Early Diagnosis of Lung Tumors by Genetically Optimized 3D-Metaball Malignancy Metric

Vitoantonio Bevilacqua
Dipartimento di Elettrotecnica ed Elettronica – Politecnico di Bari
Via Orabona, 4, Bari, 70125, Italy
bevilacqua@poliba.it

Giuseppe Filograno
Dipartimento di Elettrotecnica ed Elettronica – Politecnico di Bari
Via Orabona, 4, Bari, 70125, Italy
filograno@gmail.com

Michele Fiorentino, Antonio E. Uva
Dipartimento di Ingegneria Meccanica e Gestionale – Politecnico di Bari
{m.fiorentino, a.uva}@poliba.it

ABSTRACT
We present a novel approach to the early diagnosis of lung tumors that considers new malignancy indexes by using a metaball-based representation of this neoplasia. Starting from CT data we extract the suspected tumors represented as approximating metaballs and calculate malignancy indexes based on precise volume and surface irregularity evaluation. The mentioned approximation is performed resolving a constraint problem using a genetic algorithm whose objective function is a mathematical representation of the metaball. Compared to existing art, the metaball approximation provides a consistent reference surface for the evaluation of novel diagnosis parameters borrowed from engineering surface analysis. We have implemented the method in a demonstrator software and analyzed three different test cases.

Categories and Subject Descriptors
I.4 [**Computing Methodologies**]: Image Processing and Computer Vision

General Terms
Algorithms, Measurement.

Keywords
Lung Tumor diagnosis, Biomedical computation, Surface analysis, Genetic Algorithm Constrained Optimization Problem, GENOCOP III.

1. INTRODUCTION
In order to improve early diagnosis of tumors, researchers are working on reliable metrics to evaluate cancer growing probability in suspected areas. This is one of the most challenging research topic involving mathematicians and engineers together with biological sciences researchers. The motivation lays in the fact that, in industrial countries, cancer has now moved from the seventh to the second place in ranking of the fatal diseases, being surpassed only by cardiovascular diseases [1]. The World Health Organization estimates that, nowadays, cancer kills approximately six million people annually. Furthermore in Europe, in the near future, there will be more people over 60 than under 20, so that age related diseases, such as cancer, will impact even more the society.

This work mainly focuses on lung carcinomas which are a group of malignant cells that originates from tissues of the bronchus and lung parenchyma. This cancer represents the first death cause in men and the second in women, after the breast cancer. One of the principal risk factors of this neoplasia is the smoke. To reduce the cancer-related mortality with an early diagnosis, prevention checkups with computed tomography (CT) are strongly recommended. The images obtained are used to reconstruct patient lungs suspected carcinoma models, used to estimate the irregularities of the identified volumes and then their possible malignancy. Several medical studies have shown that tumor malignancy can be evaluated by two different growth indexes. The former index in literature is related to the whole tumors volume whilst a more recently proposed index is based on the surface irregularities. The most significant index of the first category is the Doubling Time growth factor [2, 3]. It estimates the time used by a tumor to double its mass according to an exponential growth model. Typically its evaluation requires two volume measurements of the same tumor spaced by several days. Benign lung tumors have a high value of doubling time whilst malignant could redouble even in 30-80 days [4]. For this metric, the main issue is the precise evaluation of the tumor volume. The second metric, more recent in literature [5], is related to the tumor surface irregularities because a discriminating element between benign and malignant tumors is indeed the local invasion [6]. In fact, benign masses presents an expansive behavior, while malignant ones an infiltrative one, therefore showing higher surface irregularities. The surface characterization has been widely applied in various fields of engineering, e.g. material engineering, road monitoring, remote sensing and contact design engineering. The ISO 14460-1 standards rule the surface definition and classify the surface profile into three components: roughness, waviness and form. The first has finer ridges comparing to waviness, which is smoother than roughness. This work concentrated on roughness, which can be characterized by applying several statistical parameters, such as average roughness, root mean squared roughness, maximum peak height, minimum valley depth, and so forth. In medical literature the application of 3D surface roughness has been recently used for the study of injury related to psoriasis [7, 8].

In a previous work [5] we presented a tumor shape approximation using both clusters of spheres and ellipsoids optimized by genetic algorithm. In that work we proposed a shape regularity metric

GECCO'12 Companion, July 7–11, 2012, Philadelphia, PA, USA.

using either the sphere number and radii or the distance from the ellipsoidal model. In this paper we extend and improve the previous work introducing a more effective approximation of the suspected tumor volumes based on metaball. This representation allows us to easily evaluate surface roughness parameters for early diagnosis of tumor malignancy.

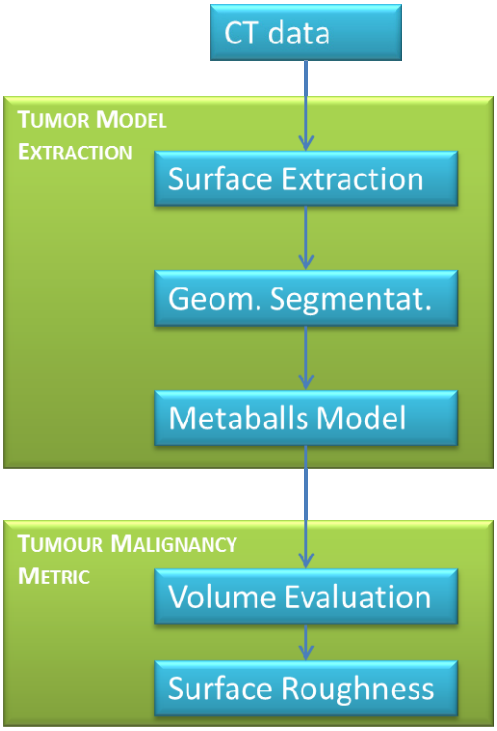


Figure 1: Workflow of the proposed approach.

Therefore the steps that characterize this work are: (i) 3D extraction of the tumor polygonal surface from axial CT data, (ii) surface segmentation to isolate tumor region from surrounding tissues, (iii) metaball approximation, (iv) evaluation of tumor volume and (v) surface roughness estimation. In Figure 1 we group these steps in two main phases. The first, called Tumor Model Extraction, processes the CT data outputting a suspected tumor volume modeled as a metaball. In the last phase, called Tumor Malignancy Metric, we analyze the metaball model to compute diagnosis parameters.

In the next section we describe the materials we used in our research.

2. MATERIALS

Three different malignant lung tumors, available for this study, are diagnosed independently by two different expert radiologists. Data are obtained using a 16 rows helical CT multi slice scanner, at a resolution of 1 mm. Each examination consists of about 500 images with resolution of 512x512 in DICOM format. They were provided by MASMEC [9], a company that implemented an innovative intra-operatory virtual guiding system called SIRIO designed provide physicians with a simple and transparent interface for reaching deep target nodules.

DICOM stands for Digital Imaging and COmmunications in Medicine. DICOM is not just an image or file format. It is an all-encompassing data transfer, storage, and display protocol built and designed to cover all functional aspects of digital medical imaging. Figure 2 depicts some slices of a lung tumor in one dataset. In our implementation the user can select a Region of Interest (ROI) starting from the entire view of the two lungs.

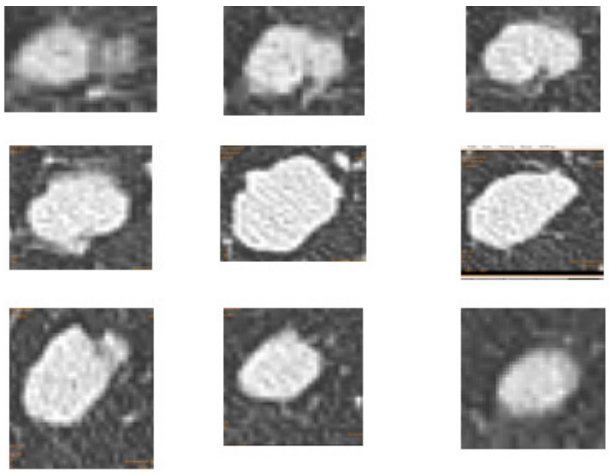


Figure 2: Several sections of a lung tumor provided in DICOM format.

In the shown images each pixel represents a X-ray level of absorption by the human tissues, typically expressed in Hounsfield Unit (HU). Bone and metallic structures completely absorb these rays so that they have a very high HU value. Instead the air surrounding lungs is dark because it does not absorb X-rays. Modern CT scanners can produce gray-scale digital images, in which pixels use 12 or more bits, showing all shades of human body. It is worth underlining that traditional monitor can only renders 256 gray levels so that reducing the values produced by CT scans is necessary. To do this, DICOM images needed to be processed using the windowing technique, that focuses on a particular range of the HU values, depending on the tissue under investigation. In our implementation we have used a so called *lung window* to better display lung tissues. In the next paragraph we analyze the algorithms used to extract the tumor model.

3. TUMOR GEOMETRY MODEL

This phase starts with a stack of DICOM images representing a section of the lung tumor and outputs a 3D mathematical model of its geometry. Our method consists of 3 main steps: (i) surface extraction, (ii) segmentation and (iii) metaball approximation. In the following sections each step will be presented in detail.

3.1 Surface Extraction

In this phase we extract the boundary of the lung tumor forming a 3D polygonal surface representation. Cancer polygonization is obtained using the "Marching Cubes" algorithm, presented in the 1987 by Lorensen and Cline [10]. This algorithm is characterized by two main steps: first, it locates the surface corresponding to a user specified value and then it creates triangles. This approach, in fact, uses a divide-and-conquer tactic to locate the surface in a logical cube created from eight pixels (from two adjacent slices, four pixels each). The algorithm determines how the surface intersects this cube, then moves, or march, to the next cube. At the end, melting consistently all polygons, the searched surface is realized and the algorithm outputs a triangular mesh. This mesh is used both for volume approximation and 3D visualization. In this phase the doctor using her/his knowledge can interactively chose

the target Hounsfield value for the surface and visualize immediately the result.

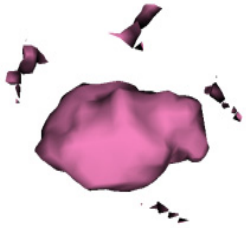


Figure 3: A 3D mesh of a lung carcinoma obtained with Marching cubes

Analyzing our lesions with expert radiologists, we have found that -150 HU is an optimal value to display lung tumor. The Figure 3 shows an example of a lung tumor surface extracted by the algorithm using the abovementioned Hounsfield value.

3.2 Segmentation

Typically a pulmonary lesion can be located inside the lung, or it can be attached to the visceral pleura. In Figure 3 we can see that the cancer is well separated so it is so simple to isolate the tumor from the surrounding tissues. Unfortunately, the marching cubes algorithm cannot distinguish the tumor if it is attached to the pleura. An example of this situation is depicted in Figure 4. To cope with these situations we use an approach based on the k-means clustering algorithm [11]. Generally speaking, this algorithm divides a group of objects into k partitions (or clusters) minimizing the total variance in each one. It is worth underlining that, in our implementation, a cluster is identified by its centroid and the k value is fixed. The algorithm follows an iterative procedure: (1) creation of k partitions and entry points assignment at each partition generated either randomly or with some heuristic information, (2) centroids computation for each cluster, (3) building of new partitions by associating each entry point to the cluster whose centroid is closest to it and (4) return to the step 2. The algorithm stops when the value of each centroid does not change.

We have used this algorithm on two different datasets of the same 3D scene, realizing two clustering phases. In the k-means it is important to define the number of cluster and with a "trial and error" approach we have chosen k=2 in each of the two phases. The idea behind the first clustering phase is that the points having a lower value along the z coordinate could belong to the pleura. These coordinates are the input of the k-means algorithm and at the end the largest cluster is used for the next phases. Starting from this set, we (i) find the polygons related to the points and (ii) run again the k-means on polygon normals. In other words the output of the first clustering stage is used to create the input of the second phase. In fact, for each point of the first phase we find the related polygon and finally compute its normal. Polygon normals are preferred to point normals for the better accuracy and are used as input of the second clustering phase. By only rendering the polygons belonging to the largest cluster (see Figure 5), our algorithm discriminates the tumor from the pleura because the latter typically has normals almost aligned.

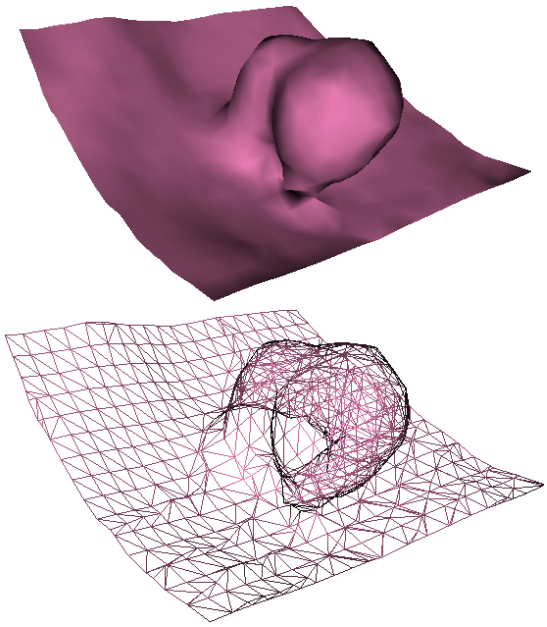


Figure 4: Solid and wireframe view of a tumor anchored to visceral pleura.

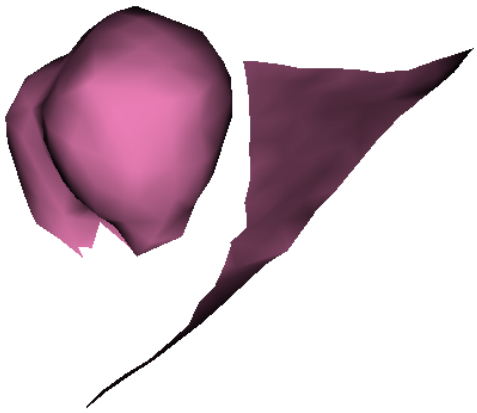


Figure 5: The tumor in Figure 4 after the k-means segmentation.

At this point, the tumor is well isolated from surrounding tissues (as in Figure 3) and, analyzing all the 3D connected component and the surface extension, it is possible to extract it (Figure 6). If the process of automatic segmentation is not satisfactory, the user can reduce the ROI on the DICOM image and then manually isolate the lesion.

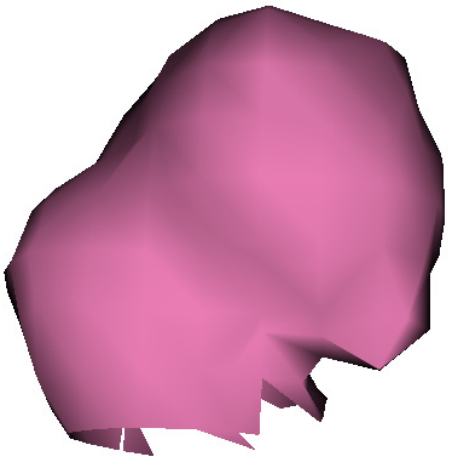


Figure 6: Automatically isolated tumor surface.

3.3 Metaball approximation

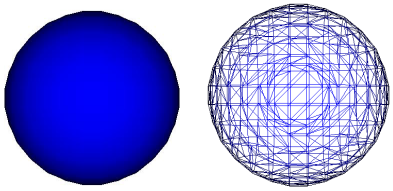
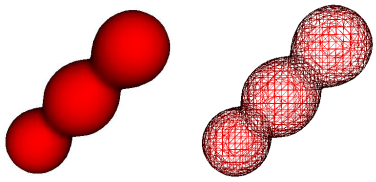
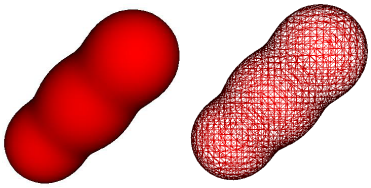
The purpose of this phase is to represent the tumor shape in a mathematical form in order to develop an effective metric for the malignancy evaluation. However, tumor geometries are difficult to be represented using eulerian geometry (primitives like spheres, cylinders and even NURBS). In this work we have used the metaball as primitive to approximate the tumor shapes. A metaball is an iso-surface in 3D space introduced for the first time by J. Blinn [12] in 1982, in order to render complex molecules. Today they are largely used to model organic objects by means of graphics tools (such as Blender, Maya, 3D Studio Max, etc.). Another application of metaball concerns the reproduction of blobby elements like animated smoke or sky clouds in modern 3D videogames. Furthermore in literature they are also used to approximate point clouds [13 - 16]. The main advantage of these primitives is their ability to represent tumor geometries in a very compact form for further computation.

A metaball can be represented by every function that satisfies two basic properties: finite support and smoothness. The finite support means that the function describing metaball goes to zero when (x, y, z) goes to infinitive, while the smoothness means that the shape does not have corners. This last property is usually verified and depends on the mathematical expression used to describe them. For our purpose we have used the Equation 1:

$$\sum_{i=0}^{N} \frac{k_i}{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2} = T \qquad (1)$$

where N is the number of balls (or spheres) that compose a metaball, $C_i=(x_i, y_i, z_i)$ represents the i^{th}-ball center coordinate, k_i is a parameter strictly related to the i^{th}-ball radius and T is the threshold of the isosurface. In Table 1 examples of metaball with different parameters values are displayed.

Table 1. Examples of Metaball varying parameters in the Equation 1.

Parameters	Metaball
N=1 T=1 C = (0, 0, 0)	
N=3 T=1 C_1 = (0, 0, 0) C_2 = (1.5,1.5,1.5) C_3 = (3, 3, 3)	
N=3 T=0.6 C_1 = (0, 0, 0) C_2 = (1.5,1.5,1.5) C_3 = (3, 3, 3)	

The algorithm performed to approximate the tumor geometry with a metaball is GENOCOP III. GENOCOP stands for GEnetic algorithm for Numerical Optimization of COnstrained Problems and blends the Genetic Algorithm theory with the Operation Research theory. GENOCOP III is the third version of this class of algorithms that was developed by Michalewicz et al. [17] in order to optimize objective functions respecting some linear and non-linear constraints. Starting from Equation 1 we have defined an objective function to minimize in the Equation 2.

$$f(X,Y,Z,K,T) = \left| \sum_{i=0}^{N} \frac{k_i}{(x-x_i)^2+(y-y_i)^2+(z-z_i)^2} - T \right| \qquad (2)$$

X, Y and Z are three vectors containing information about the coordinate of the center of each ball, K contains the k_i parameters and T is the threshold related to the iso-surface. We have to minimize the aforementioned objective function, maintaining N fixed. This parameter starts from one and it is increased in each run of the genetic algorithm. As metric to stop the iterations we

have used RMS (Root Mean Squared) between this geometric primitive and the tumor surface obtained in section 3.1. Moreover we have introduced two extra constraints. The first is that the iso-surface threshold T can range from 0.1 to 1. The second constraint is that the approximating metaball must be watertight, i.e. a closed shape, so we have imposed that each sphere center must be in a rectangular parallelepiped (or cuboid).

The position and the dimensions of this cuboid are determined running GENOCOP III using an ellipsoidal model, explained in a previous work [5]. After this, a translation of the tumor is performed in order to move it in the center of the coordinate system and a rotation is done to normalize its position in 3D scene, using the angles related to the approximating ellipsoid. Finally, the abovementioned parallelepiped is built taking as dimensions the 80% of the ellipsoid axes.

The approximating metaball having lowest RMS will be the candidate for malignancy evaluation. Figure 7 depicts an example of the best approximation of a lung tumor using 19 spheres (each center is represented as a green point). Figure 8 shows how the RMS in the previous example is not related to the fitness value of the genetic algorithm.

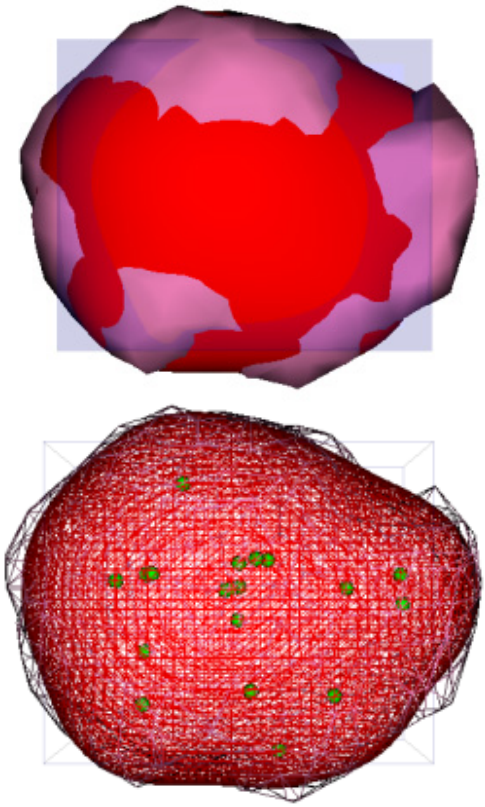


Figure 7: Best approximation of a tumor using the metaball.

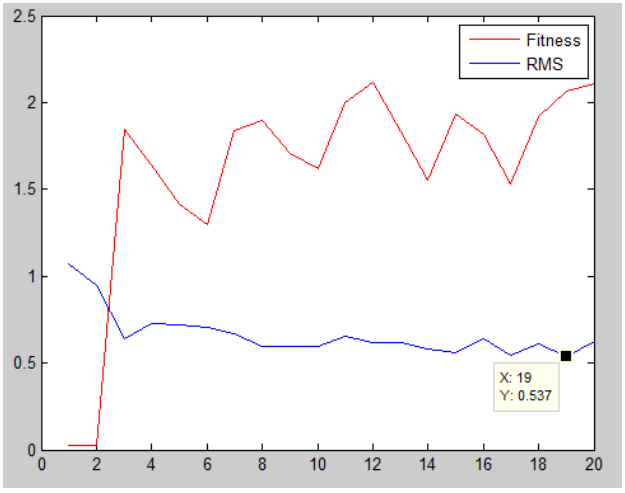


Figure 8: RMS trend based on the number of balls constituents metaball

We take advantage of the compactness of the metaball representation to calculate the metric of tumor malignancy, as detailed in the next section.

4. TUMOR MALIGNANCY METRIC

Tumor malignancy is clinically evaluated by two different growth indexes: tumor volume and the second is related to the tumor surface irregularities. However, this first index could be not appropriate for some tumors because it does not take into account that benign tumors are expansive, while malignant ones are infiltrative.

4.1 Tumor volume

One of the most important parameters to evaluate tumor malignancy [1, 5] according to its volume is the doubling time, i.e. the period of time required for a tumor to double in volume. For this reason it is critical to compute the tumor volume from CT data with precision.

It is possible to compute the volume of metaball in an analytical way [6]. The implicit surface is plunged in an upper dimension space. However, the mathematical complexity of the problem is very high. The computation depends on the number of metaball, and it is necessary to solve the system when this number changes. For this reason we prefer to evaluate the volume using voxels count and the result are comparable (less than 0,01%).

4.2 Surface Roughness

A second important indicator of malignancy is the tumor surface regularity or roughness (see Figure 9). We improved the previous metrics [5] taking advantage of the metaball approximation as more precise analytical support. We defined novel parameters borrowed from industrial standards in engineering surface analysis. All these approaches cannot be applied to 3D general surface meshes because they require the point height measured from the average surface plane, which is not univocally defined. Metaball approximation provides a consistent reference surface. Moreover the metaball implicit mathematic allows for fast computation of point to surface distance.

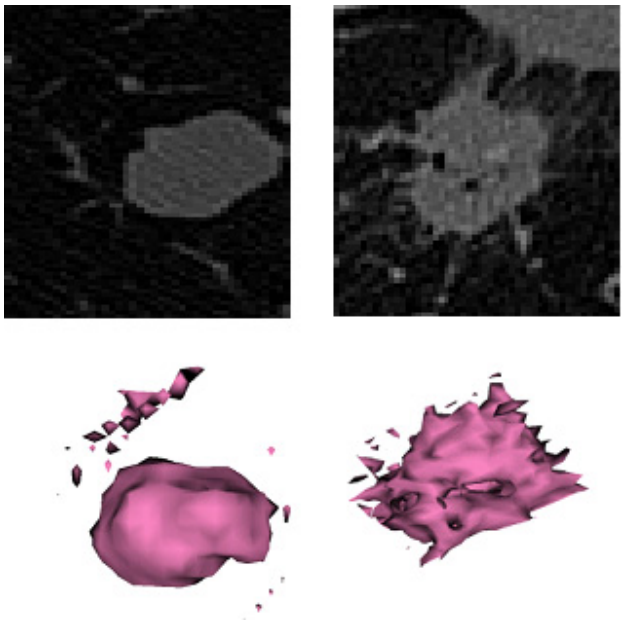


Figure 9: Tumor images and relative meshes with different regularity.

The roughness parameters are presented in Table 2: (i) maximum peak height, (ii) maximum valley depth, (iii) maximum height of the profile, (iv) root mean squared, (v) arithmetic average of absolute values, (vi) skewness, (vii) kurtosis. The y_i is the distance between the i^{th} tumor mesh point from closest point of the approximating metaball surface. Higher values mean higher probability of malignancy.

Table 2. Roughness Parameters

Parameter	Description	Formula
S_p	Maximum peak height	
S_v	Maximum valley depth	
S_t	Maximum height of the profile	$S_t=S_p-S_v$
S_q	Root mean squared	$S_q = \sqrt{\frac{1}{N}\sum_{i=1}^{N} y_i}$
S_a	Arithmetic average of absolute values	$S_a = \frac{1}{N}\sum_{i=1}^{N} \lvert y_i \rvert$
S_{sk}	Skewness	$S_{sk} = \frac{1}{N \cdot S_q{}^3}\sum_{i=1}^{N} {y_i}^3$
S_{ku}	Kurtosis	$S_{ku} = \frac{1}{N \cdot S_q{}^4}\sum_{i=1}^{N} {y_i}^4$

5. IMPLEMENTATION

The software implementation was written using C++ with the support of VTK (Visualization Toolkit) and Qt, two open source and cross-platform libraries containing helpful functions for 2D and 3D image visualization and GUI management respectively.

The software user interface has been designed for doctors and radiologists. The user can open DICOM files and browse the scanned volume using 3 section views (axial, sagittal and coronal). When s\he selects the region of interest, the system analyzes the selected area and extracts automatically the tumor surface with a 3D preview of the polygonal mesh (Figure 10). The user can interactively change the segmentation parameters until the desired result is obtained. Finally the metaball optimization is executed and the tumor malignancy parameters are computed and visualized.

To improve the perception of the three-dimensionality of the structure our rendering engine uses shading techniques by inserting virtual light sources inside the 3D scene. Moreover the system allows also geometries inspection in virtual reality using passive\active stereo visualization. We experienced that a better rendering helps the doctor to better understand the real shape of the target tumor.

6. RESULTS

In Table 3 we present the results of our algorithm applied to three significant cases of lungs carcinoma. The left column shows the DICOM section and the region of interest (ROI) inside the rectangular selection. The second column depicts the polygonal mesh after segmentation. The third shows how the metaball model well approximates the tumor shape. In particular the last case demonstrates how the segmentation algorithm can isolate a tumor region from the lung inner surface. The last two columns show the malignancy metric parameters.

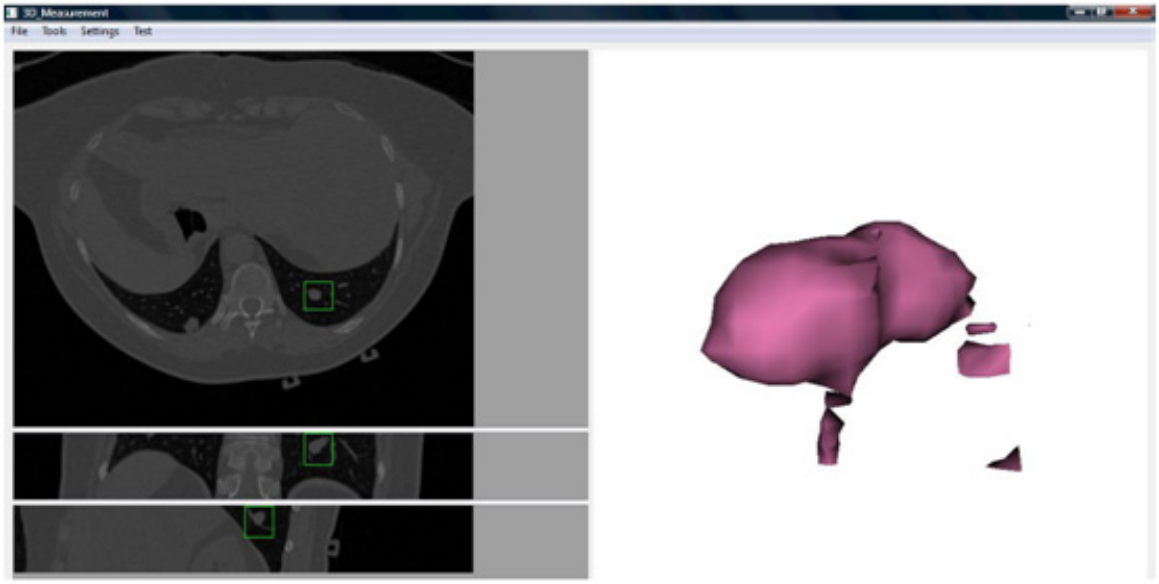
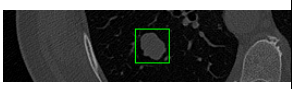
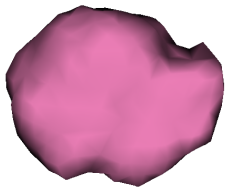
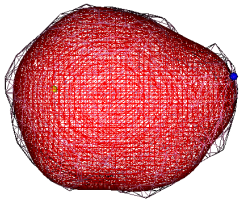
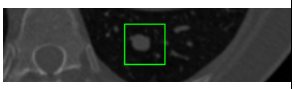
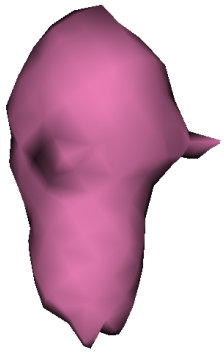
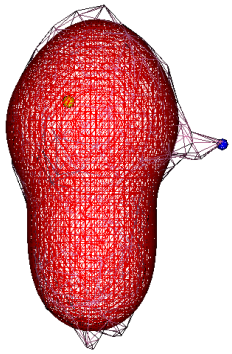
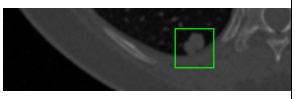
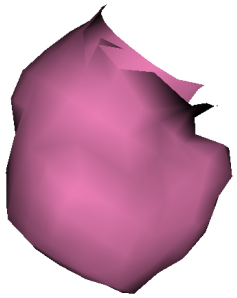
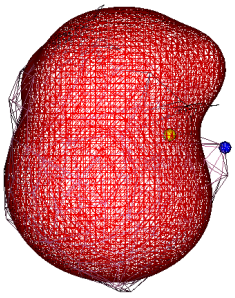


Figure 10: GUI screenshot: 2D sections and selection (left), 3D rendering (right)

Table 3. Roughness indexes on lung tumors

Case N.	DICOM	Polygonal Model	Metaball Model	Volume [mm^3]	Roughness [mm]
1				2072.1	S_p= 1.692 S_v= 1.574 S_t = 0.118 S_q= 0.539 S_a = 0.453 S_{sk}= 1.523 S_{ku}= 2.783
2				731.4	S_p = 2.573 S_v= 0.751 S_t = 1.821 S_q = 0.483 S_a = 0.328 S_{sk} = 2.997 S_{ku} = 12.564
3				1419.3	S_p = 1.241 S_v= 1.13 S_t = 0.112 S_q = 0.364 S_a = 0.293 S_{sk} = 1.691 S_{ku} = 3.492

7. CONCLUSION

This paper presents a novel approach to early diagnosis of lung tumors. The main contribution of this work the implementation of a semi-automatic method that merges 3D tumor extraction using a metaball-based representation and novel geometrical parameters for the evaluation of tumor malignancy.

The presented method executes the following steps: (i) 3D extraction of the tumor polygonal surface from axial CT data, (ii) surface segmentation to isolate tumor region from pleura, (iii) metaball approximation optimized by genetic algorithm, (iv) evaluation of tumor volume and (v) surface roughness. Compared to existing art, the metaball approximation provides a consistent reference surface for the evaluation of novel diagnosis parameters borrowed from engineering surface analysis. The novel roughness parameters are: (i) maximum peak height, (ii) maximum valley depth, (iii) maximum height of the profile, (iv) root mean squared, (v) arithmetic average of absolute values, (vi) skewness, (vii) kurtosis. The higher the indexes are, the more malignant the tumor is. In fact benign tumors and metastasis typically have a smoothed form, while primitive malignant tumor are irregular so very rough.

Starting from this considerations the calculated indexes can be used to distinguish benign from malignant tumors training a classifier. In future work we plan an extensive test with a larger database to validate the method and to find which of the extracted indexes are more effective in the diagnosis.

Furthermore another interesting application of this new malignancy indexes is the possibility to map the roughness into the TMN (Tumor-Metastasis-Node) staging system in order to measure the tumor malignancy in an accepted and standard way.

8. REFERENCES

[1] Bellomo, N. and Li, N. K. and Maini, P.K. 2008. On the foundations of cancer modelling: selected topics, speculations, & perspectives. In *Mathematical Models and Methods in Applied Sciences*, 18 (4). pp. 593-646.

[2] Arai, T., Kuroishi, T., Saito, Y., Kurita, Y., Naruke, T., Kaneko, M. 1994. Tumor Doubling Time and Prognosis in Lung Cancer Patients: Evaluation from Chest Films and Clinical Follow-up Study. In *Japanese Journal of Oncology Research*.

[3] Mehrara, E., Forssell-Aronsson, E., Ahlman, H., Bernhardt, P. 2007. Specific Growth Rate versus Doubling Time for Quantitative Characterization of Tumor Growth Rate. *Cancer Research*.

[4] Yankelevitz, D. F., Reeves, A. P., Kostis, W. J., Zhao, B., Henschke, C. I. 2000. Small Pulmonary Nodules: Volumetrically Determined Growth Rates Based on CT Evaluation. In *radiology*.

[5] Bevilacqua, V., Notarnicola, M., De Tommaso, E., Filograno, G., Mastronardi, G., Gargano, G., Bellotti, R. 2011. 3D measurements for tumors malignancies early diagnosis. In *Proc. IEEE International Workshop on Medical Measurements and Applications Proceedings (MeMeA)*.

[6] Robbins, Cotran, R., Kumar, V., Collins, T., 2000. Pathologic Basis of Disease. Piccin, 8th Edition.

[7] Fadzil, M., Prasaka, E., Fitriyah, H., Nugroho, H., Affandi, A., Hussein, S. 2010. Validation on 3D Surface Roughness Algorithm for Measuring Roughness of Psoriasis Lesion. *Biological and Biomedical Sciences*, 7(4): p. 205.

[8] Fadzil, M., Prasaka, E., Fitriyah, H., Nugroho, H., Affandi, A., Hussein, S. 2011. High Order Polynomial Surface Fitting for Measuring Roughness of Psoriasis Lesion. In *IVIC 2011*, Part I, LNCS 7066, pp. 341–351.

[9] MASMEC Sirio 2011 http://www.masmecbiomed.com.

[10] Lorensen, W., Cline, H., 1987. Marching Cubes: A High Resolution 3D Surface Construction Algorithm, *Proc. ACM SIGGRAPH 1987*, pp. 163-169.

[11] Lloyd., S. P., 1982. Least squares quantization in PCM, *IEEE Transactions on Information Theory* 28 (2): 129–137.

[12] Blinn, J. 1982. A Generalization of Algebraic Surface Drawing. *ACM Transaction on Graphics*, Vol. 1, No. 3, 235-256.

[13] Liu, S., Jin, X., Wang, C., Hui, K. 2006. Ellipsoidal-Blob Approximation of 3D Models and Its Applications.

[14] Faudot, D., Gesquière G., Garnier, L. 2004. An introduction to an analytical way to compute the volume of blobs. *International Journal of Pure and Applied Mathematics*, 11 (1), pp. 1-20, January 2004.

[15] Shen, J, Thalmann, D. 1995. Interactive Shape Design Using Metaball and Splines. In *Proceedings Implicit Surfaces 1995*, Grenoble, France.

[16] Plankers, R., Fua, P. 2003. Articulated Soft Objects for Multi-ViewShape and Motion Capture. In *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 25, Issue 9.

[16] Michalewicz, Z. 1996. Genetic Algorithms + Data Structures = Evolution Programs. *3rd edition, Springer-Verlag, Berlin Heidelberg New York*.