LBP FEATURES FOR BREAST CANCER DETECTION

Pavel Král[1,2], *Ladislav Lenc*[1,2]

[1]Dept. of Computer Science & Engineering
Faculty of Applied Sciences
University of West Bohemia
Plzeň, Czech Republic

[2]New Technologies for the Information Society
Faculty of Applied Sciences
University of West Bohemia
Plzeň, Czech Republic

ABSTRACT

Cancer is nowadays considered as one of the most dangerous diseases in the world. Especially, breast cancer represents for women the second most common type of cancer and is a main cause of cancer dead. This paper presents a novel method for breast cancer detection from mammographic images based on Local Binary Patterns (LBP). This approach successfully uses LBP based features with a classifier and thresholding. The proposed method is evaluated on a set composed of images extracted from MIAS and DDSM databases. We have experimentally shown that the proposed method is efficient and effective because the achieved accuracy is about 84%.

***Index Terms*—** Breast, Cancer Detection, Computer-Aided Detection, CAD, Local Binary Patterns, LBP, Mammography, SVM

1. INTRODUCTION

Cancer belongs to the today's most dangerous health issues in the world and its incidence is increasing. Particularly, breast cancer is for women the main cause of cancer death and the second most common type of cancer [1]. Unfortunately, there is not possible to prevent this disease, because its cause is not known. However, early detection is very important to manage and heal it. Mammography represents a very effective method for early detection of breast cancer signs [2]. It may identify several abnormalities such as masses, calcifications, asymmetries, etc. Especially, Computer-Aided Detection (CAD) systems play a pivotal role for this task and they thus can save the life of many women.

One of the most popular methods for feature extraction are Local Binary Patterns (LBP). It was originally proposed for texture classification in [3]. Lately, it is frequently used also for other tasks such as facial expression recognition, face recognition, fingerprint identification or several medical applications [4, 5].

Because LBP operator is very efficient for many image representation tasks, moreover we have good experiences in the face recognition field [6], we would like to use it for analysis of mammographic images. Therefore, we propose a novel breast cancer detection method based on LBP features. This approach computes local LBP histograms, which are further used to create feature vectors. Then a classifier is used together with thresholding to identify the cancerous tissues.

It is worth of noting that, to the best of our knowledge, only few papers [7, 8] use LBP features for breast cancer detection. However, the LBP is used significantly differently compared to the proposed approach:

- we use several separated local histograms with thresholding instead of one large vector;
- we use uniform LBP operator instead of the classical LBP.

The results of this work should be integrated to a novel computer-aided detection system to help the radiologist to discover the breast cancer in the very early stages.

The rest of the paper is organized as follows. Section 2 describes related work and is composed of the two parts. The first one summarizes the main medical analysis methods with a particular focus on mammographic screening, while the second one shows the most important methods based on LBP. Section 3 details the proposed breast cancer detection method. Section 4 first describes the database used for evaluation and then presents the results of experiments realized on this data. The last section discusses the results and proposes some future research directions.

2. RELATED WORK

The first part of this section summarizes the state-of-the-art methods in the cancer detection field. The second one presents a very short overview of local descriptors that are beneficial for this work.

2.1. Breast Cancer Detection Methods

Mammography has already a long history dated back to the late 1950s. Regardless of some controversial opinions about this procedure [9], it remains the most used method for breast cancer detection. Evaluation of mammographic images is performed by a radiologist, but since the 1990s there are efforts to find a reliable computerized method that helps the radiologists to decide automatically whether an image is malignant or not.

One of the seminal approaches was presented in [10]. This paper deals with detection of circumscribed masses in the mammographic images. Median filtering is employed as a preprocessing and a template matching approach is used for detection of suspicious regions. Finally, a correlation of the regions and the template is thresholded to determine the cancerous regions.

A sophisticated method for image preprocessing was proposed in [11]. The method is called Adaptive Density-Weighted Contrast Enhancement (DWCE). This algorithm improves visibility of important image structures so that an edge detection algorithm can be utilized for detecting object boundaries. Laplacian-Gaussian detector is used for edge detection.

Another method for mammogram enhancement called NonLinear Unsharp Masking (NLUM) was proposed in [12]. The method first applies nonlinear filtering and combines the normalized filtered image with the original one. It is a configurable algorithm that allows the user to use different filters. In addition, the parameters of the method can be set manually or adjusted automatically.

The preprocessing of original images is followed by the Regions of Interest (RoI) detection. One of methods used for this task is based on texture features, K-means clustering and Support Vector Machines (SVM) [13]. The texture features are created using a co-occurrence matrix representing spatial dependencies of gray levels. Additionally, several shape features are used together with the texture ones. K-means algorithm is used to cluster the regions into several (4 to 6) classes. Finally, the SVM classifier differentiates between mass and non-mass regions.

Another approach [14] uses Principal Component Analysis (PCA) for classification of RoI images. The experiments are performed on two subsets of MIAS and DDSM databases. Reported Az values are 0.92 and 0.83 for MIAS and DDSM datasets respectively.

Work of Oliver at al. [7] deals with the false positive reduction. The approach is based on LBP descriptors and SVM classifier is used to distinguish between true and false positives. This method is evaluated on manually created RoI subset of the DDSM database.

In [8], the authors propose a novel LBP based operator called "completed" LBP (CLBP) which considers the sign, magnitude and center gray level values in the images. Finally, extracted features are analyzed by a SVM classifier for identifying normal and cancerous images. Although the authors claim that this operator is robust for breast cancer detection, it is not clear how the CLBP features are used with SVM classifier and the detection score is not reported in the paper.

Other methods for breast cancer detection can be found in [15] or [16].

2.2. Local Binary Patterns and its Variants

Methods based on LBP usually use histograms of LBP values computed in rectangular regions [4]. The concatenated histograms constitute object representation vectors. These vectors are then compared using some distance metrics as for instance histogram intersection or Chi square distance.

An interesting LBP extension which is proposed by Ojala in [3] are so called uniform Local Binary Patterns. A pattern is called uniform if it contains at most two transitions from 0 to 1 or from 1 to 0. It was proved that approximately 90% of the patterns in facial images are uniform. The histogram can then be shortened from 256 intervals (bins) to 59, where the 59th bin is reserved for the non-uniform patterns. This LBP adaptation will be also integrated in this work.

Li et al. propose Dynamic Threshold Local Binary Pattern (DTLBP) [17]. They use the mean value of the neighbouring pixels and also the maximum contrast between the neighbouring points. Another LBP extension are Local Ternary Patterns (LTP) [18] which uses three states to capture the differences between the central pixel and the neighbouring ones. The authors claim that both DTLBP and LTP are less sensitive to the noise than the original LBP method.

Local Derivative Patterns (LDP) are proposed in [19]. The difference from the original LBP is that it uses features of higher order. It thus can represent more information than the original approach.

Davarzani et al. propose in [20] a weighted and adaptive LBP-based texture descriptor. This approach successfully handles some issues in the previously proposed LBP-based approaches such as invariance to scaling, rotation, viewpoint variations and non-rigid deformations.

For additional information about the LBP based methods, please see the surveys [21, 22].

3. LBP FEATURES FOR BREAST CANCER DETECTION

The proposed cancer detection method is composed of the four main steps:

1. image preprocessing
2. local binary patterns computation
3. image representation creation
4. cancer detection

3.1. Image Preprocesing

The whole mammographic image is composed of the breast itself and some background. Therefore, the first step consists in background removal in order to separate the breast image only.

Due to the image character where background is almost homogeneous (usually black color) and breast object differs significantly, we can do a simple thresholding technique for

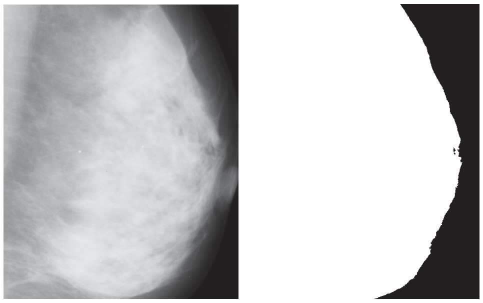

Fig. 1. Original mammographic image (left) and the result of the Otsu method (right)

this purpose. We use popular Otsu method [23] due to its very good results in many image processing areas. The result of this step is shown in Figure 1.

3.2. Local Binary Patterns Computation

Local binary patterns are used to compute our feature vectors. Therefore, this algorithm is shortly described next.

The original LBP operator uses a 3×3 square neighbourhood centred at the given pixel. The algorithm assigns either 0 or 1 value to the 8 neighbouring pixels by Equation 1.

$$N = \begin{cases} 0 & \text{if } g_N < g_C \\ 1 & \text{if } g_N \geq g_C \end{cases} \tag{1}$$

where N is the binary value assigned to the neighbouring pixel, g_N denotes the gray-level value of the neighbouring pixel and g_C is the gray-level value of the central pixel. The resulting values are then concatenated into an 8 bit binary number. Its decimal representation is used for further computation.

3.3. Image Representation Creation

We compute LBP values in all points of the breast image (background removed by Otsu). The image is then divided into a set of square cells lying on a regular grid. Feature vectors are computed for each cell as a histogram of the LBP values. We use previously described uniform patterns, therefore the feature vector size is 59. Every cell is then represented by one feature vector.

3.4. Cancer Detection

We assume that by setting an optimal LBP cell size it is possible to cover by this cell almost all breast tumors. We thus consider a binary classifier which decides whether one breast region is healthy or cancerous. These regions correspond to the optimal LBP cells which size will be set experimentally. The input of the classifier is the uniform LBP histogram. The classifier then assigns values for all square cells of the whole image. The final decision is based on the thresholding: an image with higher number of positive regions than a given threshold is classified as cancerous. Because of very good accuracy in many computer vision tasks which also includes breast cancer detection [24], support vector machines are used as a classifier.

4. EVALUATION

4.1. Corpora

4.1.1. MiniMammographic (MIAS) Database

This database [25] was created by Mammographic Image Analysis Society (MIAS) and contains 322 digitized breast images in the resolution 1024x1024. It includes radiologist's "truth"-markings on the locations of any abnormalities that may be present.

4.1.2. Digital Database for Screening Mammography (DDMS)

DDSM database [26] is another collection of mammographic images, which contains approximately 2,500 studies. Each study includes two images of each breast with associated patient information (e.g. age at time of study, ACR breast density rating, etc.). Images with suspicious areas have associated pixel-level "ground truth" information about the locations and types of suspicious regions.

4.1.3. Evaluation Set

We have created from these two databases another set of patch images containing 17,639 samples which are used to evaluate the proposed method. This set was divided into testing and training parts. Testing part is composed of 1,000 normal and 1,000 cancerous images, while all remaining images are used for training of the SVM classifier.

4.2. Experimental Set-up

We used OpenCV[1] toolkit for implementation of our experiments. As mentioned previously, we chose LBP algorithm to create our feature vectors. For cancer detection, we used OpenCV implementation of the SVM classifier with polynomial kernel (parameter $\gamma = 0.5$).

4.3. Experiments

The size of the LBP cell is one key parameter of the proposed method. This value must be set correctly to discover the majority of the tumor types. In the first experiment, we would like thus to identify its optimal value. The results of this experiment are depicted in Figure 2. This figure clearly shows that this parameter significantly influences the performance of

[1] http://opencv.org/

the whole algorithm. Based on these results, we use next the LBP cell size 32.

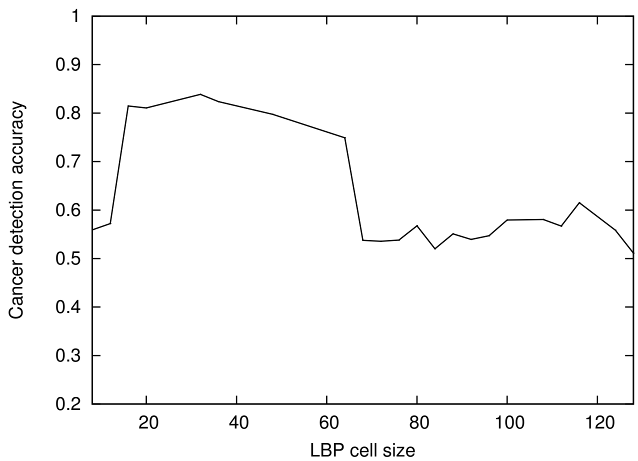

Fig. 2. Cancer detection accuracy depending on the LBP cell size

Our classification algorithm uses a threshold for cancer detection. Therefore, in the second experiment, we would like to identify its optimal value. Figure 3 shows cancer detection accuracy, precision, recall and f-measure depending on this acceptance threshold. The best threshold value is set experimentally to 0.5 where the corresponding accuracy and f-measure values are maximal, about 84%. However, it is possible to set higher value to obtain better recall which is important for radiologists.

The last experiment shows the Receiver Operating Characteristic (ROC) [27] of the proposed approach. This curve reports the true positive rate against the false positive rate with various acceptance threshold. The results of this experiment shows Figure 4. This figure confirms that the proposed approach is suitable for our task in order to identify cancerous images.

It is worth of noting that the previously reported approaches usually use different data-sets and different experimental set-ups. Therefore, it is generally not possible to compare the previous works between themselves and it is also not possible to compare the reported results of the proposed method with the previous work.

5. CONCLUSIONS AND FUTURE WORK

This paper proposes a novel breast cancer detection method which uses LBP features for breast representation. The proposed method was evaluated on a set created from MIAS and DDMS databases. We have showed that the proposed method is efficient and effective because the achieved accuracy is close to 84%.

In this work, we used the LBP operator for image classification into two classes. Therefore, our first perspective is

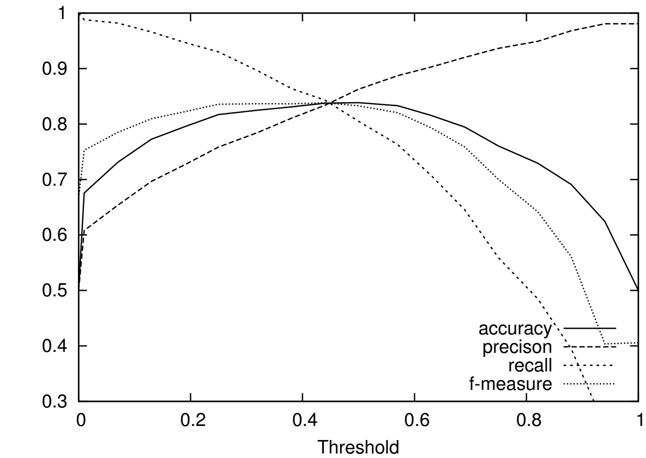

Fig. 3. Cancer detection accuracy, precision, recall and f-measure depending on the acceptance threshold

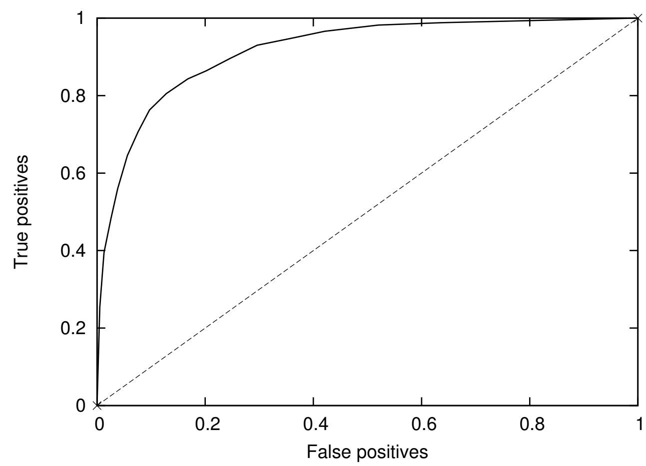

Fig. 4. ROC curve of the proposed method

to use a more sophisticated descriptor as for instance LDP or POEM together with our detection algorithm. Then, we would like to adapt the proposed algorithm to differentiate the particular cancer types.

6. ACKNOWLEDGEMENTS

This work has been partly supported by the project LO1506 of the Czech Ministry of Education, Youth and Sports. We would like to thank also to Mr. Michal Medek for some implementation work.

7. REFERENCES

[1] Jinshan Tang, Rangaraj M Rangayyan, Jun Xu, Issam El Naqa, and Yongyi Yang, "Computer-aided detection and diagnosis of breast cancer with mammography: recent advances," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 13, no. 2, pp. 236–251, 2009.

[2] Mehul P Sampat, Mia K Markey, Alan C Bovik, et al., "Computer-aided detection and diagnosis in mammography," *Handbook of image and video processing*, vol. 2, no. 1, pp. 1195–1217, 2005.

[3] Timo Ojala, Matti Pietikäinen, and David Harwood, "A comparative study of texture measures with classification based on featured distributions," *Pattern recognition*, vol. 29, no. 1, pp. 51–59, 1996.

[4] Timo Ahonen, Abdenour Hadid, and Matti Pietikinen, "Face recognition with local binary patterns," in *Computer Vision - ECCV 2004*, vol. 3021 of *Lecture Notes in Computer Science*, pp. 469–481. Springer Berlin Heidelberg, 2004.

[5] Loris Nanni, Alessandra Lumini, and Sheryl Brahnam, "Local binary patterns variants as texture descriptors for medical image analysis," *Artificial intelligence in medicine*, vol. 49, no. 2, pp. 117–125, 2010.

[6] Ladislav Lenc and Pavel Král, "Automatically detected feature positions for LBP based face recognition," in *Artificial Intelligence Applications and Innovations*, pp. 246–255. Springer, 2014.

[7] Arnau Oliver, Xavier Lladó, Jordi Freixenet, and Joan Martí, "False positive reduction in mammographic mass detection using local binary patterns," *MICCAI*, pp. 286–293, 2007.

[8] S Naresh and S Vani Kumari, "Breast cancer detection using local binary patterns," *International Journal of Computer Applications*, vol. 123, no. 16, 2015.

[9] Nikola Biller-Andorno and Peter Jüni, "Abolishing mammography screening programs? a view from the swiss medical board," *New England Journal of Medicine*, vol. 370, no. 21, pp. 1965–1967, 2014.

[10] Shuk-Mei Lai, Xiaobo Li, and WF Biscof, "On techniques for detecting circumscribed masses in mammograms," *Medical Imaging, IEEE Transactions on*, vol. 8, no. 4, pp. 377–386, 1989.

[11] Nicholas Petrick, Heang-Ping Chan, Berkman Sahiner, and Datong Wei, "An adaptive density-weighted contrast enhancement filter for mammographic breast mass detection," *Medical Imaging, IEEE Transactions on*, vol. 15, no. 1, pp. 59–67, 1996.

[12] Karen Panetta, Yicong Zhou, Sos Agaian, and Hongwei Jia, "Nonlinear unsharp masking for mammogram enhancement," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 15, no. 6, pp. 918–928, 2011.

[13] L. de Oliveira Martins et al., "Detection of masses in digital mammograms using k-means and support vector machine," *ELCVIA: electronic letters on computer vision and image analysis*, vol. 8, no. 2, pp. 39–50, 2009.

[14] Arnau Oliver, Joan Marti, Robert Marti, Anna Bosch, and Jordi Freixenet, "A new approach to the classification of mammographic masses and normal breast tissue," in *Pattern Recognition, 2006. ICPR 2006. 18th International Conference on*. IEEE, 2006, vol. 4, pp. 707–710.

[15] Rangaraj M Rangayyan, Fabio J Ayres, and JE Leo Desautels, "A review of computer-aided diagnosis of breast cancer: Toward the detection of subtle signs," *Journal of the Franklin Institute*, vol. 344, no. 3, pp. 312–348, 2007.

[16] Arnau Oliver, Jordi Freixenet, Joan Marti, Elsa Perez, Josep Pont, Erika RE Denton, and Reyer Zwiggelaar, "A review of automatic mass detection and segmentation in mammographic images," *Medical Image Analysis*, vol. 14, no. 2, pp. 87–110, 2010.

[17] Weisheng Li, Peng Fu, and Lifang Zhou, "Face recognition method based on dynamic threshold local binary pattern," in *Proceedings of the 4th International Conference on Internet Multimedia Computing and Service*. ACM, 2012, pp. 20–24.

[18] Xiaoyang Tan and Bill Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *Image Processing, IEEE Transactions on*, vol. 19, no. 6, pp. 1635–1650, 2010.

[19] Baochang Zhang, Yongsheng Gao, Sanqiang Zhao, and Jianzhuang Liu, "Local derivative pattern versus local binary pattern: face recognition with high-order local pattern descriptor," *Image Processing, IEEE Transactions on*, vol. 19, no. 2, pp. 533–544, 2010.

[20] Reza Davarzani and Saeed Mozaffari, "Robust image description with weighted and adaptive local binary pattern features," in *ICPR 2014*. IEEE, 2014, pp. 1097–1102.

[21] Di Huang, Caifeng Shan, Mohsen Ardabilian, Yunhong Wang, and Liming Chen, "Local binary patterns and its application to facial image analysis: a survey," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on*, vol. 41, no. 6, pp. 765–781, 2011.

[22] Loris Nanni, Alessandra Lumini, and Sheryl Brahnam, "Survey on lbp based texture descriptors for image classification," *Expert Systems with Applications*, vol. 39, no. 3, pp. 3634–3641, 2012.

[23] Nobuyuki Otsu, "A threshold selection method from gray-level histograms," *Automatica*, vol. 11, no. 285-296, pp. 23–27, 1975.

[24] P. Grgel, A. Sertbas, and O.N. Uan, "Computer-aided classification of breast masses in mammogram images based on spherical wavelet transform and support vector machines," *Expert Systems*, vol. 32, no. 1, pp. 155–164, 2015.

[25] John Suckling, J Parker, DR Dance, S Astley, I Hutt, C Boggis, I Ricketts, E Stamatakis, N Cerneaz, S Kok, et al., "The mammographic image analysis society digital mammogram database," in *Exerpta Medica. International Congress Series*, 1994, vol. 1069, pp. 375–378.

[26] Michael Heath, Kevin Bowyer, Daniel Kopans, Richard Moore, and W Philip Kegelmeyer, "The digital database for screening mammography," in *Proceedings of the 5th international workshop on digital mammography*. Citeseer, 2000, pp. 212–218.

[27] Christopher D Brown and Herbert T Davis, "Receiver operating characteristics curves and related decision measures: A tutorial," *Chemometrics and Intelligent Laboratory Systems*, vol. 80, no. 1, pp. 24–38, 2006.