SVM-Based Failure Detection of GHT Localizations

T. Blaffert*[a], C. Lorenz [a], H. Nickisch [a], J. Peters [a], J. Weese [a]
[a] Philips Research Laboratories, Hamburg, Germany

ABSTRACT

This paper addresses the localization of anatomical structures in medical images by a Generalized Hough Transform (GHT). As localization is often a pre-requisite for subsequent model-based segmentation, it is important to assess whether or not the GHT was able to locate the desired object. The GHT by its construction does not make this distinction. We present an approach to detect incorrect GHT localizations by deriving collective features of contributing GHT model points and by training a Support Vector Machine (SVM) classifier. On a training set of 204 cases, we demonstrate that for the detection of incorrect localizations classification errors of down to 3% are achievable. This is three times less than the observed intrinsic GHT localization error.

Keywords: Object localization, object detection, GHT shape models, model-based segmentation, SVM-Classification

1. INTRODUCTION

In this paper we study the automated coordinate determination of defined organ landmarks such as the center of the heart or the center of the aortic valve in 3D medical image data, like the example shown in Figure 1. Such a *localization* of anatomical structures is an important component inside automated image segmentation procedures[1]. Particularly in the context of model-based image segmentation the Generalized Hough Transform[2] (GHT) was proven to be a robust method solving this task with sufficient accuracy for initialization[1,3,4]. The GHT comprises a shape model containing a set of points representing image structures (Figure 2) and a voting scheme that outputs the location of the largest subset of matching model points (Figure 3).

Since GHT localization is the first step in this segmentation chain, it is relevant to detect whether the reported location of the desired object is sufficiently accurate or not. For erroneous GHT localizations it is desirable to interrupt the segmentation chain, avoid wasting computation time and switch to a fallback localization method.

Similar to the detection of localization errors, the described type of inspection can be used to determine the presence of an anatomical structure in an image, which can be referred to as anatomical *object detection*. The distinction between the localization and the object detection problem becomes, for instance, relevant for cases where images from different anatomical regions (e.g. head, chest, abdomen, pelvis, full body) need to be processed and the anatomical structures of interest in respective images are required to be segmented automatically.

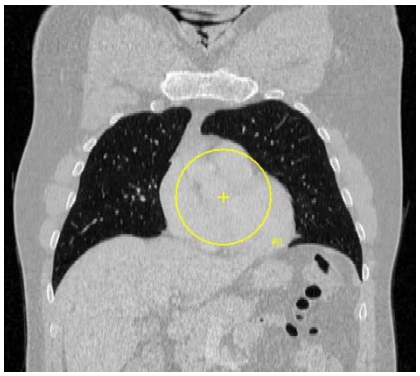


Figure 1: Localization of the heart at a central structure. The circle depicts an average shape radius.

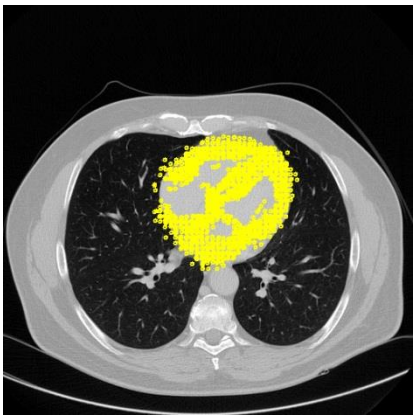


Figure 2: Position of heart GHT shape model points near the displayed axial plane, shifted by the localization in Figure 1.

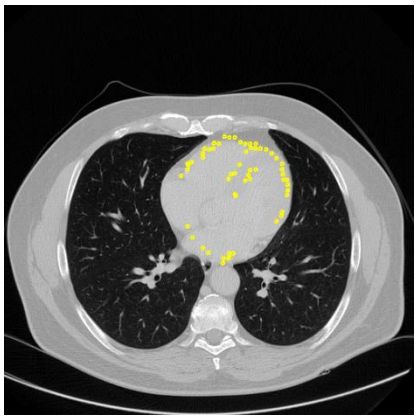


Figure 3: Heart GHT shape model points that contributed to the localization in Figure 1 by voting.

*thomas.blaffert@philips.com

For the detection of incorrect GHT localizations we are presenting an approach where collective features of contributing GHT model points are derived and used for training a Support Vector Machine (SVM) classifier. In the experiments we are going to compare this multi-feature classification with a single confidence feature threshold classification.

2. GENERALIZED HOUGH TRANSFORM (GHT)

Terminology

Before continuing with a description of the underlying localization algorithm we would like to make a clear distinction between the different entities that in the context of this paper are often referred to as 'model'. The Generalized Hough Transform (GHT) is the base localization algorithm considered here, the related framework produces and uses template features of anatomical structures that we will call *GHT shape model* or, shorter, *shape model*. In our GHT version, this shape model consists of a set of reference edge locations, which we name *shape model points*. The program that determines object locations within an image by a GHT shape model is in our software framework named the *shape finder*.

GHT localization is embedded in the broader scope of Model Based Segmentation (MBS), where the surface of an anatomical structure is modeled by a triangle mesh. Mesh models, however, are not subject of this paper and thus do not need to be distinguished.

Within the Support Vector Machine (SVM) classification method the phrase 'model' is often used to summarize the various internal settings and parameters, such as selection of kernel function or regularization values. We will refer to these as *SVM localization classifier*, or shorter as *localization classifier* if the context is more general and not bound to SVMs.

GHT Method

The Generalized Hough Transform (GHT) for object localization has been described in various papers[1,2,5]. Here, we briefly summarize the method as a template matching algorithm that yields a match score for a set of tested locations of some reference point of the object

The template that is matched is called the GHT shape model $\mathcal{M}$. It provides a description of certain image features in a neighborhood of the reference point. In the standard version of the GHT, as it is used in this study, this description is a shape outline that is represented by a discretized set of edge points with known geometric offset $\boldsymbol{d}$ from the reference point and known normalized edge gradient direction $\boldsymbol{n}$. The combination of offset $\boldsymbol{d}$ and edge orientation $\boldsymbol{n}$ will be called shape model point $\mathcal{P}_i = \{\boldsymbol{d}_i, \boldsymbol{n}_i\}$ where i indexes all model points in $\mathcal{M}$. The model points can be 'matched' with existing edge points in the image. Placing $\mathcal{M}$ at a certain location $\boldsymbol{x}$, the model points are placed with their encoded offsets in the image, and a 'match' per model point $\mathcal{P}_i$ is claimed if the image has an edge point close to $\boldsymbol{x} + \boldsymbol{d}_i$ with an edge orientation close to $\boldsymbol{n}_i$. Per match, the corresponding model point $\mathcal{P}_i$ 'votes' for the tested location $\boldsymbol{x}$, optionally with some from 1 differing weight w_i. Finally, all accumulated votes are interpreted as 'match score' or 'Hough vote' at location $\boldsymbol{x}$:

$$H(\boldsymbol{x}) = \sum_i w_i \cdot h(\boldsymbol{x} + \boldsymbol{d}_i, \boldsymbol{n}_i) \tag{2.1}$$

with the model point specific weight w_i and with the local match function

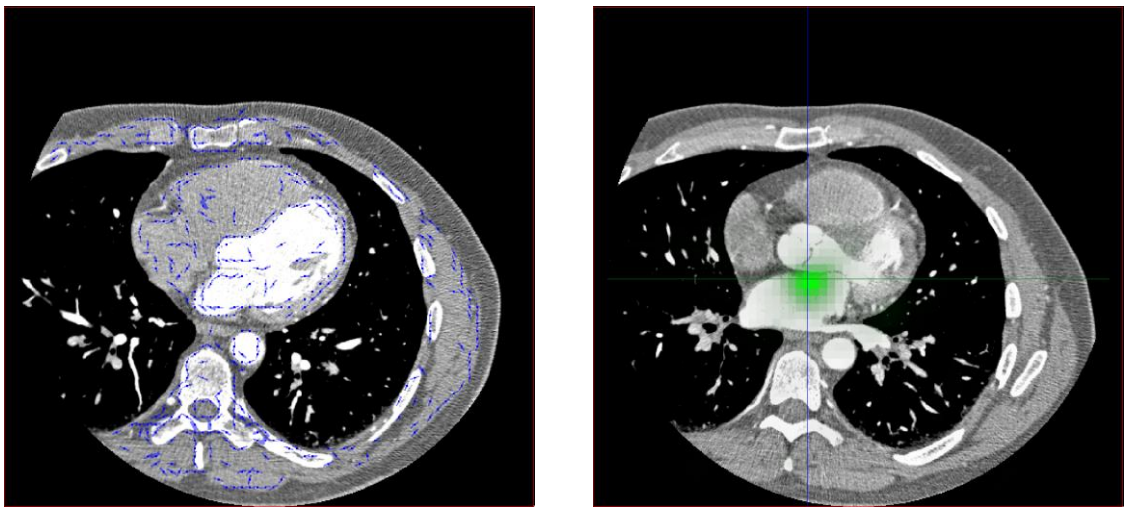
$$h(\boldsymbol{x} + \boldsymbol{d}_i, \boldsymbol{n}_i) = \begin{cases} 1 & \text{if the image has an edge point close to } \boldsymbol{x} + \boldsymbol{d}_i \\ & \text{with an edge orientation close to } \boldsymbol{n}_i \\ 0 & \text{otherwise} \end{cases} \tag{2.2}$$

Suitable edge points $\mathcal{E}_k = \{\boldsymbol{e}_k, \boldsymbol{m}_k\}$ of the image are obtained from a Canny edge detector, where $\boldsymbol{e}_k$ is a discretized location and $\boldsymbol{m}_k$ is a discretized orientation (Figure 4a). Additional processing with thresholds and filters suppresses noise edges. After calculating $H(\boldsymbol{x})$ for a set of discrete location hypotheses $\boldsymbol{x}$, the best location $\boldsymbol{x}^*$ is defined by the maximum Hough vote. By construction, placing the model $\mathcal{M}$ at $\boldsymbol{x}^*$ results in the maximum (weighted) number of model points $\mathcal{P}_i$ that match detected edges in the image (Figure 4b).

The discretization of location and orientation space into bins is the essence in the Hough Transform since it serves simultaneously two purposes: It defines what is a 'close location' or 'close orientation' by belonging to the same bin, and it permits a fast calculation of the match function by referencing shape model points via bin indices.

A practical algorithm first selects shape model points that match the orientation by a mapping $\boldsymbol{m}_k \longmapsto \{\mathcal{P}_i\}_{\boldsymbol{m}_k}$, which is commonly called the *R-table*. Then for each $\mathcal{P}_i$ the discretized location candidate $\boldsymbol{x} = \boldsymbol{e}_k - \boldsymbol{d}_i$ is calculated. These locations can be thought of as centers of a voxel, commonly called a *Hough cell*, over a voxel grid that is commonly called the *Hough space*. Finally $H(\boldsymbol{x})$ is incremented by w_i, where the increments are made in an accumulator array that covers the Hough space and is thus commonly called *Hough accumulator*.

The set of shape model points and the R-table are the main components of the shape model. Shape models are obtained for specific anatomical structures by a training procedure[2,6,7]. For the experiments in this paper all shape models are obtained from separate training data that is different from the data used for localization classifier training.



(a) (b)

Figure 4: (a) Axial slice of a cardiac CT scan (resolution 0.6x0.6x0.3mm^3) with extracted and filtered edges $\{\mathcal{E}_k\}$. (b) Detected location $\boldsymbol{x}^*$ with Hough votes $H(\boldsymbol{x})$ encoded as opacity of the green mask. Placing the model $\mathcal{M}$ at the maximum peak of $H(\boldsymbol{x})$ results in the optimum match between edges and model points.

3. NEW GHT LOCALIZATION FEATURES

The basic concept of the GHT is an independent 'voting' of all model points into one of the Hough offset bins and then decide for the best localization solution upon the majority of votes. This number of votes thus serves as a confidence measure. Our approach for the detection of GHT localization failures extends this confidence measure by considering the ensemble of all shape model points that contribute to a particular localization solution and deriving collective features from this set of voting model points.

These collective features are then used to train a decision function that determines the validity of a shape finder localization solution with a support vector machine (SVM). This section explains the new derived features; the next section describes how they are linked into SVM classification.

After the localization is determined by the highest number of votes into the Hough accumulator, those GHT shape model points that actually voted into this accumulator cell are identified. Technically a second run of the GHT procedure is performed where indices to voting shape model points are stored whenever the respective accumulator cell is hit. This process is not limited to just the 'best' localization solution offset cell with the highest number of votes, rather it can be extended to the localization solutions with the next largest number of voting counts or confidence. In our framework such a set of about 10 localizations is available, but this option is not further investigated in this paper.

The additional collective features are assembled from the m voting model points of a solution and the n shape model points, utilizing in particular the offset vector $\boldsymbol{d}_i$ from the localization center and the edge gradient direction $\boldsymbol{n}_i$ of shape model point i. The underlying assumption is that invalid solutions may be identified by a deviation from the average model point distribution, both spatially and in gradient direction. We have implemented the following features:

- Confidence (relative number of votes in percent):
  $f_c = m/n \ * 100$ (3.1)

In cases where the Hough voting includes weights, the voting score may be used rather than the number of votes. This option, however, was not investigated.

- Offset distance:
  $f_d = \|\boldsymbol{o} - \boldsymbol{r}\|$, with (3.2)
  $\boldsymbol{o} = 1/m \sum_{i=1}^{m} \boldsymbol{d}_i$ (average voting point offset),
  $\boldsymbol{r} = 1/n \sum_{j=1}^{n} \boldsymbol{d}_j$ (average model point offset)
- Gradient distance:
  $f_g = \|\boldsymbol{\omega} - \boldsymbol{\rho}\|$, with (3.3)
  $\boldsymbol{\omega} = 1/m \sum_{i=1}^{m} \boldsymbol{n}_i$ (average voting gradient),
  $\boldsymbol{\rho} = 1/n \sum_{j=1}^{n} \boldsymbol{n}_j$ (average model gradient)

When taking the average model point offset $\boldsymbol{r}$ as a center, the coordinate space around it can be divided into 8 spatial 3D octants, where the signs of the coordinate difference vector $\boldsymbol{d}_i - \boldsymbol{r}$ determine the octant boundaries. Figure 5 shows an example of an offset octant filling for a full heart model test case. By writing the difference vector with 3 coordinates as $\boldsymbol{d}_i - \boldsymbol{r} \equiv (x_i, y_i, z_i)$ and with the sign function

$$\mathrm{s}(\xi) = \begin{cases} 1 & \text{if } \xi > 0 \\ 0 & \text{if } \xi \leq 0 \end{cases}, \quad \xi = x_i, y_i, z_i \,. \tag{3.4}$$

The octants can be addressed by the index

$$k_i \equiv \mathrm{k}(\boldsymbol{d}_i - \boldsymbol{r}) = \mathrm{s}(x_i) \ + 2\,\mathrm{s}(y_i) \ + 4\,\mathrm{s}(z_i) \ , k_i = 0, \ldots, 7 \,. \tag{3.5}$$

With these indices each octant can be associated with a bin in a linear 8-bin histogram, which is filled from the occurrences of each voting model point in one of the octants. This histogram represents the spatial distribution of shape model points, an example is related to Figures 5 is shown in Figure 6.

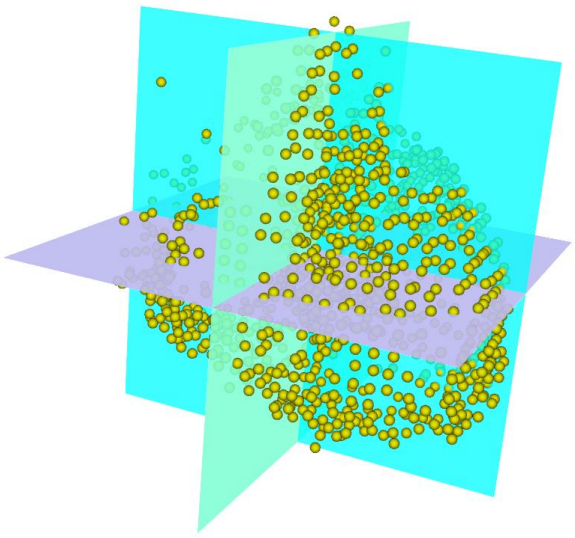


Figure 5: Voting model point offsets of a full heart localization test case, distributed over the 8 octants.

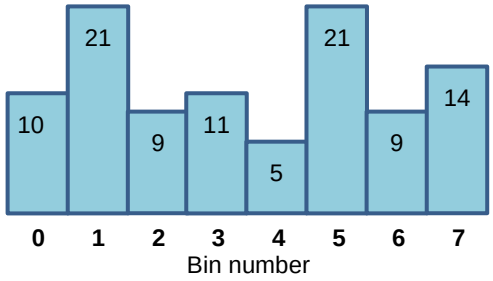


Figure 6: Histogram showing the relative octant occupancy of Figure 5, in percent.

The normalized offset distribution, which is a histogram $\boldsymbol{h_o}$ of all m voting model points, is now compared to a reference offset distribution that is calculated from all n shape model points and stored in a normalized histogram $\boldsymbol{h_r}$.

$$\boldsymbol{h_o} = (\boldsymbol{h_{ol}}) = \frac{100}{m}\left(\sum_{i=1}^{m}\begin{cases}1 & \text{if } k_i = l\\ 0 & \text{if } k_i \neq l\end{cases}\right),\ l = 0,\dots,7 \tag{3.6}$$

$$\boldsymbol{h_r} = (\boldsymbol{h_{rl}}) = \frac{100}{n}\left(\sum_{j=1}^{n}\begin{cases}1 & \text{if } k_j = l\\ 0 & \text{if } k_j \neq l\end{cases}\right),\ l = 0,\dots,7 \tag{3.7}$$

Similarly, histograms $\boldsymbol{h_\omega}$ and $\boldsymbol{h_\rho}$ can be calculated from the voting gradient vectors $\boldsymbol{n}_i$ and the shape model gradient vectors $\boldsymbol{n}_j$.

Two feature values that compare the tested and the reference histograms are calculated from their difference:

- Offset octants fill:
$$f_{od} = \sum_{l=0}^{7}|\boldsymbol{h_{ol}} - \boldsymbol{h_{rl}}| \tag{3.8}$$
- Gradient octants fill:
$$f_{og} = \sum_{l=0}^{7}\left|\boldsymbol{h_{\omega l}} - \boldsymbol{h_{\rho l}}\right| \tag{3.9}$$

4. LOCALIZATION CLASSIFIER TRAINING

In our framework, a localization classifier is a function that detects whether a GHT localization is correct or invalid. Parameters of this function, such as thresholds, are obtained by classifier training. We selected the well-studied and established approach of support vector machine (SVM) learning as our standard training method, where the localization classifier framework was implemented around the LIBSVM[8] software library functions.

For comparison purposes we also implemented a simple threshold finding procedure that was used for the single confidence feature.

A third approach was to use the special case of SVM learning for just a single confidence feature, where the main difference to threshold search is that SVM learning introduces some tolerance margin to false classified cases during training. Comparing the results of this method to the threshold search indicates the power of SVM classification, comparing the results to multi-feature SVM indicates the strength of additional features.

Single confidence threshold classifier

The simplest form of a localization classifier is a threshold on GHT votes, where the localization is considered as valid if the number of votes is above the threshold and as invalid otherwise. This holds also for a more general GHT confidence feature f_c.

We implemented a straight-forward method for the single confidence threshold localization classifier training, where a sorted list of confidence values of all training cases is constructed, this list is traversed in one direction, and the confidence value with the largest number of correctly classified training cases is selected as classification threshold.

Support vector machine classifier

In support vector machine (SVM) classification the validity of a GHT shape finder solution is predicted by calculating a prediction function, where the sign of the function value is used to predict the class. Input to the SVM is a feature vector calculated from the offset and gradient values stored in those GHT shape model points that have voted for this solution, as described in section X. The feature vector varies in length, depending on the set of features that are selected for a particular detection model.

From the different SVM variants provided by the LIBSM library the standard support vector classification with a regularization parameter C (C-SVM)[9] was best suited for our purposes. In support vector classification the given training feature vectors $\boldsymbol{x}_i \in \mathbb{R}^n, i = 1,\dots,l$ and indicator vectors $\boldsymbol{y} \in \mathbb{R}^l$ for two classes with $y_i \in \{1,-1\}$ are going to be optimally separated by the decision function

$$\mathrm{sgn}(\boldsymbol{w}^T\Phi(\boldsymbol{x}_i) + b) \tag{4.1}$$

with weights $\boldsymbol{w}, b$ and a function $\Phi(\boldsymbol{x}_i)$ that maps the feature vector $\boldsymbol{x}_i$ into a higher-dimensional space. The optimal function balances a maximal margin between the two classes against the total distance of training points lying on the 'wrong' side of the decision surface; it is found by solving a quadratic optimization problem. For efficiency reasons practical training algorithms calculate the elements of a kernel $K(\boldsymbol{x}_i, \boldsymbol{x}_j) \equiv \Phi(\boldsymbol{x}_i)^T \Phi(\boldsymbol{x}_i)$ rather than $\Phi(\boldsymbol{x}_i)$ and solve a dual optimization problem that is equivalent to the primal problem.

C-support vector classification solves the primal optimization problem

$$\begin{aligned} \min_{w,b,\xi} \quad & \tfrac{1}{2}\boldsymbol{w}^T\boldsymbol{w} + C\sum_{i=1}^{l}\xi_i \\ \text{subject to} \quad & y_i(\boldsymbol{w}^T\Phi(\boldsymbol{x}_i) + b) \geq 1 - \xi_i \\ & \xi_i \geq 0, \;\; i = 1, \dots, l \end{aligned} \tag{4.2}$$

where C is a regularization parameter that penalizes wrong classifications.

From the different kernels available from LIBSVM we selected the radial basis function since it showed consistently the best results with initial experiments:

$$K(\boldsymbol{x}_i, \boldsymbol{x}_j) = \exp\left(-\gamma\left\|\boldsymbol{x}_i - \boldsymbol{x}_j\right\|^2\right), \gamma > 0 \tag{4.3}$$

Grid search for optimal parameters and features

Since the penalty factor C and the factor γ are fixed for SVM training, optimal values for these parameters are determined with a grid search on discrete combinations of exponentially varying values $= b^i, i = s, \dots, f$, with a double type base b, the integer exponent i, the start exponent s, and the finish exponent f. Throughout our experiments we used the base $b = 2$. For each parameter set a cross validation is performed, where the optimal parameters are those resulting in the highest average accuracy over all cross validation test folds. In case of equal accuracies, the parameter closest to the center of the search range is selected.

Our implementation additionally permits an optional exhaustive search over all possible combinations of selected features, which returns the feature combination and parameter set with the highest accuracy after training or cross validation. In case of equal accuracy the combination with fewer features is selected.

5. EXPERIMENTS

In the experiments we investigated the accuracy of our method in the detection of GHT shape finder localization failures, with a particular emphasis on how additional collective shape finder solution features improve the detection accuracy over the single voting score threshold. All experiments covered cardiac structures in CT images.

SVM learning aims for achieving optimal classification accuracy, which is the sum of true positive and true negative results in percent of all cases. This classification accuracy was thus taken as success measure throughout the experiments, or alternatively the corresponding error rate 100%-accuracy.

Experiments include a 5-fold cross-validation of a C-.SVM with Gaussian kernel function, a grid search for the parameters C and γ, and a search for the optimal feature combination. The final classification accuracy for a test case is the average of classification accuracies from the test fraction of each cross validation fold.

Cardiac CT training data set

Training and test images were taken from a collection of inspected Trauma-CT cases. From the available approximately 500 data sets, 130 cases with the heart fully contained and 74 cases where the heart is not contained (head, extremities) were selected, putting aside duplicate reconstructions, reconstructions with non-standard orientation, and images where the heart is only partly contained. The set of investigated GHT shape models comprised a full heart model and models of 9 cardiac sub-structures that are shown in Table 1. The substructures were derived from the full heart model[1] and comprise heart valves and heart vessels.

For all test cases, the localization was obtained from the GHT shape finder, with shape models trained for the heart and the specific cardiac landmarks. The localization specifies a certain, usually centric position of that structure, which can be seen in Figure 7 and Figure 8. All GHT models were trained on data different from the Trauma-CT cases in this study.

Anatomical structure / Landmark
Full heart center
Aortic valve
Pulmonary valve
Mitral valve
Tricuspid valve
Left coronary artery origin
Right coronary artery origin
Right inferior pulmonary vein (RIPV) ostium
Right superior pulmonary vein (RSPV) ostium
Superior vena cava (SVC) ostium

Table 1: Set of heart and cardiac substructures investigated with the test cases.

Training for localization failure detection

All localizations from the cases with no heart belong per se to the training set of negative cases. From the images containing the heart, we manually separated those cases where the shape finder reported an incorrect localization and added them to the negative set (incorrect GHT result). The remaining locations constitute the set of positive training cases (correct GHT result). Depending on whether the positive or negative status is confirmed by the classifier, all cases then fall into the true or false positive (TP, FP) and true or false negative (TN, FN) category. Accuracy is the sum of TP and TN divided by the number of all cases, in percent.

Regarding positive cases there is the special constellation with GHT shape finders that the landmark of an anatomical structure may be present in an image but is not correctly localized by the best GHT solution. Since these error cases can again be classified as positive or negative, six different detection states need to be distinguished as follows:

Positive case: Landmark is contained in the image.

Valid case: Best GHT solution is located at the landmark, thus valid.

True valid (TV): Valid GHT solution is correctly classified as positive.

False error (FE): Valid GHT solution is incorrectly classified as negative.

Error case: Best GHT solution is not located at the landmark, thus invalid.

True error (TE): Invalid GHT solution is correctly classified as negative.

False valid (FV): Invalid GHT solution is incorrectly classified as positive.

Negative case: Landmark is not contained in the image, GHT solutions are implicitly invalid.

True negative (TN): GHT solution is correctly classified as negative.

False positive: (FP): GHT solution is incorrectly classified as positive.

The three cases and the positive and negative correctness classification state can be assembled into six entries of an extended confusion matrix:

	Positive detection	Negative detection
Valid case	TV	FE
Error case	FV	TE
Negative case	FP	TN

Table 2. Assembly of the three training cases and the two classification states into an extended confusion matrix.

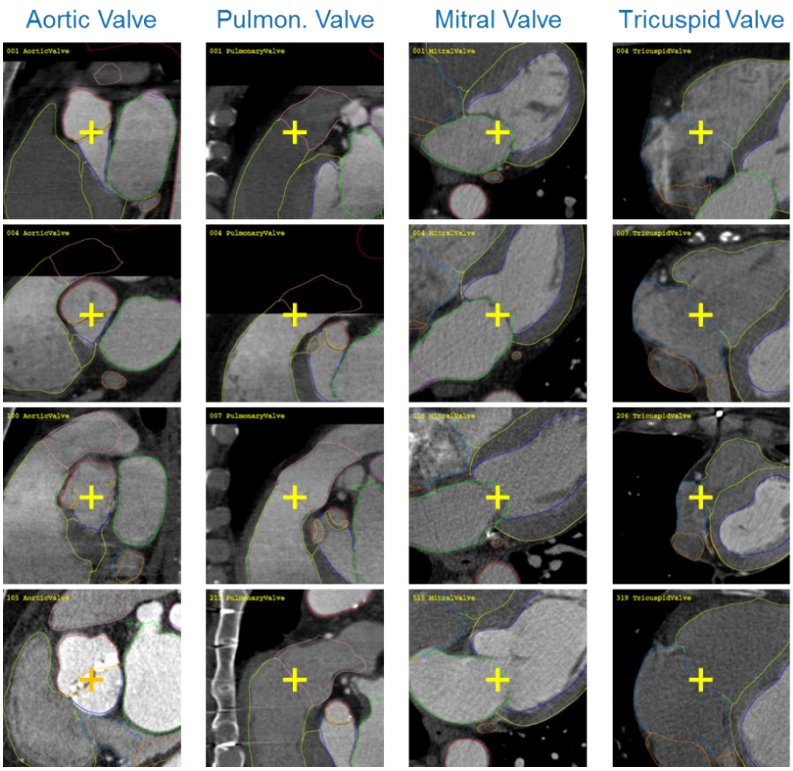


Figure 7: Examples of centric landmarks of the investigated cardiac substructures, heart valves.

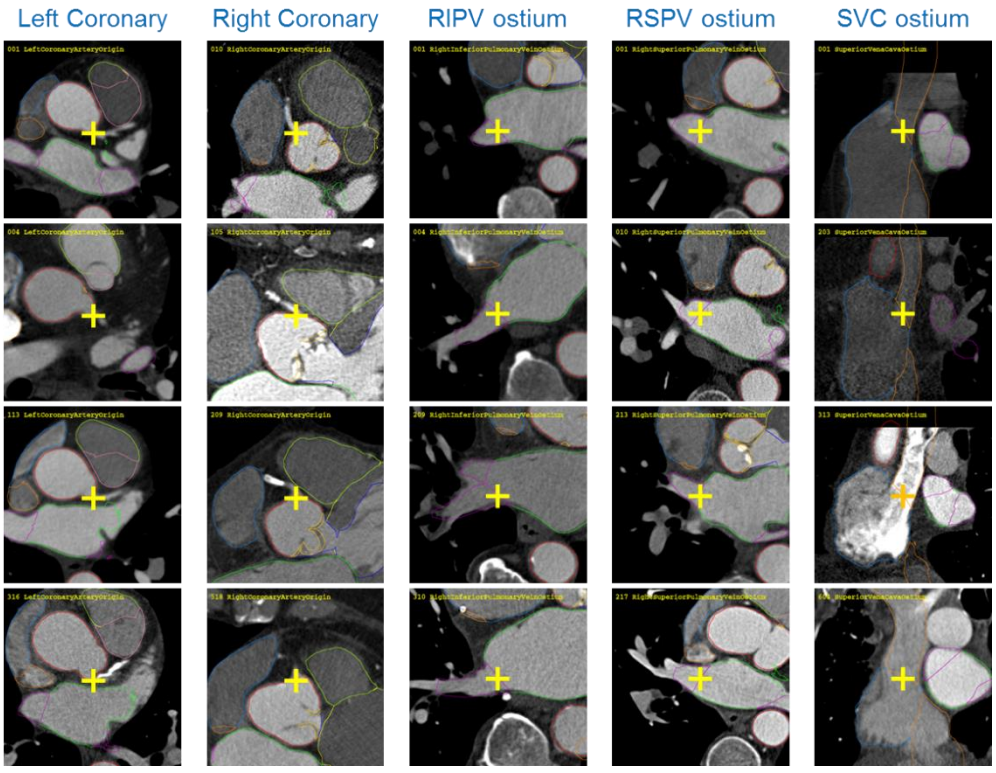


Figure 8: Examples of centric landmarks of the investigated cardiac substructures, heart vessels.

For the experiments the valid and error cases were visually distinguished, where practically all error cases allowed a clear tagging of invalid localizations as shown in Figure 9. Since for classifier training the validity of localizations is attached to the cases, we further define *positive training cases* as positive cases with valid localizations, like the examples shown in Figure 10. *Negative training cases* are either negative case with implicitly invalid localizations like those in Figure 11, or positive cases with tagged invalid best localizations like those in Figure 9. Table 3 lists the number of these different cases for the full heart and for all cardiac substructures. By comparing the sum of all positive cases and the sum of all invalid localizations we can see that approximately 11% of the GHT results are invalid.

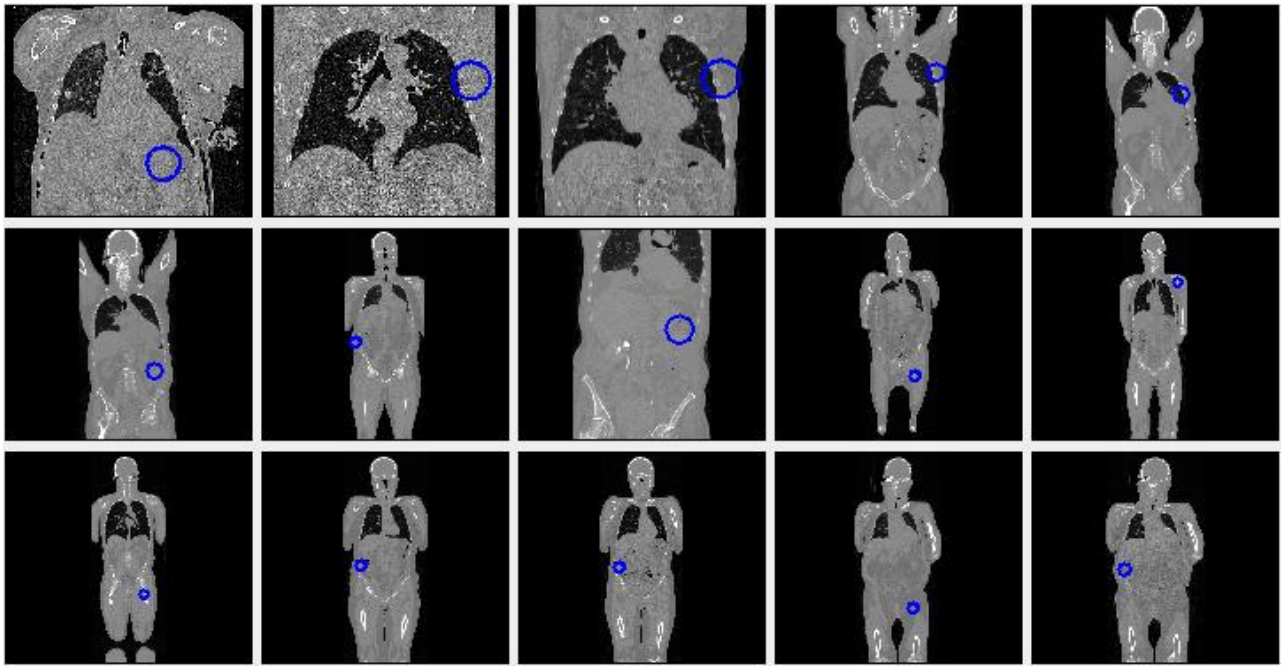


Figure 9: All 15 true error cases for the mitral valve classification in Table 4. The displayed best GHT localization solutions are clearly invalid.

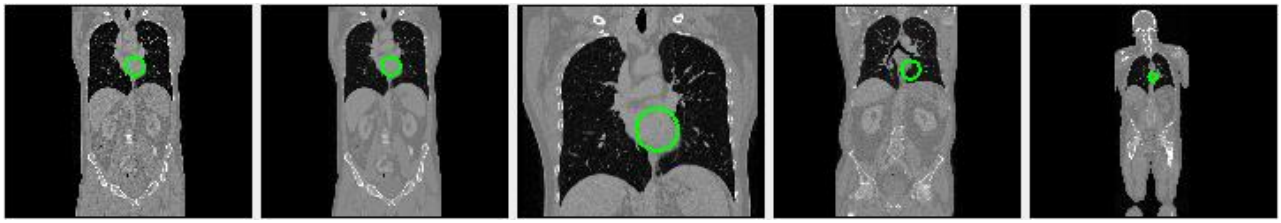


Figure 10: Examples of true valid cases for the mitral valve classification in Table 4, best GHT localization solutions.

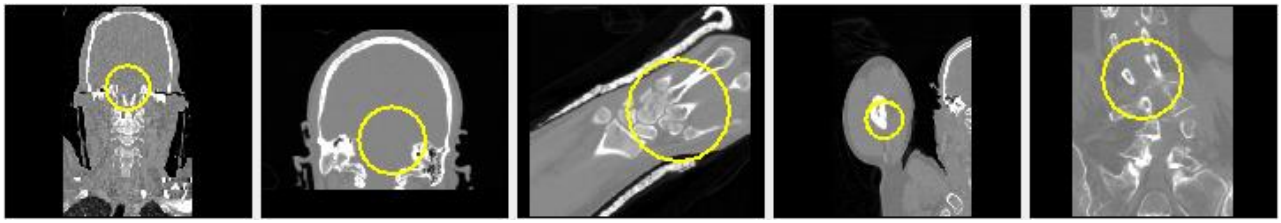


Figure 11: Examples of true negative cases for the mitral valve classification in Table 4, best GHT localization solutions.

Cardiac structure (landmark)	Positive cases, heart contained	Valid localiz. / Positive training cases	Invalid localiz.	Negative cases	Negative training cases	Total
Full heart center	130	125	5	74	79	204
Aortic valve	130	118	12	74	86	204
Pulmonary valve	130	116	14	74	88	204
Mitral valve	130	115	15	74	89	204
Tricuspid valve	130	117	13	74	87	204
Left coronary artery	130	116	14	74	88	204
Right coronary artery	130	118	12	74	86	204
Right inf. pulmon. vein	130	109	21	74	95	204
Right sup. pulmon. vein	130	115	15	74	89	204
Superior vena cava	130	124	6	74	80	204
Sum	1300	1173	127	740	867	2040

Table 3: Number of positive and negative cases, valid and invalid best GHT localization solutions, and training cases for each cardiac structure. Positive and calculated negative training cases are shaded in gray. In about 11% of the positive cases an invalid localization is reported as best GHT solution.

6. RESULTS

Table 4 and Table 5 show the results of accuracy experiments that have been executed with confidence-threshold-finder, confidence-SVM, and multi-feature-SVM. Table 4 shows the number of cases for each cardiac substructure and classification type; Table 5 shows the corresponding accuracy and error values. The average error values in Table 5 demonstrate that the error rate decreases for the single confidence feature by 50% if SVM training is used, and decreases further by 50% if down to finally 3% error if an optimal multi-feature combination is selected.

Cardiac structure (landmark)	TV			FE			FV			TE			FP			TN		
	ct	cs	ms	ct	cs	ms	ct	cs	ms	ct	cs	ms	ct	cs	ms	ct	cs	ms
Full heart center	121	122	125	4	3	0	1	1	0	4	4	5	0	0	0	74	74	74
Aortic valve	113	113	115	5	5	3	1	1	1	11	11	11	1	0	0	73	74	74
Pulmonary valve	107	107	114	9	9	2	1	0	0	13	14	14	2	0	0	72	74	74
Mitral valve	105	105	112	10	10	3	4	3	0	11	12	15	1	0	0	73	74	74
Tricuspid valve	100	103	106	14	11	8	2	3	0	14	13	16	0	2	0	74	72	74
Left coronary artery	109	109	114	7	7	2	3	1	0	11	13	14	3	1	0	71	73	74
Right coronary artery	104	107	108	14	11	10	3	3	0	9	9	12	2	1	2	72	73	72
Right inf. pulmon. vein	95	94	103	14	15	6	4	2	3	17	19	18	2	0	0	72	74	74
Right sup. pulmon. vein	105	105	112	10	10	3	2	1	0	13	14	15	1	1	0	73	73	74
Superior vena cava	119	117	118	4	6	5	7	7	4	0	0	3	4	2	2	70	72	72
Sum	1078	1082	1127	91	87	42	28	22	8	103	109	123	16	7	4	724	733	736

Table 4: Results of the cross validation accuracy experiments with threshold classification on the single confidence feature (ct), SVM classification on the confidence feature (cs), and optimized multi-feature SVM classification (ms). Accuracy is calculated from the correctly classified cases that are shaded in gray.

Cardiac structure (landmark)	Confidence thresh. (ct)		Confidence SVM (cs)		Multi feature SVM (ms)		Best Combin.
	Accuracy	Error	Accuracy	Error	Accuracy	Error	
Full heart center	95.59	4.41	98.04	1.96	100.00	0.00	c,g,og
Aortic valve	91.18	8.82	97.06	2.94	98.04	1.96	c,d,g,od
Pulmonary valve	87.75	12.25	95.59	4.41	99.02	0.98	c,og
Mitral valve	87.25	12.75	93.63	6.37	98.53	1.47	c,d,g,od,og
Tricuspid valve	85.29	14.71	92.16	7.84	96.08	3.92	c,d,og
Left coronary artery	88.24	11.76	95.59	4.41	99.02	0.98	c,og
Right coronary artery	86.27	13.73	92.65	7.35	94.12	5.88	c,d,g
Right inf. pulmon. vein	81.86	18.14	91.67	8.33	95.59	4.41	c,d,g
Right sup. pulmon. vein	87.25	12.75	94.12	5.88	98.53	1.47	c,d,g,od
Superior vena cava	92.65	7.35	92.65	7.35	94.61	5.39	c,g,og
Average	88.33	11.67	94.31	5.69	97.35	2.65	

Table 5: Accuracy and error in percent for all three experiments. The multi-feature accuracy is obtained with the listed feature combination, where the abbreviations refer to the feature subscripts in equs. (3.1), (3.2), (3.3), (3.8), and (3.9).

7. CONCLUSIONS

In this report we demonstrate that it is possible to distinguish between valid and invalid results of GHT shape finder localizations by means of classification algorithms. From the experiments, it can be seen that the number of GHT voting counts are already a strong distinguishing feature. Training a classifier for this feature with the SVM method was superior to a simple threshold search, with a reduction of the error rate by approximately 50%. However, training an SVM classifier with additional collective features derived from GHT shape finder solutions decreased the error rate again by approximately 50%. The best achievable error rate for incorrect localization classification was approximately 3%.

When comparing this value to the approximately 11% intrinsically invalid GHT localizations, this indicates that the SVM localization classification method can be used to find localizations within the Hough accumulator that are more accurate than the basic GHT localizations. Similarly the method could be used to decide whether a landmark is present in an image or not. This, however, needs to be verified on a larger number of cases.

8. ACKNOWLEDGEMENTS

The authors would like to thank Katrina Read, Dirk Müller, Amnon Steinberg, and Mark Rabotnikov from Philips HealthTech for providing the CT Data.

REFERENCES

[1] O. Ecabert, J. Peters, H. Schramm, J. Peters, V. Philomin, C. Lorenz, J. von Berg, M. J. Walker, M. Vembar, M. E. Olszewski, K. Subramanyan, G. Lavi, J. Weese, "Automatic model-based segmentation of the heart in CT images," IEEE Trans. Med. Imaging 27(9), 1189-1201 (2008).

[2] D. H. Ballard, "Generalizing the Hough transform to detect arbitrary shapes," Pattern Recogn. 13(2), 111–122 (1981).

[3] H. Schramm, O. Ecabert, V. Philomin, J. Weese, "Towards Fully Automatic Object Detection and Segmentation," Proc. SPIE Medical Imaging 2006, 6144, 614402 (2006).

[4] A. Saalbach, I. Wächter-Stehle, R. Kneser, S. Mollus, J. Peters, J. Weese, "Optimizing GHT-based Heart Localization in an Automatic Segmentation Chain," Medical Image Computing and Computer-Assisted Intervention-MICCAI 2011, 463-470 (2011).

[5] M. Brejl, M. Sonka, "Object localization and border detection criteria design in edge-based image segmentation: Automated learning from examples," IEEE Trans. Medical Imaging 19(10), 973–985 (2000).

[6] P. Beyerlein, "Discriminative model combination," Proc. Automatic Speech Recognition and Understanding Workshop (ASRU), 238–245 (1997).

[7] P. Beyerlein, "Discriminative model combination," Proc. International Conference on Acoustics, Speech, and Signal Processing (ICASSP), volume 1, 481–484 (1998).

[8] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," ACM Trans. Intelligent Systems and Technology 2 (3), 27:1-27:27 (2011).

[9] Boser, B. E., Guyon, I., and Vapnik, V, "A training algorithm for optimal margin classifiers," Proc. 5th Annual Workshop on Computational Learning Theory, ACM Press, 144-152 (1992).