
Position Prediction in CT Volume Scans

Franz Graf, Hans-Peter Kriegel, Sebastian Pölsterl, Matthias Schubert
[GRAF,KRIEGEL,SCHUBERT]@DBS.IFI.LMU.DE, POELSTERL@CIP.IFI.LMU.DE
Ludwig-Maximilians-Universität München, Oettingenstr. 67, 80538 Munich, Germany

Alexander Cavallaro
ALEXANDER.CAVALLARO@UK-ERLANGEN.DE
Imaging Science Institute Erlangen, Maximiliansplatz 1, 91054 Erlangen, Germany

Abstract

In their daily work, radiologists often need to localize and align parts of CT volume scans to perform among others differential diagnosis. In these cases, it is desired to load only the relevant sub volume of the scan and to align the sub volume automatically to the correct position of the query scan. Common techniques employ landmark detectors and interpolation to solve this problem. Yet, these techniques are not applicable in case of very small volume scans where the query scan comprises only a small amount of images. In this paper, we propose a method to use small sub volumes in CT volume scans for identifying and aligning CT Scans. Our solution employs combinations of weighted image descriptors and instance-based regression and thus demonstrates the need for machine learning techniques in the case of position prediction. The experiments show that the new method improves the mean error and standard deviation by 6% and 10%, respectively, compared to a state of the art method.

1. Introduction

Radiologists currently spend valuable time by manually localizing and aligning CT volume scans to each other in order to create differential diagnosis or to compare the temporal progress of a patient's status. Here, a problem arises by the permanently growing resolution of CT scanners which also implies a growing size of the recorded CT volume scans. Using current CT scanners, full body CT scans can consume more than

1 GB disk space. Creating a differential diagnosis for just a small part of the body currently requires to completely load the volume scan via the network from the picture archive system. Afterwards, the clinician manually navigates to the corresponding regions and aligns the scans.

Even though this very common task consumes valuable time of the physician and less important high network resources, this domain has not received much attention in the research community so far.

An approach proposed by Bürger (2008) is based on using a topogram to predict the location of a body region. The method uses a set of invariant landmarks and interpolation to create a relative coordinate system. Similarly, Haas et al. (2008) propose an elastic mapping of the slice positions to a reference scale by detecting one of eight predefined anatomic points with known position and interpolating the position of the images between them. Seifert et al. (2009) propose a Probabilistic Boosting Tree (PBT) and 2D Haar features to train multiple landmark detectors that detect up to 19 predefined landmarks. Subsequently, the detectors are incorporated into a Markov Random Field. The disadvantage of these approaches is that they require large volume scans showing a large enough amount of landmarks.

Criminisi et al. (2011) propose a method to detect and localize a set of 10 different organs in CT images. They estimate both the location and the extent for each organ by predicting the bounding box containing each organ. They use a tree-based, non-linear regression approach based on multivariate regression forests. These are similar to random forests but are able to predict continuous values instead of discrete classes.

Feulner et al. (2009; 2011) propose a method that predicts the position of each slice in a normalized coordinate system whose origin and unit length are determined by anatomical landmarks. Each slice is used to

extract an intensity histogram and a descriptor consisting of *Speeded Up Robust Features* (SURF) which are then clustered into visual words. The histogram of visual words is compared to a set of prototype volumes where the body coordinates are known to obtain the position of the slice. Nevertheless, the method is also not applicable in the case of very small query scans.

A similar concept was proposed by [Emrich et al. \(2010\)](#) which combined several feature representations. They assume a standard scale from the sole of the foot to the top of the head with a domain of $[0; 1]$. Feature extraction is based on a modified spatial pyramid kernel, where Haralick-like texture features ([Haralick et al., 1973](#)) are extracted together with histograms of oriented gradients (HoGs) ([Bosch et al., 2007](#)). For localization, the feature vectors are compared to prototype volumes with the use of k -nearest neighbor search and the Euclidean distance. Even though this method requires only a single CT Slice as query object, accuracy and precision are comparable to approaches that require larger volume scans.

In this paper, we propose a method to combine spatially co-located images in CT volume scans to further improve the quality of the localization in cases of very small query volumes. Our solution employs combinations of image descriptors and weighted combinations of spatially neighboring images and instance-based regression.

The rest of this paper is organized as follows: In section 2, we briefly describe the feature descriptor followed by the prediction in section 3. The algorithm is evaluated in section 4 followed by a conclusion and an outlook in section 5.

2. Algorithm

2.1. 2D Features

In contrast to the method proposed by [Emrich et al. \(2010\)](#), which relies on rectangular areas to extract features, our method is based on a histogram in polar space ([Belongie et al., 2000](#)) to provide an improved adaption to the shape of the human body. The final descriptor is a combination of two complementary histograms: one describing the bone structures and the other one describing the distribution of soft tissue inside the body. Due to their rather complementary nature, the descriptors perform differently in different parts of the body so that all regions of the body are covered by at least one descriptor that performs well in this region.

The histograms are constructed as follows: First the

input CT slice I is cropped to the bounding box of the patient’s body, building the image I_c . I_c is then divided into n_x sectors and n_y shells and the origin of the coordinate system is positioned to the center of I_c . The maximum distance r_{\max} to the origin is defined by half of the length of the diagonal of I_c . The angle ϕ of a sector is defined as $\phi = \frac{2\pi}{n_x}$, the width of a shell as $\frac{r_{\max}}{n_y}$. Accordingly, the histogram consists of $l = n_x \cdot n_y$ bins. The value v_i of bin i of a histogram is defined as

$$v_i = \begin{cases} -0.25 & \text{if bin } i \in I_c \\ \frac{p_i}{p_i + n_i} & \text{else,} \end{cases} \quad (1)$$

with p_i representing the amount of relevant pixels and n_i representing the amount of irrelevant pixels. A pixel (x, y) is considered to be relevant if its pixel value is $\leq -500 \text{ HU}^1$ for the soft tissue descriptor and $\geq 300 \text{ HU}$ in case of the bone descriptor. As I_c can have any aspect ratio it may occur that some bins lie completely outside of the image. This information is included into the descriptor by setting a penalty value $v_i = -0.25$ in these cases in order to increase the difference between images with different aspect ratios.

Tests showed that the bone descriptor works best when increasing the weight of the bins of the outer shells. This can be explained by the fact that the most expressive bone structures of the skeleton are located more towards the outside of the body, such as the skull, the chest or the spine. Furthermore, the method becomes less affected by contrast media which are often found in tissues inside the body towards the origin of the coordinate system. As these contrast media show a very large HU value, they also appear in the bone descriptor and would cause severe distortions of the feature vector without weighting.

Thus, we employ the quadratic weighting function (2) on the values of the bins of the descriptor, with $\text{shell}(i)$ denoting the index of the according shell of bin i .

$$v_i = v_i \cdot \text{shell}(i)^2 \quad (2)$$

After creating both histograms mentioned above, a single feature vector is created for each image by concatenating both histograms. For the purpose of dimensionality reduction, a principal component analysis (PCA) is applied to the resulting feature vector.

2.2. 3D Features

The descriptors above only use data of a single image to extract the relevant features. If a volumetric scan is

¹In computer tomography, the Hounsfield unit (HU) scale is used to describe the intensity of a pixel.

already loaded, information from the CT slices neighboring the query slice can be used to make the query vector more robust to noise.

The approach proposed in this section addresses this problem by considering m preceding and m succeeding CT slices when creating a feature vector. As it is used as a post processing step after the creation of feature vectors, it realizes an enrichment to the 2D method so that the flexibility to use a single CT slice still persists.

A new feature vector FV_i^{3D} is formed by calculating the weighted sum of the $2m + 1$ vectors. If the current vector is not preceded or succeeded by m vectors, only the existing vectors are used. Let FV_i denote the i -th feature vector in the sorted list of vectors for a single CT volume consisting of n slices. Then the FV_i^{3D} is calculated as follows:

$$FV_i^{3D} = \sum_{k=\max(0,i-m)}^{\min(i+m,n)} f(|k-i|) \cdot FV_k, \quad (3)$$

where $f(x)$ is one of the following weighting functions:

$$f_{\text{inverse}}(x) = \frac{1}{x+1} \quad (4)$$

$$f_{\text{sigmoid}}(x) = \frac{2}{1+e^{0.3x}} \quad (5)$$

$$f_{\text{polynome}}(x) = -\frac{x^3}{(m+0.1)^3} + 1 \quad (6)$$

$$f_{\text{linear}}(x) = -\frac{x}{m+0.5} + 1 \quad (7)$$

$$f_{\text{inverse-squared}}(x) = \frac{1}{x^2} \quad (8)$$

$$g(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \quad (9)$$

$$f_{\text{Gaussian}}(x) = \frac{g(x)}{g(0)} \quad ; \mu = 0; \sigma = \frac{m}{2}$$

All these functions have in common that $f(0) = 1$ and that the value of $f(x)$ decreases with increasing x , so that the weight decreases with increasing distance to the source feature vector FV_i .

3. Prediction

The task is to map the feature vector FV_i of a CT image with unknown position to a value in the standardized height model in the domain $[0; 1]$.

The prediction is based on a two-stage k -nn search: First, a k_1 -nn search is performed within each of the n volumes of the database. The resulting $n \cdot k_1$ nearest

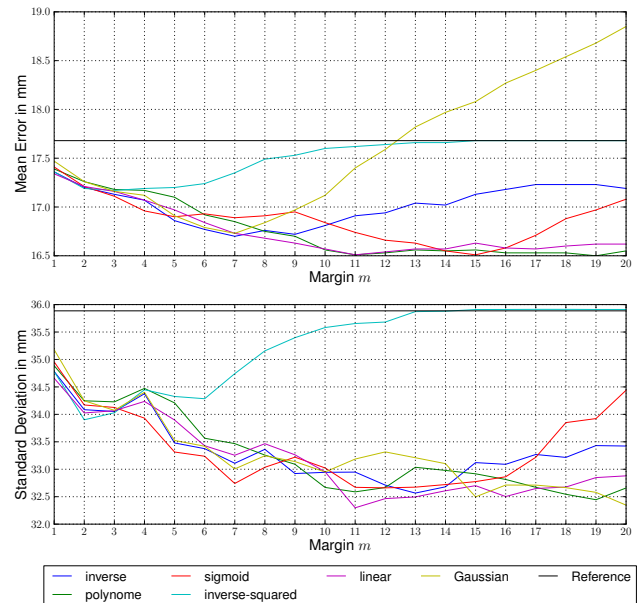


Figure 1. Mean error and standard deviation of cross-validation using 3D features (colored lines) compared to the 2D features (black line) depending on the margin m and the weighting function $f(x)$.

neighbors are added to a single list of feature vectors \mathcal{S}_1 . In the second stage, the k_2 -nn of FV_i in \mathcal{S}_1 are computed. The final position is then predicted by taking the average position value of the features returned by the k_2 -nn search.

For distance computation, we employ the cosine distance measure (10) instead of the Euclidean distance as it performs better on the high dimensional feature vectors than the Euclidean distance (Strehl et al., 2000; Qian et al., 2004).

$$\text{dist}_{\cos}(q, p) = 1 - \frac{\sum_{j=1}^d q_j \cdot p_j}{\|q\| \cdot \|p\|} \quad (10)$$

4. Evaluation

To demonstrate the improvement, we compare our results to the current state of the art method proposed by Emrich et al. (2010).

All experiments were conducted on a database of 59 CT scans of 44 different patients (27 neck and 32 thorax scans) covering the area between the top of the head to the end of the coccyx. Each patient contributed only a single scan per body region to avoid overfitting to a single patient. The data set comprises 25 965 DICOM images using more than 13 GB disk space.

The results shown in Figure 1 show that considering two adjacent slices already made a noticeable differ-

ence. The choice of $f(x)$ also becomes more significant with increasing value of m . It can be seen that all functions except f_{Gaussian} outperform the 2D feature vector (black line) for all values of m in terms of accuracy and precision. Nevertheless $f_{\text{inverse-squared}}$ converges to the performance of the 2D feature vector in case of $m \geq 5$ in both accuracy and precision, while f_{Gaussian} increases precision but loses accuracy at $m \geq 8$.

Considering both e_{mean} and the standard deviation σ , the best results were achieved with f_{linear} (magenta line) and f_{polynome} (green line). Both achieved $e_{\text{mean}} = 16.5\text{ mm}$ and $\sigma \leq 32.4\text{ mm}$ compared to $e_{\text{mean}} = 17.6\text{ mm}$ and $\sigma = 35.9\text{ mm}$ in the 2D case. Thus achieving an improvement of 6% (mean error) and 10% (standard deviation).

5. Conclusion

In this paper we demonstrated the need of machine learning in the field of medical imaging by applying weighted combinations of image features for the localization of small sub volumes in CT scans. We applied the method on a large real world data set and measured the impact by the decrease of the mean error and standard deviation by 6% and 10%, respectively.

Besides enlarging the data set, we plan to evaluate feature selection and boosting as well as more sophisticated machine learning methods for combining the feature vectors to make the localization more robust. The main goal lies in the reduction of large errors. We also plan to create a demo system that visualizes the approach discussed in this paper.

Acknowledgments

This research has been supported in part by the THESEUS program in the MEDICO and CTC projects. They are 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

- Belongie, Serge, Malik, Jitendra, and Puzicha, Jan. Shape context: A new descriptor for shape matching and object recognition. In *In NIPS*, pp. 831–837, 2000.
- Bosch, A., Zisserman, A., and Munoz, X. Representing shape with a spatial pyramid kernel. In *Proceedings of the 6th ACM international conference on Image and video retrieval*, pp. 401–408. ACM, 2007.
- Bürger, Corinna. *Automatic Localisation of Body Regions in CT Topograms*. VDM Verlag, Saarbrücken, 1st edition, 2008. ISBN 978-3-639-00565-3.
- Criminisi, A., Shotton, J., Robertson, D., and Konukoglu, E. Regression Forests for efficient anatomy detection and localization in CT studies. *Medical Computer Vision. Recognition Techniques and Applications in Medical Imaging*, pp. 106–117, 2011.
- Emrich, Tobias, Graf, Franz, Kriegel, Hans Peter, Schubert, Matthias, Thoma, Marisa, and Cavallaro, Alexander. CT slice localization via instance-based regression. In *Proceedings of SPIE*, volume 7623, pp. 762320, 2010.
- Feulner, J., Zhou, S.K., Seifert, S., Cavallaro, A., Hornegger, J., and Comaniciu, D. Estimating the body portion of CT volumes by matching histograms of visual words. In *Proceedings of SPIE*, volume 7259, pp. 72591V, 2009.
- Feulner, J., Zhou, S.K., Angelopoulou, E., Seifert, S., Cavallaro, A., Hornegger, J., and Comaniciu, D. Comparing axial CT slices next term in quantized N-dimensional SURF descriptor space to estimate the visible body region. *Computerized Medical Imaging and Graphics*, 35(3):227–236, 2011.
- Haas, B., Coradi, T., Scholz, M., Kunz, P., Huber, M., Oppitz, U., André, L., Lengkeek, V., Huyskens, D., van Esch, A., and Reddick, R. Automatic segmentation of thoracic and pelvic CT images for radiotherapy planning using implicit anatomic knowledge and organ-specific segmentation strategies. *Physics in Medicine and Biology*, 53(6):1751–71, 2008.
- Haralick, R. M., Shanmugam, K., and Dinstein, I. Textural features for image classification. *IEEE Transactions on Speech and Audio Processing*, 3(6):6103–623, 1973.
- Qian, G., Sural, S., Gu, Y., and Pramanik, S. Similarity between euclidean and cosine angle distance for nearest neighbor queries. In *Proceedings of the 2004 ACM symposium on Applied computing*, 2004.
- Seifert, S., Barbu, A., Zhou, S.K., Liu, D., Feulner, J., Huber, M., Suehling, M., Cavallaro, A., and D., Comaniciu. Hierarchical parsing and semantic navigation of full body CT data. In *Proceedings of SPIE*, volume 7259, pp. 725902, 2009.
- Strehl, A., Ghosh, J., and Mooney, R. Impact of similarity measures on web-page clustering. In *Workshop on Artificial Intelligence for Web Search (AAAI 2000)*, pp. 58–64, 2000.