Cascaded Shape Regression for Automatic Prostate Segmentation from Extracorporeal Ultrasound Images

Jierong Cheng, Wei Xiong, Ying Gu, Shue Ching Chia, and Yue Wang

Institute for Infocomm Research
1 Fusionopolis Way, #21-01 Connexis (South Tower)
Singapore 138632

Abstract. Prostate segmentation from extracorporeal ultrasound (ECUS) images is considerably challenging due to the prevailing speckle noise, shadow artifacts, and low contrast intensities. In this paper, we proposed a cascaded shape regression (CSR) method for automatic detection and localization of the prostate. A sequence of random fern predictors are trained in a boosted regression manner. Shape-indexed features are used to achieve invariance against geometric scales, translation, and rotation of prostate shapes. The boundary detected by CSR is used as the initialization for accurate segmentation by using a dynamic directional gradient vector flow (DDGVF) snake model. DDGVF proves to be useful to distinguish desired edges from false edges in ECUS images. The proposed method is tested on both longitudinal- and axial- view ECUS images and achieves Root Mean Square Error (RMSE) under 1.98 mm (=4.95 pixels). It outperforms the active appearance model in terms of RMSE, failure rate, and area error metrics. The testing time of CSR+DDGVF is less than 1 second per image.

Keywords: Cascaded regression, prostate segmentation, random ferns, dynamic directional gradient vector flow

1 Introduction

High Intensity Focused Ultrasound (HIFU) is being used throughout the world as a therapeutic procedure for prostate cancer and benign prostate hyperplasia (BPH). An important component in BPH removal using HIFU is to position and focus on the targeted prostate tissue. Extracorporeal ultrasound (ECUS) images are usually noisier than the transrectal ultrasound (TRUS) images. Therefore, accurate automatic prostate segmentation from ECUS images faces considerable challenges. Numerous prostate segmentation methods have been developed in literature, either for TRUS, MR, or CT images (see [1] for an extensive review).

According to the information used to guide the segmentation, the prostate segmentation methods can be classified into four groups [1]: contour and shape based method, region based methods, supervised and un-supervised classification methods, and hybrid methods. Since edge information is unreliable and edges

are even broken in ultrasound images, the use of the first group of methods, e.g. active contour model (ACM) [6], [5], and curve fitting [7], [8] alone are often ineffective. Region based methods such as graph partitioning [10] and regional level set [9], solve the segmentation problem in an energy minimization framework. The popular regional level set [11] relies on region homogeneity, which is often violated due to artifacts and dropouts in ultrasound images, and generates fragmented regions. Classification methods cluster [12] or classify [13] the pixels into the prostate or the background based on feature vectors. To produce accurate segmentations, the above methods are often combined into hybrid methods so that the segmentations are more robust to artifacts and noises.

A widely used approach is to match statistical shape models to images to locate points on deformable objects. Cootes et al. proposed the active shape model (ASM) [14] which maintains the principal modes of shape variations in a deformable model framework. The later active appearance models (AAM) [15] combine models of both shape and texture using an efficient parameter update scheme. One of the limitations of parametric shape model approaches is that minimizing model parameter errors in the training set is indirect and suboptimal [3]. Moreover, ACM and AAM based methods need good initialization since they are local optimization. The linear regression used in the original AAM may be insufficient to capture the variance of shape and appearance of the prostate in ECUS images. Non-linear regression based matching methods have been introduced using boosted regression [17], [18] and random forest regression [19]. Zhou proposed shape regression machine (SRM) which uses image-based boosting regression for left ventricle segmentation from echocardiogram [16]. Sequences of random fern predictors have been used in a cascaded way for face alignment [2], [3]. Recently, regression based voting approaches [20], [21] show efficiency in locating facial feature points accurately.

In this work, we propose a cascaded shape regression (CSR) method for efficient prostate detection and localization with the shape being represented by a sequence of sampled points on the prostate contours. The advantages of the CSR are:

- The alignment error is explicitly minimized during the training of regressors, instead of minimizing model parameters which is indirect.
- The regressed shapes are constrained by the linear subspace constructed by all training shapes. We need no parameter tuning to estimate the variation of shapes in the regression model.
- The initialization is fully automatic. We use the average of all training shape and the true shape of other training samples to initialize the CSR for training, assuming that the training samples well represent the possible location, rotation and scale of the prostate shapes. Afterwards, the CSR is simply initialized by the average of all training shapes during testing.

To achieve accurate prostate segmentation, the CSR results have to be refined. Following the CSR, the dynamic directional gradient vector flow (DDGVF) snakes [23] is adopted to optimize the detected shape boundary. DDGVF is a type of external force model which endows the snake/active contour model

the ability to discern edges of different orientations dynamically during the contour deformation. This property is very useful to distinguish desired edges from false edges in noisy images such as ECUS images efficiently [23]. Furthermore, DDGVF snake is faster than other existing methods such as contour or region based level set, mesh, and atlas.

2 Methodology

In this section, we present the cascaded shape regression method which is used for estimating the prostate shape and position, given a set of training data. The regressed shape is then used to initialize dynamic directional gradient vector flow snakes for accurate boundary segmentation.

2.1 Cascaded Shape Regression (CSR)

A prostate shape is represented by a sequence of M landmark points: $S = [x_1, y_1, ..., x_M, y_M]^T$. A training sample, $\{(I_i, \hat{S}_i)\}$, consists of an image I_i and a true shape $\hat{S}_i$, $i = 1, 2, ..., N$. As a landmark-based shape model, an essential requirement is that landmarks on all training samples are located at corresponding positions. Because of the ellipsoidal shape of the prostate, it is not an easy task to label the same points. To find these landmarks, we first fit the manually drawn prostate boundary in the image to an ellipse. Starting from the orientation of fitted ellipse, equally-distanced landmarks are selected automatically from the prostate boundary.

A cascaded regressor $R = (R^1, R^2, ..., R^T)$ consists of T weak regressors. Given an image I and an initial prostate shape S^0, each regressor generates a shape increment vector δS to update the previous shape:

$$S^t = S^{t-1} + R^t(I, S^{t-1}), \text{with } \delta S^t = R^t(I, S^{t-1}), t = 1, 2, ..., T. \quad (1)$$

The output of regressors depends on image I and the previous shape S^{t-1}, using random fern and shape-indexed features which will be described in the following sections. Each regressor is trained to minimize the difference between the true shape and the new shape updated by the regressor, i.e.,

$$R^t = \arg\min_R \sum_{i=1}^{N} \|\hat{S}_i - (S_i^{t-1} + R(I_i, S_i^{t-1})\|_2. \quad (2)$$

Random Fern Regressors We use random ferns as weak regressors in the cascade. The fern was firstly introduced for classification [4] and later used for regression [2], [3]. A fern regressor is created by randomly selecting s features from a vector of F features and comparing them with s thresholds randomly selected. In this way, each input feature vector is divided into one of 2^s bins. Each

bin b is associated with a regression output δS_b that minimizes the alignment error of training samples Ω_b that fall into the bin:

$$\delta S_b = \arg\min_{\delta S} \sum_{i \in \Omega_b} \|\hat{S}_i - (S_i + \delta S)\|_2. \tag{3}$$

Eqn. (3) is solved by simply taking the mean of all shape differences,

$$\delta S_b = \frac{\sum_{i \in \Omega_b} (\hat{S}_i - S_i)}{|\Omega_b| + \rho N}, \tag{4}$$

where ρ is a regularization term to overcome over-fitting when the number of training samples in the bin is insufficient. The exact solution of Eqn. 3 is given by Eqn. 4 when $\rho = 0$. At each stage in the cascaded regression, a pool of K ferns are randomly generated and the one with the lowest regression error is chosen.

In [2], single-variate regressors are trained separately for individual pose parameters. We train multi-variate regressors for all the M landmark points simultaneously: they either fall into a bin or not. As shown in Eqn. 4, each shape increment is a linear combination of certain training shapes $\{\hat{S}_i\}$. We choose the average of all training shapes as the initial estimate of shape S^0 for regression. Therefore, all intermediate shapes in the regression and the final regressed shape are always a linear combination of all training shapes [3]. Therefore, no extra constraint is used to impose smoothness on the output shape. In contrast, if we train separate single-variate regressors for each individual component of S, then the shape will become more and more irregular after each regression.

Shape-indexed Features We used simple shape-indexed features to learn each regressor. Shape-indexed features mean that a pixel is indexed relative to the currently estimated shape rather than the original image coordinates. Since the prostate shapes are mostly elliptical, we can estimate the best fit to an ellipse from a given prostate shape S, using the least-square criterion. The ellipse is parameterized by its location (t_x, t_y), major/minor axis a, b, and the orientation φ. Therefore, the current shape is reflected by the translation, scale, and rotation of the fitted ellipse.

These features are computed as the intensity difference between two pixels in the image. To compute F shape-indexed features from the current estimated shape, we first randomly sample $2F$ pixels within a circle of radius r centered at $(0, 0)$. Then F of them are randomly selected as p_1^n, and the rest of them as $p_2^n, n = 1, ..., F$. So p_1^n and p_2^n are not correlated. These points are then undergone a similarity transform according to the parameters of the best-fit ellipse $H(t_x, t_y, a, b, \varphi)$. The intensity differences at the transformed $2F$ pixels result in F shape-indexed features $I(H(p_1^n)) - I(H(p_2^n))$, which are invariant against the geometry scale, translation, and rotation of different prostate shapes. As any ellipse may be construed as an affine transformation of a circle, we use similarity transform as an approximate in order to generate randomly sampled pixel pairs nearby and within the prostate shape.

Training for CSR The training process for CSR is summarized in Algorithm 1. For each training sample S_i, we use the average of all training shape $\left(S^0 = \frac{\sum_{j=1}^{N} \hat{S}_j}{N}\right)$ and the true shape of the rest of training samples $\{\hat{S}_j | j = 1, 2, ..., N, j \neq i\}$ to initialize the CSR. The CSR is trained to move the shape to the true shape $\hat{S}_i$ even if the initial positions are far from $\hat{S}_i$. For each testing sample, CSR is only initialized for once by the average shape because it is the single shape estimate that minimize the training error before regression starts.

Algorithm 1 Training for cascaded shape regression

Require: $\{(I_i, \hat{S}_i)\}, i = 1, 2, ..., N$
  begin initialize $S_i^0 = \frac{\sum_{j=1}^{N} \hat{S}_j}{N}$ or $S_i^0 = \left\{\hat{S}_j | j = 1, 2, ..., N, j \neq i\right\}, i = 1, 2, ..., N$ for data augmentation
  for $t = 1$ to T **do**
    for $i = 1$ to N **do**
      Fit the current shape S_i^{t-1} to an ellipse and compute shape-indexed features
    end for
    Train K random ferns on all N current shapes and select the best fern which gives the lowest training error
    Apply Eqn. (4) to compute $\delta S_b = R_b^t$ for each bin b in the best fern
    for $i = 1$ to N **do**
      $S_i^t = S_i^{t-1} + \delta S_b$, suppose the features of S_i^{t-1} fall into bin $b, b \in \{1, 2, ..., 2^s\}$
    end for
  end for
return $R = (R^1, R^2, ..., R^T)$

2.2 Dynamic Directional Gradient Vector Flow

In CSR, the prostate shape is represented by a sequence of landmark points. Overfitting will occur if a large set of landmark points is used but the number of training data is limited. If using less number of landmark points, the regressed shape may miss the curvature change in prostate boundary. Given the CSR result as an initialization for the snake, we adopt the DDGVF [23] method to detect the prostate boundary more accurately in noisy and low-contrast ECUS images.

To compute DDGVF, an edge map $\mathbf{f} = [f_x^+(x, y), f_x^-(x, y), f_y^+(x, y), f_y^-(x, y)]$ is generated from image I, where f_x^+, f_x^-, f_y^+, and f_y^- are the gradients of positive step edges in x, $-x$, y, and $-y$ directions respectively. Accordingly, the DDGVF field $\mathbf{v}(x, y) = [u^+(x, y), u^-(x, y), v^+(x, y), v^-(x, y)]$ has four components, which are found by solving the following partial differential equations separately:

$$\mathbf{v}_t = \mu \nabla^2 \mathbf{v} - (\mathbf{v} - \mathrm{d}\mathbf{f})\mathrm{d}\mathbf{f}^2, \text{ initialized by } \mathbf{v}_0 = \mathrm{d}\mathbf{f} \tag{5}$$

where t is the time and $\mathrm{d}\mathbf{f} = [df_x^+, df_x^-, df_y^+, df_y^-]$. Finally, the snake is deformed under the external force $F_{ext} = [F_x, F_y]$, defined by

$$F_x = u^+ * \max\{\cos(\theta), 0\} - u^- * \min\{\cos(\theta), 0\} \quad (6)$$

$$F_y = v^+ * \max\{\sin(\theta), 0\} - v^- * \min\{\sin(\theta), 0\} \quad (7)$$

where θ is the contour's normal directional at a certain snaxel.

3 Experiments

We validate the performance of our method (CSR+DDGVF) on two datasets: 74 longitudinal view and 76 axial view ECUS images of the prostate. The resolution of the images is 488×744 pixels and (0.40 mm/pixel). 50 images are randomly selected from each dataset as the respective training sets and the rest as the testing sets respectively. Each prostate shape is described by $M = 20$ landmarks. The parameters of the CSR are set as follows: number of training data $N = 50$, number of phases in the cascade $T = 512$, fern depth $s = 5$, number of ferns $K = 128$, radius $r = 1.5$, and number of features $F = 64$.

We use average Root Mean Square Error (RMSE), failure rate, precision, recall, and Dice coefficient (DSC)[1] to evaluate the segmentation result. To calculate the failure rate, two thresholds on RMSE (ϕ = 2.4 or 3 mm) are used, which corresponds to 6.5% and 8.1% of average prostate length in longitudinal view images. The precision, recall, and DSC are averaged respectively only for those images where the segmentation result and the ground truth are at least overlapped. The results are compared with the original AAM proposed by Cootes et al. [15] using the optimized C++ implementation from [22].

During the testing of CSR, we use the average of all training shapes as the initialization by simply overlaying it on the testing image domain. In our ECUS prostate segmentation application, because the training data are from the same imaging setting as those for the testing data, their scales are considered the same. As long as the training data are representative for the size, shape and position variations, there is no need to invoke transformations in terms of scale and translation. As for rotation, the orientation of the prostate is estimated by fitting an ellipse to determine the correspondence of landmarks in the model and the object in the test image. Similarly, AAM was initialized by putting the mean shape (up to a scale factor) in the test image domain. The DSC of the average shape against true shapes is 0.79 ± 0.13 and 0.66 ± 0.24 for longitudinal and axial images respectively. In Fig. 1, the position of prostate shape at different stages is illustrated. The two initial shapes used for the two datasets respectively are shown in the first column (t=0) and the final regression results are shown in the last column (t=512).

[1] precision= $\frac{TP}{TP+FP}$, recall= $\frac{TP}{TP+FN}$, and DSC= $\frac{2TP}{2TP+FP+FN}$, where TP, FP, and FN are the number of true positive, false positive, and false negative pixels, respectively.

For the experiment, we use 5-fold cross validation (4 for training and 1 for testing) to avoid bias in such a splitting. All metrics are first averaged within each fold and the mean/standard deviation among the 5-fold are shown Table 1. It can be observed that the AAM has a much higher fail rate due to its sensitivity to initialization. The CSR achieves an average RMSE (under 3 mm) of 1.98 and 1.75 mm for the two datasets respectively. The CSR also outperforms the AAM for other performance metrics when only successful segmentation cases are counted. Our method is implemented in Matlab R2012a on a Windows machine with 3.2GHz CPU and 12GB RAM. The training and testing time is compared in Table 2. With an unoptimized Matlab code, the CSR requires 22 minutes to train 50 images, which is relatively long. However, our testing speed is only 0.92 seconds per image which is much faster than the AAM. This is desirable for real-time applications such as in-vivo experiments of BPH removal. The segmentation results of the CSR and the AAM is compared in Fig. 2.

Fig. 3 displays the final segmentation results. The prostate boundary after snake deformation using DDGVF is closer to the ground truth. We interpolate the landmark points on both the segmented boundary and the ground truth so that the distance between two neighboring points is between 0.5 to 1.5 pixels. The overlap ratios in the results of CSR+DDGVF are also shown in Table 1.

Table 1. Quantitative comparison of segmentation results for the two datasets. RMSE_ϕ: average RMSE in mm under ϕ mm, f_ϕ: failure rate (RMSE$> \phi$ mm).

Dataset	Method	$\text{RMSE}_{2.4}$	$f_{2.4}$	RMSE_3	f_3	Precision	Recall	DSC
Longi-tudinal	AAM	2.00±1.42	84.1%	2.29±0.11	72.8%	0.76±0.20	0.83±0.15	0.78±0.15
	CSR	1.89±0.17	27.0%	1.98±0.19	17.5%	0.88±0.14	0.93±0.06	0.90±0.10
	CSR+DDGVF	-	-	-	-	0.91±0.13	0.90±0.08	0.90±0.09
Axial	AAM	1.85± 0.44	96.0%	2.05±0.44	94.7%	0.87±0.11	0.82±0.06	0.84±0.08
	CSR	1.73 ±0.11	9.1%	1.75±0.14	6.5%	0.90±0.11	0.93±0.10	0.91±0.10
	CSR+DDGVF	-	-	-	-	0.94±0.11	0.89±0.10	0.91±0.10

Table 2. Training and testing time.

Method	Training (50 images)	Testing (per image)
AAM	10.9 sec	51 sec
CSR	22 min	0.11 sec
CSR+DDGVF	22 min	0.92 sec

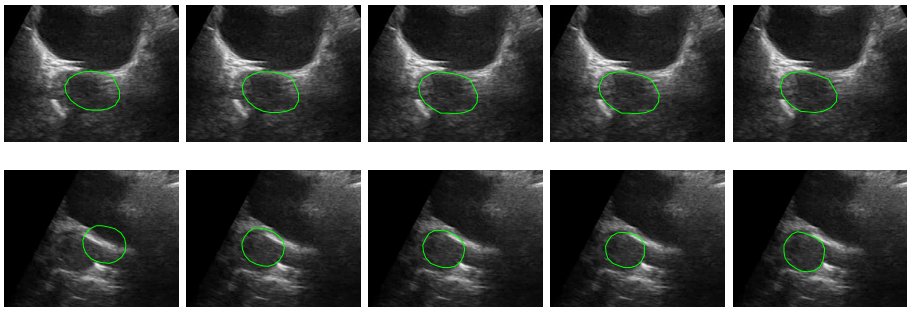


Fig. 1. Regressed shapes at different stages (from left to right) t=0, t=1, t=3, t=7, and t=512 respectively. First row is longitudinal view and the second row is axial view.

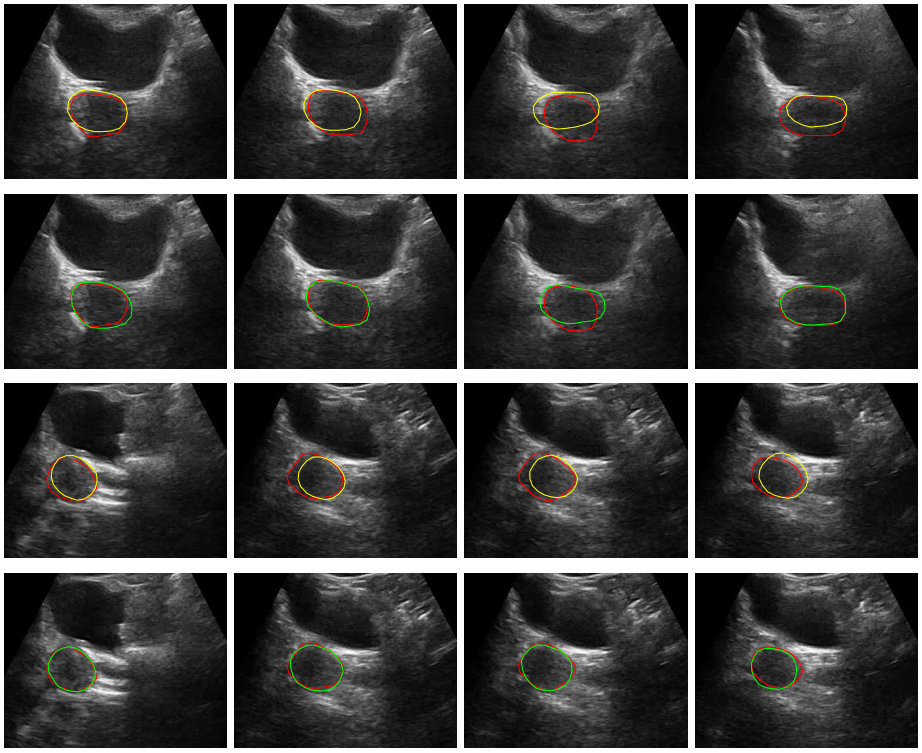
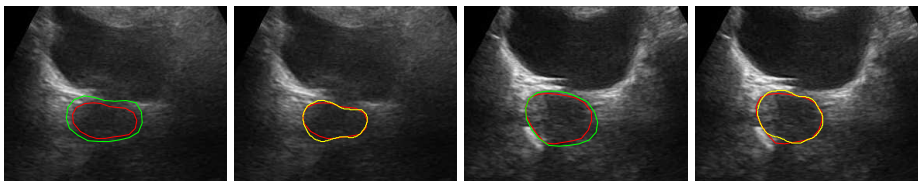


Fig. 2. Prostate segmentation results by active appearance model (first and third row) and cascaded shape regression (second and forth row). The red contour represents the ground truth and the green/yellow contour represents the segmentation results.



(a) (b) (c) (d)

Fig. 3. Segmentation results of CSR (green contour) and CSR+DDGVF (yellow contour). The red contour represents the ground truth.

4 Conclusion and Future work

A novel approach has been proposed for prostate segmentation from ECUS images. By using cascaded shape regression, our approach is able to efficiently locate the prostate boundary against shape, position, and orientation variations in ECUS images. With the help of DDGVF, efficient and accurate segmentation is achieved. Future work includes developing CSR+DDGVF into a real-time prostate tracking framework.

Acknowledgments. The project is supported by SERC Grant No. 1211480001, A*STAR, Singapore. Henry Ho and Tay Kae Jack for domain knowledge, Zhou Yufeng and Wilson Gao for experiment device, Huang Weimin and Zhou Jiayin for experiments and data.

References

1. Ghose S., Oliver A., Marti R., Llado X., Vilanova J.C., Freixenet J., Mitra J., Sidib D., Meriaudeau F.: A Survey of Prostate Segmentation Methodologies in Ultrasound, Magnetic Resonance and Computed Tomography Images. In: Comput Methods Programs Biomed. 108(1), 262–287 (2012)
2. Dollar, P., Welinder, P., Perona, P.: Cascaded pose regression. In: CVPR. 1078–1085 (2010)
3. Cao X. and Wei Y. and Wen F. and Sun J.: Face Alignment by Explicit Shape Regression. In: CVPR. (2012)
4. Ozuysal M., Calonder M., Lepetit V., and Fua P.: Fast keypoint recognition using random ferns. In: IEEE Trans. Pattern Analysis and Machine Intelligence 32(3), 448–461 (2010).
5. A. Zaim, J. Jankun: An Energy-Based Segmentation of Prostate from Ultrasound Images Using Dot-Pattern Select Cells. In: IEEE International Conference on Acoustics, Speech and Signal Processing 297C300 (2007)
6. H. M. Ladak, F. Mao, Y. Wang, D. B. Downey, D. A. Steinman, A. Fenster, Prostate Segmentation from 2D Ultrasound Images. In: Proceedings of the 22nd Annual International Conference of the IEEE Engineering in Medicine and Biology Society, 3188C3191 (2000).

7. L. Gong, S. D. Pathak, D. R. Haynor, P. S. Cho, Y. Kim: Parametric Shape Modeling Using Deformable Superellipses for Prostate Segmentation. In: IEEE Transactions on Medical Imaging 23, 340C-349 (2004)
8. S. Badiei, S. E. Salcudean, J. Varah,W. J.Morris: Prostate Segmentation in 2D Ultrasound Images Using Image Warping and Ellipse Fitting. In: Medical Image Computing and Computer-Assisted Intervention. pp. 17C-24 (2006).
9. S. Fan, L. Voon, N. Sing: 3D Prostate Surface Detection from Ultrasound Images Based on Level SetMethod. In: Medical Image Computing and Computer-Assisted Intervention, pp. 389C-396 (2002).
10. M. Zouqi, J. Samarabandu, Prostate Segmentation from 2D Ultrasound Images Using Graph Cuts and Domain Knowledge. In: Canadian Conference on Computer and Robot Vision. IEEE Computer Society Press, USA, pp. 359C-362 (2008).
11. Chan, T. F., Vese, L. A.: Active contours without edges. In: IEEE Transactions on Image Processing 10(2), 266–277 (2001).
12. W. D. Richard, C. G. Keen: Automated Texture Based Segmentation of Ultrasound Images of the Prostate. In: Computerized Medical Imaging and Graphics 20, 131C-140 (1996).
13. S. S. Mohamed, A.M.Youssef, E. F. El-Saadany, M. M. A. Salama: Prostate Tissue Characterization Using TRUS Image Spectral Features. In: Image Analysis and Recognition. Springer, pp. 589C601 (2006).
14. Cootes T.F., Taylor C.J., Cooper D., Graham J.: Active shape models - their training and application. In: Computer Vision and Image Understanding 61, 38–59(1995).
15. Cootes T.F., Edwards G.J., Taylor C.J.: Active appearance models. In: IEEE Trans. Pattern Analysis and Machine Intelligence 23, 681–685 (2001).
16. Zhou S.: Shape regression machine and efficient segmentation of left ventricle endocardium from 2D B-mode echocardiogram. In: Medical Image Analysis 14, 563C-581 (2010).
17. J. H. Friedman: Greedy function approximation: A gradient boosting machine. In: The Annals of Statistics 29(5), 1189C-1232, (2001).
18. N. Duffy and D. P. Helmbold: Boosting methods for regression. In: Machine Learning 47(2-3), 153C-200, (2002).
19. L. Breiman: Random forests. In: Machine learning (2001).
20. Valstar M., Martinez B., Binefa X., Pantic M.: Facial point detection using boosted regression and graph models. In: CVPR. (2010).
21. T.F.Cootes, M.Ionita, C.Lindner and P.Sauer: Robust and Accurate Shape Model Fitting using Random Forest Regression Voting. In: ECCV (2012).
22. Open source C++ AAM implementation, `http://www2.imm.dtu.dk/~aam/`
23. Cheng J., Foo S.: Dynamic directional gradient vector flow for snakes. In: IEEE Transactions on Image Processing 15(6), 1563–1571, (2006).