Texture-based instrument segmentation in 3D ultrasound images

Marius George Linguraru*, Robert D. Howe

Division of Engineering and Applied Sciences, Harvard University
Cambridge MA, 02138 USA
http://biorobotics.harvard.edu

ABSTRACT

The recent development of real-time 3D ultrasound enables intracardiac beating heart procedures, but the distorted appearance of surgical instruments is a major challenge to surgeons. In addition, tissue and instruments have similar gray levels in US images and the interface between instruments and tissue is poorly defined.. We present an algorithm that automatically estimates instrument location in intracardiac procedures. Expert-segmented images are used to initialize the statistical distributions of blood, tissue and instruments. Voxels are labeled of voxels through an iterative expectation-maximization algorithm using information from the neighboring voxels through a smoothing kernel. Once the three classes of voxels are separated, additional neighboring information is used to give spatial information based on the shape of instruments in order to correct for misclassifications. We analyze the major axis of segmented data through their principal components and refine the results by a watershed transform, which corrects the results at the contact between instrument and tissue. We present results on 3D in-vitro data from a tank trial, and 3D in-vivo data from a cardiac intervention on a porcine beating heart. The comparison of algorithm results to expert-annotated images shows the correct segmentation and position of the instrument shaft.

Keywords: 3D ultrasound, echocardiography, surgical instrument, segmentation, expectation-maximization, principal component analysis, watershed transform.

1. INTRODUCTION

The recent development of clinical real-time 3D ultrasound (US) enables new intracardiac beating heart procedures[2,11], avoiding the use of cardiopulmonary bypass with its attendant risks. Unfortunately, these images are difficult for the surgeon to interpret, due to poor signal-to-noise ratio and the distorted appearance of surgical instruments within the heart. Automated techniques for tracking instrument location and orientation would permit augmentation of intraoperative displays for the surgeon as well as enable image-based robotic instrument control.

Most work in cardiac imaging relates to heart visualization and diagnosis and not to cardiac interventions[1,7,13]. In particular, tissue texture analysis has been used previously to segment cardiac images[13], but little work has been devoted to correcting the distorted appearance of instruments under ultrasound imaging. Tissue and instruments have similar gray levels in US images, which makes their correct delineation difficult. Shadows, reverberations and tip artifacts can puzzle even the trained eye, and the fuzzy interference between instruments and tissue is confusing to the surgeon.

Ortmeier *et al.*[9] track instruments for visual servoing. Their experimental setup consists of various material graspers (PVC, nylon, polyurethane) in a water tank. The position and shape of their instruments are known a priori. The head of the grasper is identified using thresholding and morphological operators. Although fast and simple, the method identifies the tips of the instrument in an in-vitro setting without the presence of tissue.

*Address for correspondence: mglin@deas.harvard.edu.

Novotny *et al.*[8] localize an acetal rod in a tank setup, but in the presence of tissue. The instrument is identified from its principal component analysis (PCA) as the longest and thinnest structure in the US image. However, the instrument and tissue are not in contact, which would change the geometry of the connected components in the image.

Our work includes instruments as well as tissue and blood in a texture-based analysis and eliminated false positives (FP) using a priori shape information. Section 2 of the paper describes the methodology of the instrument segmentation, a combination of expectation-maximization (EM), principal component analysis (PCA) and watershed transform (WT) algorithms. In Section 3 we present results on detecting instruments of various materials from tank studies and clinical interventions that demonstrate the successful delineation of both the position and orientation of the instrument.

2. METHOD

The algorithm consists of three major steps to segment surgical instruments in echocardiographic images. First, it estimates from expert-segmented intracardiac images the gray-level distributions of blood, tissue and instruments. It then builds averaged probability distribution functions for the three classes and labels image voxels through an iterative EM algorithm. The next step analyzes the major axis of the labeled connected components of the image through their principal components. The results are refined by a watershed transform by immersion, which corrects the errors at the contact between instrument and tissue.

2.1 Expectation-maximization

In the EM algorithm[3], the maximum likelihood parameters are computed iteratively starting with the initial estimation. The algorithm converges to a steady state once a local maximum is reached. At each iteration, there is:

- an expectation step: the unobserved variables are estimated from the observed variables and the current parameters;
- a maximization step: the parameters are re-evaluated to maximize likelihood, assuming that the expectation is correct.

We employ the EM algorithm to solve a mixture estimation problem and separate blood, parenchyma and instruments. We express the distribution function as a sum of three Gaussians:

$$f(x) = \sum_{n=1}^{3} \pi_n \cdot G(x\,;\mu_n;\sigma_n)\,, \tag{1}$$

where the parameters of the Gaussians are:

$$\sum_{n=1}^{3} \pi_n = 1\,,\ \mu_n = \frac{1}{N}\sum_{i=1}^{N} x_i\,,\ \sigma_n^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu_n)^2\,. \tag{2}$$

The parameters of the three distributions are computed in the expectation step. Let $X = (x_1, x_2, \ldots, x_N)$ be the sequence of observations from the mixture of three Gaussians and $\theta = \{\pi_1, \pi_2, \pi_3, \sigma_1, \sigma_2, \sigma_3, \mu_1, \mu_2, \mu_3\}$ the parameters that must be estimated from X. The values of the parameters π, σ and μ will be updated with each iteration until the algorithm converges. The likelihood maximization becomes that of maximizing

$$\ln P(X \mid \theta) = \sum_{i=1}^{N} \ln P(x_i \mid \theta)\,. \tag{3}$$

In a simplified presentation, the EM algorithm tries to determine the probability of a voxel belonging to one of the three defined classes Γ_1, Γ_2, Γ_3. Using the parameters θ and Bayes's law we can compute $P(x_i \in \Gamma_n)$ as

$$P(x_i \in \Gamma_n) = \frac{\pi_n \cdot G(x_i; \mu_n; \sigma_n)}{\sum_j \pi_j \cdot G(x_i; \mu_j; \sigma_j)} \,. \tag{4}$$

The maximization step will update the values of the parameters

$$\mu_n = \frac{\sum_i P(x_i \in \Gamma_n) \cdot x_i}{\sum_i P(x_i \in \Gamma_n)} \,,\; \pi_n = \frac{\sum_i P(x_i \in \Gamma_n)}{\sum_{i,n} P(x_i \in \Gamma_n)} \,,\; \sigma_n^2 = \frac{\sum_i P(x_i \in \Gamma_n) \cdot (x_i - \mu_n)^2}{\sum_i P(x_i \in \Gamma_n)} \,. \tag{5}$$

The initial estimate of the three classes is extracted from expert-segmented images. This provides the statistical distributions of blood, parenchyma and instruments, as in Figure 1. Each observation is then approximated from the neighboring voxels through a smoothing kernel. The histograms of the three classes are not fully separated, as seen in Figure 1. The main uncertainty occurs at the overlap between tissue and instrument intensities.

As the separation between blood and the other two classes is straightforward, the final goal of the algorithm is to split the uncertainty class between tissue and instrument. Voxels are labeled iteratively until the rate of change between iterations becomes smaller than a given limit. Due to the larger number of tissue voxels in the observation, the maximum likelihood parameters tend to give priority to tissue over instrument. For that reason, we gave larger weights to the instrument distribution.

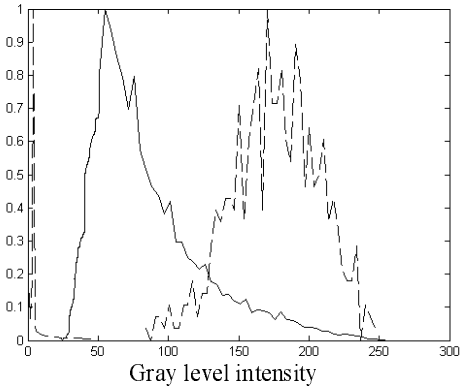

Figure 1: The normalized histograms of blood (dashed-left), parenchyma (middle) and instrument (dashed-right). The example shows the histogram of a wood instrument. Note the overlap between the tissue and instrument histograms, the main source of uncertainty in the EM algorithm.

2.2 Neighboring Information

The second stage of the method uses spatial measures based on the shape of instruments to correct for misclassified voxels. The assumption imposed here is that instruments are long and thin. We employ three kernels, as seen in Figure 2:

- μ_C : a small central kernel around the voxel (3x3x3 voxels);
- μ_N : a larger neighboring kernel (7x7x7 outside the central kernel);
- μ_B : a large background kernel (21x21x21).

The relations between the three kernels aim to correct for voxels labeled incorrectly. The situations accounted for are:

- small areas mislabeled as blood within large areas labeled as instrument or tissue;
- small areas misclassified as tissue within the instrument shaft or tip;
- small areas misclassified as instrument within large areas of tissue.

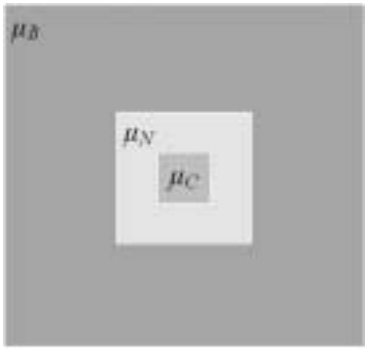

Figure 2: The kernel used for neighboring information. μ_C is the central kernel, μ_N is the neighboring kernel, while μ_B is the background kernel.

2.3 Principal Component Analysis

Next, principal component analysis (PCA)[6] is used to detect candidates for the instrument shaft. A practical way to compute principal component is by extracting the eigenvalues from the covariance matrix of the data. Eigenvalues represent the projected variances corresponding to principal components. In our application, for an image Im, the covariance matrix is computed as

$$C_{\mathrm{Im}} = \left\langle (\mathrm{Im} - \mu_{\mathrm{Im}})(\mathrm{Im} - \mu_{\mathrm{Im}})^T \right\rangle. \tag{6}$$

Under the assumption that an instrument is long and thin, its first principal component would account for most of its variation. Hence, if the instrument eigenvalues are calculated as

$$\left| C_{\mathrm{Im}} - \lambda I \right| = 0; \quad F = \frac{\lambda_1}{\lambda_2}, \tag{7}$$

the first eigenvalue λ_1 will be dominant and the ratio F would have a maximal value, where λ_2 is the second eigenvalue.

Before performing PCA, the image is separated in connected components. For each connected component, we computed the eigenvalues of the covariance matrix and discard those objects with small ratio F. The objects with the highest ratio F are instrument candidates. If there is only one instrument in the image, it will have the maximal F. Due to estimation errors in EM at the contact between instrument and tissue, we prefer to keep more candidates and discard the remaining false positives in the next stage.

2.4 Watershed Transform

To account for errors at the contact between instrument and tissue, we use a watershed (WT) by immersion algorithm[5,10,12] that assumes that there is only one instrument in every image. WT is a morphological tool that analyses the topography of the image based on its gray level. It presumes that water penetrates minima and floods all areas below the water level to form basins. Eventually basins start merging at watershed lines. Figure 4 shows an example of water flooding, where minima are presented as dark circles separated by watershed lines.

WT is applied to the instrument candidates resulting from the previous PCA and using the estimate of instrument statistics from EM. First, the Euclidean intensity distances to the upper quartile of the instrument intensity are calculated, then a gradient of the newly computed image is used to initialize local minima.

$$\mathrm{Im} = \max(\mathrm{Im}) - \left|\mathrm{Im} - \left(\mu_{instr} + \sigma_{instr}\right)\right|; \qquad \mathrm{Im} = \left|G'_{\sigma}(\mathrm{Im})\right|. \tag{8}$$

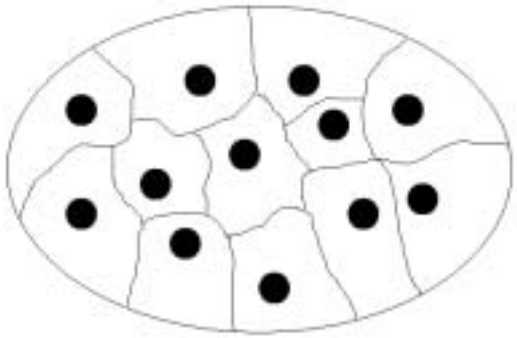

Figure 4: Watershed segmentation. Local minima are represented as dark circles. Once the basins are flooded, water merges at the watershed lines that separate local minima.

The computation of watershed lines is done through two topographic measures: the lower slope (*LS*) and the lower neighborhood (*LN*).

$$LS\left(x \in \mathrm{Im}\right) = \max_{y \in N_x}\left(\frac{\mathrm{Im}(x) - \mathrm{Im}(y)}{d(x,y)}\right); \tag{9}$$

$$LN(x \in \mathrm{Im}) = \left\{ y \in N_x \mid \frac{\mathrm{Im}(x) - \mathrm{Im}(y)}{d(x,y)} = LS(x) \right\}; \tag{10}$$

LS maximizes the relation of voxels y in the neighborhood N_x of voxel x, where $d(x,y)$ is the Euclidian distance between the locations of voxels x and y. *LN* identifies the neighborhood of x that satisfies equation (10). The process iteratively expands the water basins. For every basin, the value of ratio F (see equation (7)) is evaluated. If F decreases below a limit with the water expansion at the next level, than the process is stopped and the basin becomes an instrument candidate. Otherwise the transform continues until the entire total topography is flooded. Finally, the candidate with maximum F is labeled as instrument.

3. RESULTS

The algorithm was first tested using 3D in-vitro data from a water tank trial. The training sets and test data were acquired under similar imaging conditions using the same type of instrument. All data was acquired with a Sonos 7500 Live 3D Echo scanner (Philips Medical Systems, Andover, Mass., USA). An acetal rod was placed in the tank in the proximity of a flat bovine muscle tissue sample approximately 2 cm thick. There was no contact between the instrument and the tissue. Figure 5 shows detection results at various algorithm stages in 2D slices. In Figure 6 we present the 3D volumes of in-vitro instrument detection.

The second segmentation example detects a wooden instrument. The image was acquired during an intracardiac operation on a porcine beating heart. This in-vivo US image shows the wooden rod in contact with the cardiac tissue. Figure 7 shows the detection stages in in-vivo images. After PCA, the instrument is generally well detected, but FP occur at the contact with tissue. WT eliminates the residuals and segments the instrument correctly. Figure 8 presents the 3D US segmentation results.

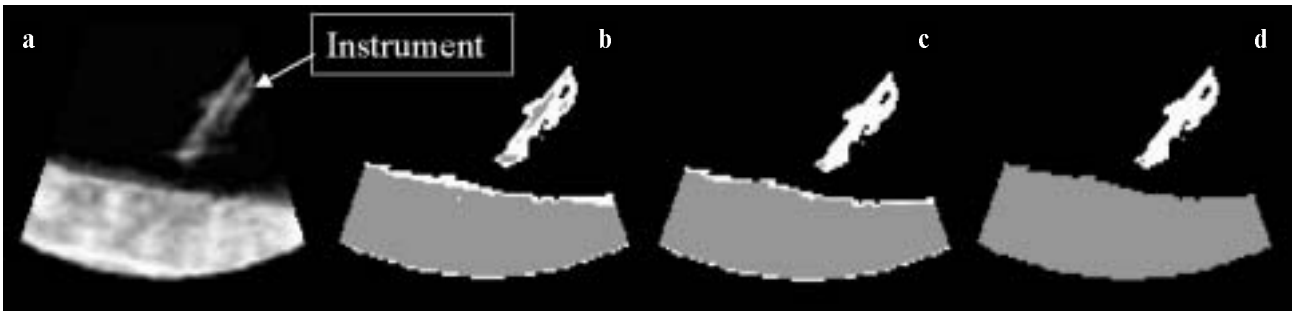

Figure 5: Tank data: (a) a 2D slice of a 3D US image of an acetal rod approaching a tissue sample in a water tank; (b) detection results after EM, where the instrument is shown in white and the tissue in gray; (c) detection results employing neighboring information; (d) detection results after PCA.

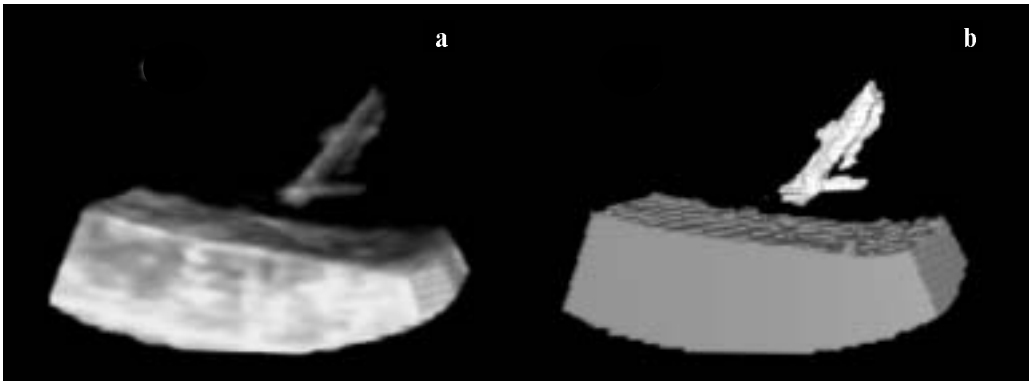

Figure 6: 3D in-vitro data: (a) the US image of an acetal rod approaching a tissue sample in a water tank; (b) detection results where the instrument is shown in white and the tissue in gray.

Comparison with expert-annotated images shows the correct segmentation and position of the instrument shaft oriented in all situations. The instrument orientation is extracted from its principal components.

4. DISCUSSION

This algorithm consists of several steps that lead to the detection of instruments in US images. It begins with finding an estimate of the statistical distributions of the classes of objects we aim to segment. Our database consists of images acquired at different times and locations, but on the same type of US machine. The method is not sensitive to image acquisition conditions. It suffices to determine the distribution functions of blood and tissue once for all the subsequent segmentations. However, the instrument probability distribution function (pdf) is unique to the type of material it is made from. Hence, for each type of instrument material, there will be a different pdf and the method requires a small database of material types and statistics. The application allows the selection of the instrument material in the user interface.

Next, the image voxels are classified and labeled into three classes: blood, parenchyma and instrument. The statistical classification converges to three stable classes of voxels, but the uncertainty at the overlap of the Gaussian distributions of instrument and tissue brings labeling errors. Hence, some spatial information is required. Using simple neighboring information proved to be very helpful for in-vitro data. For complex in-vivo data, this type of simple spatial information improves the segmentation of the instrument but introduces errors in tissue.

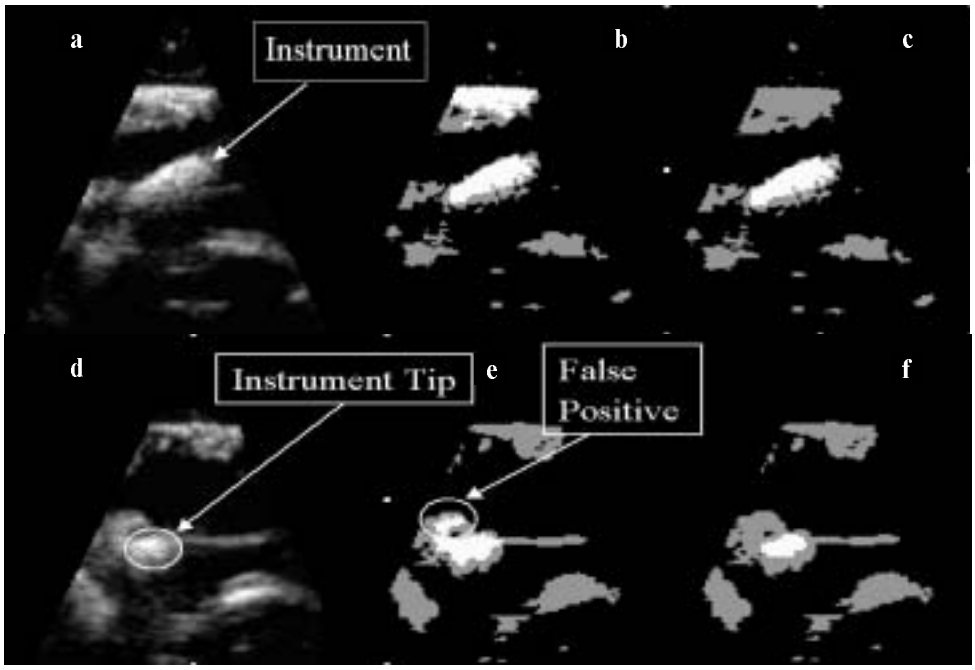

Figure 7: In-vivo data: (a) a 2D slice of a 3D US image of a porcine heart with a wooden rod inside acquired during a clinical procedure; (b) detection results after the analysis of intensity distributions, where the instrument appears white, tissue gray and blood black; (c) the segmentation of instrument and tissue after PCA: the instrument appears correctly segmented, but this is not the case for the whole 3D US image; (d) a slice of the 3D US image at a different location, where the tip of the instrument touches the tissue; (e) detection results after PCA at the new location showing FP; (f) improved results employing the watershed algorithm.

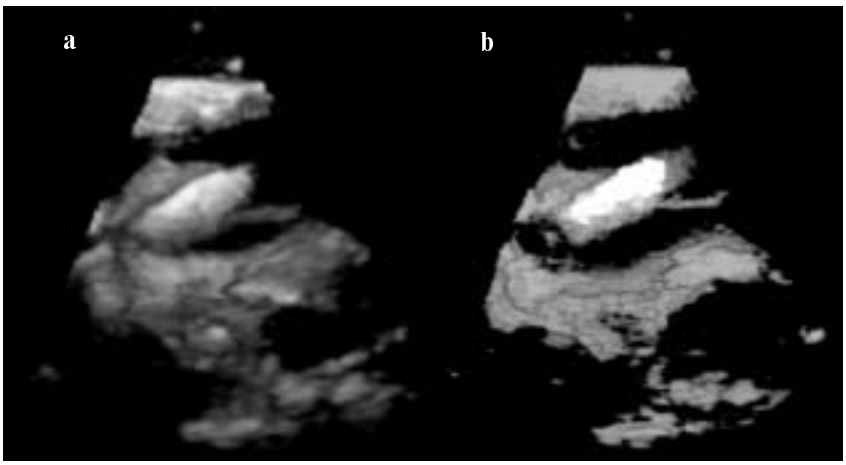

Figure 8: 3D in-vivo results: (a) a 3D US image of a porcine heart with a wooden rod inside (in contact with the tissue); (b) 3D segmentation results for image (a) with the instrument shown in white, tissue in gray and blood in black.

PCA is employed as a statistical measure. The analysis is based on the assumption that instruments are shaped like rods, long and thin. We analyze the connected components in the image and keep only the candidates that satisfy the shape requirements. The major source of remaining errors is the contact area between instrument and tissue. Finally, WT separates the instrument from tissue in critical areas on a fine to coarse approach. This is a WT-PCA hybrid approach. The previous stage built on PCA reduces the computational expense and speeds up the algorithm. Watershed reduces the segmented volume of instrument closer to the instrument shaft oriented towards the US probe. This produces

smaller errors in the approximation of the instrument axis. To this extent, we assume that there is only one instrument in every image.

Our segmentation results are 3D, but a major application results from employing temporal information of instrument movement from 4D US clinical images. A priori information on the size of the instrument (length and diameter), which combined with the closest image boundary to the instrument and an estimation of the instrument tip can be used to introduce in the image a model of a 3D rod mimicking the ideal appearance of the instrument.

Image normalization is known to reduce the deviation of the point-spread functions of object classes. We will investigate the impact of normalization to make the statistical classification more robust. Neighborhood analysis and connectivity through Markov Random Fields should correct segmentation errors.

5. CONCLUSION

We presented an algorithm for the segmentation of instruments in 3D ultrasound images. Through a combination of statistical analysis of texture and shape, we obtained very promising results in in-vivo data, where instruments appear confusing in the presence of and in contact with tissue.

For future work, more spatial and statistical information using the instrument shape and principal factor analysis[4] will be implemented for a larger variety of instrument materials. Our segmentation results can be used as an initialization step for tracking instruments in 4D echocardiography. The computational speed will be a priority for real-time interventional applications.

ACKNOWLEDGEMENTS

This work is funded by the National Institutes of Health under grant NIH R01 HL073647-01. The authors would like to thank Dr. Nikolay Vasilyev from Children's Hospital in Boston and Dr. Ivan Salgo from Philips Medical Systems for assistance with image acquisition.

REFERENCES

1. E.D. Angelini, S. Homma, G. Pearson, J.W. Holmes, A.F Laine: Segmentation of real-time three-dimensional ultrasound for quantification of ventricular function: a clinical study on right and left ventricles. In: Ultrasound Med Biol., Vol. 1(9) (2005) 31143-58
2. J.W. Cannon, J.A. Stoll, I.S. Salgo, H.B. Knowles, R.D. Howe, P.E. Dupont, G.R. Marx, P.J. del Nido: Real Time 3-Dimensional Ultrasound for Guiding Surgical Tasks. In: Computer Aided Surgery, Vol. 8 (2003) 82-90
3. C. Couvreur: The EM Algorithm: A Guided Tour. In IEEE European Workshop on Computationally Intensive Methods in Control and Signal Processing (1996) 115-120
4. M.A. Gonzalez Ballester, M.G. Linguraru, M. Reyes Aguirre, N. Ayache: On the Adequacy of Principal Factor Analysis for the Study of Shape Variability. In: Fitzpatrick, J.M. and Reinhardt, J.M. (eds.), SPIE Medical Imaging 2005 Vol. 5747 (2005) 1392-1399
5. V. Grau, A.U.J. Mewes, M. Alcaniz, R. Kikinis, S.K Warfield: Improved Watershed Transform for Medical Image Segmentation Using Prior Information. In: IEEE Transactions on Medical Imaging, Vol. 23(4) (2004) 447-458
6. J.E. Jackson: A User's Guide to Principal Components. John Wiley & Sons, Inc., New York (1991)
7. N. Lin, W. Yu, J.S. Duncan: Combinative Multi-Scale Level Set Framework for Echocardiographic Image Segmentation. In: Medical Image Analysis, Vol. 7(4) (2003) 529-537
8. P.M. Novotny, J.W. Cannon, R.D. Howe: Tool Localization in 3D Ultrasound. In: Medical Image Computing and Computer Assisted Intervention (MICCAI) (2003) 969-970

9. T. Ortmaier, M.A. Vitrani, G.Morel, S. Pinault: Robust Real-Time Instrument Tracking in Ultrasound Images for Visual Servoing. In: Fitzpatrick, J.M. and Reinhardt, J.M. (eds.), SPIE Medical Imaging 2005 Vol. 5750 (2005) 170-177
10. J.B.T.M. Roerdinck, A. Meijster: The Watershed Transform: Definitions, Algorithms and Parallelization Strategies. In: Fundamenta Informaticae, Vol. 41 (2000) 187-228
11. L.M. Shapiro, A. Kenny: Cardiac Ultrasound. Manson Publishing (1998)
12. L. Vincent, P.Soille: Watersheds in Digital Spaces: An Efficient Algorithm Based on Immersion Simulations. In: IEEE Tran. Pattern Analysis and Machine Intelligence, Vol. 13(6) (1991) 583-598
13. G. Xiao, J.M. Brady, J.A. Noble: Segmentation of Ultrasound B-Mode Images with Intensity Inhomogeneity Correction. In: IEEE Trans. Medical Imaging, Vol. 21(1), 48-57