Journal of Zhejiang University SCIENCE B
ISSN 1673-1581 (Print); ISSN 1862-1783 (Online)
www.zju.edu.cn/jzus; www.springerlink.com
E-mail: jzus@zju.edu.cn

A fast automatic recognition and location algorithm for fetal genital organs in ultrasound images

Sheng TANG[1,2], Si-ping CHEN[†‡3]

([1]*Post-Doctoral Research Station, Shenzhen University, Shenzhen 518060, China*)

([2]*Post-Doctoral Work Station, Microprofit Electronic Co. Ltd, Shenzhen 518057, China*)

([3]*Department of Biomedical Engineering, Shenzhen University, Shenzhen 518060, China*)

[†]E-mail: chensiping@szu.edu.cn

Received Mar. 19, 2009; Revision accepted July 28, 2009; Crosschecked Aug. 13, 2009

Abstract: Severe sex ratio imbalance at birth is now becoming an important issue in several Asian countries. Its leading immediate cause is prenatal sex-selective abortion following illegal sex identification by ultrasound scanning. In this paper, a fast automatic recognition and location algorithm for fetal genital organs is proposed as an effective method to help prevent ultrasound technicians from unethically and illegally identifying the sex of the fetus. This automatic recognition algorithm can be divided into two stages. In the 'rough' stage, a few pixels in the image, which are likely to represent the genital organs, are automatically chosen as points of interest (POIs) according to certain salient characteristics of fetal genital organs. In the 'fine' stage, a specifically supervised learning framework, which fuses an effective feature data preprocessing mechanism into the multiple classifier architecture, is applied to every POI. The basic classifiers in the framework are selected from three widely used classifiers: radial basis function network, backpropagation network, and support vector machine. The classification results of all the POIs are then synthesized to determine whether the fetal genital organ is present in the image, and to locate the genital organ within the positive image. Experiments were designed and carried out based on an image dataset comprising 658 positive images (images with fetal genital organs) and 500 negative images (images without fetal genital organs). The experimental results showed true positive (*TP*) and true negative (*TN*) results from 80.5% (265 from 329) and 83.0% (415 from 500) of samples, respectively. The average computation time was 453 ms per image.

Key words: Ultrasound image, Fetal genital organ, Point of interest (POI), Feature selection, Feature extraction, Class imbalance, Multiple classifier architecture

doi:10.1631/jzus.B0930162 **Document code:** A **CLC number:** TP312; R715

INTRODUCTION

Sex imbalance problem

Severe sex ratio imbalance at birth is now becoming an issue in a number of Asian countries, including China, Korea, India, etc. The sex ratio at birth (SRB) in China increased to 119.9 (male/female) in 1999 (Population Census Office under Chinese State Council, 2002), which was substantially higher than the normal level (103~107). Influenced mainly by the strong and centuries-old cultural preference for male, the leading immediate cause of the abnormally high SRB is prenatal sex-selective abortion after illegal sex identification by cheap and readily available ultrasound (US) scanning (Qi and Mason, 2005). Prenatal sex-selective abortion results in subsequent gender imbalance, and contributes to social problems including serious marriage problems, sexual crime and other social instability. While the governments of these countries are currently taking legislative actions to improve the social-security systems and to remove the culturally deep-rooted male preference tradition, their immediate aim is to stop unethical and illegal prenatal sex identification via ultrasound scanning (Li, 2007). However, it is

‡ Corresponding author

very difficult to collect related evidence while the ultrasound operator involved needs only to smile or frown at the pregnant women to skirt the law. A computer assisted system, which traces and records the abnormal operations of the technicians based on the contents of US images, would help to provide impartial and independent evidence of any malpractice. This assisted system should include a fast automatic recognition and location algorithm for fetal genital organs in US images.

Ultrasound image recognition

Because of inherent speckle noise, the absence of edge information and the complexity of the image contents, precise automatic segmentation of US images appears to be impractical. Most researchers in the field of US image recognition tend to construct recognition systems based on the following routine steps: Step 1, according to the aim of the research, predefine a region of interest (ROI) in the experimental images manually or under manual intervention; Step 2, pick up textural and/or shape features from the ROI; Step 3, refine a subset of the features which suits the classification problem; Step 4, use the selected feature set to classify examples based on appropriate supervised or unsupervised classifiers (Vince *et al*., 2000; Smutek *et al*., 2001; Lu *et al*., 2003; Zhu *et al*., 2006; Chrzanowski *et al*., 2008). Chen *et al*.(2005) distinguished solid malignant from benign breast lesions using fractal textural features. The lesions were manually segmented from breast US images by a physician, and then fractal analysis was applied to obtain the fractal textural features for each ROI. Two features which showed better results were selected from a total of eight fractal textural features and the *k*-means method was used for classification. Piliouras *et al*.(2004) proposed another classification algorithm for breast lesions. For each lesion manually segmented by a physician, 40 textural features and 8 shape features were calculated and a subset of the features was chosen using Student's *t*-test. These features were later fed to the linear-kernel support vector machine (SVM^{LSM}) classifier to distinguish simple cysts from complicated cysts, as well as benign from malignant solid masses. Strzelecki *et al*.(2006) tried to classify different intracardiac masses. For each image, an ROI was first defined by a cardiologist to mark an intracardiac mass. Then, for each ROI, 266 textural features were calculated using 5 textural measurements. The number of features was reduced based on minimizing the classification error and an average correlation coefficient, and a feedforward neural network was implemented to classify the 5 selected features. Bhanu Prakash *et al*.(2002) analyzed the maturity of fetal lungs using US image recognition. The ROIs were manually predefined and 6 features were selected using Pearson's correlation test from a total of 64 features which were extracted from 5 textural measurements, then the performances of 6 classifiers upon the selected features were compared.

The method of Buller *et al*.(1996) is quite different from common methods, which try to automatically classify and locate malignant and benign breast lesions in US images. The original breast US images were first compressed by taking the average of a 3×3 pixel window. Then, for each pixel in the 100×100 pixel preprocessed image, a total of 48 features, together with the presence of this pixel in the manually marked lesion area, formed a training example for the neural networks. After the training, the classification result of every pixel in the testing image was collected to give the synthesized conclusion, i.e., the type and location of the lesion or no lesion. Although the recognition results were not as good as those from research using ROI (Sivaramakrishna *et al*., 2002; Piliouras *et al*., 2004; Chen *et al*., 2005) and the experimental image dataset was quite small, the design concepts described in this early publication were significant, as they attempted to automate the recognition procedure.

MATERIALS AND METHODS

Overall design of the algorithm

Fig.1 illustrates eight typical US fetal images containing the genital organs (positive images), after various periods of gestation. Gestational age was determined by concurrent biparietal diameter measurements. Fetal genital organs were identified by two experienced obstetricians together. Following the descriptions of Scholly *et al*.(1980), a male fetus can be identified by visualization of the penis or scrotum. The shape of the fetal penis is visualized in the ventro-caudal abdomen (Figs.1a and 1d~1f), and the

fetal scrotum appears as two symmetrical, oval and homogeneous areas (Fig.1h). A female fetus can be identified by visualization of the labia and, in contrast to a male fetus, no other relevant structures can be visualized. The labia can be visualized either as three parallel lines on the ventral perineum, commonly called "the three-line sign" (Figs.1b and 1c), or as the dome-shaped bulges out of the buttocks, commonly called "the butterfly" (Fig.1g).

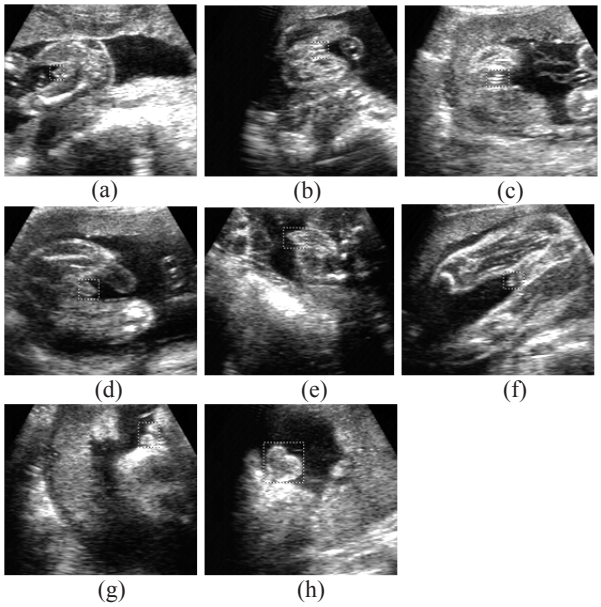

Fig.1 US images containing fetal genital organs (highlighted with a dashed square)
(a) Male at 20 weeks; (b) Female at 23 weeks; (c) Female at 25 weeks; (d) Male at 25 weeks; (e) Male at 27 weeks; (f) Male at 27 weeks; (g) Female at 31 weeks; (h) Male at 34 weeks

Representations of fetal genital organs in US images vary between different individuals, gestational ages, and fetus positions, and are affected by fetal activities and some maternal factors, including amniotic fluid volumes, placental positions, etc. Generally, the genital organ can be discriminated easily when it is surrounded by the amniotic fluid and its contour is in relief against the fluid. However, the genital organ can sometimes be obscured by excess uterine tissue in the vicinity. Furthermore, there is no clear boundary between the genital organs and other adjacent structures of the fetus; thus it is impractical to obtain the closed contour of the genital organ and recognize the genital organ as an independent object. Using the characteristics of the genital organs and inspired by the research of Buller *et al*.(1996), in this paper we construct a recognition and location algorithm for fetal genital organs based on the classification of the pixel, in the situation where no ROI can be predefined manually. Every pixel in the US images will be divided into two classes, in the fetal genital organ or out of the fetal genital organ. Moreover, constrained by the computation cost and the complexity of the task, it is impractical to execute the learning procedure on every pixel of the image, therefore this fast automatic algorithm will be globally separated into two stages. In the 'rough' stage, a few pixels in the image, which represent fetal genital organs with a high probability, are chosen according to certain salient characteristics of the genital organs, and no complicated learning procedure is adopted in this stage. The selected pixels are defined as pixels of interest (POIs). In the subsequent 'fine' stage, the selected POIs are classified by a specifically supervised learning framework, which fuses several effective ultrasound feature data preprocessing mechanisms into the multiple classifier architecture. Finally, all the classification results of the POIs are synthesized to generate the unique decision whether fetal genital organs are present in the images. Moreover, the method also locates the genital organ in the positive image.

Rough stage of the algorithm

The amniotic fluid nearby is usually the lowest intensity homogeneous region in US fetal images and provides a sharp contrast around the contour of the fetal genital organ. Therefore one of the most salient characteristics for distinguishing a fetal genital organ is the contour shape of the genital organ and its adjacent connected structures, e.g., thighs and buttocks. The contours of fetal genital organs are always smooth on a small scale since they represent the skin surface of the fetus. Furthermore, there are usually several notable high curvature regions (here referred to as corners) in and/or near the contours on a large scale, which are introduced by the intrinsic shape of the genital organ and/or the intersections of the genital organ and its adjacent structures. According to these predominant characteristics, the rough stage can be sub-divided: first, the contour of the genital organ is segmented and obtained automatically; second, the appropriate corner points on the contour

on a large scale are selected as the POIs.

The model-based approach, which includes mainly the active contour model and the active shape model, is the most successful segmentation approach for US images (Lu *et al*., 2003). However, it is still not able to remove completely the need for manual intervention and the computation cost is non-affordable using a fast algorithm. Jardim and Figueiredo (2005) presented an automatic segmentation and measurement of the contours of fetal femurs and cranial cross-sections in US fetal images based on the active shape model method. The model parameters were estimated from training images and no parameter adjustments were required in the application, but it was still necessary to mark an initial point manually at the beginning. Lu *et al*.(2005) also developed an automatic measurement of fetal head skulls in US images using an iterative randomized Hough transform (IRHT). Fetal head skulls were segmented by a *k*-mean algorithm, which classified each pixel in the image into different groups according to its intensity value, and morphologic *dilation* and *closing* were used to smooth the boundaries successively. IRHT was then used to deal with the gaps and extraneous tissues to obtain closed head contours. In these studies, the stability of the shape of the target object was used intensively for segmentation. However, fetal genital organs do not have stable shapes and even a closed contour is seldom obtained since there is usually no distinguishable boundary between the genital organ and the adjacent structures. In this paper, we propose and implement the first stage of a specific automatic segmentation algorithm based on the strong contrast in intensity between the genital organ and the surrounding amniotic fluid.

A reduced adaptive filtering algorithm based on image local statistics is first applied. Compared with the studies of Karaman *et al*.(1995) and Chen *et al*. (2003), the proposed filtering algorithm retains the analysis of speckle statistics but discards the function of region growing to reduce the computation cost. The speckle statistics of ultrasonic images in clinical US scanners are affected by the use of nonlinear processing, and the local mean might be proportional to the local variance rather than to the standard deviation (Loupas *et al*., 1989). The image-dependent characteristic value of speckle statistics can be measured in uniform speckle regions at different locations in the experimental images using different square region sizes. The pre-specified homogeneity threshold h_0 can be determined experimentally under an appropriate fixed square window W. For every pixel (i, j) in the processing US images, its local speckle statistics h_{ij} are calculated by the local variance to mean ratio in the local square window W_{ij}. When $h_{ij}<h_0$, the pixel is assumed to lie in a homogeneous background and the arithmetic mean filter is used to smooth it; otherwise, the nonlinear median filter is applied to preserve the edge information.

Although the amniotic fluid always lies in the lowest intensity area and the adjacent fetal genital organs are in much higher intensity areas in US images, the average intensity and the contrast of various fetal images from different pregnant women in different examinations are not consistent. To keep the images within a similar range of intensity levels and improve the contrast between the genital organ and the amniotic fluid, the histogram of the image is first stretched by a stepwise linear function. This sets the intensities of the 10% of the area with the lowest intensity pixels to 0, and the intensities of the 10% of the area with the highest intensity pixels to 255, and linearly stretches the intensities of the pixels occupying the remaining 80% of the area into a scale from 1 to 254. The corresponding histogram equalization is then applied to only the informative intensity scale (1, 254) for contrast enhancement. After that, the preprocessed image is binarized by the simple fixed threshold method, and morphological *closing* is operated to smooth the contours, filling the small gaps and narrow breaks in the contours (Gonzalez and Woods, 2002).

In the second sub-stage, the contours are first sampled at every five points to simulate the shape of the contours on a large scale, and then by adopting the eleven-point formula (Kadonaga and Abe, 1995), the approximate curvature of every point on the sampled contours is calculated. The corner points are selected as the POIs from the curvature extreme points under two rules: first, the curvature of the extreme point needs to exceed a threshold; second, there are no other extreme points nearby, which have larger curvature.

The ideal result of the rough stage should be that all the POIs lie in the genital organ but, of course, this is unrealistic. The performance measurement of the

rough stage is hard to define since it is not the final result of the complete algorithm. Generally, for every positive image processed by the rough stage, there must be at least one POI lying in the genital organ, because it is a basic input requirement of the subsequent POI classification stage. Therefore, the performance measurement could be defined as the ratio of the number of positive images which contain at least one POI in the genital organs to the number of all the processed positive images. An advanced reference standard can be defined as the ratio of the number of POIs that lie in the genital organs to the total number of POIs obtained. Considering the difficulty of the task, the basic standard is chosen as the performance measurement of the rough stage, in which the error rate is defined as the ratio of the number of positive images with no POIs located in the genital organs to the overall number of images testing positive.

Fine stage of the algorithm

Features are used to represent patterns with minimal loss of important information. In this paper, three classes of feature representations are selected for each POI. The first class is the contour shape of the fetal genital organ. The lengths of the contours of fetal genital organs against the amniotic fluid vary from several dozen to several hundred pixels in different individuals. For simplicity, in this paper, an appropriate fixed 150 pixel-length section of the original contour of the genital organ is set to represent the shape information. Thus, the first feature set, named CURV, is composed of the curvature values of the POI and every 15 sampled points from each side on the sampled contour, and the dimension of CURV is 31. The second class of feature representations comprises the textual and specified information on the fetal genital organ. The textual information is represented mainly by first-order statistics values, including the mean and variance, of the 60×60 pixel region around each POI. The mean and variance values of the object and background portions in the region are also recorded. The nearby crowding level is represented roughly by the area ratio of the object to the background. The skin information is represented by the average gradient value of the fixed 150 pixel-length section of the original contour around the POI. The maximum length of the successive pixels with a gradient value over 80 in the section of the contour is adopted as a useful feature. The gradient values are obtained by applying the Prewitt operator on the filtered image. The femur information is represented by the brightness and area values of the independent high intensity object nearby. The "the three-line sign" of the female fetus is represented by the length and numerical values of the parallel line-shape objects nearby, which are calculated mainly through line tracking in the gradient image. The height of the POI in the image is also a simple but helpful feature value. In total, 14 numeric values are collected to construct the second feature set (TEX). The third class of feature representations comprises the original intensity information of the fetal genital organ. The related feature set, named SAM, is composed of a cumbersome but effective sampled intensity matrix of the rectangular image region around the POI. The value of each element in SAM is derived from the average intensity of a small region near the POI. The area of the small region for sampling can be adjusted to represent the local fine structure or the global rough structure. In this paper, to represent the global structure around the genital organ (e.g., the adjacent thighs or buttocks), successive 18×18 pixel regions are sampled from an appropriate 180×180 pixel region around the POI in the original image to produce the feature element. This feature set is named SAM_L, and the dimension of SAM_L is 100. To represent the fine structure of the genital organ, successive 6×6 pixel regions are sampled from an appropriate 60×60 pixel region around the POI to produce the feature element. This feature set is named SAM_S, and the dimension of SAM_S is also 100.

From these four feature sets, a total of 245 feature values are extracted. Considering the distinct characters of the feature representations, it is unwise to handle them together in one classifier. Thus, we establish a learning system based on two-layer hierarchical multiple classifier architecture, in which the four feature sets (i.e., CURV, TEX, SAM_L, and SAM_S) are classified separately by four supervised classifiers.

Since the feature values are derived from the original intensities of the image pixels using the subjective decisions of the designers, some of them are either useless or redundant to the specific

classification problem and are removed to achieve higher classification accuracy. Also, a dimension disaster problem needs to be avoided carefully. The multiple classifier architecture, to some extent, can reduce the dimension of the input vector of the classifier. Feature selection (FS) is a widely used feature data preprocessing mechanism, which selects the subset of the features most suited to the classification problem. The filter is a typical class of FS, which evaluates one feature periodically and estimates its usefulness (Blum and Langley, 1997). To avoid overfitting to the specific filter method, in this paper we try to select the reduced feature set of TEX by combining the results of three popular filters: average mutual information, Pearson correlation coefficient (Guyon and Elisseeff, 2003) and the Kolmogorov-Smirnov test (Press and Flannery, 1992). The combining rule is defined as follows: a feature will be chosen when its performance in three filters is within the better half of the ranked feature array. However, for CURV, SAM_L, and SAM_S, it is inappropriate to select only part of the feature set since the characters of the elements in the feature set are similar and all the elements work together to represent the same feature information. Therefore principal component analysis (PCA) (Bishop, 1995; Johnson and Wichern, 1992), a popular feature extraction (FE) method, is applied to CURV, SAM_L and SAM_S. The largest 10 principal components of CURV and the largest 20 principal components of SAM_L and SAM_S are extracted to construct the reduced feature set.

Another intrinsic problem of the fetal genital organ recognition problem, which cannot be ignored, is the class imbalance problem. Class imbalance describes the situation in which one class in a dataset contains far fewer examples than the other classes (Barandela *et al*., 2003). Although it has long been known that class imbalance may cause a significant deterioration in the performance of the standard classifiers, in many studies it was not recognized as a systematic problem (Chawla *et al*., 2004). Since the positive images which contain fetal genital organs are only a minor portion of the total clinical US images, the genital organ recognition problem is a typical class imbalance issue. Moreover, the design of the POIs, which is adopted to make the algorithm automatic, will exacerbate the imbalance situation. One well-known group of techniques for solving class imbalance problems works by resampling the original training dataset (Chawla *et al*., 2002). Two simple but effective resampling algorithms are random oversampling, which oversamples the minority class training examples randomly, and random undersampling, which undersamples the majority class training examples randomly. In this paper, as another feature data preprocessing mechanism, a combination of these two resampling algorithms is adopted to reduce the impact of the class imbalance problem. The resampling rates of minority class oversampling and majority class undersampling are both chosen as 1. Therefore, the class imbalance level in the resampled training feature data will be reduced to one quarter of the original level.

The preprocessed feature sets will then be fed into the two-layer hierarchical multiple classifier architecture for classification. For each feature set, the corresponding first-layer classifier is chosen from three widely used classifiers: radial basis function (RBF) network, backpropagation (BP) network and support vector machine (SVM). The weighted sum rule is then selected as the classifier combination strategy (Kittler *et al*., 1998) to integrate the classification results of the four upper-layer classifiers.

When the classification result of every POI in the image is obtained, it is still necessary to combine all the results of the POIs to make the unique recognition decision. To restrict false positives (*FP*), one specific rule, similar to the max rule, is defined as follows: the recognition result of the image can be derived from the classification result of the POI which contains the largest output value. When the result of that POI is negative, the image will be classified as negative. When the result of that POI is positive, the image will be classified as positive and the fetal genital organ will be properly located around that POI.

EXPERIMENTS AND RESULTS

Image acquisition

US fetus examinations of volunteer pregnant women were performed at the Medical Service Center, Huidong, China. Every participating pregnant woman was informed and signed a consent

agreement. The fetal images were taken using a Toshiba Capasee SSA-220A US scanner equipped with a 3.5~5 MHz trans-abdominal convex probe under similar scanning conditions to maintain uniformity. The images were then captured and digitized to obtain the 768×576 pixel bitmap pictures using a Microview V200 video grabber card. From 2006 to 2007, a total of 658 positive images were collected from a total of 100 examinations in which practiced obstetricians were able to locate fetal genital organs. Among these images, the gestational age varied from 20 to 36 weeks and the sex ratio approximated 1:1. Typically the positive images are a coherent subset of all possible images, but the negative class is less well-defined as "everything else". In this study, another 500 fetal images without genital organs were collected as the negative images from the examinations. Since the learning procedure in the designed algorithm is executed on the POIs, the parts of POIs which do not locate the genital organ in the positive images are categorized as the negative examples in the learning procedure. Therefore, the negative images were used only as test data. Half of the positive images were randomly chosen and taken as the training set. The remaining positive images, together with all the negative images, were used as the test set. All the images were first reduced to 360×320 pixels to discard the fixed area containing meaningless non-image information. The hardware used was an Intel Pentium IV-2.4G with 512 M RAM, and the programming environment was Visual C++ 6.0 and MATLAB 7.0.

Performance measurement

Accuracy is widely used as the performance measure for classifiers. However, it is not appropriate when the classes in the dataset are imbalanced, because accuracy will yield biased conclusions by favoring the majority class (Batista *et al*., 2004). For instance, in a dataset where the majority class proportion is overwhelming at 99%, a classifier having an accuracy of 99% can be created easily by simply assigning every new example to the majority class. The class imbalance level of fetal genital organ recognition problem in the clinical situation may be expected to even surpass 99%. The ROC curve and geometric mean are two good representations of accuracy (Kubat and Matwin, 1997), whose commonality is that they are independent of the distribution of examples between classes. The geometric mean (*GM*) is used to maximize the accuracy of each of two classes by keeping the accuracies balanced (Barandela *et al*., 2003) and is defined as the square root of the product of *TP* and *TN*, where *TP* denotes the accuracy of the positive examples and *TN* is the accuracy of the negative examples.

Experimental procedure

Since the speckle statistics depend on the scanner specifications, the parameters of the reduced adaptive filter must be appropriately determined experimentally prior to its application. Speckle features are generally affected by many factors, among which measurement region size is the most important. In this study, for different window sizes, the variance and mean computations were carried out on speckle regions at different locations in a total 1158 experimental images and the results were averaged. The ratios of variance to mean for different window sizes are shown in Fig.2. The error bars in the figure represent the standard deviations of the averaged results. The results show that window sizes larger than 11×11 pixels can approximate the speckle statistics and the asymptotic value of the ratio of variance to mean of speckle is about 3. Considering the computation cost, the homogeneity threshold h_0 was chosen as 3 under the selected 11×11 pixel window size. If the pixel lay in the homogeneous region, a 5×5 pixel mean filter was applied; otherwise, a 3×3 pixel median filter was applied.

After filtering, the histogram of the image was stretched and equalized, and then the image was

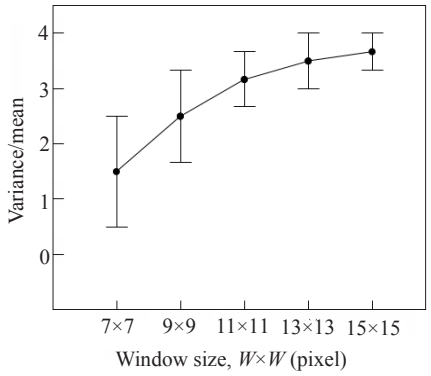

Fig.2 The local speckle statistics computed on different windows

binarized with the fixed experiential threshold value of 80. The structuring element in morphological operator *closing* was chosen as a square shape with a radius of 2 pixels. Meanwhile, considering that there are no tiny informative structures existing in the uterine region, the independent objects with an area of less than 30 pixels were discarded. Fig.3 represents the segmentation results from the eight original US fetal images in Fig.1. The contours of fetal genital organs against the amniotic fluid are well segmented.

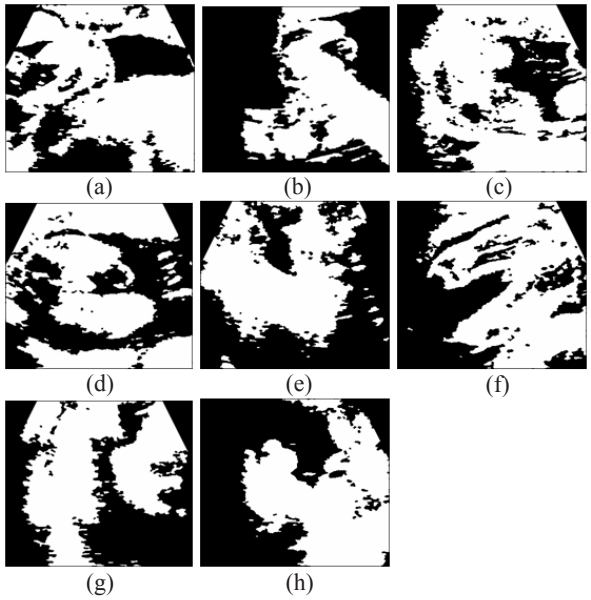

Fig.3 The segmentation results from the eight original US fetal images in Fig.1. Results from (a) Fig.1a; (b) Fig.1b; (c) Fig.1c; (d) Fig.1d; (e) Fig.1e; (f) Fig.1f; (g) Fig.1g; (h) Fig.1h

The important work of the second sub-stage of the rough stage is setting the rules for selecting the POIs from the curvature extreme points on the contours. Fig.4 represents the selection results from the rough stage, when the radius threshold was set to 5 and the distance threshold in the second rule was chosen as 10. The gray lines in the images represent the contours and the white points represent the POIs. The performance result of the rough stage was 95.0% (625 of 658) as judged by the basic standard and 1:20 as judged by the referenced advanced standard. The average computation time of the rough stage was 297 ms per 360×320 pixel image.

To search for the best configurations of the supervised learning framework in practical situations, for each feature set, we first investigated the performance results from using each of the two feature data preprocessing mechanisms and different first-layer classifiers. Five-fold cross validation was employed. The successive couples of *TP* and *TN* were collected to calculate the average *GM* value by changing the decision threshold of the classifiers under different parameters, and the largest *GM* value under certain parameters was chosen to represent the performance of the classifiers. All the experimental results are shown in Table 1, and the best performance result for each feature set is marked with a frame.

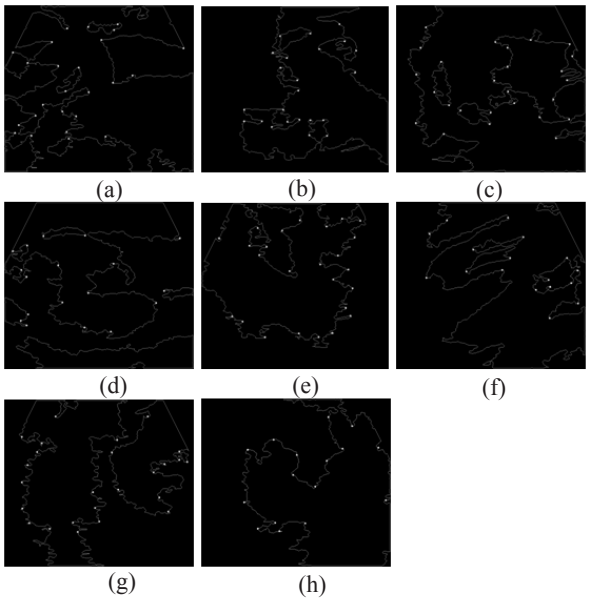

Fig.4 The selected POIs from Fig.1. Results from (a) Fig.1a; (b) Fig.1b; (c) Fig.1c; (d) Fig.1d; (e) Fig.1e; (f) Fig.1f; (g) Fig.1g; (h) Fig.1h

The configurations of the specific learning framework that gave the best performance result, together with the medium decision threshold of the corresponding classifier, were applied for each feature set. The outputs of the classifiers were further scaled to the range between 0 and 1, and were used to train the weighted sum rule. Then, based on the classification results of the POIs in the image, the recognition decision of that image was given using the combining rule defined above. The best weight rate of the four feature sets (i.e., CURV, TEX, SAM_L, and SAM_S) was (1, 4, 2, 2), when the

weight scale chosen was from 0 to 9. Testing the performance of the algorithm on all 829 test images, the value of *TP* was 80.5% (265 from 329 positive images), and the value of *TN* was 83.0% (415 from 500 negative images). Considering the error rate (17 from 329 positive images) introduced by the automatic POI selection stage, the performance of the proposed specific learning framework on the positive images should have been 84.9% (265 from 312 positive images left). The average computation time of the entire algorithm was 453 ms per image.

To validate the performance of the specific learning framework, the performance comparisons of the learning framework with the three basic classifiers, RBF, BP, and SVM, were adopted. The basic classifiers were processed on all 245 features, and 50 *GM* values were obtained by modifying the parameters of the classifiers successively; for the learning framework, 50 *GM* values were obtained by modifying the weights of the weighted sum rule. The Mann-Whitney U test was then carried out. The performance of the specific learning framework was better than that of the three basic classifiers' for this ultrasound image classification problem (Table 2).

CONCLUSIONS AND FUTURE STUDIES

In this paper, a fast automatic recognition and location algorithm for fetal genital organs is proposed as a useful and valuable tool to solve the troublesome prenatal sex-selective abortion issue. Based on an image dataset comprising 658 positive and 500 negative images, the experimental results showed *TP* and *TN* results for 80.5% and 83.0% of images respectively, and the average computation time on a normal PC was 453 ms per image, which present that the proposed algorithm is fast, automatic

Table 1 The *GM* values under different configurations of the specific supervised classification framework for each feature set

Feature set	Dimension reduction	Resampling	*GM* value		
			RBF	BP	SVM
CURV	Origin (31*)	Origin	49.30	68.27	68.44
		Resam (4**)	53.21	69.10	70.55
	FE (10)	Origin	54.83	67.51	68.36
		Resam (4)	57.59	69.02	70.87
TEX	Origin (14)	Origin	66.97	84.87	85.74
		Resam (4)	68.67	87.20	88.49
	FS (5)	Origin	68.52	88.49	88.80
		Resam (4)	70.70	88.76	89.01
SAM_L	Origin (100)	Origin	63.86	79.41	78.69
		Resam (4)	66.92	83.41	81.96
	FE (20)	Origin	63.27	78.49	78.63
		Resam (4)	65.17	80.93	79.68
SAM_S	Origin (100)	Origin	60.78	77.73	77.50
		Resam (4)	63.18	80.60	79.37
	FE (20)	Origin	61.05	75.35	76.14
		Resam (4)	63.79	77.38	77.31

* Dimension after dimension reduction; ** Resampling rate; The best performance result for each feature set is marked with a frame

Table 2 The statistical results for comparing the learning framework with BP, SVM, and RBF

Method	*GM* value			*U* value (*P*-value)*
	N	Mean	Standard deviation	
Learning framework	50	81.75	3.27	
BP	50	76.72	3.31	350 (<0.001)
SVM	50	71.95	8.78	202 (<0.001)
RBF	50	52.32	9.46	148 (<0.001)

* Compared with learning framework method

and accurate. Considering that the rough stage of the algorithm depends on the contrast between the fetal genital organ and the surrounding amniotic fluid, it is foreseen that the recognition performance will degrade when the surrounding amniotic fluid region is obscured by the crowding of other structures; however the evidence from this type of image is also difficult for obstetricians to assess accurately. Another potential problem in the rough stage is that some parameters and threshold values are image-dependent and scale-variant. This will, to some extent, prevent the proposed algorithm from being broadly applied on other US scanners. We are planning to make the algorithm more adaptive in future. In the fine stage, how to make the POIs "know" each other and "communicate" with each other earlier, before the last combining rule, is also on our to-do list. Restricted by the low quality and the complex contents of US images, even practiced US technicians cannot make an exact recognition decision based on only a single frame US image. Thus, it is unrealistic to expect that the performance of the recognition algorithm, which operates on a single image, could show a marked improvement. As a real-time medical image modality, the relative information between successive US frames provides potent evidence for technicians to recognize the image contents during examination. Understanding how to combine inter-frame relative information tracing with the proposed recognition algorithm is our most important future aim.

References

Barandela, R., Sanchez, J.S., Garcia, V., Rangel, E., 2003. Strategies for learning in class imbalance problems. *Pattern Recogn.*, **36**(3):849-851. [doi:10.1016/S0031-3203(02)00257-1]

Batista, G.E., Prati, R.C., Monard, M.C., 2004. A study of the behavior of several methods for balancing machine learning training data. *ACM SIGKDD Explor. Newsl.*, **6**(1):20-29. [doi:10.1145/1007730.1007735]

Bhanu Prakash, K.N., Ramakrishnan, A.G., Suresh, S., Chow, W.P., 2002. Fetal lung maturity analysis using ultrasound image features. *IEEE Trans. Inf. Technol. Biomed.*, **6**(1):38-45. [doi:10.1109/4233.992160]

Bishop, C.M., 1995. Neural Networks for Pattern Recognition. Oxford University Press, Oxford.

Blum, A., Langley, P., 1997. Selection of relevant features and examples in machine learning. *Artif. Intell.*, **97**(1-2):245-271. [doi:10.1016/S0004-3702(97)00063-5]

Buller, D., Buller, A., Innocent, P.R., Pawlak, W., 1996. Determining and classifying the region of interest in ultrasonic images of the breast using neural networks. *Artif. Intell. Med.*, **8**(1):53-66. [doi:10.1016/0933-3657(95)00020-8]

Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P., 2002. SMOTE: synthetic minority oversampling technique. *J. Artif. Intell. Res.*, **16**:321-357.

Chawla, N.V., Japkowicz, N., Kolcz, A., 2004. Editorial: special issue on learning from imbalanced data sets. *ACM SIGKDD Explor. Newsl.*, **6**(1):1-6. [doi:10.1145/1007730.1007733]

Chen, D.R., Chang, R.F., Chen, C.J., Ho, M.F., Kuo, S.J., Chen, S.T., Hung, S.J., Moon, W.K., 2005. Classification of breast ultrasound images using fractal feature. *Clin. Imaging*, **29**(4):235-245. [doi:10.1016/j.clinimag.2004.11.024]

Chen, Y., Yin, R., Flynn, P., Broschat, S., 2003. Aggressive region growing for speckle reduction in ultrasound images. *Pattern Recogn. Lett.*, **24**(4-5):677-691. [doi:10.1016/S0167-8655(02)00174-5]

Chrzanowski, L., Drozdz, J., Strzelecki, M., Krzeminska-Pakula, M., Jedrzejewski, K.S., Kasprzak, J.D., 2008. Application of neural networks for the analysis of intravascular ultrasound and histological aortic wall appearance—an in vitro tissue characterization study. *Ultrasound Med. Biol.*, **34**(1):103-113. [doi:10.1016/j.ultrasmedbio.2007.06.021]

Gonzalez, R.C., Woods, R.E., 2002. Digital Image Processing, 2nd Ed. Prentice Hall, New Jersey.

Guyon, I., Elisseeff, A., 2003. An introduction to variable and feature selection. *J. Mach. Learn. Res.*, **3**(7-8):1157-1182. [doi:10.1162/153244303322753616]

Jardim, S., Figueiredo, M., 2005. Segmentation of fetal ultrasound images. *Ultrasound Med. Biol.*, **31**(2):243-250. [doi:10.1016/j.ultrasmedbio.2004.11.003]

Johnson, R.A., Wichern, D., 1992. Applied Multivariate Statistical Analysis. Prentice Hall, New Jersey.

Kadonaga, T., Abe, K., 1995. Comparison of Methods for Detecting Corner Points from Digital Curves. Proceedings of the First International Workshop on Graphic Recognition. Aug. 10~11, University Park, PA, USA, p.23-34.

Karaman, M., Kutay, M., Bozdagi, G., 1995. An adaptive speckle suppression filter for medical ultrasonic imaging. *IEEE Trans. Med. Imaging*, **14**(2):283-292. [doi:10.1109/42.387710]

Kittler, J., Hatef, M., Duin, R.P.W., Matas, J., 1998. On combining classifiers. *IEEE Trans. Pattern Anal. Mach. Intell.*, **20**(3):226-239. [doi:10.1109/34.667881]

Kubat, M., Matwin, S., 1997. Addressing the Curse of Imbalanced Training Test: One-Sided Selection. Proceedings of the Fourteenth International Conference on Machine Learning. Jul. 8~12, Nashville, Tennessee, USA, p.179-186.

Li, S., 2007. Imbalanced Sex Ratio at Birth and Comprehensive Intervention in China. The 4th Asia Pacific Conference on Reproductive and Sexual Health and Rights,

Oct. 29~31, Hyderabad, India. Available from: http://www.unfpa.org/gender/docs/studies/china.pdf

Loupas, T., McDicken, W.N., Allan, P.L., 1989. Adaptive weighted median filter for speckle suppression in medical ultrasonic images. *IEEE Trans. Cir. Syst.*, **36**(1):129-135. [doi:10.1109/31.16577]

Lu, C., van Gestel, T., Suykens, J.A.K., van Huffel, S., Vergote, I., Timmerman, D., 2003. Preoperative prediction of malignancy of ovarian tumors using least squares support vector machines. *Artif. Intell. Med.*, **28**(3):281-306. [doi:10.1016/S0933-3657(03)00051-4]

Lu, W., Tan, J., Floyd, R., 2005. Automated fetal head detection and measurement in ultrasound images by iterative randomized Hough transform. *Ultrasound Med. Biol.*, **31**(7):929-936. [doi:10.1016/j.ultrasmedbio.2005.04.002]

Piliouras, N., Kalatzis, I., Dimitropoulos, N., Cavouras, D., 2004. Development of the cubic least squares mapping linear-kernel support vector machine classifier for improving the characterization of breast lesions on ultrasound. *Comput. Med. Imaging Graph.*, **28**(5):247-255. [doi:10.1016/j.compmedimag.2004.04.003]

Population Census Office under Chinese State Council, 2002. Tabulation on the 2000 Population Census of the People's Republic of China. China Statistics Press, Beijing (in Chinese).

Press, W.H., Flannery, B.P., 1992. Numerical Recipes in FORTRAN: The Art of Scientific Computing, 2nd Ed. Cambridge University Press, Cambridge, p.617-620.

Qi, Y., Mason, W.M., 2005. Prenatal sex-selective abortion and high sex ratio at birth in rural China: a case study in Henan Province. California Center for Population Research, On-Line Working Paper Series. Available from: http://repositories.cdlib.org/ccpr/olwp/ CCPR-057-05

Scholly, T.A., Sutphen, J.H., Mackey, S.C., Langstaff, L.M., 1980. Sonographic determination of fetal gender. *AJR*, **135**(6):1161-1165.

Sivaramakrishna, R., Powell, K.A., Leiber, M.L., Chilcote, W.A., Shekhar, R., 2002. Texture analysis of lesions in breast ultrasound images. *Comput. Med. Imaging Graph.*, **26**(5):303-307. [doi:10.1016/S0895-6111(02)00027-7]

Smutek, D., Tjahjadi, T., Sara, R., Svec, M., Sucharda, P., Svacina, S., 2001. Image Texture Analysis of Sonograms in Chronic Inflammations of Thyroid Gland. Research Report CTU-CMP-2001-15. April 2001, Czech Technical University, Prague, Czech Republic, ISSN 1213-2365. Available from: ftp://cmp.felk.cvut.cz/pub/cmp/articles/sara/Smutek-TR-2001-15.pdf

Strzelecki, M., Materka, A., Drozdz, J., Krzeminska-Pakula, M., Kasprzak, J.D., 2006. Classification and segmentation of intracardiac masses in cardiac tumor echocardiograms. *Comput. Med. Imaging Graph.*, **30**(2):95-107.

Vince, D.G., Dixon, K.J., Cothren, R.M., Cornhill, J.F., 2000. Comparison of texture analysis methods for the characterization of coronary plaques in intravascular ultrasound images. *Comput. Med. Imaging Graph.*, **24**(4):221-229. [doi:10.1016/S0895-6111(00)00011-2]

Zhu, Y., Williams, S., Zwiggelaar, R., 2006. Computer technology in detection and staging of prostate carcinoma: a review. *Med. Image Anal.*, **10**(2):178-199. [doi:10.1016/j.media.2005.06.003]