

## INJURED LUNG VOLUME ESTIMATION ON CT DATA USING LINEAR METRICS

N.A. Smelkina<sup>1</sup>, A.V. Kolsanov<sup>2</sup>, S.S. Chaplygin<sup>2</sup>, P.M. Zelter<sup>2</sup>, A.G. Khrarov<sup>1</sup>,  
A.V. Nikonorov<sup>1</sup>

<sup>1</sup>Samara National Research University, Samara, Russia

<sup>2</sup>Samara State Medical University, Samara, Russia

**Abstract.** The article proposes a method for measuring the volume of the injured regions of human lung tissue on CT scan based on the volume of healthy lung tissue subtraction from the total lung volume in the absence of pathology. The total lung volume is calculated on the base of the regression model using a combination of linear metrics produced manually. The volume of healthy lung tissue is determined using adaptive thresholding. Experimental studies have shown that the error of the proposed technique does not exceed 10% in comparison with ground true evaluation made by expert. This technique is useful for the case of large injury, when accurate lung segmentation is impossible.

**Keywords:** diagnostic images, the image linear metrics, image processing, data mining, linear regression model

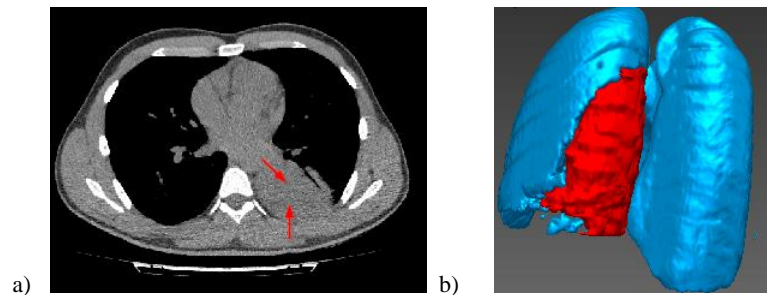
**Citation:** Smelkina NA, Kolsanov AV, Chaplygin SS, Zelter PM, Khrarov AG, Nikonorov AV. Injured lung volume estimation on CT data using linear metrics. CEUR Workshop Proceedings, 2016; 1638: 401-410. DOI: 10.18287/1613-0073-2016-1638-401-410

### 1 Introduction

Historically medicine considered being a kind of art, but nowadays it is a science, that's why tendency to make medical diagnose more objective, sometimes quantitative is very popular. Segmentation based on computed tomography – CT-scans is the best accurate method for lung volume evaluation, quantitative principles were used in patients with emphysema and cancer. These parameters would help doctor to predict effectiveness of surgery or sometimes avoid it because of high risk. Unfortunately, segmentation of contused or injured tissue cannot be done accurate enough.

Currently, the main methods for the segmentation of lung [1] are based on the voxel-wise analysis of the intensity function, texture features of the CT-scans. In any of these methods presence of pulmonary contour is mandatory. In case of interstitial lung lesions (pneumonia, alveolar infiltration) (fig.1) or severe mechanical injury the contour of the lung is often unexpressed or even absent. Various methods to segment the lungs with this pathology and calculate their volume exist [1, 2, 3], but they are com-

putationally expensive, while achieving error of 10%, and require a lot of initial information and manual image processing, such as: the location of the damage, main points that characterize the lung, etc.



**Fig. 1.** Alveolar infiltration in the lung: a) a 2D slice; b) a 3D polygonal model (red color – pathology, blue color – lungs)

Anyway, all of these methods of segmentation aimed at diagnosis lungs disease and the determining of its quantitative characteristics. As the methods for finding the volume of injuries are very complicated and slow, reached insufficient accuracy of calculations, some papers addressed the methods of quantification the volume of lung disease, avoiding organs segmentation. For example, in [4] proposed a method for determining the amount of pleural effusion, having the characteristic form, using simple measurements on the plane and in space, without segmenting pathology. The accuracy of this method varies 4.5-5.1%, depending on the selected measurements. In this paper, we propose to estimate the total lung volume with the help of a simple linear metrics, without using segmentation tools.

## 2 Linear metrics for abnormal lungs volume estimation

We propose the following linear metrics for abnormal lungs volume estimation. Metric a (2a) is Euclidian distance from the maximum lung point in the vertical projection on the lower slice to the minimum lungs point in the vertical projection on the upper slice. Metrics b and c (2b, 2c) represent the “depth” and “thickness” of the lungs. Metric d (2d) is measured between the extremes, the easiest points to select. Metric e (2e) contains a diametral characteristic of the pleural cavity and represents a Euclidian distance from the minimum lung point to the maximum lung point in the horizontal projection.

Thus, estimating the total lung volume, we can find the volume of the pathological areas. It should also be noted that the anatomical structure of the lung (the position, shape, texture) is unique, and in this case, in contrast to the [4], it is impossible to use lung's shape prior information to increase the model accuracy.

Here we evaluated linear regression models on the base of CT dataset consisting of 49 patients with healthy lungs. For each of these patients were found the metrics and volume of the lungs. The test sample consisted of 8 patients. Two patients had severe

mechanical injury to the lungs as a result of the accident, and one patient with alveolar infiltration in the lung. As the best model was chosen with minimum sum of squared residuals coupled with the cross-checking. The cross-validation has been used as an additional criterion of the quality of regression models.

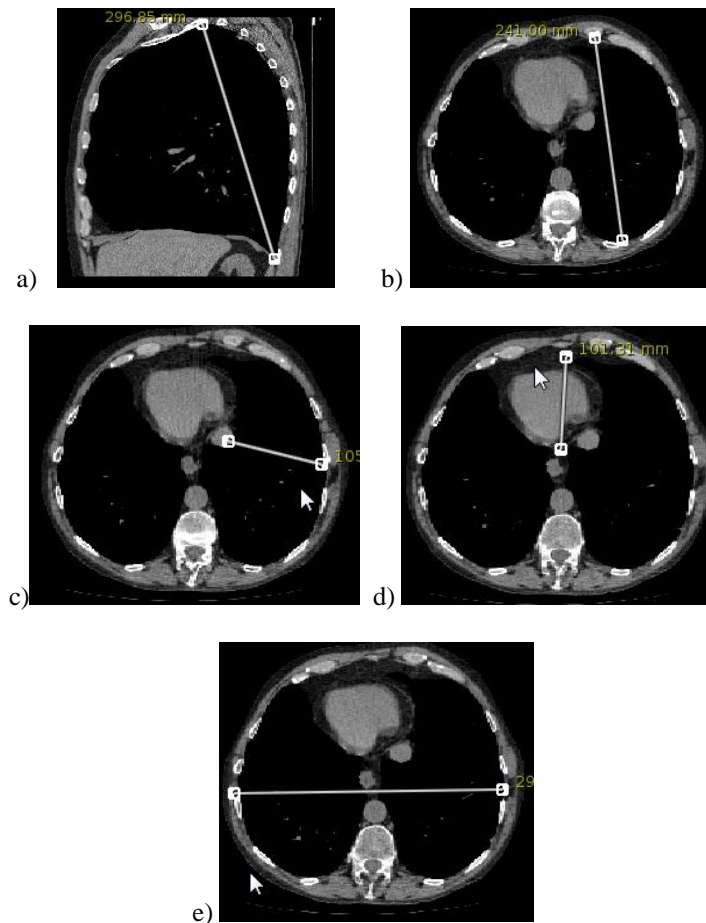


Fig. 2. Lungs metrics: a) metric a; b) metric b; c) metric c; d) metric d; e) metric e

### 3 Linear regression model for lung volume estimation

Here we used metrics a-e and their products as regressors to estimate lungs volume using linear regression model. The classical linear regression model with the parameter vector  $\theta$  is given by [5]:

$$Y = F\theta + \xi,$$

where  $F$  - is an input regressors,  $Y$  - observation vector,  $\xi$  - observation noise vector.

One of the quality criteria of this model is the minimum sum of squared residuals:

$$R_0 = \|\mathbf{y} - \mathbf{F}\boldsymbol{\theta}\|_2^2 \rightarrow \min_{\boldsymbol{\theta}}. \quad (1)$$

The cross-validation procedure has been chosen as an additional quality criterion. This procedure is a standard method of testing and comparing of regression models and is an empirical evaluation of generalization capability of algorithms. In contrast with the average error on the training set, used earlier as a criterion, the average error of cross-validation is unbiased [6].

In our research we used the following cross-validation procedure – the CT dataset randomly divided into training and control samples. The criterion (1) was calculated in the control sample to obtain a more accurate estimation.

Stepwise regression [7] was also used for the selection of the most informative metrics combination, with the largest contribution to the variation of the dependent variable. The forward stepwise regression consists of successive inclusion of factors in the equation of regression (forward stepwise regression), checking their significance. The criterion of informativeness of different metrics is also criterion (1).

The best model of stepwise regression is following:

$$Y = 7.4646 - 5.2426 \times a + 3.2191 \times b - 6.63 \times e + 3.3367 \times (a \times e), \quad (2)$$

$$R_0 = 0.4928.$$

Thus, the most informative metrics were a, b, e. These measurements more fully describe the lungs in three basic dimensions. Note, that models containing information about the "thickness" of lungs (metric c) and the distance between the extreme points of lungs (metric d) - showed less accurate result.

Best value of the criterion (1), was obtained for the model depending on the product of the three most informative metrics and has the following form:

$$Y = 0.1608 + 0.2496 \times (a \times b \times e) + 0.0097 \times (a \times b \times e)^2, \quad (3)$$

$$R_0 = 0.1635.$$

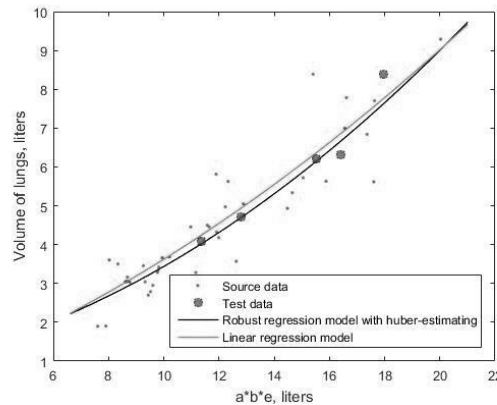
The graph of this model is shown in fig. 3.

The source data are sufficiently large variation along the regression curve. That's why we also used more robust criterion for constructing regression using Huber's M-estimator [8]. The feature of the method of robust regression is as follows: points located far from the main regression curve, make a linear contribution to the model. On the criterion (1) this method showed slightly more accurate results than the ordinary regression. The model is as follows:

$$Y = 0.8243 + 0.1135 \times (a \times b \times e) + 0.0148 \times (a \times b \times e)^2, \quad (4)$$

$$R_0 = 0.1445.$$

The graph of this model is also shown in fig. 3.



**Fig. 3.** Graphs of regression models (3) and (4)

The best value of criterion (1) obtained using the cross-validation and depending of only one metric, is significantly larger, the model has following form:

$$Y = 0.3049 + 0.6282 \times (b \times b) + 0.1582 \times (b \times b)^2, \quad (5)$$

$$R_0 = 1.5215.$$

Respectively, the robust regression model:

$$Y = 4.0058 - 1.5649 \times (b \times b) + 0.4493 \times (b \times b)^2, \quad (6)$$

$$R_0 = 1.7005.$$

The graphs of models are shown in fig. 4.

#### 4 Estimation of the healthy lung tissue volume

Estimation of the volume of healthy lung regions is by using the Otsu's method. We divide the image into two regions according to the intensity. One region includes the area of healthy lung tissue and the air around the body, and the second region consists of the body. And then we set a point to the lung area in first region, leaving areas only associated with this point. So we find the area of healthy lung tissue.

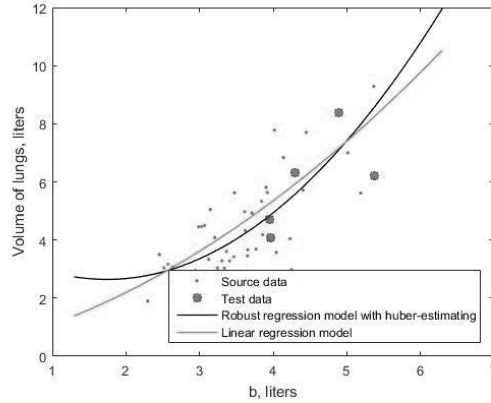


Fig. 4. Graphs of models (5) and (6)

#### 4.1 Otsu's method

Otsu's method, due to its universality, is widely used in image processing and pattern recognition. The algorithm is based on dividing the pixels in the grayscale image on "useful" and "background" by finding the threshold reducing intraclass dispersion which determined as a weighted sum of the dispersions of the two classes [9]:

$$\sigma_w^2(t) = w_1(t)\sigma_1^2(t) + w_2(t)\sigma_2^2(t),$$

where weights  $w_i$  - the probability of two classes, separated by threshold  $t$ ,  $\sigma_i^2$  - dispersions on these classes.

The resulting binary image is as follows:

$$y(i, j, k) = \begin{cases} 1, & x(i, j, k) \in D \\ 0, & x(i, j, k) \notin D \end{cases},$$

where  $D$  - lung area and air,  $x(i, j, k)$  - pixel of source grayscale image,  $y(i, j, k)$  - pixel of resulting binary image.

#### 4.2 Determination of connected region

After dividing image into 2 regions, we have an image with lungs and air around body. Thus, we set a point in the lungs, defining the area of interest and cast away the air. Thereafter processed area connected with this point. Thus, the resulting image is given by:

$$y(i, j, k) = \begin{cases} 1, & x(i, j, k) \in D \cup x(i, j, k) \in O \\ 0, & x(i, j, k) \notin D \cap x(i, j, k) \notin O \end{cases},$$

where  $O$  - connected region with a given initial point.

Then we calculate the amount of lungs received by the formula binary image:

$$V = \frac{1}{10^6} \sum_{i=1}^n \sum_{j=1}^m \sum_{k=1}^l x_1(i, j, k) \times size(i) \times size(j) \times size(k),$$

where  $x_1(i, j, k)$  – pixels with value 1,  $size(i) \times size(j) \times size(k)$  - the sizes of the corresponding voxels in cubic millimeters ( $mm^3$ ),  $[V] = liters$ .

## 5 Experimental results

Cross-validation of proposed technique was made by dividing dataset of 49 CT-scans into train and control samples. The control sample consists of 5 patients with normal lungs and 3 patients with pathology. Pathology volume is calculated as a difference between total volume of healthy lungs, determined by the formula of linear regression, and the volume of healthy lung tissue:

$$V_{path} = V_{total} - V_{healthy}.$$

For CT-scans of patients with normal lungs difference should be close to 0. Accuracy of the method for patients with pathology is calculated as follows:

$$eps = \left| \frac{V_{total} - (V_{healthy} + V_{path})}{V_{path}} \right|. \quad (7)$$

For patients with normal lungs the formula calculation error is as follows:

$$eps = \left| \frac{V_{total} - V_{healthy}}{V_{healthy}} \right|.$$

Table 1 shows the results of a study of the model (2).

Dashes in the table indicate that the model gives an unacceptable error value, and in this case model cannot be used.

As shown in table 2, the error of this method does not exceed 10% and some cases as low as 0.3%. Ground true value of pathology volume was evaluated manually for estimation of the criterion (7).

A model based on a single metric is less accurate as compared to the model, depending on the composition of different metrics.

**Table 1.** The experimental results for the model (2)

Patient number	1	2	3	4	5	6	7	8
Pathology	no	no	no	no	no	yes	yes	yes
$a, dm$	2.452	2.210	2.681	2.362	2.641	2.613	2.302	2.722
$b, dm$	2.073	1.990	2.211	1.988	2.319	1.563	1.446	1.756
$e, dm$	3.233	2.588	3.031	2.725	2.535	2.623	2.392	2.607
$V_{healthy}, l$	6.305	4.09	8.388	4.705	6.200	3.14	1.974	4.003
$V_{total}, l$	6.299	4.383	7.792	5.086	6.819	4.484	2.732	5.456
$V_{path}, l$	-	-	-	-	-	0.596	0.697	0.584
$eps$	<b>0.001</b>	<b>0.072</b>	<b>0.071</b>	<b>0.081</b>	<b>0.099</b>	-	<b>0.087</b>	-

**Table 2.** The experimental results for the models (3), (4)

Patient number	1	2	3	4	5	6	7	8
Pathology	no	no	no	no	no	yes	yes	yes
$a, dm$	2.452	2.210	2.681	2.362	2.641	2.613	2.302	2.722
$b, dm$	2.073	1.990	2.211	1.988	2.319	1.563	1.446	1.756
$e, dm$	3.233	2.588	3.031	2.725	2.535	2.623	2.392	2.607
$V_{healthy}, l$	6.305	4.09	8.388	4.705	6.200	3.14	1.974	4.003
$V_{total}^{(3)}, l$	6.869	4.253	7.760	4.933	6.363	3.948	2.742	4.777
$V_{total}^{(4)}, l$	6.675	4.029	7.627	4.691	6.145	3.739	2.666	4.537
$V_{path}, l$	-	-	-	-	-	0.596	0.697	0.584
$eps^{(3)}$	<b>0.09</b>	<b>0.04</b>	<b>0.075</b>	<b>0.049</b>	<b>0.026</b>	<b>0.212</b>	<b>0.101</b>	<b>0.326</b>
$eps^{(4)}$	<b>0.059</b>	<b>0.015</b>	<b>0.091</b>	<b>0.003</b>	<b>0.009</b>	<b>0.004</b>	<b>0.007</b>	<b>0.086</b>

According to table 3, error of models (5), (6) significantly exceeds the error of models (3), (4). In some cases, the model is inapplicable due to a very large calculations error. This is due to lack of knowledge about the form of lungs and lack of spatial dimensions. Enough high error due to the fact that the dispersion of metrics is rather large.



**Table 3.** The experimental results for the models (5), (6)

Patient number	1	2	3	4	5	6	7	8
Pathology	no	no	no	no	no	yes	yes	yes
$b, dm$	2.073	1.990	2.211	1.988	2.319	1.563	1.446	1.756
$V_{healthy}, l$	6.305	4.09	8.388	4.705	6.200	3.14	1.974	4.003
$V_{total}^{(5)}, l$	5.922	5.275	7.154	5.255	8.258	1.840	2.310	3.746
$V_{total}^{(6)}, l$	5.573	4.856	7.090	4.835	8.584	2.864	2.698	3.452
$V_{path}, l$	-	-	-	-	-	0.596	0.697	0.584
$eps^{(5)}$	<b>0.061</b>	<b>0.29</b>	<b>0.147</b>	<b>0.117</b>	<b>0.332</b>	-	<b>0.518</b>	-
$eps^{(6)}$	<b>0.116</b>	<b>0.187</b>	<b>0.155</b>	<b>0.028</b>	<b>0.384</b>	-	<b>0.039</b>	-

## 6 Conclusion

This work aims to determine the pathological lungs amount for the purpose of diagnosing the severity of injuries. We studied the method of simple calculation pathological lung volume using linear metrics. Estimation error, obtained for the proposed method less than 10%, when compared with ground true value. There are two main causes of model inaccuracy. First is human intervention - the arrangement most accurately describing the simple lung's measurement. Second cause is lung's anatomical structure. Also, the accuracy of the regression model somewhat reduced because quality of images in the training sample is very different. The main error of the algorithm associated with a large spread of metrics. This problem will be addressed in future works. Currently we are researching the possibility of automatic measurement arrangement for minimizing computation errors.

## Acknowledgements

The work was performed as part of the project "Autoplan" (automatic system of planning and monitoring of the operation) under the state contract to perform research and development work "Development of technology and organization of production of the automated systems of planning, management and monitoring of results of surgical treatment" ("4.3-Autoplan-2014").

## References

1. Mansoor A, Bagsi U, Foster B, Xu Z, Papadakis GZ, Folio LR, Udupa JK, Mollura DJ. Segmentation and Image Analysis of Abnormal Lungs at CT: Current Approaches, Challenges, and Future Trends. *RadioGraphics*, 2015; 35(1): 1056-1076.
2. Birkbeck N, Sofka M, Kohlberger T, Zhang J, Wetzl J, Kaftan J, Zhou SK. Robust Segmentation of Challenging Lungs in CT Using Multi-stage Learning and Level Set Optimization. *Computational Intelligence in Biomedical Imaging*. Springer, New York, 2014; 185-208.
3. Mansoor A, Bagsi U, Foster B, Xu Z, Papadakis GZ, Folio LR, Udupa JK, Mollura DJ. A Generic Approach to Pathological Lungs Segmentation. *IEEE Transactions on Medical Imaging*, 2014; 33(12): 2293-2310
4. Hazlinger M, Ctvrtlik F, Langova K, Herman M. Quantification of pleural effusion on CT by simple measurement. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub*, 2014; 158(1): 107-111.
5. Draper NR, Smith H, Pownell E. *Applied regression analysis*. New York: Wiley, 1966; 3.
6. Vorontsov KV. Combinatorial approach to the assessment of the quality of the trained algorithms. *Mathematical questions of Cybernetics*, 2004; 13: 5-36. [in Russian]
7. Efron MA. *Multiple regression analysis. Mathematical methods for digital computers*, 1960; 1: 191-203.
8. Huber PJ. *Robust statistics*. – Springer Berlin Heidelberg, 2011: 1248-1251.
9. Otsu N. A threshold selection method from gray-level histograms. *Transactions on systems, man, and cybernetics*, 1979; 9(1): 62-66.