LUNG NODULE DETECTION USING EYE-TRACKING

*Michela Antonelli, Guang-Zhong Yang**

Dipartimento di Ingegneria dell'Informazione: Elettronica, Informatica, Telecomunicazioni
University of Pisa Via Diotisalvi 2, 56122 Pisa, Italy.
* Royal Society/Wolfson Foundation Medical Image Computing Laboratory,
Imperial College London, 180 Queens Gate, SW7 2Az London U.K.

ABSTRACT

This paper describes a decision support system for determining salient features for CT lung nodule detection using an eye-tracking based machine learning technique. The method first analyses the scan paths of expert radiologists during normal examination. The underlying features are then used to highlight salient regions that may be of diagnostic relevance by merging visual features learned from different experts with a weighted probability function. The framework has been evaluated using data from CT lung nodule examination and the results demonstrate the potential clinical value of the proposed technique, which can also be generalized to other diagnostic applications.

Index Terms— Eye tracking, feature selection, image processing, image region analysis, decision support system.

1. INTRODUCTION

Lung cancer represents the most common cause of death from malignancy in the world due to the late appearance of symptoms, which only become apparent when the cancer has developed to an advanced stage when further treatment is rarely effective. If malignant pulmonary nodules were detected earlier, the survival rate could be dramatically improved [1]. To increase the survival rate, patients at risk (usually smokers older than fifty) are recommended to undergo periodic screening programs with low-radiation-dose spiral computed tomography (LDCT) [2]. LDCT scans are highly sensitive for detecting nodules as small as 2 to 3mm, much smaller than that can be viewed on chest X-rays. The use of LDCT, however, generates a large amount of images [3]. For this reason, Computer Aided Diagnosis (CAD) based on pattern recognition and image enhancement techniques to distinguish nodules from normal anatomic structures has received increasing attention in recent years [4].

The pre-requisite of developing an effective CAD system is to understand the human diagnostic process. In radiology, eye-gaze tracking has been used to improve tumour detection [5] and provide information on nodule misclassifications [6]. Eye-tracking systems can also be used to study the common features which attract the radiologist's attention during the diagnosis process.

The purpose of this paper is to provide a framework which can help the radiologists during the diagnosis process by automatically highlighting areas of interest. Our method analyses the scan path, which consists of a sequence of saccades (*i.e*. fast and ballistic eye movements) from one fixation to another during normal radiological diagnosis [7]. This analysis allows the extraction of a set of salient features from a generic features library. They can then be used to identify hot spots (*i.e.* salient regions in the images, where most likely there could be a nodule) during routine LCDT examination to provide automatic decision support.

In this preliminary study, we applied the framework to four sets of 10 pulmonary nodule images extracted from three multi-slice LDCT exams (all with medical report, three of which containing one nodule) and analysed by five radiologists. All images were first segmented to extract only the lung parenchyma [8].

2. METHOD

The basic structure of the proposed framework is illustrated in Fig. 1, which consists of four modules. First, the eye-tracker records spatial-temporal information during the radiologist's diagnosis; the data is then used as the input to the second module which calculates the fixation points. In the following module, salient features based on eye fixation are extracted to determine hot spots. These tasks are performed by expert radiologists and the last module fuses salient regions identified in the previous phase.

2.1. Eye-movement tracking

For eye-tracking, a Tobii ETx50 eye tracker (Tobii, Sweden) is used. During the diagnosis process, DICOM images are shown on a monitor with a resolution of 1280x1024; radiologists are able to navigate through the 3D LDCT data by using the keyboard. The eye tracking system measures the relative position of the pupil and the corneal reflection in order to identify the gaze direction. The

accuracy of this system is 0.5°, the minimum dwell time for recording a fixation is 20ms and the system sample rate is 50 Hz. At the end of the tracking process, two files from the raw data are extracted: one with time-stamped x-y coordinates of the gaze points (CF) and the other with the time-stamped keyboard events (KF).

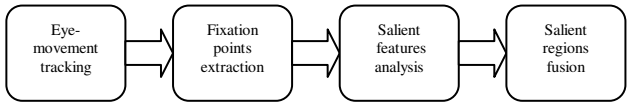

Fig. 1. A schematic illustration of the proposed processing framework.

2.2. Fixation points extraction

As mentioned earlier, a scan path is a sequence of saccades between points. A saccade is a rapid voluntary eye movement from one point to another and the purpose of which is to foveate a particular area of interest in a search scene. To calculate the fixations in x-y coordinates, we use the CF file. Due to the size of the fovea vision, there could be in a range of 23 pixels on the coordinates determined by the experimental setup. We then cluster all the points which are less then 23 pixels apart in one single fixation; the coordinates of this single fixation are equal to the centre of mass of all the points. The KF file is used to calculate the z-coordinate: each time the radiologist presses the key to move to and from the slice the z value of the current fixation is increased or decreased accordingly. At the end of this module, a sequence of fixations is collected and expressed by x-y-z coordinates and time duration.

2.3. Salient features analysis

Using fixation points, we extract a set of salient features from a features library consisting of 21 texture features. The computational framework is largely based on the visual saliency algorithm described in [11][12]. These are then used to identify important areas in the lung lobes where nodules could be found.

2.3.1. Feature Library

The feature library is divided in five main groups. The first-order statistics group consists of mean, standard deviation, absolute deviation, skewness, and kurtosis, which are used to analyse the distribution of the pixels' grey level.

In the second group, the spatial dependences of the CT value are taken into account by analysing second order statistics. The features derived from a set of 20 co-occurrence matrices of grey level; from the image we extract a circular region of 5 pixels radius centred on each fixation. This area is first down sampled to $p = 16$ grey level, the co-occurrence matrices $\{C_{d,l}\}_{1\le d\le 4, 1\le l\le 5}$ are then calculated using $d = 4$ directions and a distance l ranging from one to five pixels. From each co-occurrence matrix, a set of five scalar proprieties are computed [9]: energy, entropy, maximum, contrast and homogeneity. The mean values over all the 20 matrices represent the five features of the second group.

The third group analyses the texture coarseness: a large number of neighbouring pixels of same grey level represent a coarse texture; a small number indicates a fine texture. Run-length parameters are computed to represent the lengths of texture primitives (*i.e.* maximum contiguous set of constant grey level pixel on a line) [9]. Using the quantized circular regions calculated above we compute for each fixation four parameters: short primitive emphasis (spe), long primitive emphasis (lpe), grey-level uniformity (glu), and primitive length uniformity (plu) as shown below

$$\text{spe} = \frac{1}{K}\sum_{a=1}^{L}\sum_{r=1}^{N_r}\frac{B(a,r)}{r^2} \quad (1)$$

$$\text{lpe} = \frac{1}{K}\sum_{a=1}^{L}\sum_{r=1}^{N_r}B(a,r)r^2 \quad (2)$$

$$\text{glu} = \frac{1}{K}\sum_{a=1}^{L}[\sum_{r=1}^{N_r}B(a,r)]^2 \quad (3)$$

$$\text{plu} = \frac{1}{K}\sum_{r=1}^{N_r}[\sum_{a=1}^{L}B(a,r)]^2 \quad (4)$$

where $B(a,r)$ is the number of all the primitives of all directions having length r and grey level a, L the number of image grey levels, N_r the maximum primitive length in the image, and K the total number of runs.

The fourth group contains only the Fractal Dimension (FD) feature: if we consider the pixel intensity as the height above a plane, the intensity surface of the image can be viewed as a rugged surface, and FD represents the raggedness of this surface. There are many different definitions of fractal dimension; in our approach we use the fractional Brownian motion model [10].

The groups of features discussed so far use the spatial frequencies to describe textures, while the last group studies the edge frequencies; any edge detection operator can be employed for this purpose. We use the Robert's operator to compute the gradient for all pixels belonging to the selected circular region and afterwards the first and the second order statistics of edge elements distribution are calculated. In particular, the mean and the entropy of the edge strength measure the contrast, and the randomness of the distribution. The co-occurrence matrix, calculated on the edge direction, defines linearity, periodicity and size of the texture [9]

2.3.2. Salient feature extraction

This step extracts the features of every pixel covered by the scan path. The vector $\langle x_i, y_i, z_i, t_i\rangle$ $(1 \le i \le K)$ represents a scan path, x_i, y_i, z_i are the fixation coordinates and t_i the corresponding dwell time.

	Subject 1		Subject 2		Subject 3		Subject 4		Subject 5	
Images	1	2	1	2	1	2	1	2	1	2
Mean	0.38	0.11	0.28	0.76	0.77	0.36	0.43	1.00	0.07	0.10
Standard deviation	0.46	0.37	0.46	0.46	0.46	0.46	0.13	0.10	0.24	0.46
Absolute deviation	0.30	0.30	0.19	0.19	0.30	0.39	0.24	0.22	0.34	0.30
Skewness	0.44	0.18	0.19	0.45	1.74	0.30	0.25	0.16	0.81	0.29
Kurtosis	0.25	0.12	0.14	0.23	0.25	0.23	0.25	0.23	0.25	0.23
Entropy	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01	0.02	0.01
Co-energy	0.81	0.16	0.29	0.17	**1.78**	0.59	0.81	0.58	0.81	1.03
Co-entropy	0.73	0.26	0.32	0.16	1.49	0.29	**0.91**	0.61	1.04	1.29
Co-maximum	1.76	0.17	0.35	0.31	1.65	0.78	0.58	0.37	1.02	0.87
Co-contrast	0.25	0.21	0.25	0.21	0.25	0.00	0.25	0.20	0.25	0.21
Co-homogeneity	0.51	0.34	0.43	0.42	0.51	0.10	0.30	0.30	0.18	0.47
spe	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02	0.02
Lpe	1.44	0.01	0.05	0.12	0.66	0.14	0.08	0.00	0.06	0.42
Glu	0.13	0.13	0.10	0.03	0.13	0.13	0.13	0.13	0.13	0.13
Plu	0.95	0.04	0.10	0.04	1.09	0.26	0.41	0.83	0.95	1.16
Fractal dimension	**3.31**	**0.89**	**0.53**	**1.26**	1.64	**1.28**	0.48	**1.30**	**2.51**	**1.29**
Contrast	0.31	0.15	0.19	0.68	0.74	0.59	0.19	1.21	0.11	0.25
Randomness	0.89	0.28	0.14	0.37	0.35	0.37	0.35	0.37	0.00	0.37
Linearity	0.02	0.02	0.41	0.11	0.24	0.31	0.17	0.63	0.61	0.01
Periodicity	0.00	0.00	0.10	0.02	0.40	0.05	0.05	0.37	0.40	0.37
Size	0.05	0.01	0.05	0.08	0.05	0.08	0.05	0.08	0.05	0.08

TABLE I - KL divergences computed in two images by five radiologists, the underlined values are the largest in the column signifying visual saliency related to the feature extractor.

Each fixation f_k has a feature vector associated $f_k = (p_1, p_2, ..., p_N)$, where N is the number of features belonging to the features library.

Due to the basic characteristics of foveal vision, the value of the features on a given fixation has to take into account all pixels falling within the foveal field (*i.e.* a circle centred on the fixation with radius equal to 23 pixels). Because the visual acuity drops off dramatically from the centre of the focus, the feature value on a fixation is weighted with a Gaussian function centred on the fixation point where sigma is set to 15 pixels so that the value on the farthest point from the centre is 0.15. To extract the salient features, we analyse the fixation distribution in the feature space by taking into account the projection bias which is independent of visual search strategies and it is only affected by the features distribution on the background. When there is an abundance of certain feature value is quite possible that the fixation points will land on this value. Such projection bias can be eliminated by comparing the distribution of the features on the background to that of fixation points; a feature is salient when those two distributions differ. First of all, the features are quantized to sixteen levels. Subsequently, $p_i(t)$ is defined as the value of the *i-th* feature of the point *t*, and $T(p_i)$, $T^o(p_i)$ as the two distribution vectors with a number of components equal to the number of quantization levels. Each component is computed as

$$T_j(p_i) = \sum_{t=1}^{N_f} nf_j(t),\ T_j^o(p_i) = \sum_{t=1}^{N_p} np_j(t) \qquad (5)$$

where:

- N_f is the number of fixations belonging to the scan path,
- $nf_j(t)$ is the dwell time of the fixation *t* if $p_i(t) = j$, and 0 otherwise,
- N_p the number of image points,
- $np_j(t)$ is equal to 1 if $p_i(t) = j$, and 0 otherwise.

There are many methods one can use to measure the difference between distributions; in this paper the Kullback-Leibler (KL) divergence is used, which is defined as

$$KL(T, T^o) = \sum_{i=1}^{NF} \sum_{j=1}^{NL} T_j'(p_i) \log_2 \left(\frac{T_j'(p_i)}{T_j^{o'}(p_i)} \right) \qquad (6)$$

where NF is the number of the subset of features used, NL the number of quantization level, $T_j'(p_i)$ and $T_j^{o'}(p_i)$ the normalized value of the two distributions. $KL(T, T^o)$ represents the distance between the distributions, it is nonnegative and has a zero value, if and only if $T_j'(p_i) = T_j^{o'}(p_i)$.

After computing the distributions, a Genetic Algorithm [13] is used to find out the minimum subset of features, with the largest KL value, which can unambiguously identify the visual attention. A bit string is used as chromosome where the *i-th* bit corresponds to the *i-th* feature; the bit value is 1 if this feature is selected, otherwise it is 0. The fitness function of a chromosome is the KL divergence computed by only using the features with the corresponding bit set to 1. The algorithm uses a population of 40 organisms. Crossover, which is used as recombination operator, which takes copies of two selected organisms and swaps substrings of equal length between their chromosomes, creating two new organisms in the process. Mutation is applied with a small probability to randomly alter the value of single position in the chromosome string. The optimization process, which ends after 50 generations, produces a near optimal features set that will be used in the next phase of salient region construction.

2.4. Salient regions fusion

After identifying the features space used during the visual search, we select as hot spots the set of points which satisfy the equation $T_j(p_i) > T_j^o(p_i)$ for each feature *i* and level *j*. If

the scan path is collected from a group of subjects, we can use the map of the features derived from each scan path to predict salient regions for other images. To this end, we join the hot spots originated from different subjects. The probability of a pixel being hot is the ratio between the number of subjects for whom these pixels belong to a hot region, and the total number of subjects. By doing so, the skills difference of the subjects is ignored. To overcome this, we weight the above ratio so as to take into account both the subjects skills and the confidence given to the nodule detected during the visual search.

3. EXPERIMENTS AND RESULTS

The proposed framework was evaluated on four sets of images extracted from LDCT scans. Two consultants and three registrars were recorded during the assessments of each set. At the end of the session, the subjects pointed on a lung picture where the nodule was and also gave their confidence level. For each image, we calculate the value of $T(p_i)$, $T^o(p_i)$ and $KL(T,T^o)$, and the result are shown in TABLE I, the larger the $KL(T,T^o)$ value, the more discriminating is the feature. It can be seen that there is a clear preference on the fractal dimension feature, in which case will be the salient feature resulting from the Genetic Algorithm optimization process. The results above also indicate how the features preference is more evident when the subjects locate the nodule during the assessment, less noticeable in the first one/two images. This is due to the different strategies used by radiologists during their diagnosis: first they use a pre-attentive search using a parallel visual search to take a first glance of the image; next they use an attentive serial search to find out relationships among features.

After extracting the salient features from the images, hot spots were identified for each radiologist. An example is shown in Fig. 2. By using the weighted probability function, the saliency map is built; this map can be used to predict salient regions for other images in the same diagnostic contest.

4. CONCLUSIONS

In this paper, we have described a framework based on visual search scan-paths to identify both the set of salient features utilized and the strategies involved in the experts' assessment. The radiologist's attention (fixation points), together with information on the background, are used to analyse the images without being affected by the projection bias. The salient features are then used to build salient maps for analyzing other similar images. The method was evaluated on 4 sets of CT scans and the results have shown that there is an evident features preference especially when the nodule is hit. After the salient map is built using the 4 sets of CT, we have applied it to 5 different sets of CT scans, each containing a nodule. In this way, we have confirmed that all the 5 nodules belong to the area highlighted as salient. This framework can be used both to help the experts in their diagnosis by highlighting areas where there could be nodules and to train novices to look at the right features during the diagnosis process. The results show promising strength in identify intrinsic visual attention. Being a general framework, it can also be employed in other clinical applications for computer assisted diagnosis.

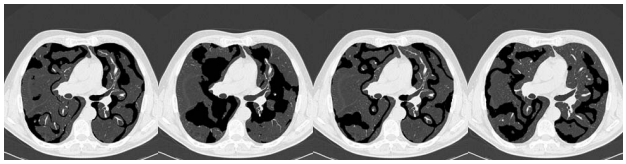

Fig 2. Salient regions of the four radiologists superimposed on the images

5. REFERENCES

[1] M.J.R. Dalrymple-Hay, N.E. Drury, "Screening for lung cancer", Journal of the Royal Society of Medicine, vol. 94, pp. 2-5, January 2001.

[2] C.I. Henschke, et.al, "Early Lung Cancer Action Project: overall design and findings from baseline screening", The Lancet, vol. 354, pp. 99-105, July 1999.

[3] S.J. Swensen, et. al., "Screening for lung cancer with low-dose spiral computed tomography", Am J Respir Crit Care Med, vol. 165, n. 4, pp. 508-513, February 2002.

[4] S.G. Armato III, "CAD dissects growing volume of data from lung CT exams", http://www.diagnosticimaging.com/.

[5] C.F. Nadine, H.L. Kundel, "Using eye movements to study visual search and to improve tumor detection", Radiographics, vol 7, n. 6, pp. 1241-1250, November 1987.

[6] H.L. Kundel, C.F. Nadine, D. Carmody, "Visual scanning, pattern recognition and decision-making in pulmonary nodule detection", Inv. Rad., vol. 13, n. 3, pp. 175-181, 1978.

[7] G.Z. Yang, L. Dempere-Marco, X. P. Hu, A. Rowe, "Visual search: psychophysical models and practical applications", Image and Vision Computing, vol. 20, pp. 291-305, 2002.

[8] M. Antonelli, B. Lazzerini, F. Marcelloni, "Segmentation and reconstruction of the lung volume in CT images", 20th Annual ACM Symposium on Applied Computing, vol. 1, pp. 255-259, 2005.

[9] M. Sharma, S. Singh, "Evaluation of texture methods for image analysis", Intelligent Information Systems Conference, vol. 1, pp. 117-123, 2001.

[10] C.C Chen, J.S. Daponte, M.D. Fox, "Fractal feature analysis and classification in medical imaging", IEEE Trans. Med. Imag., vol. 8, n. 2, pp. 133-142, June 1989.

[11] X.P. Hu, L. Dampere-Marco, G.Z. Yang, "Hot spot detection based on feature space representation of visual search", IEEE Trans. Med. Imag., vol. 22, n. 9, September 2003.

[12] L. Dempere-Marco, X-P. HU, S.M. Ellis, D.M. Hansell, G-Z. Yang, "Analysis of visual search patterns with EMD metric in normalized anatomical space", IEEE Trans. Med. Imag, vol. 25, n. 8, pp. 1011-1021, 2006.

[13] D. Beasley, D.R. Bull, R.R. Martin "An Overview of Genetic Algorithms: Part1, Fundamentals" University Computing, vol. 15, no. 2, pp. 58-69, 1993.