Fully Automatic Feature Localization for Medical Images using a Global Vector Concentration Approach

Tatsuo Kozakaya, Tomoyuki Shibata, Tomoyuki Takeguchi and Masahide Nishiura
Corporate Research and Development Center, Toshiba Corporation
1, Komukai-Toshiba-cho, Saiwai-ku, Kawasaki, 212-8582, Japan
tatsuo.kozakaya@toshiba.co.jp

Abstract

In this paper, we propose a novel feature localization method based on a global vector concentration approach. Our approach does not rely on the detection of local salient features around feature points. Instead, we exploit global structural information of the object extracted by calculating the concentration of directional vectors from sampling points. Those vectors are combined with local pattern descriptors of a query image and selected from preliminarily trained extended templates by nearest neighbor search. Due to the insensitivity of local changes, our method can handle partially occluded and noisy objects. We apply the proposed method to fully automatic feature localization of the left ventricular in echocardiograms. The results show the effectiveness of our method in comparison with a conventional edge-based method in terms of accuracy and robustness.

1. Introduction

Accurate localization of feature points of an object is a base technology for solving object detection and classification problems. A number of feature localization methods have been proposed to detect meaningful and reproducible points on the object. Most conventional methods use local structural information such as a local pattern or edge intensity as a keypoint to extract candidates and localize the feature points. The Harris corner detector [7] and the SIFT descriptor [10] are representative methods using local salient features. Although these methods have shown good results in most cases, they may be unstable due to noise in the image acquisition process or partial occlusion of the local features.

Figure 1 shows examples of our feature localization of a left ventricular (LV) boundary in an echocardiogram. The feature localization of the endocardial boundary helps in the diagnosis of heart disease. However, fully automatic and accurate boundary extraction is still a challenging problem because the quality of echocardiographic images varies significantly in each case and the boundary may disappear from the imaging area because of the motion of the heart. In this case, feature localization based on *global* structural information is more suitable to represent the noisy and partially occluded features than that of local salient features.

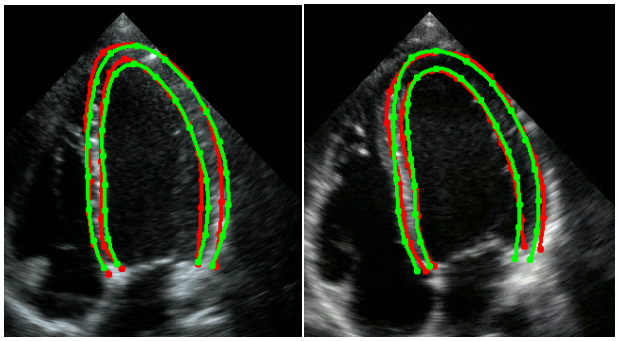

Figure 1. Example results of our fully automatic feature localization of the LV boundary in an echocardiogram. The red points are manually annotated by an expert. As shown in green, our method accurately estimates the feature points. We also present the boundaries interpolated by spline curves.

In this paper, we propose a novel feature localization method which extracts global structural information of the object using a vector concentration approach. The proposed method represents a feature point as concentration of directional vectors from several sampling points in a query image. Since the feature point is defined by the relative positions from a number of the sampling points, we can suppress the influences of local changes around the feature points being detected. The directional vectors are combined with a local pattern around the sampling point to form an extended template. We learn the extended templates using all sampling points from the training images. In the detection phase, the best matched extended template is selected by nearest neighbor search (NNS) using the local pat-

978-1-4244-2243-2/08/$25.00 ©2008 IEEE

terns around sampling points of a query image. The feature points are then localized by calculating the concentration of directional vectors of the selected extended templates.

We apply the proposed method to fully automatic feature localization of the LV boundary. To demonstrate the effectiveness of the proposed method, comparisons have been performed with a conventional method which localizes the feature points based on the local salient features such as edge information.

The main contributions are as follows. First, we propose a novel feature localization framework. Unlike energy minimization and voting-based methods, our method requires neither an initial shape, iterative processing nor cumulative voting. Second, the global shape information is implicitly obtained by the vector concentration approach with no prior information about the target object. Finally, we show the effectiveness of the proposed method through experiments. Our method outperforms the conventional methods with respect to accuracy and robustness against partial occlusions.

2. Related Work

Numerous algorithms have been proposed for feature localization in echocardiograms. Active contour models (snakes) [8] were applied to boundary detection, which minimize an energy function formed by weighting multiple active contour models [3]. Cootes *et al*. introduced a contour detection method [4] based on active shape models (ASM). The ASM incorporates global shape information by using a statistical prior to constrain inappropriate shapes. However, since both the methods require an energy minimization process based on the local features such as edge intensity, they may cause a serious error if the local features cannot be reliably extracted. Furthermore, these iterative processes require a reasonable initial position. For another approach, Zhou and Comaniciu proposed a feature localization method using boosting regression [13]. The regression-based algorithm generally suffers from the curse of dimensionality, and regression requires training data in proportion to the number of the feature points being localized to prevent overfitting.

From the viewpoint of the feature localization method integrating multiple local information, the proposed method is similar to the generalized Hough transform [2] and more recently the bag-of-words models [6, 12, 9]. These methods detect a feature point by voting all possible Hough parameters or probabilistically represented positions. The voting-based methods are robust against local changes due to majority decision, but they may be unstable when not enough votes are obtained and they often have a high computational cost for the voting.

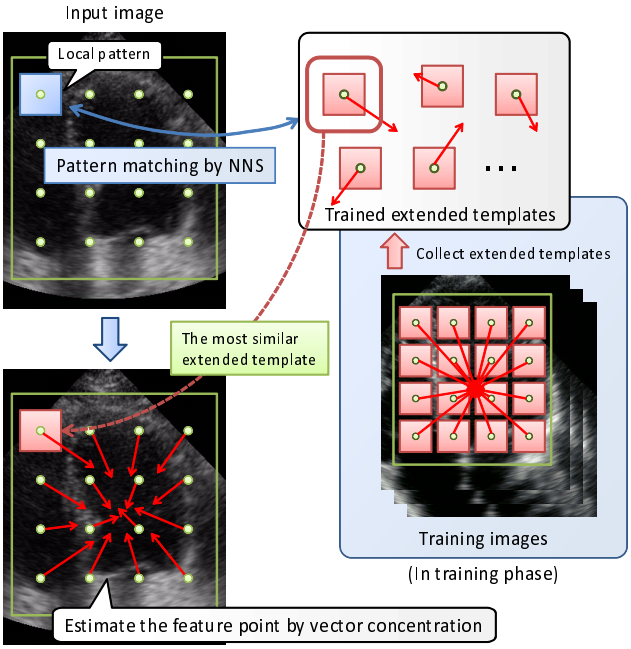

Figure 2. Overview of our method. Sampling points are located on an image and local pattern descriptors are extracted for each sampling point. We obtain directional vectors by matching with trained extended template and extract feature points by vector concentration.

3. Overview of our approach

The goal of this paper is fully automatic localization of the feature points on the LV boundary. We first give an overview of the proposed method (figure 2).

In the training phase, we place sampling points on training images and extract local pattern descriptors around each sampling point. An extended template for each sampling point is defined as combination of a local descriptor and directional vectors from the corresponding sampling point to the given feature points. If there are N feature points, an extended template also has N vectors which represent the relative positions to each feature point. We then extract a set of the extended templates from all sampling points in the training images.

In the localization phase, we similarly place sampling points and extract local pattern descriptors from a query image. The local pattern descriptors are used as queries for the NNS and the search result gives the most similar extended template to each query descriptor. Since the extended template contains the directional vectors, we obtain N *observations* for each sampling point, which suggest the directions to the feature points. We estimate the N feature points based on the vector concentration that yields the least mean square error from the directional vectors for each feature point.

4. Feature localization using global vector concentration

4.1. Setting of sampling points

For robust feature localization, we need repeatable keypoints on a target object. Most previous works utilize a keypoint detector such as the Harris corner detector [7] and the difference of Gaussians [10] but it is a difficult problem to extract stable keypoints from a noisy echocardiographic image. By contrast, our approach does not constitutively depend on keypoint detection methods.

In this paper, we define *sampling points* as the keypoints, which are placed at regular grid nodes [6, 12]. We first define a region of sampling (ROI) from an average position of training LV boundaries. The sampling points are then placed at regular intervals in the ROI. The ROI is common to all input images, so an initial detection is not required.

Note that we tried to use Harris corner detector for locating the sampling points, but the final localization accuracy was worse than that of using regular grid sampling points.

4.2. Local pattern descriptors

A local pattern descriptor is used as a code which represents relative positional relationship between the corresponding sampling point and all feature points being detected. The descriptor is required to have a tolerance of small changes of the local pattern due to heart motion or partial occlusions.

We use a localy normalized histogram of oriented gradients (HOG) descriptor [5]. Dalal *et al*. applied the HOG to pedestrian detection and they indicated the HOG has the potential to extract suitable features for robust visual object recognition. The parameters of HOG descriptors are determined through a preliminary experiment, such that 3 orientation bins and 144 blocks in 30×30 pixel regions, hence the total dimension of the descriptor is $3 \times 144 = 432$.

We have performed a preliminary experiment to determine the parameters of the HOG and the result shows that the orientation resolution of the descriptor is less effective than the spatial resolution. We think this is because the LV boundary has a smooth shape and a detailed orientation cannot be extracted from noisy echocardiogram images.

4.3. Training extended templates from images

In the training phase, we extract extended templates from images and use them as training samples for the NNS. An extended template consists of combination of a local pattern descriptor and directional vectors from the corresponding sampling point to the feature points. We define the extended template f_i as follows:

$$f_i = (\boldsymbol{p}_i, \boldsymbol{v}_{i1}, \boldsymbol{v}_{i2}, \ldots, \boldsymbol{v}_{iN}), \tag{1}$$

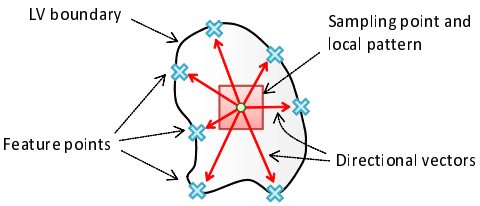

Figure 3. Illustration of an extended template. The other extended templates are also defined as a local pattern descriptor and directional vectors to the feature points for each sampling point.

where $\boldsymbol{p}_i$ represents the local pattern descriptor, $\boldsymbol{v}_{ij}$ is the directional vector to the j-th feature point and N is the number of the feature points being localized.

Figure 3 shows an example of the extended template. For training, we assume that feature points are manually localized on the LV boundary in advance. The sampling points and local pattern descriptors are provided by the same procedure as described in 4.1 and 4.2. The directional vectors are defined as starting from the sampling point to all feature points. We repeat this process for all sampling points and training images and then obtain the set of the training samples.

4.4. Matching with the extended templates

We perform matching with the local pattern descriptors and the extended templates and retrieve the closest extended template for each descriptor. We use approximate nearest neighbor search (ANNS) [1] as the matching method. The ANNS partitions the feature space into a number of clusters by a kd-tree to make the NNS much faster. We define the distance function of the ANNS as follows:

$$d_{ij} = d(\boldsymbol{p}_i, f_j) = ||\boldsymbol{p}_i - f_j(\boldsymbol{p})||^2, \tag{2}$$

where $\boldsymbol{p}_i$ is the i-th local pattern descriptor of the query image and f_j and $f_j(\boldsymbol{p})$ are the j-th extended template and its local pattern descriptor of the training samples. The ANNS needs an upper bound ϵ as the allowable distance error. We empirically set $\epsilon = 10.0$ in the experiments.

4.5. Localizing the feature points

The ANNS provides correspondences between the local pattern descriptors and the extended templates. The retrieved extended template gives the directions to the feature points. Therefore, we know the direction to the feature points at each sampling point according to directional vectors of the extended templates.

We now calculate coordinates of the feature points by considering the vector concentration. Figure 4 illustrates the calculation of the vector concentration. In the following paragraphs, we explain the procedure to localize the position of a feature point.

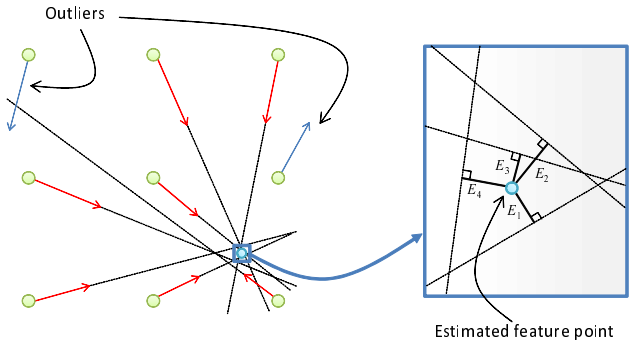

Figure 4. Calculation of the vector concentration with directional vectors. The feature point is estimated by minimizing the sum of the square distances E_i^2.

First, we define the error function for the vector concentration. We assume that a line is written as $a_i x + b_i y + c_i = 0$, which passes through the i-th sampling point and directs to the vector derived from the extended template (figure 4). The error E_c for a point (x, y) is defined as the sum of square distances between the point and the lines and represented by a simple quadratic form:

$$E_c = \sum_{i=1}^{N_p} ||a_i x + b_i y + c_i||^2 = \boldsymbol{x}^T \mathbf{A} \boldsymbol{x}, \tag{3}$$

where N_p is the number of sampling points, $\boldsymbol{x} = (x, y, 1)^T$ and $\mathbf{A}$ is a 3×3 matrix. We minimize E_c by calculating the partial derivatives as follows:

$$\frac{\partial E_c}{\partial \boldsymbol{x}} = \mathbf{A}' \begin{pmatrix} x \\ y \end{pmatrix} + \mathbf{C} = 0, \tag{4}$$

where $\mathbf{A}'$ and $\mathbf{C}$ are constants derived from the position of the sampling points and the directional vectors.

We localize the feature point by solving Eq. (4). However, since the ANNS does not always give the correct extended template, we have to take into account outliers. While several methods are available for rejecting outliers, we employ the least median of squares (LMedS) method [11]. The algorithm of outlier rejection by LMedS is summarized as follows.

1. Randomly choose F directional vectors.
2. Estimate the vector concentration using the F directional vectors and obtain model parameters (x, y) by Eq. (4).
3. Evaluate the model parameters using the LMedS criterion:

$$E_M = \min_i \operatorname{med} ||a_i x + b_i y + c_i||^2, \tag{5}$$

where med is a median operator.

4. Iterate q times from the step 1 and find the parameters corresponding to the minimum E_M.

We calculate the standard deviation $\hat{\sigma}$ from the minimum E_M, and reject the outliers based on $\hat{\sigma}$. In practice, we introduce binary weights w_i to the error function E_c of Eq. (3).

$$w_i = \begin{cases} 1 & \text{if } ||a_i x + b_i y + c_i||^2 < (2.5\hat{\sigma})^2, \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

We then calculate the final position of the feature point with the weights w_i by the least squares method. We also introduce a confidence measure of the localization result, which is represented by the ratio of the outliers with the total number of sampling points.

$$Conf = \frac{1}{N_p} \sum_{i=1}^{N_p} w_i, \tag{7}$$

For accurate localization, we reject the localization result if the confidence value is less than a predefined threshold.

If there is more than one feature point, we sequentially perform the localization process for each feature point. Note that we do not need to conduct the matching process again because the closest extended templates are common with respect to all feature points. Furthermore, the localization process is performed independently, so we can avoid a combinatorial explosion in the number of the feature points.

4.6. Recovery of the rejected feature points

We discuss an algorithm to recover the rejected feature points in the robust estimation process using the directional vectors utilized for the normally detected feature points.

We obtain the weight $w_i^{(j)}$ of the i-th sampling point in the j-th feature point localization from Eq. (6). We assume that the weights $w_i^{(j)}$ is set to zero when the j-th feature point is rejected. The reliability of the extended template is defined as:

$$C_i = \frac{1}{N} \sum_{j=1}^{N} w_i^{(j)}. \tag{8}$$

The reliability C_i represents the percentage of frequency adopted for feature point detection for each sampling point. If this reliability is greater than a predetermined threshold, the corresponding extended template is *reliable* for the feature point detection.

We estimate the rejected feature point using only those reliable extended templates. We define the weight w_i' for the rejected feature points as follows:

$$w_i'^{(k)} = \begin{cases} 1 & \text{if } C_i \geq th, \\ 0 & \text{otherwise,} \end{cases} \tag{9}$$

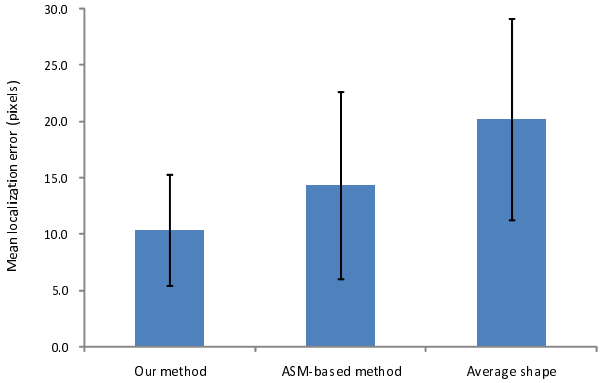

Figure 5. Result of the fully automatic feature localization with real echocardiographic images. The graph shows the RMS error and the standard deviation for each method.

where k is an index of the rejected feature point and th is the predetermined threshold. The rejected feature point is then calculated with these weights in the same as described in 4.5. This recovery process is useful when the feature point is rejected in the case that the confidence is less than the threshold by a small margin.

5. Experiments

In this section, we conduct some experiments with real echocardiographic images to investigate the effectiveness of the proposed method.

5.1. Experimental setup

The dataset consists of 82 cases (images) of end-diastolic in the apical four-chamber view. The ground truth is provided by a single expert who outlined the inner endocardial boundaries with several points per image. We interpolated these points for each boundary by spline curves, and obtained uniformly sampled 17 feature points.

For comparison, we implemented an ASM-based feature localization method as the conventional edge-based approach and we also provide an average shape as a baseline. The experiments were performed using leave-one-out cross validation for all methods. We give an initial shape of the ASM by the average shape, whose position is estimated using the difference of Gaussian blob detector. In the proposed method, a constant fixed ROI is used for all images and cropped to 80×80 pixels. The sampling points are allocated to 10×10 grid nodes. For fair comparison, we do not apply any preprocessing to the images for all methods.

5.2. Localization result

Figure 5 describes the localization performance in form of the root mean square (RMS) error between the localized feature points and the ground truth data. The bar graph represents the RMS error in pixels for the proposed method, the ASM-based method and the average shape respectively. The error bars shows the standard deviation of the localization results.

The experimental result shows that the proposed method provides more accurate localization in comparison with the other two methods. The ASM-based method is better than the average shape result, but the standard deviation of the error is almost equivalent to that of the average shape. This means the ASM-based method may be unstable depending on the input images due to misalignment of the initial shape or failure of the iteration process. In contrast, since our method does not require any initial shape and iterative processes, its standard deviation of the error is much smaller than the others.

We also measured the computational time for the localization process with a 2.66GHz Intel Core 2 Extreme QX6700 processor. The average time of our method and the ASM-based method are 42.4 and 46.9 milliseconds per image respectively. This shows that our method is equivalent to the ASM in terms of speed and much faster than conventional voting-based methods like the generalized Hough transform.

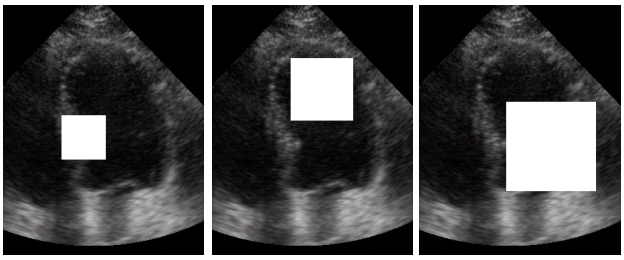

Figure 6. Examples of partially occluded images whose sizes of the masked area are (a) 5%, (b) 10% and (c) 20%.

5.3. Partially occluded images

To evaluate the tolerance of the proposed method against occlusion, we perform the experiment with partially occluded images. A part of the query image is masked with a white box which is randomly located in the ROI and the size of the masked areas is varied from 1% to 20%. Figure 6 illustrates examples of the occluded images whose sizes of the masked area are 5%, 10% and 20% respectively.

We conducted the feature localization with the occluded query images under the same condition described in 5.1. Figure 7 shows the result in terms of the RMS error of the proposed method, the ASM-based method and the average shape, where the x-axis of the graph represents a percentage of the masked area. The result of the average shape is not affected by the size of the masked area because it does

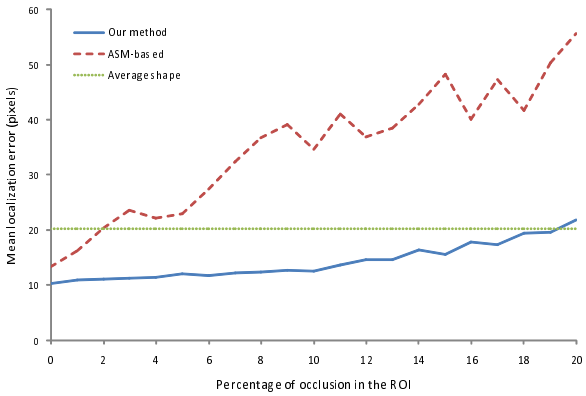

Figure 7. Result in terms of the RMS error of the three method. The x-axis of the graph represents a percentage of the masked area.

not utilize information derived from the query images. The results of the other two methods are getting worse due to increasing the occlusion size. However, the rate of increase of the RMS error is significantly different for our method and the ASM-based method. The RMS error of the ASM-based method increases to the same level as that of the average shape in the 2% masked area and rapidly becomes worse. On the other hand, our method keeps better results than the average shape until that of the 19% masked area. This result denotes the proposed method exhibits better performance against the partial occlusions in comparison with the conventional local search method.

5.4. Discussion and future work

According to the experimental results, our method effectively works and accurately localizes the feature points by extracting the global shape information in the case that a part of the query image is occluded. However, since our approach does not adjust the position with respect to local information around the feature points being localized, the extracted boundary is not exactly fitted to the actual boundary if we can see sharp edges on the boundary. The ASM-based method may present better results in that situation. The other issue of our approach is the size of the training extended templates used by NNS, which increases linearly as training samples increase. A feature-merging method may be required such as codebook clusters described in [9] for large training sets.

The proposed approach presented in this paper is a very basic framework for feature localization using vector concentration. It is not limited to the LV boundary localization, and so we plan to apply it to various feature localization problems and improve it by introducing scale and rotation invariant features. Furthermore, since our approach can be straightforwardly extended to higher dimensions, we also plan to localize feature points on 3D objects.

6. Conclusions

In this paper, we proposed a novel feature localization method which exploits global shape information. The feature points are represented as the concentration of directional vectors from sampling points. Unlike voting-based methods, such as the generalized Hough transform, our approach extracts the feature points by minimizing the distances from lines defined by the directional vectors and the sampling points.

We applied the proposed method to fully automatic localization of the left ventricular boundary in echocardiograms. The experimental results show the proposed method outperforms the conventional edge-based localization method in accuracy and has strong robustness against partially occluded images.

References

[1] S. Arya, D. Mount, R. Silverman, and A. Y. Wu. An optimal algorithm for approximate nearest neighbor searching. *Journal of the ACM*, 45:891–923, 1998.

[2] D. H. Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981.

[3] V. Chalana, D. Linker, D. Haynor, and Y. Kim. A multiple active contour model for cardiac boundary detection on echocardiographic sequences. *IEEE Trans. on Medical Imaging*, 15(3):290–298, June 1996.

[4] T. F. Cootes, A. Hill, C. J. Taylor, and J. Haslam. The use of active shape models for locating structures in medical images. *Image and Vision Computing*, 12(6):355–366, 1994.

[5] N. Dalal and B. Triggs. Histograms of oriented gradients for human detection. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pages 886–893, 2005.

[6] L. Fei-Fei and P. Perona. A bayesian hierarchical model for learning natural scene categories. In *Proc. of IEEE Computer Vision and Pattern Recognition*, pages 524–531, 2005.

[7] C. Harris and M. Stephens. A combined corner and edge detector. In *Proc. of Alvey Vision Conference*, pages 147–151, 1988.

[8] M. Kass, A. Witkin, and D. Terzopoulos. Snakes: Active contour models. *Int. J. of Computer Vision*, 1(4):321–331, January 1988.

[9] B. Leibe and B. Schiele. Interleaved object categorization and segmentation. In *Proc. of British Machine Vision Conference (BMVC'03)*, pages 759–768, 2003.

[10] D. G. Lowe. Distinctive image features from scale-invariant keypoints. *Int. J. of Computer Vision*, 60(2):91–110, 2004.

[11] P. J. Rousseeuw and A. M. Leroy. *Robust regression and outlier detection*. John Wiley & Sons, Inc., New York, NY, USA, 1987.

[12] J. Vogel and B. Schiele. Performance evaluation and optimization for content-based image retrieval. *Pattern Recognition*, 39(5):897–909, 2006.

[13] S. Zhou and D. Comaniciu. A boosting regression approach to medical anatomy detection. In *Proc. of Information Processing in Medical Imaging*, pages 13–25, 2007.