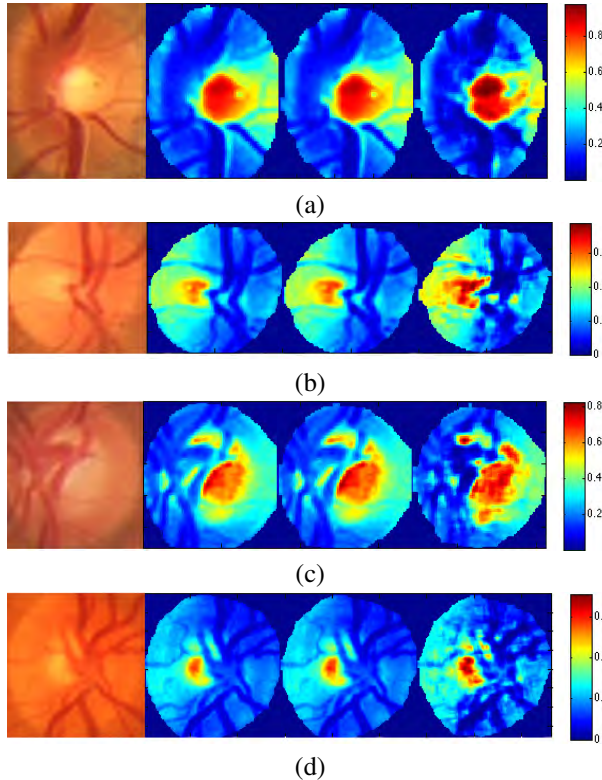

(a)

(b)

(c)

(d)

Fig. 4. Probability maps with different block sizes on 4 images (a-d). First column: Original image. Third column: 3×3 blocks. Third column: 5×5 blocks. Forth column: 8×8 blocks.

Finally, in Fig. 6, several resulting probability maps, for multivariate model and 5×5 block size, of the optic cup can
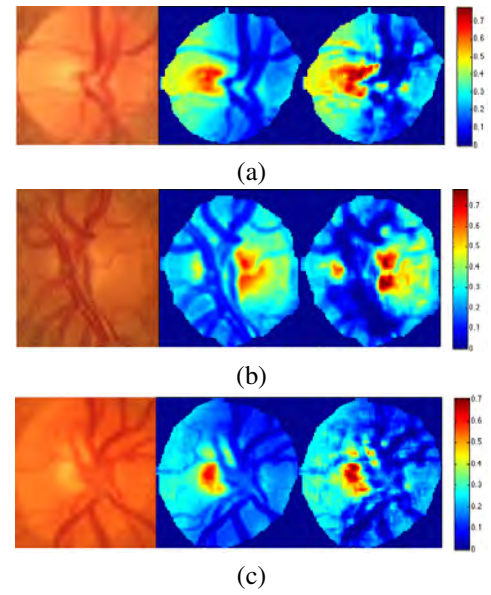

(a)

(b)

(c)

Fig. 5. Probability maps with different models on 3 images (a-c). First column: Original image. Second column: F_i model. Third column: $F_i + F_t$ model.

be observed and compared with the manual cup segmentation provided by an expert (ground truth). In addition, Fig. 7 demonstrates how the proposed method is able to identify the optic cup in healthy and glaucomatous images despite the fact that in some of the healthy images the optic cup is barely visible on the original images.
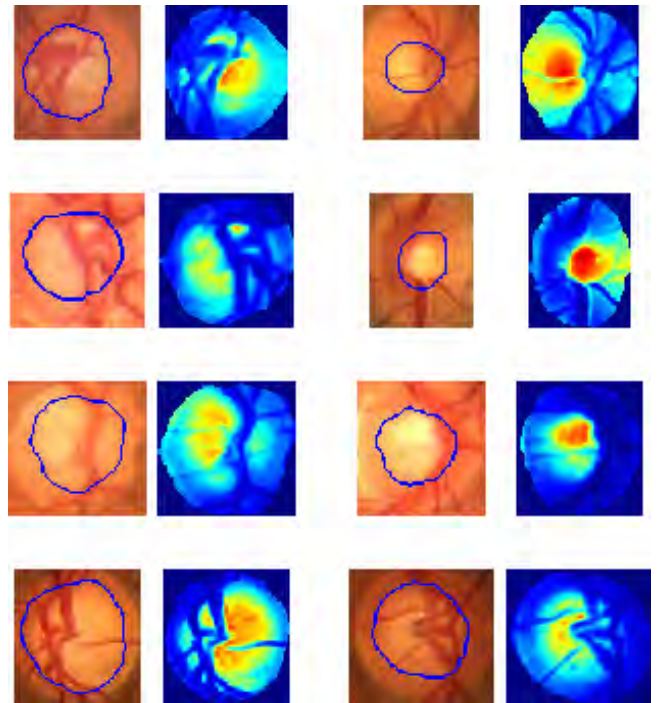

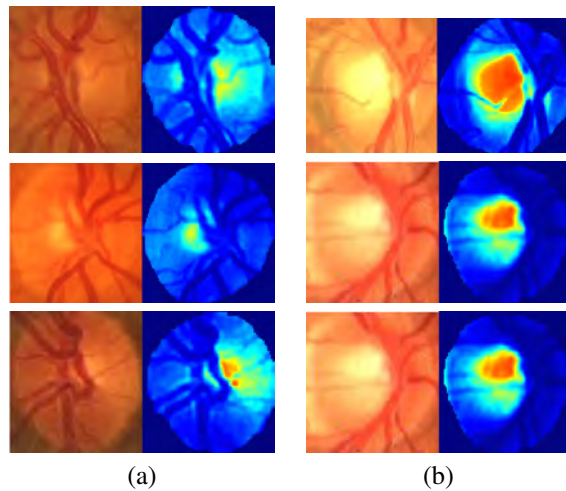Fig. 6. Manual segmentation and probability maps of the optic cup.

Fig. 7. Optic cup probability maps on different images: (a) Healthy images and (b) Glaucomatous images.

V. Conclusion

In this paper, a new approach for optic cup characterization was presented. It is based on using sparse representation, dictionary learning and machine learning to determine the probability that a pixel within the optic disk belongs to the optic cup. Distinguishing between optic disk and optic cup pixels is required for example for computing the cup-to-disk ratio used for glaucoma diagnosis.

The main characteristic of this work is the combination of intensity and texture information for optic cup identification which allows to enhance the cup although in the RGB image was not too visible.

As future work, the probability map of the optic cup obtained with the proposed method will be used as a feature in a glaucoma detection algorithm. Moreover, other dictionary learning algorithms will be used to compare all of them and choose the most efficient.

Acknowledgment

References

[1] Y. H. Kwon, J. H. Fingert, M. H. Kuehn, and W.L.M. Alward, "Primary open-angle glaucoma," *New England Journal of Medicine*, vol. 360, no. 11, pp. 1113–1124, 2009.

[2] M. K. Nath and S. Dandapat, "Techniques of glaucoma detection from color fundus images: A review," *International Journal of Image, Graphics and Signal Processing*, vol. 4, no. 9, pp. 44–51, 2012.

[3] D. W. K. Wong, J. Liu, J.H. Lim, X. Jia, F. Yin, H. Li, and T. Y. Wong, "Level-set based automatic cup-to-disc ratio determination using retinal fundus images in argali," in *30th Annual International IEEE EMBS Conference*, Vancouver, British Columbia, Canada, 2008.

[4] Jun Cheng, Jiang Liu, Yanwu Xu, Fengshou Yin, Damon Wing Kee Wong, Ngan-Meng Tan, Dacheng Tao, Ching-Yu Cheng, Tin Aung, and Tien Yin Wong, "Superpixel classification based optic disc and optic cup segmentation for glaucoma screening," *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 6, no. 32, 2013.

[5] Gopal Datt Joshi, Jayanthi Sivaswamy, and S. R. Krishnadas, "Optic disk and cup segmentation from monocular color retinal images for glaucoma assessment," *IEEE TRANSACTIONS ON MEDICAL IMAGING*, vol. 30, no. 6, 2011.

[6] R. Román Morán, R. Barea Navarro, L. Boquete Vázquez, E. López Guillén, J. Campos Pavón, L. de Pablo Gómez de Liaño, D. Escot Bocanegra, L. de Santiago, and M. Ortiz, "Color analysis in retinography: Glaucoma image detection," in *XIII Mediterranean Conference on Medical and Biological Engineering and Computing 2013*, vol. 41 of *IFMBE Proceedings*, pp. 325–329. 2014.

[7] I. T. Jolliffe, *Principal Component Analysis*, Springer, second edition, 2002.

[8] J. C. Russ, *Image Processing Handbook*, CRC Press, Inc., 5th edition, 2007.

[9] Robert Tibshirani, "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 267–288, 1996.

[10] J. Mairal, F. Bach, J. Ponce, G. Sapiro, R. Jenatton, and G. Obozinski, "Spams v2.5 (sparse modeling software)," http://spams-devel.gforge.inria.fr/index.html, 2014.

[11] K. Engan, S.O. Aase, and J. Hakon Husoy, "Method of optimal directions for frame design," in *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, 1999, vol. 5, pp. 2443–2446 vol.5.

[12] M. Aharon, M. Elad, and A. Bruckstein, "k -svd: An algorithm for designing overcomplete dictionaries for sparse representation," *Signal Processing, IEEE Transactions on*, vol. 54, no. 11, pp. 4311–4322, 2006.

[13] K. Engan, K. Skretting, and J. Håkon Husoy, "Family of iterative ls-based dictionary learning algorithms, ils-dla, for sparse signal representation," *Digital Signal Processing*, vol. 17, no. 1, pp. 32 – 49, 2007.

[14] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proceedings of the 26th Annual International Conference on Machine Learning*, 2009, ICML '09, pp. 689–696.

[15] K. Skretting and K. Engan, "Recursive least squares dictionary learning algorithm," *Signal Processing, IEEE Transactions on*, vol. 58, pp. 2121–2130, 2010.

[16] L. P. Kotu, K. Engan, K. Skretting, S. Orn, L. Woie, and T. Eftestol, "Segmentation of scarred myocardium in cardiac magnetic resonance images," *ISRN Biomedical Imaging*, vol. 2013, 2013, Article ID 504594.

[17] L. P. Kotu, K. Engan, K. Skretting, F. Maloy, S. Orn, L. Woie, and T. Eftestol, "Probability mapping of scarred myocardium using texture and intensity features in cmr images," *BioMedical Engineering OnLine*, vol. 12, no. 91, 2013, Article ID 504594.