AUTOMATIC RECOGNITION OF FETAL STANDARD PLANE IN ULTRASOUND IMAGE

Baiying Lei[1], *Liu Zhuo*[1], *Siping Chen*[1], *Shengli Li*[2],*Dong Ni*[1*], *and Tianfu Wang*[1*]
[1]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.
[2] Department of Ultrasound, Affiliated Shenzhen Maternal and Child Healthcare
Hospital of Nanfang Medical University, 3012 Fuqiang Rd, Shenzhen, P.R.China.
(Email: *nidong@szu.edu.cn, *tfwang@szu.edu.cn)

ABSTRACT

Detection and recognition of standard plane automatically during the course of US examination is an effective method for diagnosis of fetal development. In this paper, an automatic algorithm is developed to address the issue of recognition of standard planes (i.e. axial, coronal and sagittal planes) in the fetal ultrasound (US) image. The dense sampling feature transform descriptor (DSIFT) with aggregating vector method (i.e. fish vector (FV)) is explored for feature extraction. The learning and recognition of the planes have been implemented by support vector machine (SVM) classifier. Experimental results on the collected data demonstrate that high recognition accuracy is obtained.

Index Terms—Ultrasound image, Standard plane, Detection and recognition, Dense SIFT, Aggregating vector.

1. INTRODUCTION

Due to the widely used portable, non-invasive, low-cost ultrasound (US) probes [1, 2], US imaging techniques have become more and more prevalent in the industrial and academic field compared to CT and MRI imaging. In US imaging, experienced clinicians can handle with US diagnosis very effectively, but imaging experts and advanced imaging equipment is very scarce in the underprivileged country. Automatic diagnosis technology is very beneficial to assist nonexperts. Moreover, it is worthwhile to develop this technology for its potential application in underprivileged countries for both experienced and inexperienced examiner. Actually, identifying and recognizing the standard plane such as axial, coronal and sagittal planes from the US images/videos are crucial for accurate diagnosis and biometric measurements. Therefore, some researches have been attracted by the automatic detection of fetal standard planes (facial or abdominal) in US images [3-5].

To detect standard plane, local binary pattern (LBP), AdaBoost [4, 6], conditional random forest [2] have been widely applied due to its effectiveness. These detection methods have been explored to identify the standard plane from US data in the literature. For example, in [7], fetal brain from 3D US data has been proposed by Siemens researcher. The fetal abdominal standard plane in US volume has been learned and classified by AdaBoost approach [3]. In [5], standard plane of gestational cancer has been identified and detected by Zhang et al. using local context information and cascade AdaBoost. The stand plane in the 3D ultrasound volume is automatically selected in [8] by Rahmatullah et al. Apart from detection method alone, there is some hybrid detection and classification scheme for automatic diagnosis too [9].

Though the above mentioned methods are promising and effective in detection, the recognition after detection of the fetal facial standard plane extracted from consecutive 2D US videos/images is still an undeveloped area. To the best of our knowledge, there is still no automatic recognition algorithm for these planes after detection to reduce the diagnosis time. This system is the first step toward automated recognition of fetal facial standard plane in the US fetal images after detection, which paves the way for the prenatal care and development too.

Scale invariant feature transform (SIFT) and semantic attributes are very promising for image representation. Nowadays, dense features play an essential role in the state-of-the-art object recognition task [10, 11], which extracts feature densely instead of sparse and possible unreliable points from interest point detector. To further enhance the discriminative power, dense features are encoded into a single feature vector by a histogram of occurrence. The most popular encoding methods for learning and recognition are vectors of locally aggregated descriptors (VLAD), super vector coding (SVC) and fish vectors (FVs) [10], which reduce the information loss of quantization. Aggregating vectors are essentially an extension of the bag of visual words (BoVW). Therefore, DSIFT descriptor is integrated with aggregating vectors such as VLAD and FV. Besides, the dimensionality reduction can be done by principal component analysis (PCA) projection to reduce processing time and improve the discriminative learning ability. In view of this, the standard plane recognition is implemented by dense descriptor with local aggregated feature vectors.

978-1-4673-1961-4/14/$31.00 ©2014 IEEE

2. METHODOLOGY

2.1. System framework and dense features

For an input fetal facial US image, the preprocessing of the original US images including removal of the script, noise reduction and image enhancement is first applied. After pre-processing, the region containing axial, coronal and sagittal plane (namely region of interest, ROI) is detected by the weak AdaBoost classifier [6] to reduce the search range. The ROI region is then computed by dense sampled DSIFT and encoded by aggregating vectors for feature extraction. Since aggregating vector [10] is very effective in encoding the sparse feature structure, the spatial distribution of the signal is often ignored. To compensate this, spatial pyramid coding was applied to divide the image into cells, and then feature vectors of each cell are concatenated together. The procedure for feature vector construction and representation is shown in Fig.1. Finally, a strong and efficient classifier is designed to identify each plane. In our work, the recognition task is completed by a conventional linear SVM classifier. The overall framework of our proposed method is shown in Fig.2.

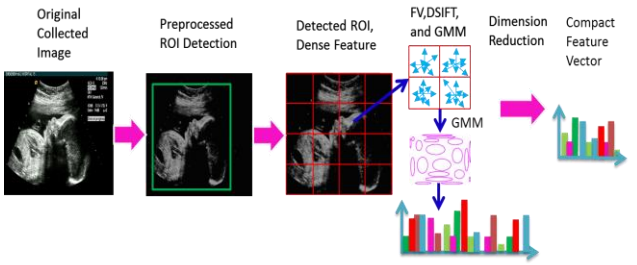

Fig.1. Diagram for feature extraction.

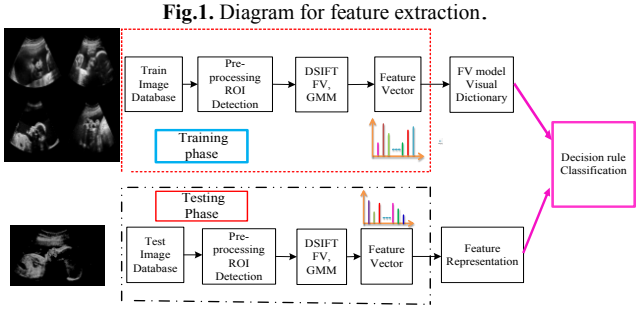

Fig.2. The overall framework of the proposed system.

SIFT has been commonly used for image descriptor in the recent year. It is observed that different US images have different spatial organizations of image gradients, and hence the spatially dense computed DSIFT descriptor is very suitable for this task [11]. It is assumed that spatial relationships between local appearances play an important role in recognition of underlying structure of US image. Therefore, contextual spatial feature (i.e.augmented feature), that is, the appearance of neighborhood around each pixel is also included. A histogram using K-means algorithm by clustering DSIFT representatives further reduces the dimensionality (as opposed to using all DSIFT descriptors). The feature vector dimension can be further reduced by PCA projection.

2.2. Feature vector and image representation

In image representation, a generative Gaussian mixture model(GMM) for spatial location combined for encoding feature and location can improve the performance greatly as claimed in [10], which has been explored for recognition of the standard plane in US image data. Specifically, given a codebook learned by the K-means: $\{\mu_k, k=1,...,T\}$, and a set of local descriptors, $X=\{x_p, p=1,...N\}$, the steps to extract the feature vector are denoted as follows:

1) Assign neighboring:

$$NN(x_t)=\underset{\mu_k}{\operatorname{argmin}}\|x_k-\mu_k\| \tag{1}$$

2) Computes v_k:

$$v_k=\sum_{x_k:NN(x_k)=\mu_k} x_k-\mu_k \tag{2}$$

3) Concatenate v_k and normalize all the feature vectors.

For a graphical representation, v_k, the dimension of a fixed length vector is dependent on the number of parameters, which arouse the question to optimize parameter to better fit the data, that is, results can be optimal by adding higher order statistics. Therefore, the aggregating vector is implemented by a GMM to fit the feature vectors, and then the derivatives of the log-likelihood of the model are encoded using its parameters. Gaussian mean and variances for the average first and second order differences between dense features and GMM center are computed by:

$$\Phi_k^{(1)}=\frac{1}{N\sqrt{w_k}}\sum_{p=1}^{N}\alpha_p(k)\left(\frac{x_p-\mu_k}{\sigma_k}\right) \tag{3}$$

$$\Phi_k^{(2)}=\frac{1}{N\sqrt{2w_k}}\sum_{p=1}^{N}\alpha_p(k)\left(\frac{(x_p-\mu_k)^2}{\sigma_k}-1\right) \tag{4}$$

where $\{w_k,\mu_k,\sigma_k\}$ are the GMM mixture weights, means, and diagonal covariance. $\alpha_p(k)$ is the soft assignment weight of the p-th feature x_p to the k-th Gaussian. By concatenating the difference vectors together: $\phi=[\Phi_1^{(1)},\Phi_1^{(2)},...,\Phi_k^{(1)},\Phi_k^{(2)}]$, FV ϕ is obtained. The main purpose of the encoding is to discriminate distribution difference between a specific test image and all fitted training image. In order to remove the correlation of dense features, PCA is utilized since uncorrelated feature and GMM covariance matrices of diagonal assumption are consistent. The whitening and low energy dimension has been explored for performance boosting.

Let the total dimension of FV is $2Kd$, where K means the GMM Gaussian number, and d is the feature vector patch

dimension. Feature vector of BoVW is n dimension. Essentially, FV is a higher order soft assigned bag of visual words with high-order statistics and dimension. Though the FV dimension is high, but it is still significantly lower than all the dense features combined together. Actually, the vocabulary generated by FV is the probabilistic visual vocabulary. In our system, a total of 128 feature vector is generated per image pixel. PCA can reduce the feature dimensions from 128 to 64.

The recognition performance is further improved by the square rooting. L_2 normalization is applied to remove the dependence on image-specific content. Furthermore, the variance stability is improved by the transform:

$$f(x) = \text{sign(x)} \times |x|^p, 0 \le p \le 1 \quad (5)$$

A default parameter $p = 0.5$ has been applied in our work.

2.3. Learning and recognition

PEGASOS SVM algorithm using a linear SVM solver [11] is utilized to train one versus all classifiers. The scoring function for hyperplane H in SVM classifier is defined as:

$$H : \mathbf{w}\mathbf{x}_i^t + b = 0, \quad (6)$$

where $\mathbf{w} = (w_1, w_2 ... w_N)$ is an adaptable weighting parameter, $b \in R$ is a bias parameter, and t denote the transpose operation. Non-negative variables, $\xi_1, ..., \xi_N$ (i.e., $\xi_i > 0$, for each i), are employed to generalize the derivation of the decision function, and hence the above function can be transformed as below:

$$d_i(\mathbf{w}\mathbf{x}_i^t + b) \ge 1 - \xi_i, \quad i = 1, 2, ..., N \quad (7)$$

Commonly, $\mathbf{w}$ is minimized to obtain optimal values by:

$$\min: \frac{1}{2}\mathbf{w}\mathbf{w}^t + C\sum_{i=1}^{N}\xi_i$$

$$\text{s.t. } d_i(\mathbf{w}\mathbf{x}_i^t + b) \ge 1 - \xi_i, \quad \xi_i \ge 0, \ i = 1, 2, ..., N \quad (8)$$

where C is regularization parameter which has an effect on the tradeoff between margin maximization and the number of misclassified input vectors. 10-fold cross validation is adopted for the parameter. SVM transform the primal problem to the dual problem for easily solving the problem with Lagrange function.

3. EXPERIMENTAL RESULTS

The training set is composed of 87 images of axial plane, 87 images of coronal plane, 112 images of sagittal plane, and 200 other images without containing any standard plane randomly extracted from the US videos. All images were extracted from US videos acquired by an ultrasound scanner from Siemens Acuson Sequoia 512 from Shenzhen Maternal and Child Health Hospital. Fetal gestational age was between 20 and 36 weeks. Conventional US sweep was performed to obtain the videos on pregnant women in the supine position by a radiologist with more than five years of experience in obstetrics US. Fetal images sample for the training standard planes are shown in Fig.3.

The system is implemented using Matlab2010b and only 0.6 seconds are taken on a single CPU core (in the case of 2 pixel SIFT density). The mean average precision (mAP), concept based precision and recall curve, true and false positive curve are provided for performance evaluation.

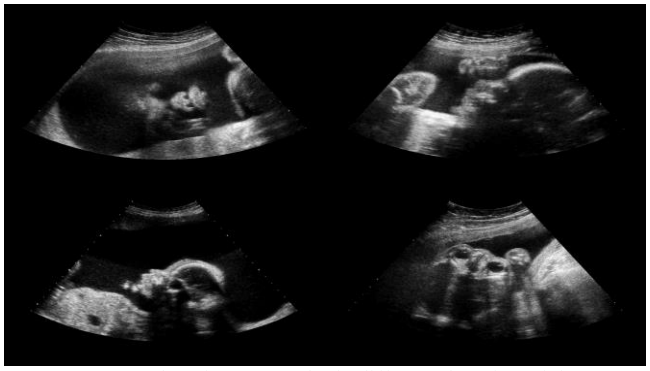

Fig.3. Fetal US image samples in different plane for training.

The recognition confusion matrix of the standard plane is shown in Fig.4. The rows represents the actual standard plane labels, the column denotes the predicted labels. The diagonal elements mean that recognition accuracy for each label. As can be seen from the confusion matrix, the overall recognition accuracy for each class is very good. The mis-recognition ratio is very low too.

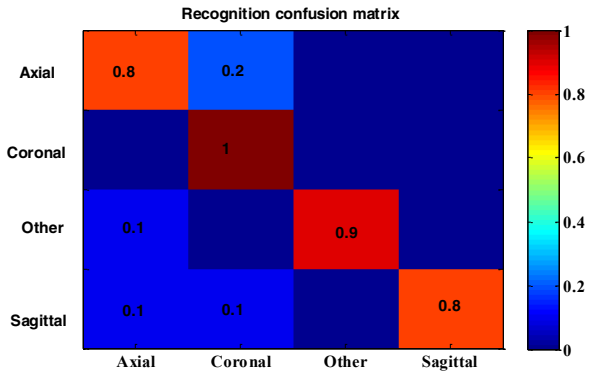

Fig.4. Confusion matrix for the standard plane recognition.

Fig.5 shows recognition results for these three standard planes based on true positive and false positive results. We can see that sagittal and coronal plane are more easily to be discriminated than axial plane. Axial plane is more confused in US imaging, which requires more power to separate it. The recognition performance is further confirmed with precision and recall curve shown in Fig.6. The precision and recall results demonstrate that the highest recognition error could happen in axial plane, the results are consistent with the recognition result in Fig. 5 as well.

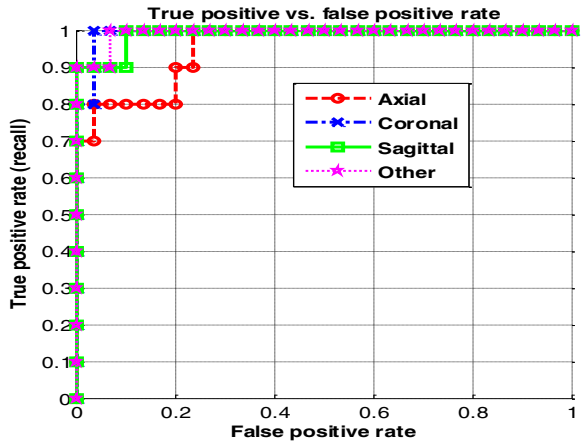

Fig.5. True positive and false positive curve.

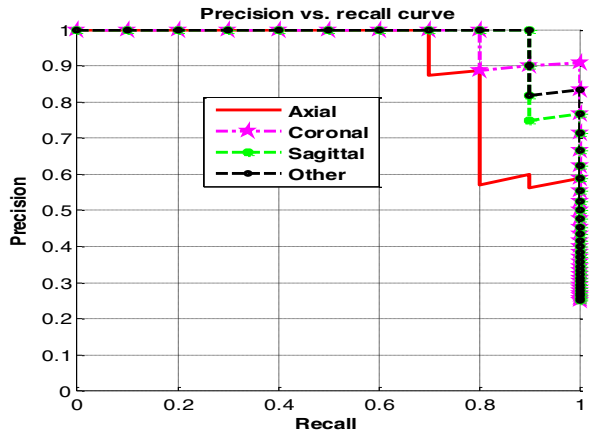

Fig.6. Precision and recall curve.

Table 1 shows recognition results for these 4 classes based on mAP (%) results with different feature encoding algorithm (suffix Aug means additional geometrical feature). It can be seen that very high recognition results have been obtained for each class. Generally, aggregating features (VLAD and FV) outperform the traditional BoVW feature. Augmented feature like geometrical feature is able to improve the recognition accuracy of the standard plane too. FV-Aug algorithm has obtained the best recognition performance among all algorithms, which should be applied in our system with possible extension to other applications.

Table 1. Recognition results (mAP (%)).

Algorithm	Axial	Coronal	Sagittal	Other
BoVW	84.46	97.52	96.59	95.69
VLAD	91.61	98.35	98.48	97.9
FV	88.54	99.17	96.99	96.97
FV-Aug	90.48	96.25	98.35	98.48

4. CONCLUSIONS

In this paper, the first automatic recognition of standard plane in fetal US images is proposed by using a novel image representation based on aggregating vectors. Experimental results show that the proposed method can recognize the important standard planes successfully with high mean accuracy precision. The future direction for this work is to extend this work to identify other anatomical structures in fetal or nonfetal US images. Another possible future direction is to design the hybrid detection and recognition system for practical application. This work can be investigated further for cancer prediction and recognition.

5. ACKNOWLEDGEMENT

This work was supported partly by National Natural Science Foundation of China (Nos. 61101026, 61372006, 60871060 61031003 and 81270707), Shenzhen Key Basic Research Project (Nos. 201101013, JCYJ20130329105033277, and JSE201109150013A), National Natural Science Foundation of China Postdoc (No. 2013M540663) and National Natural Science Foundation of Guangdong Province (No. S201304001448).

6. REFERENCES

[1] A. Abuhamad, P. Falkensammer, F. Reichartseder, and Y. Zhao, "Automated retrieval of standard diagnostic fetal cardiac ultrasound planes in the second trimester of pregnancy: a prospective evaluation of software," *Ultrasound in Obstetrics and Gynecology,* vol. 31, pp. 30-36, 2008.

[2] G. Carneiro, B. Georgescu, S. Good, and D. Comaniciu, "Detection and measurement of fetal anatomies from ultrasound images using a constrained probabilistic boosting tree," *IEEE Transactions on Medical Imaging,* vol. 27, pp. 1342-1355, 2008.

[3] D. Ni, T. Li, X. Yang, J. Qin, S. Li, C.-T. Chin, S. Ouyang, T. Wang, and S. Chen, "Selective Search and Sequential Detection for Standard Plane Localization in Ultrasound," in *Abdominal Imaging. Computation and Clinical Applications*. vol. 8198, 2013, pp. 203-211.

[4] D. Ni, Y. Yang, S. Li, J. Qin, S. Ouyang, T. Wang, and P. A. Heng, "Learning based automatic head detection and measurement from fetal ultrasound images via prior knowledge and imaging parameters," 2013, pp. 772-775.

[5] L. Zhang, S. Chen, C. T. Chin, T. Wang, and S. Li, "Intelligent scanning: Automated standard plane selection and biometric measurement of early gestational sac in routine ultrasound examination," *Medical Physics,* vol. 39, pp. 5015-5027, 2012.

[6] P. Viola and M. Jones, "Robust Real-Time Face Detection," *International Journal of Computer Vision,* vol. 57, pp. 137-154, 2004.

[7] F. Shaolei, S. K. Zhou, S. Good, and D. Comaniciu, "Automatic fetal face detection from ultrasound volumes via learning 3D and 2D information," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 2488-2495.

[8] B. Rahmatullah, A. Papageorghiou, and J. A. Noble, "Automated Selection of Standardized Planes from Ultrasound Volume," in *Machine Learning in Medical Imaging*. vol. 7009, 2011, pp. 35-42.

[9] L. Gorelick, O. Veksler, M. Gaed, J. Gomez, M. Moussa, G. Bauman, A. Fenster, and A. Ward, "Prostate Histopathology: Learning Tissue Component Histograms for Cancer Detection and Classification," *IEEE Transactions on Medical Imaging,* pp. 1804-1818, 2013.

[10] H. Jegou, F. Perronnin, M. Douze, J. Sanchez, P. Perez, and C. Schmid, "Aggregating Local Image Descriptors into Compact Codes," *IEEE Trans. on Pattern Analysis and Machine Intelligence,* vol. 34, pp. 1704-1716, 2012.

[11] A. Vedaldi and A. Zisserman, "Efficient additive kernels via explicit feature maps," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 34, pp. 480-492, 2012.