SEMANTIC-ORIENTED SYNTACTIC ALGORITHMS FOR CONTENT RECOGNITION AND UNDERSTANDING OF IMAGES IN MEDICAL DATABASES

Marek R. Ogiela and Ryszard Tadeusiewicz

University of Mining and Metallurgy, Institute of Automatics
Al. Mickiewicza 30., PL-30-059 Krakow, POLAND
mogiela@agh.edu.pl, rtad@biocyb.ia.agh.edu.pl

ABSTRACT

An important feature to be considered in the design of medical databases is content-based retrieval of images showing particular case of illness. Our work has focused on semantic and feature-based retrieval, which include existence of morphological signs characteristic for some particular lesions. An analysis of results of the experimental methods proposed indicates that the syntactic methods allow creating an efficient index structure for semantic-based retrieval of images from multimedia specialist medical databases.

1. INTRODUCTION

With the advances in medical and multimedia databases and the popularization of the telemedicine, it is possible to access large image and diagnosis repositories distributed throughout the world. One of the most important problems in such an access is how the information in the respective medical databases can be summarized to enable an intelligent selection of interested case of illness based on visual queries. This paper presents an approach to solve this problem based on image semantic indexing of specialist database.

In the paper was proposed the general methodology of structural pattern recognition methods used for creation of semantic description of pathological lesions visible on selected medical images. Such description may be than used as indexing key for medical images in multimedia databases. Proposed methods allow for context and semantic oriented searching a special kind of images with visible particular pathological changes characteristic for illness process in selected organs.

Proposed approach is consisting on finding semantic description using special kind of grammars, for entering to the database images. Depending on various kinds of morphological shapes of described lesions, for created description are used grammars with appropriate generating power i.e. context-free grammars or graph grammars. Make in such way description is further treated as metrics and short diagnosis, and is written in database in the form of index key. Searching of images in the moment of coming queries, also in form of image or morphological description, consist on making for input image semantic description and finding in database the description, which characterized similar pathology.

We proposed grammars defined diagnostic description of selected disease symptoms visible on x-ray images showing morphological changes in the form of local stenoses of the lumen of coronary arteries, and pathological signs in upper parts of ureter ducts and renal calyxes. In the case of the analysis of coronary artery images, the main objective is a semantic description of the different forms of cardiac ischaemic diseases. In the case of kidney radiograms, the main goal is to describe of local irregularities in ureter lumens, and an examination of the morphology of renal pelvis and calyxes.

The main advantage of using syntactic approach for indexing is effectiveness of these methods. Searching in the database consisting in comparisons and evaluation of the conformability degree of the descriptions generated by syntactic parser is faster (due the small capacity of description strings) than simple comparison of images. Such searching is moreover independent from many deformations and transformations of images.

An example of a coronary artery with stenosis and image with stenosed ureter are presented in Fig. 1.

For a proper description of the above mentioned changes, and for a verification of how advanced their level, an attributed context-free grammar of type LR (1) and a graph grammar of type EDT [1] have been proposed.

Before coming to the creation of description of the changes in question, it is necessary to preserve the sequence of operations, which are included in the image preprocessing. The goal of this analysis is obtaining width graphs, which show the pathological changes occurring in these structures [2].

2001 IEEE International Conference on Multimedia and Expo
ISBN 0-7695-1198-8/01 $17.00 © 2001 IEEE

2. SYNTACTIC DESCRIPTION OF IMAGES WITH CORONARY ARTERIES STENOSES

The detection and description of changes occurring in the form of different kinds of lumen stenoses in coronary arteries was carried out on width graphs obtained during the preprocessing of coronary artery angiograms. These graphs show the straightened profiles of investigated coronary vessels.

For a proper description of the original components in the images, which fully permit us to create a linguistic description of changes, the obtained profile graphs of coronary arteries are submitted to a linear approximation method [2]. As a result of this operation, every investigated graph obtains a representation in the form of an approximating segments sequence, which is next inputted to the succeeding terminal symbols which create the new linguistic representation. This representation sets up the input information to the syntax analyzer, which is based on the grammar prepared by the authors. Such an analyzer creates a proper program for recognizing changes in the coronary artery lumen.

For recognition of the different kind of stenosis shapes, the following attributed grammar is suggested:

$$G_{CA} = (V_N, V_T, SP, STS)$$

where

V_N = {SIGN, STENOSIS, H, U, D} – is a set of non-terminal symbols

$V_T = \{u, h, d, \lambda\}$ – is a set of terminal symbols (λ is an empty symbol)

for $u \in (11°, 90°)$, $h \in (-10°, 10°)$, $d \in (-11°, -90°)$

STS = SIGN - starting symbol of the grammar

SP – is a set of production defined as follows:

SP:

1. SIGN $\rightarrow$ STENOSIS
2. STENOSIS $\rightarrow$ D H U
3. D $\rightarrow$ d | d D $w_{sym}:=w_{sym}+w_d$; $h_{sym}:=h_{sym}+h_d$
4. U $\rightarrow$ u | u U $w_{sym}:= w_{sym}+w_u$; $h_{sym}:=h_{sym}+h_u$
5. H $\rightarrow \lambda$ | h H $w_{sym}:=w_{sym}+w_h$; $h_{sym}:=h_{sym}+h_h$

In the presented grammar, the second of the suggested productions defines the potential shapes of the lumen of coronary vessels stenoses. The succeeding steps in this grammar the resulting specific linguistic formula, characterize the descending and the ascending part of the analyzed stenosis, and the last of the productions defines the horizontal segment, which can appear between both parts of the vessel. Semantic variables h_e and w_e determine the height and length of the terminal segment with label e. These attributes are used to determine the numerical parameters of the detected stenosis, which allows us to characterize the degree of lumen stenosis of the coronary artery as a percentage, which has an essential meaning for the diagnosis of the state of a patient health.

3. SYNTACTIC DESCRIPTION OF URETER STENOSES

Description of morphological changes in the form of stenoses and dilatations of ureters is undertaken using the following attributed grammar:

$$G_u = (V_N, V_T, SP, STS)$$

where

V_N = {SIGN, STENOSIS, DILATATION, H, A, D}- non-terminal symbol set

V_T = {h, a, d} – terminal symbol set

for $h \in (-8°, 8°)$, $a \in (9°, 180°)$, $d \in (-9°, -180°)$

STS = SIGN – starting symbol

SP –set of production:

1. SIGN $\rightarrow$ DILATATION Sign=Dilatation
2. SIGN $\rightarrow$ STENOSIS Sign=Stenosis
3. DILATATION $\rightarrow$ A H D | A D | A H
4. STENOSIS $\rightarrow$ D H A | D A | D H
5. A $\rightarrow$ a | a A $w_{sym}:=w_{sym}+w_a$; $h_{sym}:=h_{sym}+h_a$
6. H $\rightarrow$ h | h H $w_{sym}:=w_{sym}+w_h$; $h_{sym}:=h_{sym}+h_h$
7. D $\rightarrow$ d | d D $w_{sym}:=w_{sym}+w_d$; $h_{sym}:=h_{sym}+h_d$

This grammar permits us to detect different forms of stenosis and dilatation, which characterize the different disease units. Using attributes as before permits us to calculate the numerical parameters of detected morphological change.

4. GRAPH GRAMMAR BASED DESCRIPTION OF RENAL PELVIS SHAPE

As graph, methods of syntactical pattern recognition are generally designed for the description of slightly more complex objects in the image, in the case of nephrogram analysis these methods are used to check regularities of renal pelvis and calyx morphology. Although these structures are characterized by unusual shape variation, it is possible to distinguish certain common features which characterize all proper structures, using the number of smaller and bigger calyxes revealed in the renal sinus [1]. To describe shapes of these structures we have suggested an expansive graph grammar defined with the aim of analyzing the skeleton morphology of investigated renal pelvises and renal calyxes. A graph grammar describing the skeletons of renal pelvises and renal calyxes is defined

2001 IEEE International Conference on Multimedia and Expo
ISBN 0-7695-1198-8/01 $17.00 © 2001 IEEE
0-7695-1198-8/01/$10.00 (C) 2001 IEEE

in such a way that root of the graph is defined by the location bifurcation in the bigger calyxes. Next, its consequents are determined by branch points of the second degree, which is the beginning of the smaller calyxes. The last layer of vertices is defined by branch points of the third degree, that is branches which appear in a case when the renal papillae has a concave shape.

To diagnose morphological changes in renal pelvises the following graph grammar was used:

$$G_{EDT}=(\Sigma, \Gamma, r, P, Z)$$

where

$\Sigma = \Sigma_N \cup \Sigma_T$ is a set of terminal and non-terminal vertex labels defined in the following way:

Σ_T = {pelvis_renalis, calix_major, calix_minor, papilla_renalis}

Σ_N = {PELVIS_RENALIS, CALIX_MAJOR,CALIX_MINOR}

r - is a function which assigns to the graph vertex the number of its consequents

$\Gamma = \{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}\}$ - is a set of edge labels
for $\boldsymbol{y} \in (-30°, 30°)$, $\boldsymbol{x} \in (30°, 180°)$, $\boldsymbol{z} \in (-30°, -180°)$

Z={PELVIS_RENALIS} - is a finite set of starting graphs

P- is a set of production:

1. PELVIS_RENALIS → pelvis_renalis(***x***CALIX_MAJOR ***y***CALIX_MAJOR ***z***CALIX_MAJOR)
2. PELVIS_RENALIS → pelvis_renalis(***x***CALIX_MAJOR ***y***CALIX_MAJOR)
3. PELVIS_RENALIS → pelvis_renalis(***x***CALIX_MAJOR ***z***CALIX_MAJOR)
4. PELVIS_RENALIS → pelvis_renalis(***y***CALIX_MAJOR ***z***CALIX_MAJOR)
5. CALIX_MAJOR → calix_major(***x***CALIX_MINOR ***y***CALIX_MINOR ***z***CALIX_MINOR)
6. CALIX_MAJOR → calix_major(***x***CALIX_MINOR ***y***CALIX_MINOR)
7. CALIX_MAJOR → calix_major(***x***CALIX_MINOR ***z***CALIX_MINOR)
8. CALIX_MAJOR → calix_major(***y***CALIX_MINOR ***z***CALIX_MINOR)
9. CALIX_MINOR → calix_minor(***x***papilla_renalis)
10. CALIX_MINOR → calix_minor(***y***papilla_renalis)
11. CALIX_MINOR → calix_minor(***z***papilla_renalis)
12. CALIX_MINOR → calix_minor(***x***papilla_renalis ***y***papilla_renalis)
13. CALIX_MINOR → calix_minor(***x***papilla_renalis ***z***papilla_renalis)
14. CALIX_MINOR → calix_minor(***y***papilla_renalis ***z***papilla_renalis)

The first group of production allow to describe different kinds of shapes of renal pelvis i.e. having two or three smaller calyxes. The succeeding productions define the form of bigger calyxes formed from two or more smaller calyxes. The last group defines the proper form of renal papillae, which obtains a fork form during the skeletonization, which means it finishes with short branches, which arise only when it is concave to the interior of a smaller calyx. Convex forms during skeletonization are thinned to the line without end branches, which results from the properties of skeletonization algorithms.

5. RESULTS

Thanks to the application of the presented context-free grammars, it is possible quite precisely to describe different kinds of irregular shapes in the form of stenoses, or a dilatation of investigated structures.

The detection of irregularities appearing in the set of test data is carried out by syntax analyzers, generated on the basis of the formal description of grammars. The analyzing program formed in this way was tested during a series of tests whose aim was to recognise stenoses on a few dozen coronography images showing major stenoses both of the right and the left coronary arteries, while for different projections of registered images there are several representatively selected images of coronary arteries. Analysis of the morphological changes in upper urinary tracts was carried out on the basis of a set of a dozen or so urograms and kidney roentgenograms containing images which show both congenital changes in renal pelvises and ureter morphology, and the acquired changes which show the existence of a disease process. The efficiency of recognition algorithm, at the same time of creating a semantic representation for symptoms described with context-free grammars is very high and amounts to about 95%. This is the result of a small number of pathological lesions (which could take on even very complicated shapes) described in this way. At the same time this gives a possibility to direct closely the recognizing system and dedicating the grammar constructed for the needs of analysis of concrete disease symptoms.

The efficiency of analysis of shapes of renal calyx with the use of graph grammar is at a lower level. So far it was conducted on a test-group composed of a dozen most representative disease cases.

For the presented graph grammar, an automaton with output above the set of vertex labels was constructed to serve the syntactical analysis of the graph languages defined by these types of grammars. Because of analysis of these images, in each case we have obtained the kind of recognized symptom and a sequence of production numbers, which lead to grammar derivation of shape description of such lesions. Such sequences create the proper description of analyzed shapes and have been storing in indexing record.

2001 IEEE International Conference on Multimedia and Expo
ISBN 0-7695-1198-8/01 $17.00 © 2001 IEEE
0-7695-1198-8/01/$10.00 (C) 2001 IEEE

In Figure 1 we present an examples which show description of the changes in question for coronary artery (Fig. 1A) and urinary tract (Fig 1B) images. The recognition process is based on production's numbers generated as output string by syntax analyzers. Recognized symptoms are marked by the bold line.

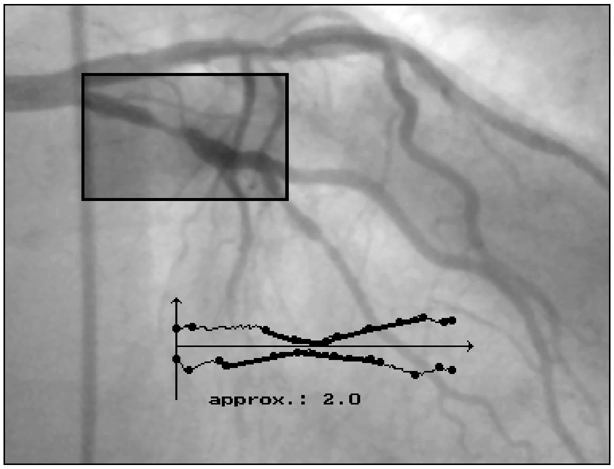

Fig. 1 A. An image of coronary artery with visible stenosis. Result of disease symptom description using syntactical methods of pattern recognition. Width diagram shows the place in which the looked-for lesion was diagnosed. Its structural description with the use of terminal symbols and the proper production sequences constitutes its unambiguous description; it is the indexing key to searching and archivisation of a given diagnosed case.

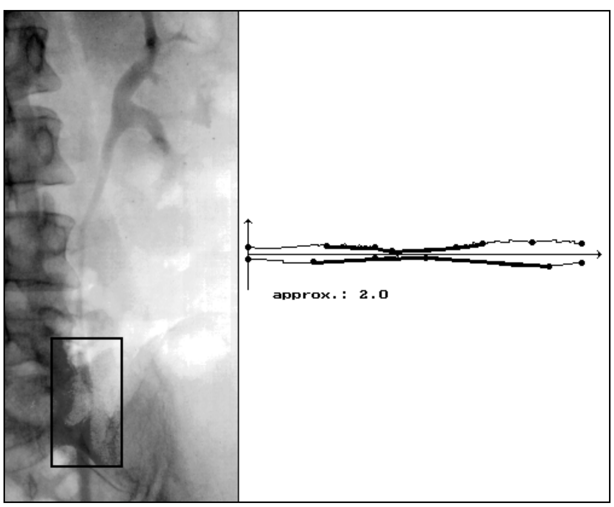

Fig. 1 B. An image of left renal pelvis with stenosed ureter. Result of disease symptom descriptions using syntactical methods of pattern recognition. Width diagram shows the place in which the looked-for lesion was diagnosed. Its structural description created to recognize it constitutes the key used for indexation and categorization of database for individual disease symptoms.

6. CONCLUSIONS

Despite lack of efficient searching technique in databases of images that resemble the query image, we proposed methods of defining semantic description for a few pathological lesions. These algorithms are based on syntactic methods of pattern recognition that are very useful in the task of indexing images in medical databases.

In all cases, proposed grammars are generated univocal and representative diagnostic description, adequate to the semantic contents of images. This indicates that descriptions generated by different contents are also different in some way, but descriptions of images with similar meanings are congruent.

Proposed schemes of indexing and creation of semantic description are moreover independent from geometrical transformation or noise disturbances of archiving images.

ACKNOWLEDGEMENTS

Supported by the University of Mining and Metallurgy under Grant No. 11.11.120.249

REFERENCES

[1] M. R. Ogiela and R. Tadeusiewicz, "Syntactic methods of shape feature description and its application in analysis of medical images", in Visual Data Exploration and Analysis VII, Robert F. Erbacher, Philip C. Chen, Jonathan C. Roberts, Craig M. Wittenbrink, Editors, Proceedings of SPIE vol. 3960, pp 326-333, 2000.

[2] M. R. Ogiela and R. Tadeusiewicz, "Syntactic analysis and languages of shape feature description in computer aided diagnosis and recognition of cancerous and inflammatory lesions of organs in selected X-ray images", *Journal of Digital Imaging*, vol. 12, no 2, Suppl 1, pp. 24-27, 1999.

[3] Y. Niu, Ozsu Mtamer, and Li Xiaobo, "2-D-S tree, an index structure for content-based retrieval of images", The 1999 Multimedia Computing and Networking, Proceedings of SPIE, vol. 3654, pp. 110-121, 1999.

[4] W. Chang, G. Sheikholeslami, J. Wang, and A. Zhang, "Data resource selection in distributed visual information systems", *IEEE Transactions on Knowledge and Data Engineering*, vol. 10, no. 6, pp. 926-946, 1998.

[5] M. R. Ogiela and R. Tadeusiewicz, "New aspects of using the structural graph-grammar based techniques for recognition of selected medical images", *Journal of Digital Imaging*, vol. 14, no 2, suppl 1, pp. 231-232, 2001.

2001 IEEE International Conference on Multimedia and Expo
ISBN 0-7695-1198-8/01 $17.00 © 2001 IEEE
0-7695-1198-8/01/$10.00 (C) 2001 IEEE