Discriminative Generalized Hough Transform for Localization of Lower Limbs

Heike Ruppertshofen · Cristian Lorenz · Sarah Strunk · Peter Beyerlein · Zein Salah · Georg Rose · Hauke Schramm

Received: date / Accepted: date

Abstract A fully automatic iterative training approach to generate discriminative shape models for usage in the Generalized Hough Transform (GHT) is presented. The method aims at capturing the shape variability of the target object contained in the training data as well as identifying confusable structures (anti-shapes) and to integrate this information into one model. To distinguish shape and anti-shape points and to determine their importance, an individual positive or negative weight is estimated for each model point with means of a discriminative training technique. The model is built from edge points surrounding the target point and the most confusable structure as identified by the GHT. Through an iterative approach, the performance of the model is gradually improved by extending the training dataset with images, where the current model failed to localize the target point. The proposed method is successfully tested on a set of 670 long-leg radiographs, where it achieves a localization rate of 74-97 %.

Keywords Object Localization · Generalized Hough Transform · Discriminative Training · Optimal Model Generation

1 Introduction

Many applications in the area of computer-aided diagnosis are in need of object localization prior to execution in order to run fully automatic. In this paper, we will describe a general method for automatic object localization utilizing the GHT; thereby focusing on the task of joint localization in lower limb radiographs for orthopedic applications. Among these are automatic measurements of length and angles of bones or automatic segmentations, which are necessary for bone density estimation as well as pre-operative planning, as described by Gooßen et al (2010). For most applications, the required positions are still determined manually, which is time-consuming and often presupposes expert-knowledge. These challenges are evaded by the employment of automatic localization procedures, which furthermore have the advantage that the results are reproducible, more reliable and independent of the operator.

In literature, different methods for automatic object localization exist. Many of these are tailored for the specific localization task at hand making use of anatomical knowledge in combination with e.g. gray-value thresholding and morphologic operators (Heimann et al, 2009), which in many cases still require expert-knowledge. More general approaches are given by an atlas-based registration (Seghers et al, 2007) or a global search of the target object conducted with e.g. evolutionary algorithms (Heimann et al, 2007), template matching (Lee et al, 2001) or the GHT (Ballard, 1981). However, the
vast global search has the disadvantage that it is time-
consuming. Another option is the usage of marginal
space learning (Zheng et al, 2009), which determines
the object position and further parameters like scaling
or rotation iteratively keeping multiple transformation
candidates after each iteration.

Despite the presumed long processing times, we want
to focus on the GHT as means for object localization
with the aim to obtain the object position after the
first iteration. The advantages of the GHT are that it
is robust towards image noise and occlusions or missing
object parts, which renders it interesting for medical
applications, especially in case of pre- and post-operative
image acquisitions. The execution of the GHT can be
separated into two types based on the kind of model,
which is used to represent the target object. Different
groups (Leibe et al, 2008; Gall and Lempitsky, 2009;
Maji and Malik, 2009) use appearance models based
on codebooks of image patches. The second approach
is to employ point models representing the shape of
the object (Schramm et al, 2006; Ruppertshofen et al,
2010). Codebooks of image patches are mainly used for
the localization of objects in cluttered scenes. Medical
images are clearly arranged and the objects displayed
are often well represented by their shape, therefore we
prefer the faster shape models.

In order to improve the localization accuracy of the
GHT and to speed up the procedure, we employ the
training of slim, discriminative models, which contain
the variability of the object expressed in the data as
well as information about rivaling object (anti-shapes),
which resemble the object of interest. To this end, each
model point is equipped with an individual weight, rel-
ative to its importance. This approach has also been
followed by Deselaers et al (2005) and Maji and Ma-
lik (2009); yet they do not allow for negative weights,
which we employ to represent anti-shapes. This ap-
proach results in models, which are discriminative for
the target object, and reduces false-positive rates. Through
an iterative training of model points and weights, it is
possible to capture the variability of the target object
such that the sole estimation of object position is suf-
ficient for the localization. Rotation and scaling of the
model do not need to be considered; thus decreasing
processing time. Furthermore, the procedure is not in
need of prior shape or appearance information about
the target object, whereby it runs fully automatic and
is applicable to detect arbitrary objects.

2 Methods

The proposed method for model generation consists of
two main modules: the GHT for object localization to
establish the necessary shape information and a dis-
criminative training procedure (DMC) to estimate model
point weights for an ideal combination of this informa-
tion. The building process of the models is run iter-
atively to successively identify and include the shape and
anti-shape variability contained in the dataset into the
model. Thereby the set of training images is extended
in each iteration by adding images, where the current
model performed badly. A general overview of the pro-
cedure is given in Fig. 1; the individual modules are
explained in the following sections.

2.1 Model Generation and Iterative Model Training

The aim of the iterative supervised training is to cre-
ate models, which capture the variability of the target
object in the dataset, and to determine the importance
of shape points without the incorporation of expert-
knowledge or user interaction.

To this end a number of images are chosen, forming
the development dataset, which is used to generate a
model and to test its performance. This dataset should
contain different characteristics of the target object to
be able to create a discriminative, but yet general model
of the target object. Usually taking a large number of
images is sufficient to achieve this. Part of the devel-
opment images make up the training dataset, on which
the model point weights are determined. This dataset
contains a further subset, which is utilized to gener-
ate new model points. Not all training images are in-
tegrated into the model generation to avoid overfitting
effects and to reduce model size.

At the beginning of the procedure an initial mean-
ingful model is created using the approach shown in
Fig. 2. A region of interest of given size is extracted
around the target point from the images of the model
dataset. By employing these extracts for model cre-
ation, the neighborhood of the target object is included
in the model as well, which in many cases contains significant information about object location. On these extracts, edge points are computed and fused to form the initial shape model. Through this approach shape variability of the target object is included, as becomes obvious in Fig. 2, here especially visible for the angle of the femur. In the last step, the model is thinned to guarantee that each model point carries exclusive shape information to keep its size small while at the same time increasing the entropy of the model. This step is to ensure that the generated model is compatible with the training algorithm as described in Sec. 2.3, which weights the points according to their importance. Points, which carry the same information content, are considered less important and are thereby downgraded. As thinning criteria, the distance and difference in gradient directions of model points are considered, such that they do not leave the same trace in the Hough space.

Since the processing time of the GHT depends on the size of the model, a further point selection is performed after each training step consisting of the GHT and DMC. Thereby the model size is reduced by rejecting points, which obtained low absolute weights and therefore do not have a large impact in the GHT.

The obtained weighted model is finally tested on the development dataset. Images where the current model performed poorly are added to the set of training and model images. Using these new images, the model is extended by adding new points, again using the method shown in Fig. 2. By this method, the model would consist exclusively of shape information. To be able to incorporate anti-shape information as well, a second region is extracted around the location, which obtained the highest vote in the GHT and therefore contains the strongest rivaling object. During the next training iteration, these points will most likely obtain negative weights, repelling the model from the anti-shape location.

Once the model is able to successfully localize the target object in all images of the development dataset, the procedure ends.

### 2.2 Generalized Hough Transform

The GHT is a model based method for object localization, which has been introduced by Ballard (1981) as an extension of the standard Hough Transform for lines or circles. The method is capable of localizing objects of known arbitrary shape and is insusceptible against occlusion or image noise, which makes it very interesting for medical image processing.

The localization procedure operates by transforming the given image in a parameter space, the so-called Hough space, where each cell represents a certain model transformation. In our case, only translation is considered; however, a full affine transformation would be conceivable, but is computationally expensive.

For the localization procedure, a point model representing the shape of the target object is needed. This model consists of the coordinates of model points relative to a reference point (as shown in Fig. 2) and the gradient (or surface normal) direction at each point. During the next training iteration, these points will most likely obtain negative weights, repelling the model from the anti-shape location.

Once the model is able to successfully localize the target object in all images of the development dataset, the procedure ends.

![Fig. 2 Processing chain for model generation. First a region of interest is extracted from the images from which the edges are extracted. These intermediate models are fused to create the final model, which is furthermore thinned as to increase the entropy of each model point. The reference point is shown in red.](image)
model, as described in the previous section, to reduce computational effort.

In the process of model training, the GHT is first performed using an unweighted model, where each model point casts a vote of 1, to obtain the information content of the current model. This information is exploited in the training procedure described in the next section to determine model point weights for an optimal combination of the available information. When employing the GHT for object localization a weighted model is utilized, where each model point votes with its individual weight such that they have unequal impact on the localization result. The determined model point weights can have positive as well as negative values. Thereby the positive weights belong to shape points, while anti-shape points are allocated negative weights.

2.3 Training of model point weights

The DMC (Beyerlein, 1998) employed for weight training is a machine learning technique, which aims at an optimal combination of different knowledge sources into one model.

To deduce the training procedure, we will take a probabilistic view of the Hough space obtained with the unweighted model by transforming it into a probability mass function. This is achieved through normalization of the number of votes \( N_i \) in each Hough cell \( c_i \) by the total number of votes \( N \). Thereby we obtain a posterior probability for each possible object location (represented by the Hough cells) given an image \( X \):

\[
p(c_i|X) = \frac{N_i}{N},
\]

Instead of searching for the cell with the highest number of votes, the result of the GHT can now also be obtained by determining the Hough cell with the highest posterior probability.

Dividing the Hough space into the separate contributions from the individual model points, a model point dependent posterior probability is established for each model point \( m_j \):

\[
p_j(c_i|X) = \frac{N_{ij}}{N_j}
\]

The posterior probabilities stated in (2) cannot solely be used for object localization. Therefore, a suitable combination of these knowledge sources needs to be found. Following the Maximum-Entropy principle introduced by Jaynes (1957) the optimal incorporation of model point information is given via a long-linear combination:

\[
p_A(c_i|X) = \frac{\exp \left( \sum_j \lambda_j \cdot \log p_j(c_i|X) \right)}{\sum_k \exp \left( \sum_j \lambda_j \cdot \log p_j(c_k|X) \right)}.
\]

The coefficients \( A = \{ \lambda_j \}_{j} \) regulate the influence of each model point posterior probability \( p_j(c_i|X) \) on the weighted posterior probability \( p_A(c_i|X) \). The value of \( \lambda_j \) is therefore related to the importance of the model point \( m_j \) and will be used as model point weight.

To estimate the model point weights from (3) with respect to a minimal localization error, an error function \( E \) is defined:

\[
E(A) = \sum_n \sum_i \varepsilon(\hat{c}_n, c_i) \cdot \frac{p_A(c_i|X_n)^\eta}{\sum_k p_A(c_k|X_n)^\eta}.
\]

The function accumulates the weighted error over all images \( X_n \). Thereby the Euclidian distance \( \varepsilon \) of the correct object position \( \hat{c}_n \) and a Hough cell \( c_i \) is chosen as error measure, which is weighted with an indicator function comprised of the posterior probabilities. The exponent \( \eta \) in the indicator function regulates the influence of rivaling hypotheses \( c_k \) on the error measure. The error function is designed in a way that in a cell distant of the true solution a large probability is penalized stronger than a small one, while the true cell, which holds no error, can have the highest probability.

For the determination of optimal \( \lambda_j \), which minimizes (4), a gradient descent scheme is explored. Due to the high dimensional search space and the most likely not convex error function, the existence of a global minimum and therefore its determination cannot be guaranteed. However, the usage of model point weights resulting from a local minimum already significantly increase the localization accuracy of the model as can be seen in Sec. 4.

3 Experimental Setup

3.1 Material and Task

The procedure is tested on a dataset of 670 long-leg radiographs of adult patients with differing field of view. Most images cover both legs from hip to ankle, while some images show only one leg or certain joints. Varying artificial replacements of the femur or knee are visible as well as fractures or further medical conditions. The images were stitched together from up to three images using the procedure described by Gooszen et al (2008). In Fig. 3 a few examples of the database are displayed. To demonstrate the diversity of the images, Fig. 4 shows exemplary extracts of the right knee.
Fig. 4 Exemplary extracts of the right knee to show the diversity visible in the dataset. The images exhibit differences in the rotational angle and size of the joint and the visibility of the fibula. Furthermore, disturbing objects like the ruler or implants may occur.

Fig. 3 Examples from the dataset of long-leg radiographs.

The images have an isotropic resolution of 0.143 mm. In order to reduce processing time the images were down-sampled to a resolution of 1.14 mm.

The localization is performed for the three joints: femur, knee and ankle. The results will be integrated in the segmentation procedure described by Gooßen et al (2010) for an initial placements of the models.

To evaluate the accuracy of the automatic localization, one observer annotated target points for all joints, which are used as ground-truth. The mean intra-observer error for the annotation adds up to 2.3 mm for the femur, 1.3 mm for the knee and 2.6 mm in case of the ankle.

3.2 Design of Experiments

The training procedure will be employed to generate discriminative models for the right joints only. For the localization of the left joints the mirrored model can be used, since there are no significant differences between right and left joints.

Prior to the training procedure, about 60 images without pathological findings for the joint of interest were chosen as development dataset. From these, three images were taken, which form the initial training dataset. One of these images is furthermore utilized for model generation.

After each training step, the three images with the largest localization error are determined and added to the set of training images. New model points are extracted from the image with the largest error.

The training stops if an error of less than 5 Hough cells, which have a spacing of 2.29 mm, is obtained on all images of the development dataset or if no further improvement is achieved.

In the end, the models are tested on the left and right joints of the complete dataset yielding the results shown in the next section.

4 Results

The evolution of the models is shown exemplary in Table 1 for the case of the knee, which needs 4 iterations to converge. As can be seen there, the number of model points increases in every iteration resulting in a more and more discriminative model demonstrated by the decreasing error rate and number of misclassifications on the development dataset.

In case of the knee, the strongest anti-shape is the knee of the opposite leg, which appears quite similar, when regarded without the fibula. While the initial model localized only 44 of 51 knees, from which only 32 are right knees, the final model has a localization rate of 100%. Such that in the end, a model evolved which represents the target object, captures its variability and is capable of distinguishing it from anti-shapes.
In case of the femur 5 iterations are needed to create a meaningful model, while the ankle, which seems to be the most difficult object, probably due to its rotational freedom, needs 6 iterations. The final models are shown in Fig. 5. To relate the model points to the object structures, they are superimposed on an image. Furthermore, the model points are color-coded relative to their weight to be able to distinguish target and anti-shape points and to determine their importance. Model points with negative weights make up about 25-40% of the models.

The femur model contains mainly strong target points, which represent different sizes of the femur head and different angles of the shaft position. The same is true for the ankle model, which concentrates on the gap between tibia and talus, which is the most robust part with only little inter-patient variance. The model points, which extend below the image border, belong to anti-shapes, among others the upper part of the tibia. In case of the knee, the model focuses on the gap between femur and tibia, while also containing strong anti-shapes, in the region where the fibula of the opposite leg would be.

In all models, horizontal or vertical chains of model points can be seen, which most likely originated from the ruler, the image boundary or artifacts introduced by the stitching algorithm. These model points do not contain important information for the object location nor are they reliable anti-shape points. Therefore, further research will aim at the elimination of these points to obtain slim models without redundant information.

The localization performance of the generated models can be seen in Table 2 and the chart in Fig. 6. For a fair comparison, the images are sorted into three categories, namely images with artificial replacements, with pathological findings and the remaining images, where the joints do not exhibit any abnormalities.

The knee has the best localization rate of 97% in case of the normal joints, followed by the ankle with 87%. The femur achieved the lowest rate with only 74%. The low localization rate of the femur is due to an often very low image contrast in that region, which impairs its delineation even for a human observer.

Since the models were trained on healthy images, the correct localization of joints containing artificial replacements and abnormalities cannot be premised. Yet, in case of the knee a localization rate of 85% and 71% is achieved on these images, respectively. In Fig. 7 extracts of example images are displayed where the localization of the right knee was successful, although the task was difficult. Figure 8 shows examples where the localization failed.

To motivate the usage of model point weights, further experiments have been conducted for the knee localization to compare our approach with a standard unweighted GHT. For this purpose, 100 random images without pathological findings of the knee were chosen. The results of the following two experiments are com-

---

**Table 1** Evolution of the knee model. In the top part of the table the number of training images and the size of the trained model is specified. The middle part states the mean error in mm on the training and development images. In the bottom part the number of right or left legs, which were localized by the current model, are listed.

<table>
<thead>
<tr>
<th>iteration</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>no. of training images</td>
<td>3</td>
<td>6</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>no. of model points</td>
<td>75 555 1817 1923</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>error on training data</td>
<td>1.7 2.1 3.3 1.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>error on development data</td>
<td>97.5 25.2 24.8 3.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>right knees localized</td>
<td>32 48 48 51</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>left knees localized</td>
<td>12 2 1 0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

---

**Table 2** Results of the localization task for the three joints. Stated is the size of the respective model and the mean error of the successful localizations in mm divided by the state of the joint.

<table>
<thead>
<tr>
<th>model</th>
<th>size</th>
<th>normal</th>
<th>replacement</th>
<th>anomaly</th>
</tr>
</thead>
<tbody>
<tr>
<td>femur</td>
<td>1608</td>
<td>12.5</td>
<td>14.6</td>
<td>17.4</td>
</tr>
<tr>
<td>knee</td>
<td>1923</td>
<td>4.3</td>
<td>8.5</td>
<td>6.8</td>
</tr>
<tr>
<td>ankle</td>
<td>1187</td>
<td>9.8</td>
<td>-</td>
<td>13.3</td>
</tr>
</tbody>
</table>

---

**Fig. 6** Number of successfully or failed localizations of the femur, knee and ankle. Results are separated based on the state of the joint, which can be normal, with artificial replacement or with pathological findings; results for right and left joints are combined.
In the first experiment, an initial model was created from the four model images, which were determined by the iterative procedure. This model should contain sufficient information to localize all joints correctly since it is a superset of shape points of the weighted knee model created by the previous experiment, which had a localization rate of 100%. Yet, when employing this unweighted model only in 45% of the images the correct knee was localized with a mean localization error of 7.1 mm.

In the second experiment, the trained weighted model was employed, with anti-shape points and weights excluded. This model achieves a localization rate of 91%, but with a larger mean error of 7.1 mm compared to the weighted model.

Apart from yielding a higher localization rate and a lower mean error, the weighted model also generates a clearly arranged Hough space with a definite maximum as can be seen in Fig. 9. This result is more robust and reliable than the other two, although all experiments yield about the same result in this case.

### 5 Discussion

We presented an approach for generation of discriminative shape models and object localization with means of the GHT. The attractiveness of the introduced procedure lies in the fully automatic application flow and the
The necessity of model point weights was proven in further experiments, which revealed that the sole usage of contour information is not sufficient for object localization. In fact, the determination and weighting of robust model points is of high importance to increase localization accuracy and reduce false-positive rates. Establishing this information by hand is a difficult task, which would require expert-knowledge.

The current experiments were run on down-sampled images to reduce processing time. The achieved accuracy is sufficient for the given task of model initialization for segmentation procedures as utilized by Gooßen et al (2010). If a higher accuracy is needed, the procedure could be embedded into a multi-level setting.

Acknowledgements The authors would like to thank General X-Ray, Philips Healthcare for providing the radiographs used in this study. This work is partly funded by the Innovation Foundation Schleswig-Holstein under the grant 2008-40 H.

References


