Semantic Localization-driven Partial Image Retrieval in CT Series

A. Cavallaro1; H.-P. Kriegel2; M. Petri2; M. Schubert2

1 Imaging Science Institute, University Hospital Erlangen, Erlangen, Germany; 2 Institute for Informatics, Ludwig-Maximilians-Universität München, Munich, Germany

1. Introduction

In commercial PACS, querying CT series is currently restricted to retrieving complete series. Simultaneously loading multiple CT series from a centralized server seriously strains the capacities of the available resources and the patience of the waiting users. In many cases, it is not necessary to display the complete series. Some users are only interested in a certain organ whereas others may want to see a body region corresponding to a focused region in another series. All of them want to see their region of interest (ROI) within a minimum waiting time.

Limited bandwidth is also a problem for medical image analytics. Sofka et al. [1] lately introduced a fast solution for landmark detection, which minimizes communication costs by applying a pyramid of varying image resolutions. This strategy is well-suited for divisible machine learning problems and visual presentation [2]. However, seeing a coarse resolution image while waiting for the complete image to load has often no additional value for a radiologist.

An alternative approach is to include a method into the PACS that quickly presents the high-resolution ROI to the user during the loading process, while optionally loading the remaining image material in the background. In [3], Ström and Cosman propose to use compression methods with lossless ROIs for decreasing waiting times. If the presented image region does not completely contain the ROI no harm has been done, except that the user has to wait for the complete series to load. The ROI is usually defined by the image content and not by its explicit coordinates. Hence, it is necessary to employ content-based methods to locate the CT slices intersecting the ROI.

The most established solution to locate ROIs is to annotate CT images with visible anatomical concepts [4]. Thus, a query can be computed by retrieving all images being labeled with the particular concept. To automatically generate these labels, organ and landmark detectors are applied. However, these methods are working on particular areas of the body which might not be displayed in all CT series. Furthermore, organ detectors try to locate the exact spatial boundaries and thus, they usually require an unnecessary computational overhead.

This is an approach of context-based image retrieval and it is not to be confused with content-based image retrieval, which usually relates to a similarity ranking of complete images [5–7]. The actual goal of content-based image retrieval in medicine is a combination of the two approaches: when given a query image, the query system should be smart enough to automatically determine the query context and therefore the locations of potential resulting similar images. In [8], Deserno et al. name this problem the Semantic Gap and identify it as one of the major remaining issues in content-based image retrieval. Consequently, a lot of work has been invested in the area of image registration, the mapping of images into a standardized anatomical reference system [9, 10]. These approaches are usually specialized to a local body part and they are expensive to compute. There-
Therefore, classical image registration approaches are not suitable for our purpose. Deserno et al. call this problem the Indexing Gap [8].

Recently, Criminisi et al. [11] applied regression forests to directly locate bounding boxes around certain organs in CT series. Though this approach does not result in exact organ outlines, it was shown to be considerably faster than more complex detection methods. Further extensions of the method [12] showed a very high robustness for varying body proportions. The main application of these methods is called semantic localization and it aims at the efficient real-time annotation in large PACS. However, the approach still predicts the positions of a fixed set of concepts and it cannot be easily extended to queries being specified by marking a query region in another CT series.

In this paper, we present methods for answering ROI queries retrieving the relevant slices from a CT series with a minimal overhead of irrelevant slices. We propose to map CT series to a height atlas describing body proportions. To summarize, the contributions of this paper are:

i. A method for learning a body atlas as an anatomical model, containing distribution parameters for landmarks and organs on the z-axis.
ii. A framework for mapping concept-based and example-based ROI queries into a standardized height space.
iii. A system for answering ROI queries allowing incomplete landmark annotations.

2. Method

In the following section, we specify our query types and give an overview of the proposed query pipeline. In Section 2.2, we describe two methods for registering a query ROI into a standardized height space, followed by our algorithm for learning a body atlas from semantically annotated example series in Section 2.3. The results of our experimental evaluation are presented in Section 3.

2.1 Workflow Overview and Notation

A PACS in our setting contains multiple CT volumes \( v_i \), represented as stacks of \( z(i) \) 2D-images or slices. The standardized height space \( H = [h_{\text{min}}, h_{\text{max}}] \) is a real-valued range representing the extension of the human body in the z-axis. A mapping function \( h_i : \mathbb{N} \to H \) maps slices of volume \( v_i \) to a standardized height value \( h \in H \). Correspondingly, the reverse mapping function \( s(h) : H \to \mathbb{N} \) maps a position \( h \) in the standardized height space to a slice number \( s \) in \( v_i \). The particular functions will be described in Section 2.2. As a link between the two height spaces, we define matching points as triples \( p_{ij} = (s_i, h, w_j) \in \mathbb{N} \times H \times \mathbb{R} \). One point \( p_{ij} \in P \) associates a slice number \( s_i \in \{0, \ldots, z(i)-1\} \) of volume \( v_i \) to a standardized height value \( h \) in \( H \) and a reliability weight \( w_j \).
Table 1 displays an overview of the most frequently used parameters including some additional notations that will be introduced in the following sections. Figure 1 illustrates the complete workflow of query processing for concept-based queries and example-based ROI queries where the user provides an example ROI.

In our system, a concept-based ROI query retrieves a consecutive sequence of CT slices showing an anatomical content specified by the named concept C ∈ C, e.g., an organ like the heart or the liver. C is the set of all considered query concepts. A body atlas contains one probability distribution for each concept c. This allows us to specify a standardized query range \([h_{lb}, h_{ub}] \subseteq H\) as

\[
\arg\max_{h} \left\{ P_i (H \leq h) \leq t \right\}, \arg\min_{h} \left\{ P_i (H \geq h) \geq t \right\}
\]

for a threshold quantile t and a concept c. (In Section 3.1, we demonstrate a good trade-off between completeness and retrieval speed for t = 0.05).

In example-based ROI queries, the queried ROI is described by the content of a user-selected set of consecutive slices \((s_{lb_1}, \ldots, s_{ub}) \subseteq \{0, \ldots, z(l) - 1\}\) in an example series \(v_x\). The standardized query range \([h_{lb}, h_{ub}] \subseteq H\) can be determined using a mapping function \(h_i(s)\) on the slices of the given example ROI.

With the query interval \([h_{lb}, h_{ub}]\) given, the server then applies \(s_i(h)\) to determine the result set of corresponding CT slices \((s_{lb_1}, \ldots, s_{ub}) \subseteq \{0, \ldots, z(i) - 1\}\). Afterwards, the CT slices from \(s_{lb_1}\) to \(s_{ub}\) are transferred to the client computer.

### 2.2 Mapping Functions

Our interpolation approach maps a series \(v_i\) to the height space \(H\) and back based on a set of available matching points \(P_i\). Inspired by the work of Haas et al. [15], we tested various non-linear interpolation approaches for dealing with varying body proportions and imprecise matching points. Function \(h_i(s_{lb_1}, p)\) maps slice number \(s_{lb_1}\) to a height value \(h_{lb}\) ∈ \(H\). It is adapted to the series \(v_i\) by the set of matching points \(P_i\) and the slice spacing \(\delta_i\), describing the thickness of a slice in the target space \(H\) (Fig. 2), where \(rel_p(s, q)\) is a relevance term for the matching point \(p\) with respect to the input slice \(s_{lb_1}\). The simplest relevance function, \((1n) rel_{lin}(s_{lb_1}) = 1\), is equivalent to applying a weighted linear interpolation. In [13], we used the inverse absolute difference, \((ABS) rel_p^{inv}(s_{lb_1}) = \min(1, |s_{lb_1} - s_p|^{-1})\) as a relevance term, as it showed to be more stable than the squared inverse difference \((SQR) rel_p^{sqr}(s_{lb_1})\) used in [15]. Other, more complex relevance functions showed to be even less suited for obtaining a reliable height interpolation. In accordance with [16], we define a radial basis function kernel (RBF) as

\[
rel_p^{rbf}(s, q) = \exp\left( -\frac{(s - q)^2}{\sigma_p^2}\right)
\]

being the standard deviation of a matching point \(p\) over the training dataset.

For mapping atlas positions in \(H\) to slices in \(\mathbb{N}\), we vary the above interpolation function to a reverse interpolation model \(s_i(h)\) (Fig. 3). The transformation inverts the spacing \(\delta_i\) and swaps the mapping spaces \(\mathbb{N}\) and \(H\).

Additionally, a relevance function \(rel'\) now uses position differences in the atlas scale instead of slice differences like the relevance functions rel.

In order to answer ROI queries on a series \(v_x\), function \(s_i(h)\) requires at least two matching points as \(P_i\). For this publication, we use the landmark detector of [17], which detects up to 22 landmarks within the thorax region. This detector, however, does not return any or not enough matching points for CT series with a blurred image quality or limited anatomical ranges. Thus, we recommend using the instance-based regression technique of [14] for generating matching points if none are available. This approach maps a single slice \(s_{lb_1}\) into a height scale via \(k\)-nearest neighbor regression using an image descriptor for \(s_{lb_1}\). We already used this method in a partial retrieval algorithm detailed in [13], which at the time of writing did not yet make use of a semantic localization atlas.

---

**Table 1** Notation of frequently used parameters

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Explanation</th>
</tr>
</thead>
<tbody>
<tr>
<td>(H = [h_{lb}, h_{ub}] \subseteq \mathbb{R})</td>
<td>Standardized height space</td>
</tr>
<tr>
<td>(h_i)</td>
<td>One height value in (H)</td>
</tr>
<tr>
<td>(s_{lb_1})</td>
<td>One slice number of volume (v_i) in ({0, \ldots, z(i) - 1})</td>
</tr>
<tr>
<td>(p = p_i = (h_i, w_i))</td>
<td>Matching point of (v_i) with reliability weight (w_i)</td>
</tr>
<tr>
<td>(h_i(s_{lb_1}, \ldots, s_{ub}))</td>
<td>Mapping functions (\mathbb{N} \rightarrow H) and (H \rightarrow \mathbb{N}) using matching points (p_i)</td>
</tr>
</tbody>
</table>

![Fig. 2](image-url) \[
\begin{align*}
    h_i(s_{lb_1}, \ldots, s_{ub}) & = \delta_i \cdot s_{lb_1} + \frac{\sum_{p \in P_i} w_p \cdot rel_p(s_{lb_1}) (h_p - \delta_i \cdot s_{lb_1})}{\sum_{p \in P_i} w_p \cdot rel_p(s_{lb_1})} \\
    & = h_i(s_{lb_1}) - \frac{\sum_{p \in P_i} w_p \cdot rel_p'(h_p) (h_p - s_{lb_1})}{\sum_{p \in P_i} w_p \cdot rel_p'(h_p)}
\end{align*}
\]
2.3 Height Atlas Definition

The accuracies of our partial retrieval queries are mainly dependent on the quality of the matching points $P_i$. When using landmark detection for defining matching points, we require a standardized height atlas $(\mathbf{h}, \mathbf{w}) \in H^{[C]} \times \mathbb{R}^{[C]}$, mapping the set of available concepts $C$ to standardized heights and reliability weights.

There are two principal ways to generate such an atlas: using the real-world landmark positions of a template or example series or by forming a consensus model from a set of manually-aligned series. Since these approaches are biased, we additionally propose a refinement algorithm, which improves an existing atlas by iterative adaptation to a larger dataset.

2.3.1 Model by Example (EXMP)

Our example series for using the first approach EXMP is a publicly available full body scan showing a 1.75 m tall woman. We average the output of the landmark detector over varying resolution scales as $h$ and derive the weights $w$ from the variation of the detected coordinates for each landmark.

2.3.2 Manually Aligned Examples (ALIGN)

As an alternative, we form a consensus model ALIGN using a set of 33 CT series (12 female, 21 male patients): all series are manually mapped into a standard [0, 1] scale ($0 =$ feet, $1 =$ head) by defining the start and end positions of the series and linearly interpolating the positions in between. For 12 volumes, all landmarks were found, two vol-

---

**Input:** $s \in \mathbb{N}^{n \times |C|}$: Slice numbers for $n$ volumes and $|C|$ concepts, $M_{\text{seed}} \in H^{[C]} \times \mathbb{R}^{[C]}$: seed atlas $(\mathbf{h}, \mathbf{w})$, $h$: interpolation function

1. **function** $\text{ADAPT}$.$\text{MODEL}(s, M_{\text{seed}}, h)$
2. $M_0 \leftarrow M_{\text{seed}}$
3. $P_i \leftarrow P_i \mid M_{\text{seed}}$
4. for $t \in 1 : t_{\text{max}}$ do
5. \hspace{1em} $P_i \leftarrow P_i \mid M_{\text{seed}}$
6. \hspace{2em} $h_c \leftarrow \text{median}_{i \in \{1, \ldots, n\}} \{h_i(s_{c,e}) \mid \{s_{c,e}, h, w\})\}$
7. \hspace{2em} $w_{c} \leftarrow 1 \left(\text{stddev}_{i \in \{1, \ldots, n\}} \{h_i(s_{c,e}) \mid \{s_{c,e}, h, w\})\} \right)$
8. \hspace{2em} $w_{c} \leftarrow \text{size} (\{i \in \{1, \ldots, n\} | s_{c,e} \text{ was found})$
9. \hspace{1em} $M_{t} \leftarrow (h, w)$
10. $\text{err} \leftarrow \text{LOSSBASE}(s, M_t, h)$
11. if $\text{err}_{t-1} < \text{err}_t$ then $\text{return} M_{t-1}$
12. $\text{return} M_{t}$

**Output:** New height atlas adapted to the dataset $s$

---

![Fig. 4](image-url) Atlas refinement algorithm

![Fig. 5](image-url) The distributions of landmark and organ boundaries of [17] when mapped to the atlas scale via $h(s, \cdot)$ using abs interpolation and EXMP seeding. The bars correspond to the probability (density in [0, 1]) to observe a landmark or organ within a range of 0.225 cm.
umes only resulted in one detected landmark and on average, 16 of the 22 landmarks were detected (with a standard deviation of 7). The standardized height $h_c$ of a concept $c$ is defined as the median of the annotated height values, and the weight $w_c$ is formed from a combination of the inverse standard deviation and the number of series $c$ occurs in. This model is less biased than EXMP, but it can be subject to annotation imprecision.

### 2.3.3 Improving a Standardized Height Atlas

Both atlases can be further improved by using a refinement procedure (c.f. the algorithm in Figure 4) starting with the initial atlas $M_{seed} = (h, w)$. The algorithm requires a set of slice annotations $(s_i, q_i) \in \mathbb{N}^{n \times |C|}$ for the set of detectable concepts $C$ within a database of $n$ volumes which may contain missing values as not all volumes will show all concepts and there may be false negatives of the detector.

In iteration step $i$, the concept positions $s$ are combined with the current model $M_i$ to generate a set of model-induced matching points $\hat{P}_i = P_{h_i, M_{i-1}} = (s_i, h, w)$ for each training series $v_i$. We apply $h_i(s_i)$ to determine a new height value in $H$ for each series $v_i$ and each concept $c$ by excluding the current concept $c$ from the set $P_i$. Then, a new atlas model is generated in steps 4–10 in a similar way as in the ALIGN model. The quality of the new model is evaluated using a least squares error function, $\text{LOSS}_Y(s, M_i, h)$, accumulating the squared difference of the predicted concept positions $h_i(s_i)$ from the newly defined atlas position $h_i$. If the new model is an improvement, the algorithm proceeds with the next iteration $i + 1$, else it returns the previous model $M_{i-1}$.

Let us note that the $H$ coordinates of the matching points generated in step 4 also define the standardized anatomical distributions used for determining the retrieval range $[h_{lb}, h_{ub}]$ for a query by concept. Figure 5 displays these observed density distributions in $H$ for the 22 landmarks in a set of 1,360 CT series. It also shows nine additional organ boundaries; hence these distributions have natural spread. The organ detectors used here were provided by [17].

### 3. Results

In this paper, we use the landmark detector of [17]. Any other landmark detector or even a combination of several detectors could be used as well. Our experiments were conducted on a heterogeneous database of 1,360 CT series from the THESEUS MEDICO data repository provided by the university of Erlangen where at least 20 landmarks could be detected, containing $7 \times 10^5$ 2D images with a total size of 360 gigabytes.

The dataset is characterized in Figure 6: it consists of a varying number of slices covering body ranges of only 40 centimeters (cm) up to one meter of the scanned patient. The slice spacing and slice thickness are at most 0.5 cm.
3.1 Runtime Evaluation

We now demonstrate the advantages with respect to retrieval times of our solution over common PACS. Our goal is partial retrieval of CT series. Therefore, we could not use a real PACS. Instead, we resorted to a remote file system with a custom-designed indexing structure. While mapping into the height atlas with available matching points takes only 3 ms (milliseconds), the detection of the 22 landmarks of [17] usually requires between 3 to 10 seconds. In our group's 100 megabit Ethernet, it takes 110 ms to load one slice of 512 × 512 pixels. Thus, this time investment of 3 seconds pays off if at least 30 slices can be excluded from the series to be loaded. Since the expected number of slices in our repository is 548, it is very likely that in most partial retrieval queries at least 30 slices can be omitted.

In Figure 7, we see histograms over the pruned number of slices for our used database of 1,360 CT series with four predefined retrieval ranges. The retrieval range is the part of the series to be loaded in centimeters. For a retrieval range of 50 cm, we already have a chance of 71% that at least 30 slices can be pruned. For smaller retrieval ranges this probability is by far higher. A 10 cm retrieval range, for instance, has a 95% chance of pruning at least 100 slices. Since Figure 7 was created for the homogeneous dataset summarized in Figure 6, the overall distributions are divided into bulks of CT series with a lower number of slices and CT series with many slices. The retrieval range can be tuned with the \( \tau \) parameter, introduced in Section 2.1.

We examine the effect of \( \tau \) on the retrieval ranges of the average landmark, the average organ and the heart as a single organ in Figure 8. Our chosen setting of \( \tau = 0.05 \) implies a certainty of 90% that the requested range is actually loaded. Figure 8 shows that on the average, 2 cm need to be loaded for a landmark query and that 34 cm need to be loaded for an average organ query, respectively. The heart, which is more compact than, for instance, the lungs has a retrieval range of 15.5 cm. Taking into account the information of Figure 7, even the retrieval ranges of organs allow a clear runtime improvement.

Using the online matching point generating method of [13], both the data reduction and the time gain are even stronger, as one artificial matching point via regression can be generated in only 42 ms. If, for instance, only a certain transversal view is requested, the retrieval range of a volume of 1000 slices and more can quickly be narrowed down to only 10 slices, therefore resulting in a speed gain of up to 99%.

3.2 Retrieval Accuracy

For validating the quality of our atlas generation method, we evaluate all combinations of atlas seedings and interpolation functions. The atlas is built and validated within a 10-fold cross-validation setting on 1,360 CT series providing at least 20 landmarks, the same dataset used for the visualization in Figure 5. Consequently, no tested volume has any influence on its training set. The average localization errors of a single land-
Fig. 9 Leave-one-out localization errors (in cm) of $hi$ (series to atlas) and $si$ (atlas to series positions) for the seeding approaches EXMP (model by example), ALIGN (manually aligned examples) and RAND (random) using 20 matching points with the relevance functions ABS (inverse absolute difference), SQR (inverse square difference), RBF (RBF kernel) and LIN (linear interpolation). The number of iterations required for the training algorithm is displayed at the bottom of the bars. The RAND experiments do not show any seed errors for the anatomically meaningless initial atlases. Since the SQR and RBF interpolation approaches do not converge to any interpretable atlas, they are not displayed as well. SQR interpolation even got caught in an endless loop, denoted by Inf at the position of the bars.

Fig. 10 Leave-one-out localization errors (in cm) for a variable number of used matching points (EXMP seeding) with the same notation as in Figure 7. The solid bars display the performance of $hi$ (series to atlas), the white dashed bars display the performance of $si$ (atlas to series positions).

mark (predicted by using only the remaining landmarks as matching points) are displayed in Figure 9. We transformed the error values to the range of the initial EXMP mapping of 0–175 cm for a correct comparison of the different mappings.

Figure 9 shows that ABS and LIN interpolation are comparably successful with average errors around 1 cm, while the locally more adaptive mappings SQR and RBF (and other, more complex interpolation approaches not shown here) have a stronger susceptibility to outliers, and therefore generate worse mapping results. Atlases based on EXMP seedings converge to comparable errors like those based on ALIGN seedings. However, the EXMP seedings really profit more from the atlas refinement algorithm (cf. the original seed error in dashes). The original ALIGN atlas formed on a set of manually-aligned series is not significantly improved by using our refinement algorithm. When testing random seedings (RAND: uniformly sampled height positions and weights in [0, 1]), atlases of comparable quality resulted for ABS and LIN, whereas SQR and RBF reordered the landmarks into disjoint sub-clusters due to local minima. The most surprising result of these experiments is the good performance of the simple weighted linear interpolation LIN. Let us note that for the pre-dominant spacing in our database of 0.5 cm, an average precision of 1 cm is only a deviation of two slices.

Figure 10 demonstrates the effect of the number of available matching points for EXMP-seeded models. Other seedings show the same trend: there is a strong decrease of the prediction errors when using five instead of two matching points. Using even more matching points results in further but smaller prediction improvements.

Finally, Figure 11 shows the stability of our retrieval approach by displaying extreme values and the standard deviation for EXMP seeding. Note that this figure corresponds to the first set of bars in Figure 10. Once, the mapping function can use at least five matching points, even the maximum error stays below 20 cm for both the mapping from the series to $Hi$, $hi$, and the reverse direction, $si$. Five matching points are consequently a good compromise between the effort of generating matching points and a good localization.
Note that matching points generated by regression can be validated by their leave-one-out prediction error in a set of height-labeled CT series. In [13], we reported an error of 1.88 cm for a query time of 42 ms. Any further increase in accuracy comes at the price of longer query times due to more detailed slice descriptors (e.g., an error of 1.21 cm for 3 seconds per matching point).

Therefore, even though the precision of regression matching points is not as good as the precision based on landmark detection (with errors of 1 cm), using the regression approach allows queries on CT series with missing or error prone landmark annotations.

4. Conclusion

In this paper, we proposed an efficient method for processing partial image retrieval queries using a semantic localization approach on a repository of CT series with the goal of decreased I/O costs. By naming a retrieval context either verbally as anatomical location or by providing an example region, the requested region of interest is inferred from a general height atlas of the human body. This allows the focused retrieval of arbitrary height ranges within the query series even without requiring prior annotation.

Our system is based on interpolation functions using so-called matching points to link CT series to the height atlas. Additionally, the proposed system allows generating matching points based on imaging techniques if no landmarks are available. In our experimental evaluation, we demonstrate the accuracy and runtime costs of our approach.

We conclude that with low runtimes of our height mapping, we can drastically decrease the retrieval times of partial CTs compared to the retrieval of a complete series. A clinical retrieval system can thus save or prioritize valuable time as well as hardware resources by allowing the users to focus on volume portions which are actually relevant to the current anatomical context. For future work, we plan to extend our system to partial CT retrieval in all three dimensions and to test further landmark types.

Acknowledgments

This research has been supported in part by the THESEUS program in the MEDICO project. It is funded by the German Federal Ministry of Economics and Technology under the grant number 01MQ07020. The responsibility for this publication lies with the authors.

References


