35th Annual International Conference of the IEEE EMBS
Osaka, Japan, 3 - 7 July, 2013

BENEFITS OF TEXTURE ANALYSIS OF DUAL ENERGY CT FOR COMPUTER–AIDED PULMONARY EMBOLISM DETECTION

Antonio Foncubierta–Rodríguez[⋆], *Óscar Alfonso Jiménez del Toro*[⋆], *Alexandra Platon*[†], *Pierre–Alexandre Poletti*[†] *Henning Müller*[⋆†] *and Adrien Depeursinge*[⋆†]

[⋆]University of Applied Sciences Western Switzerland (HES–SO)
[†]Department of Radiology, University and University Hospitals of Geneva (HUG)

ABSTRACT

Pulmonary embolism is an avoidable cause of death if treated immediately but delays in diagnosis and treatment lead to an increased risk. Computer–assisted image analysis of both unenhanced and contrast–enhanced computed tomography (CT) have proven useful for diagnosis of pulmonary embolism. Dual energy CT provides additional information over the standard single energy scan by generating four–dimensional (4D) data, in our case with 11 energy levels in 3D. In this paper a 4D texture analysis method capable of detecting pulmonary embolism in dual energy CT is presented. The method uses wavelet–based visual words together with an automatic geodesic–based region of interest detection algorithm to characterize the texture properties of each lung lobe. Results show an increase in performance with respect to the single energy CT analysis, as well as an accuracy gain compared to preliminary work on a small dataset.

Index Terms— Texture analysis, pulmonary embolism, dual energy CT, visual words, 4D image analysis

1. INTRODUCTION

Acute pulmonary embolism (PE) is a common condition, particularly in emergency medicine, that consists of the obstruction of one or more arteries in the lungs as a complication of deep vein thrombosis. Studies have shown that acute pulmonary embolism mortality rates can reach 75% during initial hospital admission [1] and after the hospital discharge 30% within 3 years [2]. Although it can be succesfully treated with anticoagulants, delays in diagnosis have shown to increase the risk of death [3]. There is evidence that 3D texture features correlate with ventilation and vascularization of the lung parenchyma [4] and that pulmonary embolism induces wedge–shaped pleura–based regions of heterogeneous increased attenuation in unenhanced computed tomography (CT) scans that are also visible on contrast–enhanced CT [5].

Dual energy computed tomography (DECT) produces 4D data in a single scan. Information is sampled at the three spatial coordinates and the level of x–ray energy between 40 and 140 keV derived from the two energy levels used for image acquisition (80 and 140 keV). One of the properties of DECT is that materials have different energy–attenuation curves, making it possible to distinguish materials with similar densities. Several studies showed the value of DECT to quantify perfusion defects of the lung parenchyma [6, 7] using iodine components, which can be derived from CT attenuation at two energy levels of 80 and 140 keV.

In this work a 4D texture analysis method for PE detection is presented. Automatically detected regions of interest (ROI) at multiple scales are used for describing each lung lobe, using wavelet–based features summarized with the bag of visual words approach. By using these regions of interest, local patterns are better described at multiple scales and results show performance improvement with respect to preliminary efforts on a smaller dataset without region detection [8]. Comparison to conventional single energy CT highlights the benefits of using a combination of energy levels for Computer–Aided Detection (CAD) of PE.

2. MATERIALS AND METHODS

This section presents the dataset of 4D DECT images and our computer–assisted detection method based on three main ideas: (1) an automatic multiscale region of interest detector, (2) a 3D wavelet transform for multi–scale texture descriptors for each scale and (3) visual words [9, 10] to obtain discriminative visual features based on patterns actually occurring in the data.

Dataset Pulmonary parenchyma ischemia in 4D dual energy CT (DECT) images of 25 patients were identified in collaboration with the emergency radiology of the University Hospitals of Geneva. The images in the dataset contain approximately 300 slices per patient and energy level. Energy levels are sampled from 40 to 140 keV in steps of 10 keV. The total amount of data per patient is approximately $512 \times 512 \times 300 \times 11 = 865.08$ million voxels. For each patient, the five pulmonary lobes were manually segmented

and the Qanadli index [11] was manually computed as a measure of the obstruction on a lobe basis. The Qanadli index is calculated by adding a score per artery in the lobe: 0 if there is no obstruction, 1 if there is partial obstruction and 2 if the artery is completely obstructed.

Automatic region of interest detection As mentionned, pulmonary embolism induces wedge–shaped heterogeneous attenuation regions in the lung lobes. Therefore, image analysis needs to be sufficiently local to accurately characterize these texture changes. On the other hand, an extremely local analysis might fail to capture meaningful texture patterns at larger scales. In addition to the *local versus global* constraint, the enormous amount of data contained in a single DECT scan suggests the use of data reduction techniques.

In 2D image processing, key–point detection has been widely used to produce a set of *salient* points where local analysis is performed, thus reducing the amount of areas to analyze [12]. However, in 3D there is currently no standard method for finding salient volumes. The superpixel [13] approach was extended to supervoxels [14] and provides a region–based descriptor of the image. However, supervoxels are an exhaustive partition of the complete image and therefore do not reduce the amount of voxels to be analyzed. To fullfill the requirements, a multiscale key–region detection was developed [15]. It is designed to reduce the number of points to be considered and to detect regions at different scales.

The multiscale region detection is based on the wavelet transform. At each wavelet scale, the difference of Gaussians (DoG) of the image is obtained (see Definition 1). Then, the regional maxima of the DoG are computed using the fill–hole geodesic algorithm [16]. Small unconnected regions are removed performing an opening operation on the regional extrema image. Regions computed for every scale yield a set of k connected components or regions of interest $R_1, R_2, \ldots, R_k$. Figure 1 shows an example of the regions detected.

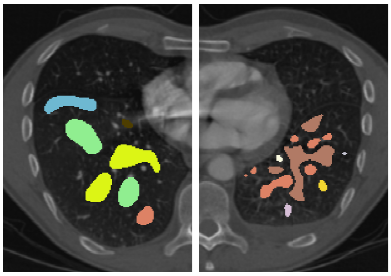

Fig. 1. Examples of regions detected in the lower lobes for scales 4 (left) and 3 (right).

Definition 1 *Let $I(\boldsymbol{x})$ be a n–dimensional image indexed by the coordinates $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$ and j a non–negative integer value. The difference of Gaussians wavelet at scale $s = 2^j$, $I^s_\psi(\boldsymbol{x})$ is defined as:*

$$\sigma_1 = s;\ \sigma_2 = 1.6\sigma_1 \tag{1}$$

$$\psi_s(\boldsymbol{x}) = \frac{e^{-\frac{|\boldsymbol{x}|^2}{2\sigma_1^2}}}{\sigma_1^n\sqrt{(2\pi)^n}} - \frac{e^{-\frac{|\boldsymbol{x}|^2}{2\sigma_2^2}}}{\sigma_2^n\sqrt{(2\pi)^n}}. \tag{2}$$

Then, the wavelet coefficients of an image $I(\boldsymbol{x})$ at scale s are obtained as:

$$I^s_\psi(\boldsymbol{x}) = I(\boldsymbol{x}) * \psi_s(\boldsymbol{x}). \tag{3}$$

Wavelet–based descriptors To describe each region of interest detected, the energies of the wavelet coefficients are used. For each region detected the energies of the wavelet coefficients within the region are computed as a descriptor of the local texture.

Definition 2 *Let $R_1, R_2, \ldots, R_k$ be k regions of interest. Then, an image $I(\boldsymbol{x})$ can be described by the set of k feature vectors f_i, $i = 1 \ldots k$ calculated as:*

$$E_i(I, s) = \sum_{\boldsymbol{x} \in R_i} I^s_\psi(\boldsymbol{x})^2. \tag{4}$$

$$f_i = (E_i(I, 2^0)), E_i(I, 2^1), \ldots, E_i(I, 2^{N-1})). \tag{5}$$

Bag of visual words Visual words [17, 9] have been widely used in image retrieval and image classification for describing image content. The approach is similar to the bag–of–words approach used for text retrieval or text similarity matching [18]. For each region, this technique maps a set of continuous low–level features (the region–level descriptors) into a compact discrete representation consisting of a visual word. Visual words are cluster centers in the low–level feature space. This guarantees to have a set of visual features actually corresponding to discriminative patterns that do occur in the database. Since each region is described by only one visual word, an image–level descriptor is defined as the histogram of visual words assigned to each of the regions in the image.

Definition 3 *Let $\mathcal{F} = \{f_1, f_2, \ldots, f_m\}$ be the set of m descriptors, $f_i \in \mathbb{R}^N$, describing visual characteristics of a given set of images. A visual vocabulary $W_{\mathcal{F},K} = \{w_1, w_2, \ldots, w_K\}$, with $w_c \in \mathbb{R}^N$, is constructed by grouping the elements of $\mathcal{F}$ into K disjoint subsets or words, and selecting their K centroids w_c with $c \in \{1, \ldots, K\}$.*

The bag–of–visual–words *of an image I, described by m_I visual descriptors $\{f_1, f_2, \ldots, f_{m_I}\}$, is defined as a vector $h_I = \{S_1, S_2, \ldots, S_K\}$:*

$$S_c = \sum_{i=1}^{m_I} g_c(f_i) \quad \forall c \in \{1, \ldots, K\}$$

where

$$g_c(f) = \begin{cases} 1 \text{ if } d(f, w_c) \leq d(f, w_d) & \forall d \in \{1, \ldots, K\} \\ 0 \text{ otherwise} \end{cases}$$

being $d(f, w)$ *the distance between two vectors* f *and* w.

Experimental configuration Given the different nature of the four coordinates in the data, i.e., three spatial coordinates and one energy coordinate, for each patient one 3D image was obtained per energy level. The region–level descriptors were computed independently for each energy level. Once the region–level descriptors were obtained for each of these 3D images, they were regrouped to construct the complete feature space that describes the images. Since regions were detected for each image, the number of regions of interest varied across patients and also across different energy levels from the same patient.

Regions produced at various energy levels were considered as additional instances for the same patient. One patient was described by the total number of regions independently of the energy level. The region–level descriptors were 4–dimensional ($N = 4$ in Equation 5). Eleven energy level–specific vocabularies were constructed and concatenated into eleven visual word histograms. This allowed enabling only some of the energy–levels to find the best combination of energy–levels.

Visual words were computed using a leave–one–patient–out (LOPO) cross–validation. The lungs contain five lung lobes (lower right, lower left, middle right, upper left, upper right) that have differing shapes and sizes and contain specific texture patterns independently of being healthy or not. Therefore, to perform an accurate evaluation of the system, the experimental configuration described in this section was repeated independently for each of the lobes, in order to learn the healthy/pulmonary embolism patterns for each of the lobes and not the patterns that distinguish one lobe from another. To achieve this, the region detection was limited to find regions in each of the lobes independently.

3. RESULTS

For the configuration described in Section 2, the classification accuracy with two classes: *healthy* and *pulmonary embolism* (for Qanadli index equal or above 0) was measured using a nearest–neighbor classifier in the histogram of visual words space.

In order to evaluate the accuracy of the proposed method and specifically if using Dual Energy CT improves the results, an exhaustive evaluation of all possible energy levels combinations was performed. Table 1 shows the best combinations of energy levels for each lobe compared to the correponding accuracy value for conventional single energy CT (70 KeV). In all cases the best accuracy was obtained when using 4D data (DECT).

Lobe	DECT	Words	Energy levels	SECT
LR	84 %	5	(50,130)	52 %
LL	84 %	5	(100,140)	48 %
MR	80 %	5	(40,50,130,140)	52 %
UL	76 %	25	(40,70,80,90)	60 %
UR	80 %	25	(90,120)	56 %

Table 1. Parameters for the best DECT accuracies obtained compared to single energy CT (SECT) for the same number of words for each lobe (LL= lower left, LR= lower right, MR= middle right, UL = upper left, UR = upper right).

4. DISCUSSION AND CONCLUSIONS

In this paper 4D texture analysis for pulmonary embolism detection is presented. Results show that when using an appropriate combination of energy levels from dual energy CT, detection accuracy enormously increases when compared to using single energy CT. The current work has only been validated on a small number of patients, which is a limitation of the work. On the other hand it is expected that results improve once a larger and more representative set of patients will exist as lung patterns can change significantly based on age, and environmental conditions, also for healthy tissue. We are currently working on increasing the size of the database to have a larger set of patients and textures to compare and model.

Future work will include the use of supervised clustering based on non–linear transformations of the feature space to obtain more stable and meaningful words and the use of richer texture features such as directional wavelets. We are also working with clinicians at the moment to enlarge the database, which will ensure better validation.

5. ACKNOWLEDGMENTS

This work was partially supported by the Swiss National Science Foundation (FNS) in the MANY project (205321–130046), the EU 7^{th} Framework Program in the context of the Khresmoi project (257528).

6. REFERENCES

[1] Samuel Z Goldhaber, Luigi Visani, and Marisa De Rosa, "Acute pulmonary embolism: clinical outcomes in the international cooperative pulmonary embolism registry (icoper)," *The Lancet*, vol. 353, no. 9162, pp. 1386 – 1389, 1999.

[2] Jr Anderson, Frederick A., H. Brownell Wheeler, Robert J. Goldberg, David W. Hosmer, Nilima A. Patwardhan, Borko Jovanovic, Ann Forcier, and James E. Dalen, "A population-based perspective of the hospital incidence and case-fatality rates of deep vein thrombosis and pulmonary embolism: The worcester dvt study,"

Archives of Internal Medicine, vol. 151, no. 5, pp. 933–938, 1991.

[3] Savas Ozsu, Funda Oztuna, Yilmaz Bulbul, Murat Topbas, Tevfik Ozlu, Polat Kosucu, and Asiye Ozsu, "The role of risk factors in delayed diagnosis of pulmonary embolism," *The American Journal of Emergency Medicine*, vol. 29, no. 1, pp. 26 – 32, 2011.

[4] Balaji Ganeshan, Kenneth A. Miles, Rupert C. D. Young, and Chris R. Chatwin, "Three–dimensional selective–scale texture analysis of computed tomography pulmonary angiograms," *Investigative Radiology*, vol. 43, no. 6, pp. 382–394, June 2008.

[5] Heidi C. Schwickert, Franz Schweden, Hans H. Schild, Rolf Piepenburg, Christoph Düber, Hans-Ulrich Kauczor, Christian Renner, Stein Iversen, and Manfred Thelen, "Pulmonary arteries and lung parenchyma in chronic pulmonary embolism: preoperative and postoperative CT findings.," *Radiology*, vol. 191, no. 2, pp. 351–357, 1994.

[6] Sven F. Thieme, Christoph R. Becker, Marcus Hacker, Konstantin Nikolaou, Maximilian F. Reiser, and Thorsten R. C. Johnson, "Dual energy CT for the assessment of lung perfusion—correlation to scintigraphy," *European Journal of Radiology*, vol. 68, no. 3, pp. 369–374, 2008.

[7] Sven F. Thieme, Thorsten R. C. Johnson, Christopher Lee, Justin McWilliams, Christoph R. Becker, Maximilian F. Reiser, and Konstantin Nikolaou, "Dual–energy CT for the assessment of contrast material distribution in the pulmonary parenchyma," *American Journal of Roentgenology*, vol. 193, no. 1, pp. 144–149, 2009.

[8] Antonio Foncubierta-Rodríguez, Pierre-Alexandre Poletti, Alexandra Platon, Alejandro Vargas, Henning Müller, and Adrien Depeursinge, "Retrieval of 4D dual energy CT for pulmonary embolism diagnosis," in *Medical Content–based Retrieval for Clinical Decision Support*, Oct. 2012, MCBR–CDS 2012.

[9] Antonio Foncubierta-Rodríguez, Adrien Depeursinge, and Henning Müller, "Using multiscale visual words for lung texture classification and retrieval," in *Medical Content–based Retrieval for Clinical Decision Support*, Hayit Greenspan, Henning Müller, and Tanveer Syeda Mahmood, Eds. Sept. 2012, vol. 7075 of *MCBR–CDS 2011*, pp. 69–79, Lecture Notes in Computer Sciences (LNCS).

[10] Uri Avni, Hayit Greenspan, Eli Konen, Michal Sharon, and Jacob Goldberger, "X–ray categorization and retrieval on the organ and pathology level, using patch–based visual words.," *IEEE Transactions on Medical Imaging*, vol. 30, no. 3, pp. 733–746, 2011.

[11] Salah D. Qanadli, Mostafa El Hajjam, Antoine Vieillard-Baron, Thierry Joseph, Benoit Mesurolle, Vincent L. Oliva, Olivier Barré, Frédéric Bruckert, Olivier Dubourg, and Pascal Lacombe, "New CT index to quantify arterial obstruction in pulmonary embolism," *American Journal of Roentgenology*, vol. 176, no. 6, pp. 1415–1420, 2001.

[12] David G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[13] Greg Mori, "Guiding model search using segmentation," in *Proceedings of the Tenth IEEE International Conference on Computer Vision - Volume 2*, Washington, DC, USA, 2005, ICCV '05, pp. 1417–1423, IEEE Computer Society.

[14] Andreas Burner, Rene Donner, Marius Mayerhoefer, Markus Holzer, Franz Kainberger, and Georg Langs, "Texture bags: Anomaly retrieval in medical images based on local 3D–texture similarity," in *Medical Content-based Retrieval for Clinical Decision Support*, Hayit Greenspan, Henning Müller, and Tanveer Syeda-Mahmood, Eds. September 2011, vol. 7075 of *MCBR-CDS 2011*, Lecture Notes in Computer Sciences (LNCS).

[15] Antonio Foncubierta-Rodríguez, Henning Müller, and Adrien Depeursinge, "Region–based volumetric medical image retrieval," in *SPIE Medical Imaging: Advanced PACS–based Imaging Informatics and Therapeutic Applications*, submitted.

[16] Pierre Soille, *Morphological Image Analysis: Principles and Applications*, chapter 6, Springer, second edition edition, 2003.

[17] Josef Sivic and Andrew Zisserman, "Video google: A text retrieval approach to object matching in videos," in *Proceedings of the Ninth IEEE International Conference on Computer Vision - Volume 2*, Washington, DC, USA, 2003, ICCV '03, pp. 1470–1477, IEEE Computer Society.

[18] C. J. van Rijsbergen, *Information Retrieval*, Prentice Hall, Englewood Cliffs, New Jersey, USA, 1979.