Discriminative Learning for Automatic Staging of Placental Maturity via Multi-layer Fisher Vector

Wanjun Li, Dong Ni, Siping Chen, Baiying Lei*, Tianfu Wang*

Department of Biomedical Engineering, School of Medicine, Shenzhen University, National-Regional Key Technology Engineering Laboratory for Medical Ultrasound, Guangdong Key Laboratory for Biomedical Measurements and Ultrasound Imaging, Shenzhen, China

Yuan Yao Shengli Li

Department of Ultrasound, Affiliated Shenzhen Maternal and Child Healthcare, Hospital of Nanfang Medical University, Shenzhen, China

(Email:{leiby, tfwang}@szu.edu.cn)

***Abstract*—In this paper, a new method is proposed to automatically stage the placental maturity from B-mode ultrasound (US) images based on multi-layer Fisher vector (MFV) and densely sampled visual features. The proposed method first densely extracts visual features at a regular grid based on dense sampling instead of a few unreliable interest points. These features are clustered using generative Gaussian mixture model (GMM) to have soft clustering ability, and then learned discriminatively by Fisher vector (FV), which incorporates high-order statistics to enhance the staging accuracy. Differing from the previous studies, a multi-layer FV instead of a single layer FV is adopted in our method to exploit the spatial information of the features. Experimental results show that the proposed method achieves an area under the receiver of characteristics (AUC) of 96.77%, sensitivity of 98.04% and specificity of 93.75%, respectively, for staging placental maturity. Moreover, experimental results also demonstrate that the proposed MFV outperformed traditional methods for placental maturity staging.**

Keywords-Placental maturity staging, multi-layer Fisher vector;discriminative learning

I. INTRODUCTION

Over the past decade, ultrasound (US) imaging has been extensively applied in prenatal diagnosis and prognosis [1] since it is radiation-free, direct-use, and low-cost. B-mode US imaging is one of the most frequently used US imaging especially for placental maturity staging [2-7]. As one of the most frequently used functional evaluation of placental abnormalities such as fetal death, still birth, small gestational age, and various pregnancy complications, placental maturity evaluation has attracted increasing interest over the years. One of the limitations of current methods is their highly dependence on visual observation, which may suffer from doctor's misjudgment. Also, placental maturity staging based on B-mode gray-scale US images are subjected to complex illumination modification, exposure time change, and specular reflection by the imaging process, which makes staging challenging. The calcification degree variations and image quality constraints render the subjective measurement unappealing as well. Hence, evaluation based on subjective methods is not desirable.

To address the limitations of the current diagnosis and prognosis, computer-assisted staging methods have been proposed to make diagnosis and prognosis decision based on the doctor's subjective diagnosis experience and decision scores from developed models. Such methods provide more accurate interpretation and diagnosis. Automatic US placental maturity staging algorithms have been widely developed, which not only reduce the occurrence of judgment error, but also standardize medical tests and alleviate the doctor's workload. For example, Grammum *et al.* first proposed to automatically grade the placental maturity into four grades in 1979 [2] based on the chorionic, substance, and basal plates. However, this method ignored the visual features for the US images, and the staging performance was not satisfactory. In [3], Liu *et al.* suggested to stage the placental US images based on calcification degree and support vector machine (SVM) classifier. Essentially, this method is not fully automated as it still requires the subjective judgment of the operator. Also, a classification rate of 90% is still not desirable via the features of gray variance, distortions, and kurtosis features. Currently, none of the existing methods has been applied in clinical practice, and developing a practical computer-assisted placental maturity staging method would be timely.

High accuracy and precise interpretation are extremely desirable in staging algorithm for placental maturity evaluation. Conventionally, most methods detect the points of interest first, and then invariant feature descriptors are calculated using these points. It is argued that the interest point detector method is unreliable for placental maturity staging. Meanwhile, dense sampling is proved to be effective for the classification tasks. In view of this, dense sampling method is utilized for placental maturity staging. Although there are numerous features and descriptors available for the classification task, it remains a challenging task to find relevant features for the placental staging task due to the complex and invisible differences in the US images.

Recently, visually discriminative features have attracted increasing interest. The widely used scale invariant Fourier transform (SIFT) [8-10] achieves superior performance compared to other descriptors such as color and intensity values in the computer vision field. For example, a US image retrieval method had been proposed in [11] and this method was based on features extracted by SIFT, local binary pattern (LBP), and image intensity value. This method produced quite promising results with publicly available dataset. Meanwhile, local intensity order pattern (LIOP) [12] and DAISY [13] have been demonstrated as highly effective methods for recognition tasks due to their robustness to numerous variations and distortions. Inspired by the promising performance of these visual discriminative features, we believe that the placental maturity staging can be solved effectively using visual features due to the proved desirability and suitability of these features for medical image classification [11].

It is known that most extracted features from the original image space are redundant, indiscriminative, and in high dimension. Feature encoding is an effective way to reduce the feature dimension, which can increase the discriminability as well. One of the most popular feature encoding methods is the bag of visual words (BoVW) [14], which can be further extended by vector of locally aggregated descriptor (VLAD) [15], and Fisher vector (FV) [8]. The recent developed FV is capable of obtaining higher discriminative subspace with the complementary high order information exploration compared with the conventional methods. Motivated by promising performance of FV [5, 6], the extracted features are encoded by FV method to increase the discriminative power. Unlike the traditional methods, a multi-layer FV (MFV) method is devised instead of a single layer FV to incorporate spatial information, which in turn improves the staging performance. Since BoVW framework is simple and effective, our study is based on the BoVW framework using MFV. To the best of our knowledge, MFV has never been applied for fetal placental maturity evaluation using US images. This research is of vital importance in establishing optimal placental evaluation. The development and use of such a tool could consistently increase the diagnostic information used for placental maturity intervention in various conditions.

II. PROPOSED METHOD

Figure 1 shows the flowchart of the placental maturity staging. In the training stage, the visual features (i.e., SIFT, LIOP or DAISY) are first extracted after pre-processing. To further boost the staging performance, these low level visual features are divided into a distinctive group, and then the group representatives are identified by the clustering method. Namely, the low-level visual features are modeled by GMM to obtain the statistics' parameters such as clustering mean and standard deviation of each cluster. After GMM, FV is exploited to encode features based on the parameters learned by GMM. The main purpose of encoding is to discriminate the distributional difference between a test image and all fitted training images. After obtaining low level information and high-order statistics by FV, a histogram in terms of the number of visual word occurrence is formed for image representation. In the testing stage, the same procedure of training is applied for extracting features. For stage decision of the placental maturity, a supervised learning method by the popular SVM is applied since SVM is capable of handling high-dimensional data. SVM can address the over-fitting problem as well.

Essentially, BoVW is a simple counter of feature distribution and represented by the first moment information (i.e., cluster means), and VLAD keeps both mean and residual information. Compared with these two methods, FV is a special case of these two traditional methods by incorporating high order information. As defined in [16], FV is a special case of the Fisher kernel construction. It is designed to encode local image features in a format that is suitable for learning and comparison with simple metric such as the Euclidean distance. In fact, FV is derived from an approximate and improved case of the general Fisher kernel framework. The key advantage of FV is it has the best representation of the feature descriptor distribution for discriminative classification. Another advantage is the alternative soft assignment of feature descriptor to the visual words.

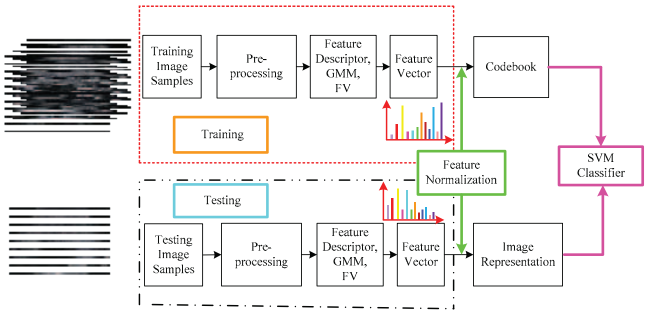

Figure 1. System overview of the automatic placental maturity staging method. Input images are first pre-processed. Features are extracted based on the dense sampling on the placental images. Extracted features are clustered by GMM and encoded by FV, a high level histogram representation is formed in terms of feature occurrence and fed into SVM classifier for staging the placental maturity.

To enhance the staging performance, a generative model based on GMM is first implemented to obtain the posterior probability to improve the discrimination ability of FV. Specifically, for a graphical representation, the dimension of a fixed length vector is dependent on the number of parameters. The first and second order derivatives between dense features and GMM centers are obtained as:

$$\Phi_k^{(1)} = \frac{1}{N\sqrt{w_k}} \sum_{m=1}^{N} \gamma_m(k) \left(\frac{x_m - \mu_k}{\sigma_k} \right), \tag{1}$$

$$\Phi_k^{(2)} = \frac{1}{N\sqrt{2w_k}} \sum_{m=1}^{N} \gamma_m(k) \left(\frac{(x_m - \mu_k)^2}{\sigma_k^2} - 1 \right), \tag{2}$$

where w_k, μ_k, σ_k are the GMM mixture weights, means, and diagonal covariance, $\gamma_m(k)$ is soft assignment weight of the m-th feature x_m to the k-th Gaussian. By concatenating them together, we obtain: $\phi_{FV}(x_m) = [0,...,0,\underbrace{\Phi_k^{(1)},\Phi_k^{(2)}}_{2D\ non-zeros},0,...,0]$ for feature representation via FV, which include high order statistics. Accordingly, high order statistics such as the derivatives of the log-likelihood of the model are encoded.

It is noted that spatial relationships among local appearances play an essential role in recognizing the US image underlying structures. To further increase the discriminability of the descriptor and to take advantage of spatial information, the original image is divided into different divisions and using multi-layer strategy, which is detailed in Figure 2.

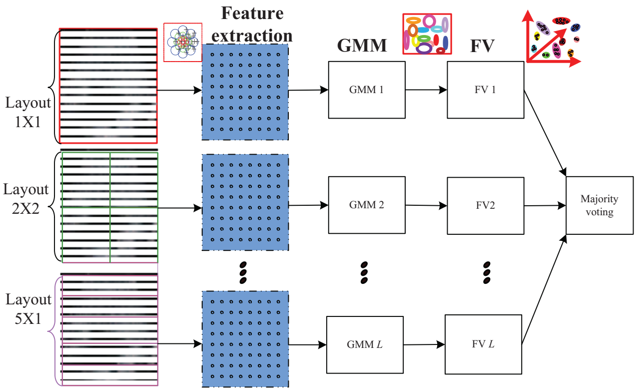

Figure 2. Pipeline of the proposed multi-layer FV (MFV) algorithm. Original US image is partitioned into sub-divisions according to different layouts. Features are extracted by the visual descriptor (i.e., DAISY) from each division, and then concatenated to form the feature matrix for each input placental image. In each layer, extracted features are clustered by GMM to obtain cluster means and deviations, which are then encoded by FV independently. Final rule is based on the majority voting.

It is also proved that dense feature descriptor and FV feature encoding method obtain high descriptive power for image representation [8]. Inspired by the remarkable performance of the spatial pyramid model [14], the spatial pyramid strategy is also applied to divide the image into sub-divisions to incorporate spatial distribution information. Therefore, the densely sampled feature is partitioned into subdivisions using the spatial pyramid model. The feature vectors are built by concatenating all features in every division. The means and standard deviations of each visual word occurrence are concatenated to form the multi-layers of FV. Each layer concatenates all subdivisions in each layout. MFV is a concatenation of different layers to incorporate all information. The majority voting, concatenation or averaging strategy can be utilized to incorporate the information from all layers. In our method, majority voting method is adopted.

III. EXPERIMENTAL RESULTS

We conducted an extensive analysis to assess the proposed method for addressing the issue of placental function evaluation. The US images used in this study were acquired by a commercial US scanner from Shenzhen Woman and Children Hospital. The representative images of each stage are shown in Figure 3. Our database is composed of a total of 443 placental images, where 187, 135, 85, and 36 images are in stage 0, stage 1, stage 2, stage 3, respectively. The fetal gestational age in these placental images ranges from 18 to 40 weeks. Conventional US sweep was conducted to obtain the images of pregnant women in the supine position by a radiologist with more than five years of experience in US obstetrics. The written informed sheets are obtained by all the participants. The performance of the placental maturity staging is quantified by classification metrics such as mean average precision (mAP), area under receiver operation characteristics (AUC), sensitivity, specificity, and receiver operating characteristic (RoC) curves. All the experiments are 10-fold cross-validated to avoid any bias. The experiments are repeated at least 10 times and average results are reported. Quantitative assessment of the placental maturity in the four distinctive stages is used to evaluate the effect of different feature descriptors, encoding methods, and the proposed MFV method.

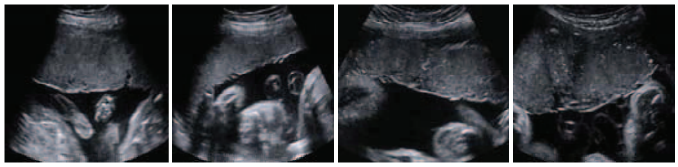

Figure 3. From left to right, image samples of 4 stages are stage 0,1,2, and 3 images, respectively.

Figure 4a shows the staging performance of different feature descriptors for four different stages in terms of mAP, sensitivity, and specificity based on the dense sampling method, where Combine denotes combining SIFT and intensity information together. Figure 4b demonstrates the ROC curves of different methods. It is noteworthy that DAISY feature outperforms the traditional SIFT feature in staging the placental maturity. Given the comparison result, the highest staging results are achieved by DAISY feature. The Combine feature obtains better performance than each individual feature, which means that feature concatenation is effective for the placental maturity staging. From Figure 4, it is demonstrated that visual feature especially DAISY feature is suitable for placental staging. The main reason is that visual feature is able to capture the complicated changes of the US placental images by using the gradient method. Meanwhile, the scale and rotation invariance property of the visual feature descriptors has advantage for placental staging. The brightness and calcification degree change can be better represented by the visual features, and hence remarkable staging performance has been obtained.

Figure 5a shows mAP, sensitivity, and specificity of the staging results, and Figure 5b shows AUC and RoC curves obtained by various feature encoding methods based on DAISY feature since DAISY has obtained the best performance. It can be seen that the aggregating vectors such as VLAD and MFV methods can boost staging performance.

It is also shown that MFV outperforms traditional FV, VLAD and BoVW methods due to its high order statistics in the feature encoding method. The AUC result for the FV feature encoding based on dense sampling is 96.77%, and this value is dropped to 95.3% and 90.62% with VLAD and BoVW feature encoding methods, respectively. The sensitivity and specificity results of MFV method are 98.04%, 93.75%, respectively. Generally, aggregating vectors obtain better staging performance in terms of sensitivity, specificity, mAP results, than the commonly used BoVW method, whereas MFV method obtains the best staging performance among all methods due to the introduced discriminative learning and high order information in the GMM model. It also indicates that the introduced discriminative learning by FV is quite suitable and effective for the placental staging and promising in the clinical application as well. The primary explanation is that discriminative learning and high order statistics are quite effective to boost placental maturity staging performance among various descriptors. The obtained high staging result of the proposed method also demonstrates its potential in clinical practice.

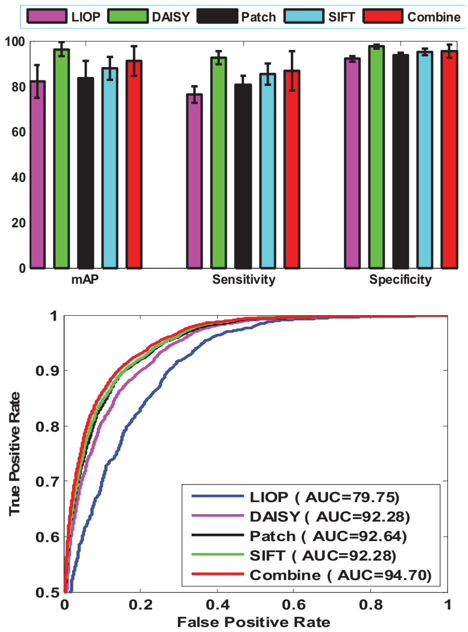

Figure 4. Staging results of different feature descriptors using dense sampling method; (a) mAP, sensitivity, specificity results; (b) RoC curves results.

IV. Conclusions

In this paper, an automatic and highly accurate staging algorithm based on the dense descriptor and MFV for placental maturity evaluation is presented. Dense features have an advantage of extracting informative features, whereas MFV is able to boost the staging performance by introducing high order statistics. Our experiments showed that the proposed method achieves promising performance for the placental maturity staging. The automatic staging algorithm is able to provide diagnosis assistance to evaluate the placental maturity evaluation, and has extensive potential applications. It will be useful in expediting the manual work of doctors and reduce time-consuming visual surveillance of clinician in clinical practice. In addition, the methodologies utilized in this study are quite general and can be extended to classification and staging in other fields.

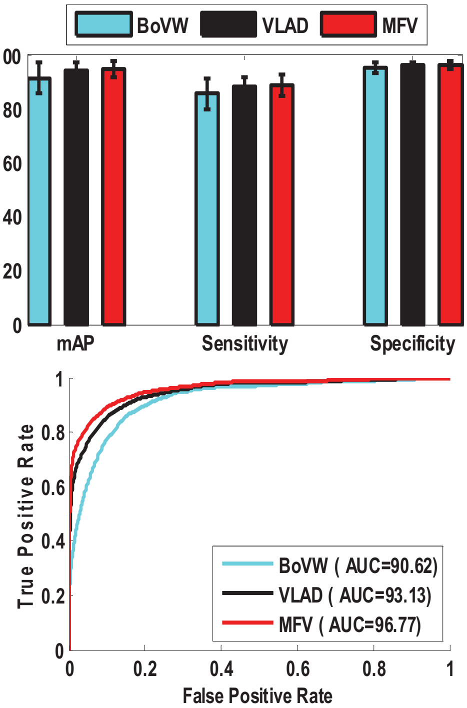

Figure 5. Staging results of feature encoding method using DAISY descriptor; (a) mAP, sensitivity, specificity results; (b) RoC curves results.

ACKNOWLEDGEMENT

This work was supported partly by National Natural Science Foundation of China (Nos. 61402296, 81571758, 61571304, 61101026, 61372006, 81270707 and 61427806), the 48th Scientific Research Foundation for the Returned Overseas Chinese Scholars, National Natural Science Foundation of Guangdong Province (No.S2013040014448), Shenzhen Key Basic Research Project (Nos. JCYJ20130329105033277, JCY20150525300004O, and JCYJ20140509172609164), and Shenzhen-Hong Kong Innovation Circle Funding Program (No. JSE201109150013A).

References

[1] B. Lei, L. Zhuo, S. Chen, S. Li, T. Wang, and D. and Ni, "Automatic Recognition of Fetal Standard Plane in Ultrasound Image," in *Proceedings of 2014 IEEE International Symposium on Biomedical Imaging (ISBI 2014)*, 2014, pp. 85-88.

[2] P. A. Grannum, R. Berkowitz, and J. C. Hobbins, "The ultrasonic changes in the maturing placenta and their relation to fetal pulmonic maturity," *American Journal of Obstetrics & Gynecology,* vol. 133, pp. 915-22,1979

[3] Z. Liu, H. Zheng, and S. Lin, "Application of Multi-Classification Support Vector Machine in the B-Placenta Image Classification," in *Proc. of International Conf. on Computational Intelligence and Software Engineering*, 2009, pp. 1-4.

[4] B. Lei, E.-L. Tan, S. Chen, L. Zhuo, S. Li, D. Ni, and T. Wang, "Automatic Recognition of Fetal Facial Standard Plane in Ultrasound Image via Fisher Vector." vol. 10, 2015, p. e0121838.

[5] B. Lei, Y. Yao, S. Chen, S. Li, W. Li, D. Ni, and T. Wang, "Discriminative Learning for Automatic Staging of Placental Maturity via Multi-layer Fisher Vector," *Scientific Reports,* vol. 5, pp. 2015

[6] X. Li, Y. Yao, D. Ni, S. Chen, S. Li, B. Lei, and T. Wang, "Automatic staging of placental maturity based on dense descriptor," *Bio-medical materials and engineering,* vol. 24, pp. 2821-2829,2015

[7] X. Li, Y. Yao, D. Ni, S. Chen, S. Li, B. Lei, and T. Wang, "Automatic staging of placental maturity based on dense descriptor," in *36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, 2014, pp. 2821-2829.

[8] J. Sánchez, F. Perronnin, T. Mensink, and J. Verbeek, "Image classification with the fisher vector: theory and practice," *International Journal of Computer Vision,* 1-24,2013

[9] B. Lei, E.-L. Tan, S. Chen, D. Ni, and T. Wang, "Saliency-driven image classification method based on histogram mining and image score." vol. 48, pp. 2567-2580.

[10] B. Lei, T. Wang, S. Chen, D. Ni, and H. Lei, "Object recognition based on adapative bag of feature and discriminative learning," in *20th IEEE International Conference on Image Processing*: IEEE, 2013, pp. 3390-3393.

[11] H. Shen, D. Tao, and D. Ma, "Multiview Locally Linear Embedding for Effective Medical Image Retrieval," *PloS one,* vol. 8, pp. e82409,2013

[12] Z. Wang, B. Fan, and F. Wu, "Local intensity order pattern for feature description," in *Proc. of ICCV*, 2011, pp. 603-610.

[13] E. Tola, V. Lepetit, and P. Fua, "Daisy: An efficient dense descriptor applied to wide-baseline stereo," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 32, pp. 815-830,2010

[14] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: spatial pyramid matching for recognizing natural scene categories," in *Proc.of CVPR*, 2006, pp. 2169-2178.

[15] H. Jégou, F. Perronnin, M. Douze, and C. Schmid, "Aggregating local image descriptors into compact codes," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 34, pp. 1704-1716,2012

[16] F. Perronnin and C. Dance, "Fisher kernels on visual vocabularies for image categorization," in *Proc. of CVPR*, 2007, pp. 1-8.