A ROBUST FEATURE SELECTION APPROACH USING LOW RANK MATRICES FOR BREAST TUMORS IN ULTRASONIC IMAGES

Tao Wan[1*], *Renjie Liao*[2], *Zengchang Qin*[1,2,]

[1]Robotics Institute, Carnegie Mellon University, Pittsburgh, USA
[2]Intelligent Computing and Machine Learning Lab,
School of ASEE, Beihang University, Beijing, 100191, China

ABSTRACT

The primary use of breast ultrasound today is to help diagnose breast abnormalities detected by a physician during a physical exam and to characterize different types of breast conditions, including both benign and malignant lesions. For this purpose, a large number of features are computed to determine the nature of a breast abnormality. This paper aims to focus on the feature selection problem for classifying benign and malignant breast tumors to assist the clinical diagnosis. We formulate the problem of choosing discriminative features as a decomposition of the computerized feature matrix into a low-rank principal matrix and a sparse error matrix. The low-rank principal matrix contains the best distinctive features for determining the benign and malignant cases whereas the sparse error matrix has the features with a less identification capability. By identifying and selecting essential features, the low-rank matrix based feature selection method can improve the classification outcomes.

Index Terms— feature selection, robust principal component analysis, low-rank matrices, breast sonography.

1. INTRODUCTION AND MOTIVATION

Breast ultrasound (US) is frequently used to evaluate breast abnormalities that are found during a physician performed clinical breast exam. It can capture several different types of breast tumor conditions, including both benign (non-cancerous) and malignant (cancerous) lesions. For example, benign tumors have a relevantly uniform growth, usually producing images with round, smooth and well-defined boundaries. On the other hand, malignant tumors tend to generate an irregular pattern of impedance discontinuities, which will be represented as irregular, spiculate or ill-defined boundaries [1]. With the aim of distinguishing the nature of breast lesions, a large number of features, including sonographic and textural features, are computed and quantify the characteristics of tumor contours and textures. However, the issue for effectively choosing the essential features still remains a challenging problem due to the large variance of normal/abnormal lesion differences and the intrinsic limitations of the imaging process. Therefore, it is desirable to select the most discriminative features that yield better quantifications of tumor characteristics. There are many feature selection methods, such as principal component analysis (PCA) [2], independent component analysis [3] and linear/non-linear discriminant analysis [4] have been extensively used in literature. Among them, PCA is the most widely used statistical tool for dimensionality reduction today. However, its performance and applicability in real scenarios are limited by a lack of robustness to outlying or corrupted observations [5].

In recent decades, a number of approaches to strengthening PCA have been developed in literature. A new approach referred to as robust principal component analysis (RPCA) was proposed by Candès and co-researchers [5, 6], in which data matrix is the superposition of a low-rank component and a sparse component. In theory, under certain assumptions, it is possible to recover both the low-rank and the sparse components exactly by solving the principal component pursuit (PCP). The theoretical and empirical results suggested that the principal components of a data matrix can be restored even though a fraction of its entries are arbitrarily corrupted. There are many important applications can naturally be modeled using this methodology, such as video surveillance, face recognition, bioinformatics and web search [6]. In this work, we utilize this technique to establish a feature selection algorithm. We formulate the problem of choosing principal features as a low-rank matrix plus a sparse contribution. The low-rank principal matrix represents the best distinctive features for differentiating malignant tumors from benign cases, and the sparse error matrix contains the features with a less distinguishability due to the intrinsic and extrinsic corruption.

The paper is organized as follows. Section 2 formulates the feature selection problem based on the RPCA model. The ultrasound image acquisition and the developed methods are introduced in Section 3. The simulation results and discussions are presented in Section 4. Finally, we conclude the paper in Section 5.

* Corresponding author's email: tao.wan.wan@gmail.com. This work is partially funded by the NCET and the China Scholar Council.

2. PROBLEM SETTING

Motivated by the latest research on RPCA [6], the feature selection problem can be formulated as a RPCA model. The computerized tumor features are subject to the low-rank property. Meanwhile, the error matrix is sparse due to the two reasons: 1) the number of the principal features is essentially sparse compared with the relatively large feature set, and 2) the features with less discriminative power are in small number.

Here we use the notion adopted in [5]. Let the feature matrix $D \in \mathbb{R}^{m \times n}$ be a $m \times n$ matrix generated by the feature extraction method. It can be decomposed as:

$$D = A + E \tag{1}$$

where matrix A is the principal feature matrix known to be low rank, and E is the sparse error matrix. Given the feature matrix D, the essential purpose of the feature selection is to uncover the principal components in matrix A. Although under general conditions this problem is intractable to solve, recent studies [6] have discovered that the principal component pursuit, a convex program, can effectively solve this problem under broad conditions. The low-rank matrix A can be computed by solving the following convex optimization problem:

$$\min_{A,E} \|A\|_* + \lambda \|E\|_1 \quad \text{s.t.} \quad A + E = D \tag{2}$$

where $\|\cdot\|_*$ represents the nuclear norm of a matrix, $\|\cdot\|_1$ is the l_1 norm denoting the sum of the absolute values of matrix entries, and $\lambda > 0$ is a weighting parameter. In our approach, the low-rank and error sparsity are modeled by nuclear norm and l_1 norm according to the definition of Eq. 2. The proposed methods are described in the following section.

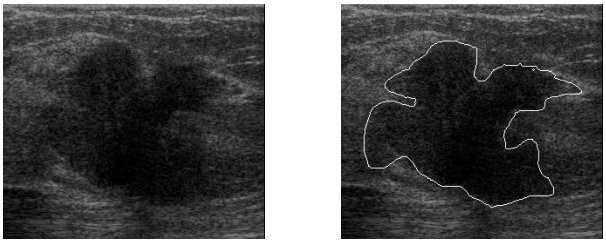
3. MATERIALS AND METHODS

3.1. Data Acquisition

The digital ultrasound image database is provided by the Harbin Institute of Technology and the Second Affiliated Hospital of Harbin Medical University. Ultrasound images were performed using a high-resolution Vivid7 sonography system (GE Healthcare, Milwaukee, WI) and 7.5 14 MHz liner transducer. The tumor boundaries are marked by five radiologists with more than ten years experience. 321 pathologically proven benign and malignant cases are selected and categorized into four classes. The training datasets comprise 92 benign cases and 172 malignant cases, and the test datasets have 21 benign cases and 36 malignant cases. An example of supplied image and its marked result are shown in Fig. 1.

3.2. Feature Extraction

According to [7], the features of breast US images can be divided into four categories: texture, morphologic, model-based and descriptor features. In this work, we focus on the morphological and textural features due to their remarkable discrimination power. More implementation details are given in [8].



(a) (b)

Fig. 1. (a) The original US image. (b) The marked US image.

3.2.1. Morphologic features

With widespread use of ultrasound, the American College of Radiology developed a BI-RADS lexicon for breast ultrasound to standardize the characterization of ultrasonic lesions [9]. Based on the description of these BI-RADS features, we compute the contour moment, improved chain code, substantial depression, and three mathematic ratios to determine benign breast nodules from malignant cases.

Contour moment is a computerized feature by summing over all the pixels of the contour. In general, the image moments M_{pq} are calculated by:

$$M_{pq} = \sum_x \sum_y I(x,y)\, x^p y^q \tag{3}$$

where $I(x,y)$ is the pixel intensity at location (x,y), and $p, q = 0, 1, 2, ...$ We adopted the 7 Hu's invariant moments defined in [10] as the descriptors to characterize the solid breast nodules.

Freeman chain code is one of the most effective chain code representations to describe the contour. In order to capture the unevenness of the contour, we introduce a new angle chain by weighting the edge between two adjacent nodes. By using the angle chain, five features are computed as:

- Normalized sum angle
- Maximum length of consecutive subsequence
- Average length of consecutive subsequence
- Most frequent length of consecutive subsequence
- Normalized number of most frequent length

The consecutive subsequence denotes as the longest subsequence in the angle chain containing none zero values.

The substantial depression feature is calculated to characterize the abnormal lesion contour. Here, we only consider the substantial depression features since protuberance and depression normally appear concurrently.

In addition, three mathematic ratios are considered, and they are defined as follows:

• *D:W ratio* — the ratio of the depth to the width of a lesion.
• *L:S ratio* — the length ratio of the major (long) axis to the minor (short) axis of the equivalent ellipse of the lesion.
• *C:O ratio* — the ratio of the convex hull area to the whole area of a lesion.

3.2.2. Textural features

The textural variation between normal and abnormal tissues is an effective feature for classifying breast tumors. The spatial gray-level dependence (SGLD) matrices are widely adopted to measure the textures in images. In this work, totally 140 SGLD features are extracted for distinguishing benign and malignant nodules.

3.3. Feature Selection

In total, we compute 1285 features with different parameter settings, including morphological and textual features. With so many features available, the crucial task is to find an optimal set of features with relative low dimension. As we previously mentioned, the 1285-dimentional feature matrix D can be formulated as a sum of a low-rank matrix A and a sparse error matrix E. As a result, the problem of feature selection is converted to the problem of recovering a complete low rank matrix from its noisy observation through a convex optimization program. Lin *et al.*[11] recently addressed this problem by using an augmented Lagrange multiplier (ALM) algorithm. Following [11], the Lagrangian function of Eq.2 is:

$$L(A, E, Y, \mu) = \|A\|_* + \lambda\|E\|_1 + \langle Y, D - A - E\rangle$$
$$+\frac{\mu}{2}\|D - A - E\|_F^2 \quad (4)$$

where Y and μ are Lagrangian multipliers, μ is a positive scalar, $\langle P, Q\rangle = tr(PQ^T)$ is trace of matrix multiplication between P and Q, and $\|\cdot\|_F$ is Frobenius norm. Eq. 4 would solve PCP by repeatedly setting $(A_k, E_k) = \arg\min_{A,E} L(A, E, Y_k, \mu_k)$, and then updating the Lagrange multiplier matrix via $Y_{k+1} = Y_k + \mu_k(D - A_k - E_k)$. k is the iteration index. For the particular low-rank and sparse decomposition problem, Lagrange multiplier should be chosen with the constraint in Eq. 2, thus leading to simple update rules for Y and μ. It turns out that updating A_k and E_k once when solving $(A_k, E_k) = \arg\min_{A,E} L(A, E, Y_k, \mu_k)$ is sufficient for A_k and E_k to converge to the optimal solution for the RPCA. This is the idea of the inexact ALM method proposed in [11] which requires one singular value decomposition step per iteration, making the algorithm very efficient for large scale data. In the algorithm, the stopping criterion is $\|D - A - E\|_F <= \delta\|D\|_F$, with $\delta = 10^{-7}$.

Algorithm 1 Feature selection using the RPCA

1. Take the training feature matrix $D \in \mathbb{R}^{264\times1285}$.
2. Apply the RPCA to D with pre-specified values of λ.
3. Take the low rank components of A as the selected training features for the SVM classifier.
4. Repeat Step 2 and Step 3 for the test datasets, and take the low rank components as the selected test features.
5. Classify the tumors into benign/malignant category.

3.4. Classification

Support vector machines (SVM), a well-known classifier, is applied to evaluate the performance of the selected features. In SVM, kernel functions are used to map the input data into a higher dimension space where the data are supposed to have a better distribution, and then an optimal separating hyperplane is chosen. Here we utilize the linear kernel and tune the parameters to yield the best classification results.

4. EXPERIMENTAL RESULTS AND DISCUSSIONS

In the experiments, 321 clinically diagnosed benign and malignant images are used to train and test the performance of the proposed method. The RPCA is employed to choose the best distinctive feature set as an input to the SVM classifier. The proposed algorithm is described in Algorithm 1. The parameter λ is used to adjust the sparsity of the error matrix E and assigned the values ranging from 0.06 to 0.08.

In Table 1 and Table 2, the classification results are compared to the ones obtained by using the classical PCA for feature selection in terms of accuracy, specificity, sensitivity, positive predictive value (PPV) and negative predictive value (NPV). The computation of these evaluation criteria can be found in [7]. For the sake of fair comparison, the number of the principal components in the classical PCA is chosen as same as the rank of the low-rank matrix computed from the RPCA. Fig. 2 shows the receiver operating characteristic (ROC) curve evaluation of the classification accuracy using these two feature selection methods. By examining the experimental results, we can see that the RPCA clearly achieves superior classification results comparing to the classical PCA in all defined evaluation criteria. We notice that in the classical PCA, by increasing the number of principal components in small step, it does not much affect the outputs unless a larger step applied. However, the RPCA generates more vibrating results by tuning the parameter λ to generate different low-rank matrices.

The experiments indicate that the RPCA provides an effective means to select the relevant and important features from a large set of features, resulting in an improved performance for the breast tumor classification. The low-rank based feature selection approach could be interesting for enhancing the computer-aided diagnosis system to further increase the

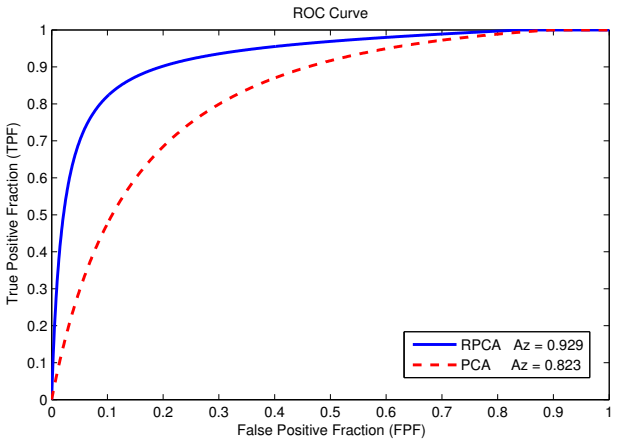


Fig. 2. The ROC curves of classification using the RPCA and the classical PCA.

Table 1. The classification performance using the RPCA

Experiment	1	2	3	4	5	6
Rank	65	66	68	69	70	71
Accuracy(%)	82.5	84.2	84.2	85.9	85.9	85.9
Sensitivity(%)	72.0	73.1	74.9	75.9	78.3	78.3
Specificity(%)	90.6	93.6	90.9	93.8	91.2	91.2
PPV(%)	85.7	90.5	85.7	90.5	85.7	85.7
NPV(%)	80.6	80.6	83.3	83.3	86.1	86.1

Table 2. The classification performance using the PCA

Experiment	1	2	3	4	5	6
N_p*	65	66	68	69	70	71
Accuracy(%)	66.7	66.7	66.7	66.7	66.7	66.8
Sensitivity(%)	53.1	53.1	53.1	53.1	53.1	52.9
Specificity(%)	83.9	83.9	83.9	83.9	83.9	86.9
PPV(%)	80.9	80.9	80.9	80.9	80.9	85.7
NPV(%)	58.3	58.3	58.3	58.3	58.3	55.6

* N_p is the number of principal comonents

diagnostics accuracy and decrease the number of unneeded biopsies.

5. CONCLUSIONS AND FUTURE WORK

In this paper, we have presented a robust feature selection method based on a newly proposed RPCA technique of recovering a low-rank matrix from a high-dimensional data matrix with corrupted observations. The feature selection problem can be formulated as a low-rank principal matrix plus a sparse error matrix. The classification results have been compared to the classical PCA and shown the potentiality of using this feature selection approach to improve breast cancer computer-aied diagnosis systems. In future, we would like to investigate more effective visual features to be integrated in the current framework for measuring the variance of abnormality in breast tumors. Also, we are interested in extending of our work to other medical image modalities.

6. REFERENCES

[1] A.T.Stavros, *Breast Ultrasound,* Lippincott Williams & Wilkins, Philadelphia, PA, 2004.

[2] A.R.Jamieson and M.L.Giger and K.Drukker and H.Li and Y. Yuan and N.Bhooshan, "Exploring nonlinear feature space dimension reduction and data representation in breast CADx with Laplacian eigenmaps and t-SNE," in *Medical Physics*, vol. 37, pp. 339–352, Jan. 2010.

[3] H.Neemuchwala and A.Hero and P.Carson, "Feature coincidence trees for registration of ultrasound breast images," in *Proc. IEEE Conf. Image Process.*, vol. 3, pp. 10–13, Oct. 2001.

[4] W.A.Pereira and A.V.Alvarenga and A.C.Infantosi and L.Macrini and C.E.Pedreira, "A non-linear morphometric feature selection approach for breast tumor contour from ultrasonic images," in *Computers in Biology and Medicine*, vol. 40, pp. 912–918, Oct. 2010.

[5] J.Wright and A.Ganesh and S.Rao and Y.Peng and Y.Ma, "Robust principal component analysis: exact recovery of corrupted low-rank matrices via convex optimization," in *Proc. of Advances in Neural Info. Process. Systems*, 2009.

[6] E. Candès and X.Li and Y.Ma and J.Wright, "Robust principal component analysis?," in *website: http://arxiv.org/abs/0912.3599v1*, 2009.

[7] H.D.Cheng and J.Shan and W.Ju and Y.Guo and L.Zhang, "Automated breast cancer detection and classification using ultrasound images: A survey," in *Pattern Recognition.*, vol. 43, pp. 299–317, Jan. 2010.

[8] R.Liao and T.Wan and Z.Qin, "Classification of benign and malignant breast tumors in ultrasound image based on multiple sonographic and textural features," *BUAA Technical Report*, 2010.

[9] American College of Radiology Breast Imaging Reporting and Data System Atlas (BI-RADS Atlas). Reston, VA, 2003.

[10] M.K.Hu, "Visual pattern recognition by moment invariants," in *IRE Trans. Info. Theory.*, vol. 8, pp. 179–187, Feb. 1962.

[11] Z.Lin and M.Chen and L.Wu and Y.Ma, "The augmented Lagrange multiplier method for extract recovery of corrupted low-rank matrices," *UIUC Technical Report*, Oct. 2009.