CASCADED REGRESSION FOR CT SLICE LOCALIZATION

Olga C. Avila-Montes[a], *Uday Kurkure*[a], *Ryo Nakazato*[b], *Daniel S. Berman*[b], *Damini Dey*[b], *Ioannis A. Kakadiaris*[a]

[a]Computational Biomedicine Lab, Dept. of Computer Science, University of Houston, Houston, TX
[b]Department of Imaging, Cedars-Sinai Medical Center, Los Angeles, CA

ABSTRACT

Automated computational tools are needed to estimate the position of a slice of interest within a contiguous stack of slices. Such estimation is useful to retrieve relevant slices from a volume of slices in clinical analysis or it can be used as an initialization step to other post-processing and image analysis techniques. In this paper, we present a novel method to determine the location of a slice of interest within a given volume by formulating it as a regression problem. The input variables for the regression are obtained from simple intensity features computed from a pyramid representation of the slice. We assess the performance of the proposed method by comparing the estimated positions of slices of interest in CT data with manual annotations. Our method was validated on a dataset of 45 volumes and promising results were obtained for 5 different target slices, the average error being 2 slices.

Index Terms— non-contrast CT, regression, image retrieval

1. INTRODUCTION

As part of the common practice of scanning large parts of a patient's body for diagnosis and therapy, large amount of data [1] is obtained in each computed tomography (CT) scan in clinics and imaging centers around the country everyday. Non-contrast CT has been used as a tool for cardiovascular risk assessment, especially for coronary artery calcium scoring (CACS). These cardiac CT scans might possess other valuable information. For example, pericardial fat has been shown to be associated with coronary heart disease (CHD) events [2]. Also, Eisen *et al.* [3] showed an association of thoracic aorta calcium with an increased risk of CHD.

The assessment of these CT scans is generally performed manually or by semi-automated/automated methods. To perform manual assessment, slice(s) of interest, on which measurements are made, are first identified. For example, measurement of the diameter of the thoracic aorta is performed either on the slice at the *lower level of the pulmonary artery bifurcation* [4] or at the *mid-level of right pulmonary artery* [5]. Furthermore, to apply a semi-automated/automated method for thoracic fat measurement [6, 7] or calcium scoring [8, 9], the upper and lower limit of heart need to be specified if the scans contain slices beyond the *base* and *apex of the heart*. Similarly, the ascending aorta segmentation method by Avila-Montes *et al.* [10] requires a lower limit as the slice closest to the *root of the aorta*. Thus, there is a critical need for the automated localization of such key slices in a CT volume.

Feulner *et al.* [11] presented a method for estimating the body region of an input CT volume based on 1D registration of histograms of visual words. These are obtained by computing intensity histograms and SURF descriptors which are then classified into visual words and accumulated in a histogram. However, as the authors state, their method has difficulty localizing single slices. Seifert *et al.* [12] used PBT classifiers and HAAR features to detect salient slices. All the slice detectors are connected into an network which has the advantage of preserving the slice ordering.

In this paper, we propose a novel framework for estimating the position of a slice of interest within a CT scan. Unlike other methods that require a similarity function to retrieve a slice of interest, we use regression techniques to directly estimate the position of the target slice without the need for a similarity criterion. Thus, given any slice in a volume, the location of the target slice can be estimated using the learned regression model. However, to robustly estimate the precise position of the target slice, we have to learn a very complex regression function with a large number of independent variables. Instead, we use simple intensity moments as independent variables to learn a less complex regression function and use its estimate to initialize another function, successively. Thus, each function refines the estimates of the previous function by compensating for the errors of its predecessor. We evaluate the algorithm by comparing the predicted positions with manual annotations of the following slices of interest: the lower level of the pulmonary artery bifurcation, the mid-level of right pulmonary artery, the lower level of right pulmonary artery, the base of the heart and the root of the aorta.

2. METHODS

Given a CT volume, we want to predict the location of the slice of interest using the learned regression model. The outline of the method is presented in Algorithms 1 and 2.

Algorithm 1 Learning Phase

L1: Extract the region of interest (ROI)
L2: Compute features in the spatial pyramid representation of the ROI
L3: Perform dimensionality reduction via Principal Component Analysis
L4: Estimate the model parameters using grid search
L5: Learn an embedding function via Cascaded Regression

Algorithm 2 Testing Phase

T1: Extract the region of interest
T2: Compute features in the spatial pyramid representation of the ROI
T3: Project features onto the lower-dimensional space obtained by PCA
T4: Predict the position for all the slices in the input volume
T5: Compute a density estimate using a kernel-smoothing method
T6: Select the position with the highest density in the position space

2.1. Learning Phase

L1: To include only relevant information, we extract the inner thoracic cavity using a graph-based method [13]. To reduce the effect of rotational variations in the slices at different body positions of a subject, all slices were aligned with each other along the major axis of the obtained thoracic cavity. Next, we fit a bounding box around the thoracic cavity as our region of interest. However, since the features computed from the thoracic cavity contain a lot of redundant information from non-relevant regions (e.g., the lungs in our case) that decreases the descriptiveness of the features, we further reduce the region of interest, which adapts to the relevant regions within the thoracic cavity. The reduced region of interest is obtained by fitting a bounding box on the heart region obtained by thresholding and connected-component analysis (Fig. 1).

L2: The target slices are generally identified by the presence/absence or (dis)appearance of one or more anatomical structures from those slices with respect to the neighboring slices. Also, these anatomical structures have well-defined anatomical positions relative to other structures with some inter-subject variations. To capture all the information from the region of interest and also keep the spatial ordering of the features with some tolerance to the inter-subject variations, we use a spatial pyramid representation [14] to aggregate local statistical features as a slice descriptor.

The descriptor is computed in the following manner: (i) a pyramid representation is built by dividing the region of interest successively into sub-regions of decreasing size, where at each level the number of sub-regions are doubled along each dimension (Fig. 2); (ii) a feature vector is computed by aggregating various statistical measurements (mean, standard deviation, skewness, kurtosis, and entropy) of the intensity values from each sub-region at each level of the pyramid; and (iii) the slice descriptor is obtained by concatenating the features vectors from all the sub-regions. Thus, the final slice descriptor characterizes the region of interest and also captures the local information without disregarding its spatial layout.

L3: The dimensionality of the slice descriptor is further reduced by using Principal Component Analysis (PCA).

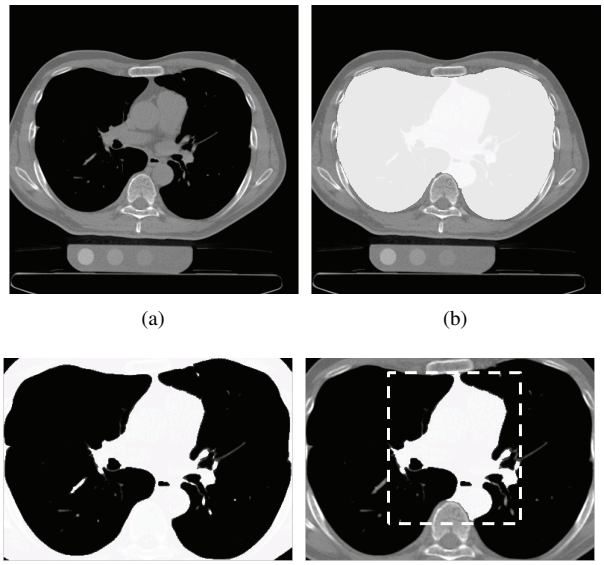

Fig. 1: Depiction of the extraction of the region of interest (dashed line). (a) Original CT, (b) thoracic cavity mask, (c) threshold mask of tissue of interest, and (d) heart region (obtained by keeping the largest component after applying the AND operation between the masks (b) and (c)).

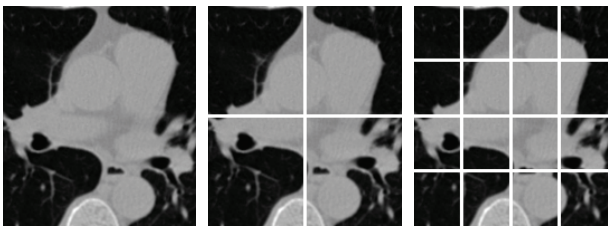

Fig. 2: Image grid obtained at three different levels of resolution (a) level 0, (b) level 1, and (c) level 2.

L4: Any regression technique can be used to learn the model. In this work, we use support vector machine based regressors (SVR). Therefore, we perform a grid search in the space of parameter C, which indicates the weight of the error penalty, and γ, which determines the width of the RBF kernel used. To create the predictors, for each pair (C, γ), we conduct a five-fold cross validation over the learning data set, and select the parameters leading to the highest accuracy.

L5: We formulate the problem of slice localization as a regression problem where the dependent variable to be learned is the distance (in mm) of a particular slice from the target slice. We use the slice descriptors as independent variables to learn a regression function and use its estimate to initialize its successor. Moreover, unlike a single regression function that uses a fixed set of variables, the variables of the cascaded functions are updated at each level of the cascade based on the current estimate of the position of the target slice [15].

We construct a cascaded regressor $\mathcal{R} = (R^1, \cdots, R^T)$, such that each component regressor is trained to learn the embedding function of the dataset $\mathcal{D} = \{(\vec{x}_1, y_1), \ldots, (\vec{x}_N, y_N)\}$, where N is the total number of slices in the dataset, $\vec{x}_i$ is the extracted feature vector for the i^{th} slice, and y_i is the label to be predicted for the i^{th} slice. With each regressor level, the objective is to minimize the difference between the target location and the prediction obtained by the previous regressor.

To obtain the regressor model R^1, we define the label y_i by $E(\hat{\theta}, \theta_i)$ where $E(\cdot)$ is a measure of relative error between the location of the target slice $\hat{\theta}$ and the initial location of the i^{th} slice. In subsequent levels, $t > 1$, the label y_i is defined by $E(\hat{\theta}, R^{t-1}(\vec{x}_i))$, where $R^{t-1}(\vec{x}_i)$ is the resulting prediction from previous regressor R^{t-1}. The process is repeated until either the regressor R^t is incapable of reducing the prediction error compared to the previous level or when a set number of T levels is reached.

2.2. Testing Phase

T1-2: Given an unseen volume, the region of interest extraction and feature computation of each slice in the volume is performed in the same manner as steps L1 and L2.

T3: Each feature descriptor obtained in the previous step is then mapped onto the low-dimensional subspace, determined during the learning phase described in step L3.

T4: Since each slice of the volume contributes to determine the location of the target slice, we evaluate the cascaded regressor $\mathcal{R}$ obtained in the step L5, and predict the position θ^T for each slice. Each predicted position is obtained by:

$$\theta^t = \theta^{t-1} + R^t(\vec{x}_{\theta^{t-1}}) \tag{1}$$

from $t = 1, \cdots, T$, where $\vec{x}_{\theta^{t-1}}$ is the feature vector computed from the slice at position θ^{t-1}, and $R^t(\vec{x}_{\theta^{t-1}})$ is the resulting prediction from the regressor at level t. The initial position θ^0 of each slice is its location (in *mm*) within the volume.

T5-6: We determine the location $\hat{\theta}$ of the target slice within the unseen volume by as follows:

$$\hat{\theta} = \arg\max K([\theta_1^T | \cdots | \theta_M^T]) \tag{2}$$

where M is the total number of slices in the scan, θ_i^T is the resulting prediction from the cascaded regressor for the i^{th} slice of the volume and $K(\cdot)$ is the kernel-smoothing density estimate. This process is depicted in Fig. 3.

3. RESULTS AND DISCUSSION

The performance of the proposed method was assessed on 45 non-contrast cardiac CT datasets. The images were acquired at the Department of Imaging, Cedars-Sinai Medical Center, as part of the EISNER registry; the acquisition protocol has been described in detail by Dey *el at.* [7]. Each scan consisted of a stack of 49 to 79 axial slices with a resolution of 0.68 *mm* $\times$ 0.68 *mm* and slice thickness of 3 *mm*, resulting in a total of 2701 CT slices. The dataset was split into two groups: a learning set with 15 randomly selected volumes, resulting in a total of 894 CT slices; and a testing set with the remaining 30 volumes, with a total of 1807 CT slices. We used the LIBSVM [16] package to learn the regression functions.

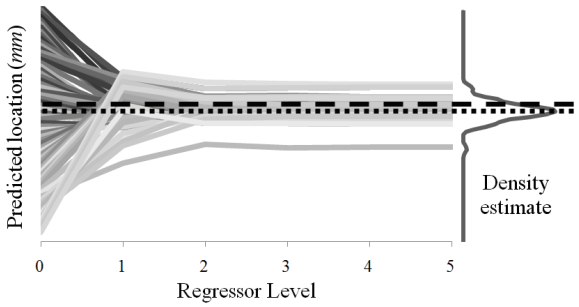

Fig. 3: Schematic depicting the process of slice position prediction. The dashed line represents the localization of the target slice, while the dotted line represents the predicted position.

In order to quantitatively assess the quality of the proposed method, our measure of error was defined as the distance between the target slice location and the predicted location in voxel space. The descriptors were computed from a three-level pyramid. The functions were trained using 40 dimensional descriptors obtained after dimensionality reduction. The number of principal components to be used for the descriptor was determined empirically.

We compared the performance of the method on different slices from the cardiac CT scan: lower level of the pulmonary artery bifurcation, mid-level of right pulmonary artery, the base of the heart and the root of the aorta. We also experimented with the lower level of right pulmonary artery slice and found it to be more discriminatory, which was corroborated by the localization results. Table 1 provides descriptive statistics of the error measures. Note that the performance of the proposed method improved significantly when the reduced adaptive bounding box (Sec. 2.1) was used. In addition, the lower level of right pulmonary artery was the slice most robustly detected by our method. In future work this slice can be used to find the position of other target slices by using a locally driven cascaded regression. We are currently using a 2D descriptor, but including additional information from neighboring slices would be beneficial to improve the quality of the model. We consider this a topic for future research.

The main limitation of our method is that it cannot preserve the ordering of slices. For example, in a cardiac CT scan the mid-level of right pulmonary artery slice is always above the slice at the root of the aorta, but our method cannot ensure that the relative position of these slices will be maintained. While the accuracy on localizing slices from a CT scan covering the thoracic region has been assessed, further experiments need to be carried out to evaluate the performance of our slice localization method in other regions (e.g., abdomen).

Table 1: Descriptive statistics of the error (mean $\pm$ std) in number of slices. BB_1 and BB_2 corresponds to the results using the reduced adaptive bounding box and the thoracic cavity bounding box, respectively.

Target Slice	BB_1	BB_2
Lower Level of Pulmonary Artery Bifurcation	1.9 ± 1.8	3.2 ± 2.9
Mid-level of Right Pulmonary Artery	2.6 ± 2.2	4.5 ± 3.1
Lower Level of Right Pulmonary Artery	1.4 ± 1.5	5.3 ± 3.7
Base of the Heart	2.4 ± 1.6	5.6 ± 5.1
Root of the Aorta	2.0 ± 1.7	3.5 ± 2.7

4. CONCLUSION

In this paper, we presented a regression-based method for position estimation of a slice of interest within a volume. We proposed to use a series of regression functions in a cascaded manner, where each function refines the estimate of the previous function. We evaluated the performance of our method by comparing its predictions with the manual annotations and obtained promising results.

Acknowledgment

This work was supported in part by NIH 1R21EB006829-01A2 and the UH Eckhard Pfeiffer Endowment Fund. Any opinions, findings, conclusions or recommendations expressed in this material are of the authors and may not reflect the views of the NIH or UH.

5. REFERENCES

[1] G.D. Rubin, "Data explosion: The challenge of multidetector-row CT," *European Journal of Radiology*, vol. 36, no. 2, pp. 74–80, 2000.

[2] V.Y. Cheng, D. Dey, B. Tamarappoo, R. Nakazato, H. Gransar, R. Miranda-Peats, A. Ramesh, N.D. Wong, L.J. Shaw, P.J. Slomka, and D.S. Berman, "Pericardial fat burden on ECG-gated noncontrast CT in asymptomatic patients who subsequently experience adverse cardiovascular events," *JACC Cardiovascular Imaging*, vol. 3, no. 4, pp. 352, 2010.

[3] A. Eisen, A. Tenenbaum, N. Koren-Morag, D. Tanne, J. Shemesh, M. Imazio, E.Z. Fisman, M. Motro, E. Schwammenthal, and Y. Adler, "Calcification of the thoracic aorta as detected by spiral Computed Tomography among stable angina pectoris patients: Association with cardiovascular events and death," *Circulation*, vol. 118, no. 13, pp. 1328, 2008.

[4] A. Wolak, H. Gransar, L.E.J. Thomson, J.D. Friedman, R. Hachamovitch, A. Gutstein, L.J. Shaw, D. Polk, N.D. Wong, R. Saouaf, S.W. Hayes, A. Rozanski, P.J. Slomka, G. Germano, and D.S. Berman, "Aortic size assessment by noncontrast cardiac Computed Tomography: Normal limits by age, gender, and body surface area," *Journal of the American College of Cardiology*, vol. 1, pp. 200–209, 2008.

[5] S.S. Mao, N. Ahmadi, B. Shah, D. Beckmann, A. Chen, L. Ngo, F.R. Flores, Y.L. Gao, and M.J. Budoff, "Normal thoracic aorta diameter on cardiac Computed Tomography in healthy asymptomatic adults: Impact of age and gender," *Academic Radiology*, vol. 15, no. 7, pp. 827–834, 2008.

[6] R. Yalamanchili, U. Kurkure, D. Dey, D.S. Berman, and I.A. Kakadiaris, "Knowledge-based quantification of pericardial fat in non-contrast CT data," in *Proc. Society of Photographic Instrumentation Engineers Medical Imaging Conference*, San Diego, CA, Feb. 13-18 2010, vol. 7623, p. 76231X.

[7] D. Dey, N.D. Wong, B.K. Tamarappoo, R. Nakazato, H.G. Gransar, V.Y. Cheng, A. Ramesh, I.A. Kakadiaris, G. Germano, P.J. Slomka, and D.S. Berman, "Computer-aided non-contrast CT-based quantification of pericardial and thoracic fat and their associations with coronary calcium and metabolic syndrome," *Atherosclerosis*, vol. 209, no. 1, pp. 136 – 141, 2010.

[8] U. Kurkure, D.R. Chittajallu, G. Brunner, Y. Le, and I.A. Kakadiaris, "A supervised classification-based method for coronary calcium detection in non-contrast CT," *International Journal of Cardiovascular Imaging*, vol. 26, no. 7, pp. 817–828, March 2010.

[9] I. Isgum, A. Rutten, M. Prokop, and B. van Ginneken, "Detection of coronary calcifications from Computed Tomography scans for automated risk assessment of coronary artery disease," *Medical Physics*, vol. 34, no. 4, pp. 1450–61, 2007.

[10] O.C. Avila-Montes, U. Kurkure, and I.A. Kakadiaris, "Aorta segmentation in non-contrast cardiac CT images using an entropy-based cost function," in *Proc. Society of Photographic Instrumentation Engineers Medical Imaging Conference*, San Diego, CA, Feb. 13-18 2010, vol. 7623, p. 76233J.

[11] J. Feulner, S.K. Zhou, S. Seifert, A. Cavallaro, J. Hornegger, and D. Comaniciu, "Estimating the body portion of CT volumes by matching histograms of visual words," in *Proc. Society of Photographic Instrumentation Engineers Medical Imaging Conference*, Lake Buena Vista, FL, Feb. 7-12 2009, vol. 7259, p. 72591V.

[12] S. Seifert, A. Barbu, S.K. Zhou, D. Liu, J. Feulner, M. Huber, M. Suehling, A. Cavallaro, and D. Comaniciu, "Hierarchical parsing and semantic navigation of full body CT data," in *Proc. Society of Photographic Instrumentation Engineers Medical Imaging Conference*, Lake Buena Vista, FL, Feb. 7-12 2009, vol. 7259, p. 725902.

[13] D.R. Chittajallu, P. Balana, and I.A. Kakadiaris, "Automatic delineation of the inner thoracic region in non-contrast CT data," in *Proc. International Conference of the IEEE Engineering in Medicine and Biology Society*, Minneapolis, MN, Sep. 2-6 2009, pp. 3569 – 3572.

[14] S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Washington, DC, USA, 2006, pp. 2169–2178.

[15] P. Dollar, P. Welinder, and P. Perona, "Cascaded pose regression," in *Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, Jun. 13-18 2010, pp. 1078 –1085.

[16] C.-C. Chang and C.-J. Lin, "LIBSVM: A library for support vector machines," 2001, Software available at http://www.csie.ntu.edu.tw/.