ELECTRONICS AND ELECTRICAL ENGINEERING
ISSN 1392 – 1215 *2010. No. 2(98)*
ELEKTRONIKA IR ELEKTROTECHNIKA

MEDICINE TECHNOLOGY
T 115
MEDICINOS TECHNOLOGIJA

Establishing Medical Diagnosis using Pattern Semantic Rules

S. C. Udristoiu, A. L. Ion
Department of Software Engineering, Faculty of Automation, Computers and Electronics, University of Craiova, Bvd. Decebal, Nr. 107, 200440, Craiova, Romania, e-mail: anca_soimu@yahoo.com

Introduction

An impressive amount of medical images is daily generated in hospitals and medical centers. Consequently, the physicians have an increasing number of images to analyze manually. Computational techniques can be provided to assist the physician's work, as CAD (computer assisted diagnosis) systems, which support physicians in analyzing digital images and to find out possible diseases [2, 3]. In the medical domain, content-based image retrieval applications could offer new opportunities. At a first look, it seems easily to annotate or interpret, for example, a landscape picture, but the semantic concepts in the medical domain are much better defined for any specialty areas in medicine [1].

In this paper, we propose methods that employ pattern semantic rules to support CAD systems. The rules reflect how humans learn and memorize new knowledge.

Also in [5, 6], relevant methods using association rules for image analysis are developed. In this paper, we will show that pattern rules can be successfully applied to support medical image diagnosis.

The remainder of this paper is structured as follows. Section 2 presents the selection of visual features; section 3 presents the mapping between visual features and diagnosis; section 4 presents the generation of semantic rules and details the process of medical diagnosis based on pattern semantic rules. Finally, section 5 discusses the experiments and summarizes the conclusions of this study.

The selection of visual features

The selection of the visual feature set and the image segmentation algorithm are the definitive stage for establishing the diagnosis of medical colour images [4]. The diagnosis of medical images is directly related to the visual features (colour, texture, shape, position, dimension, etc.) because these attributes capture the information about the semantic meaning. A set of dominant colours obtained from each image by segmentation after colour characteristic. The ability and efficiency of the colour feature for characterizing the colour perceptual similitude is strongly influenced by the colour space and quantization scheme selection. The HSV colour space quantized to 166 colours is used to represent the colour information. Before segmentation, the images are transformed from RGB to HSV colour space and quantized to 166 colours. The colour regions extraction is realized by the colour set back projection algorithm [7]. The results of the segmentation algorithm applied on two image diagnosed with esophagitis, respectively with gastric cancer, can be visualized in Fig.1 and Fig. 2.

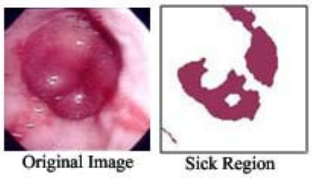

Fig. 1. Segmentation results from an image diagnosed with esophagitis

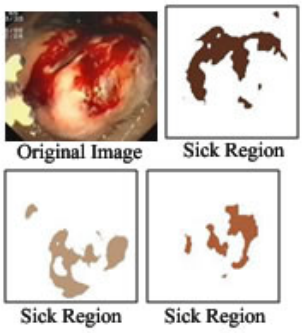

Fig. 2. Segmentation results from an image diagnosed with gastric cancer

In conformity with the defined characteristics, a region is described by:

- The colour characteristics are represented in the HSV colour space quantized at 166 colours. A region is represented by a colour index which is, in fact an integer number between 0...165.

- The spatial coherency represents the region descriptor, which measures the spatial compactness of the pixels of same colour.
- A seven-dimension vector (maximum probability, energy, entropy, contrast, cluster shade, cluster prominence, correlation) represents the texture characteristics.
- The region dimension descriptor represents the number of pixels from region.
- The spatial information is represented by the centroid coordinates of the region and by minimum bounded rectangle.
- A two-dimensional vector (eccentricity and compactness) represents the shape feature.

Mapping image visual features to semantic indicators

The developed vocabulary is based on the concept of semantic indicators, and the syntax captures the basic models about patterns and diagnosis. The proposed representation language is simple, because the syntax and vocabulary are elementary. The language words are limited to the name of semantic indicators. Being visual elements, the semantic indicators are, by example, the colour (colour-light-red), spatial coherency (spatial coherency-weak, spatial coherency-medium, spatial coherency-strong), texture (energy-small, energy-medium, energy-big, etc.), dimension (dimension-small, dimension-medium, dimension-big, etc.), position (vertical-upper, vertical-center, vertical-bottom, horizontal-upper, etc.), shape (eccentricity- small, compactness-small, etc.).

The syntax is represented by the model, which describes the images in terms of semantic indicators values. The values of each semantic descriptor are mapped to a value domain, which corresponds to the mathematical descriptor.

The values domains of visual characteristics were manually experimented on images of WxH dimension.

A value of colour semantic indicator is associated to each region colour in the HSV colour space quantized at 166 colours. The colour correspondence between the mathematical and semantic indicator values is determined basis on the experiments effectuated on a training image database. The colour correspondence is illustrated by the following examples: light-red (108), medium-red (122), dark-red (139), light-yellow (109), medium-yellow (125), and dark-yellow (141).

Similarly, a hierarchy of values, which are mapped to semantic indicator values, is also determined for the other visual characteristics.

At the end of the mapping process, a figure is represented in Prolog by means of the terms *figure(ListofRegions)*, where *ListofRegions* is a list of image regions.

The term *region(ListofDescriptors)* is used for region representation, where the argument is a list of terms used to specify the semantic indicators. The term used to specify the semantic indicators is of the form:

descriptor(DescriptorName, DescriptorValue).

The model representation of the image from Fig. 1 can be observed in the bellow example:

```
figure([
region([descriptor(colour,dark-red),
        descriptor(dimension,big),
        descriptor(horizontal-position, center),
        descriptor(vertical-position,center),
        descriptor(shape-eccentricity, medium),
        descriptor(texture-probability, medium),
        descriptor(texture-inversedifference, medium),
        descriptor(texture-entropy,small),
        descriptor(texture- energy,big),
        descriptor(texture-contrast,big),
        descriptor(texture-correlation, small)]),
region([descriptor(colour,light-red), …
```

Determining image medical diagnosis using pattern semantic rules

Being given an image database *DB*, let be *U= {S1,…,Sn}* a subset of *DB* that contains *n* image-examples, labelled by a diagnosis.

In the learning phase, the scope is to automatically generate semantic rules *R* based on diagnosed image-examples set *U*. A rule determines the set of semantic indicators, which identify a diagnosis.

In the testing/annotation phase, for each image of the subset *DB-U* (namely the images from *DB*, but that are not in *U*), the generated semantic rules are used to diagnose them. Since the images and rules are represented in Prolog, a Prolog interpreter is used for rules inference to recognise the diagnosis.

A pattern semantic rule is of form:

"semantic indicators -> diagnosis"

The pattern semantic rules have the body composed by conjunctions of semantic indicators, while the head is the diagnosis.

The stages of the learning process are:

- relevant images for diagnosis are used for learning it.
- each image is automatically processed and segmented and the primitive visual features are computed, as it is described in section 2.
- for each image, the primitive visual features are mapped to semantic indicators, as it is described in section 3.
- the rule generation algorithms are applied to produce rules, which will identify each diagnosis from the database.

In our system, the learning of semantic rules is continuously made, because when a categorized image is added in the learning database, the system continues the process of rules generation.

The scope of image pattern rules is to find semantic relationships between image objects and diagnosis. The proposed method uses the Apriori algorithm for discovering the pattern semantic rules between primitive characteristics extracted from images and diagnosis, which images belong to [6].

The algorithm is based on „region patterns" and necessitates some computations, being necessary a pre-processing phase for determining the visual similitude between the image regions from the same diagnosis.

In the pre-processing phase, the region patterns,

which appear in the images, are determined. So, each image region Reg_{ij} is compared with other image regions from the same diagnosis. If the region Reg_{ij} matches other region Reg_{km}, having similar the features from the positions n_1, n_2, …, n_c, then the generated region pattern is *SRj (-, -, -, n_1, n_2,…,n_c,-,-)*, and the other features are ignored.

A database with five images, relevant for a diagnosis, is considered. An example of images representations using region patterns is presented in the Table 1:

Table 1. Relevant images for a certain diagnosis

ImgID	Image Regions
1	R(a,b,c,d), R(a',b',c',d')
2	R(a,b,c,f), R(a',b', c',d')
3	R(a,b,c,f''),R(a',b',c',i)
4	R(a,b,c,f'),R(a',b',c',i')
5	R(a,b,e,d), R(m,b',c',d')

By determining the region patterns, the following results are obtained:

Table 2. Region patterns of images from a certain diagnosis

ID	Region patterns
1	R(a,b,c,-),R(a,b,-,d),R(a',b',c',d'), R(a',b',c',-), R(-,b',c',d')
2	R(a,b,c,-),R(a,b,-,-),R(a',b',c',d'), R(a',b',c',-), R(-,b',c',d')
3	R(a,b,c,-), R(a,b,-,-), R(a',b',c',-), R(-,b',c',-)
4	R(a,b,c,-), R(a,b,-,-),R(a',b',c',-), R(-,b',c',-)
5	R(a,b,-,d),R(a,b,-,-), R(-,b',c',d'), R(-,b',c',-)

The image modelling in terms of itemsets and transactions is the following:

- the transactions represent the set of region patterns, determined by the previous algorithm.
- the itemsets are formed by region patterns of the images laying in the same diagnosis.
- the frequent itemsets represent the itemsets with the support greater than the minimum support.
- the itemsets of cardinality between 1 and k are iteratively found.
- the frequent itemsets are used for rule generation.

The algorithm for rules generation based on region patterns is described in pseudo-code:

Algorithm 3.4: rules generation based on region patterns.
Input: the set of images represented as: I = (RS_1, …, RS_k), where RS_m is the region pattern.
Output: the set of pattern rules.
Method:
Ck: the set of region patterns of k-length
Lk: the set of frequent region patterns of k-length
Rules: the set of rules constructed from frequent itemsets for k>1.
L1= {frequent region patterns};
for(k=1; Lk!=null; k++) do begin
  Ck+1= candidates generated from the set Lk;
  for each transaction t in the database do
    *Increment the number of all candidates that appear in t.
  end.
  Lk+1=candidates from Ck+1 that has the support greater or equal than suport_min
end.
Rules = Rules + {Lk+1->diagnosis}
end.

Applying the algorithm on the transaction set from Table 2, the diagnosis is determined by the following semantic pattern rules:

R(a,b,c,-) and R(a',b',c',-) ->diagnosis,
R(a,b,c,-) and R(a,b,-,-)->diagnosis,
R(a',b',c',-) and R(a,b,-,-) ->diagnosis,
R(a,b,-,-) and R(-,b',c',-) ->diagnosis,
R(a,b,c,-) and R(a',b',c',-) and R(a,b,-,-) ->diagnosis.

The image testing/annotation phase has as scope the automatic diagnose of images.

- each new image is processed and segmented in regions,
- for each new image the low-level characteristics are mapped to semantic indicators,
- the classification algorithm is applied for identifying the image diagnosis.

Being given a new image, the classification process searches in the rules set for finding its most appropriate diagnosis. Images are processed and are represented by means of semantic indicators as Prolog facts. The semantic rules are applied on the set of images facts, using the Prolog inference engine.

A semantic rule matches an image if all characteristics, which appear in the body of the rule, also appear in the image characteristics.

Experimental results

In the experiments realized through this study, two databases are used for learning and testing process. The database used to learning the correlations between images and diagnosis contains 200 images from digestive diagnosis. The database used in the learning process is categorized into the following diagnosis: ulcer, polyps, gastric cancer and rectocolitis. The system learns each concept by submitting about 20 images per diagnosis.

For each diagnosis, the following metrics (accuracy-A, sensitivity-S, specificity-SP) are computed:

$$A = \frac{TP + TN}{TP + FP + FN + TN}, \; S = \frac{TP}{TP + FN}, \; SP = \frac{TN}{TN + FP}, \qquad (1)$$

where TP represents the number of true positives (images correctly diagnosed with the searched diagnosis), FP represents the number of false positives (images incorrectly diagnosed with the searched diagnosis), TN represents the number of true negatives (images correctly diagnosed with a different diagnosis, FN represents the number of false negatives (images incorrectly diagnosed with a different diagnosis.

The results of the presented methods are very promising, being influenced by the complexity of endoscopic images as can be observed in Table 1. Improvements can be brought using a segmentation method with greater semantic accuracy.

Table 3. Results recorded by each diagnosis

Diagnosis	Accuracy (%)	Sensitivity (%)	Specificity (%)
Ulcer	96.5	92.5	71
Polyps	96.5	92	71.5
Esophagitis	96.0	90.5	71
Gastric Cancer	96.1	90.5	71.5
Rectocolitis	97.1	93.7	72.9

Conclusions

In this study, the proposed and developed methods could assist physicians by doing automatic diagnosis based on visual content of medical images. For establishing correlations with diagnosis, we experimented and selected some low-level visual characteristics of images. So, each diagnosis is translated into visual computable characteristics and terms of sick regions. On the other hand, images are represented as a single colour regions list and they are mapped to semantic descriptors. The annotation procedure starts with the semantic rules generation or each image diagnosis. The language used for rules representation is Prolog. The advantages of using Prolog are its flexibility and simplicity in representation of rules.

Actually, the results of experiments are very promising, because they show a good accuracy and sensitivity and a medium precision for the majority of the database categories, making the system more reliable.

References

1. **Zhou X. S., Zillner S., Moeller M., Sintek M., Zhan Y., Krishnan A., and Gupta A.** Semantics and CBIR: a medical imaging perspective // Proceedings of the 2008 International Conference on Content-Based Image and Video Retrieval. – 2008.
2. **Smeulders A. W. M., Worring M., Santini S., Gupta A., Jain R.** Content-based image retrieval at the end of the early years // IEEE Trans. Pattern Anal. Machine Intelligence. – 2000. – Vol. 22(12). – P. 1349–1380.
3. **Duncan J., Ayache N**. Medical image analysis: Progress over two decades and the challenges ahead // IEEE Trans. Pattern Anal. Machine Intelligence. – 2000. – Vol. 22 (1). –P. 85–106.
4. **Ribeiro M. X., Traina A. J., Rosa N. A., and Marques P. M.** How to Improve Medical Image Diagnosis through Association Rules: The IDEA Method // Proceedings of the 21st IEEE international Symposium on Computer-Based Medical Systems IEEE Computer Society. – 2008.
5. **Antonie M. L., Zaane O. R., and Coman A.** Associative classifiers for medical images // LNAI 2797, MMCD. – Springer-Verlag. – 2003. – P. 68–83.
6. **Pan H., Li J., Wei Z.** Mining interesting association rules in medical images // Advance Data Mining and Medical Applications. – 2005.
7. **Smith J. R. and Chang S.-F.** VisualSEEk: a fully automated content-based image query system // The Fourth ACM International Multimedia Conference and Exhibition. – Boston, MA, USA. – 1996.

Received 2009 09 09

S. C. Udristoiu, A. L. Ion. Establishing Medical Diagnosis using Pattern Semantic Rules // Electronics and Electrical Engineering. – Kaunas: Technologija, 2010. – No. 2(98). – P. 75–78.

The developed methods are based on pattern rules to support medical image diagnosis. They have important characteristics that make them different from other CAD methods: the process is completely automatic, with the possibility to define a great number of diagnosis; the methods could be applied to any medical domain, because the visual features, the semantic indicators remain unchangeable, and the semantic rules are generated by learning from labeled images-examples; the selection of the visual characteristics set is based on their retrieval accuracy; the spatial information of the regions is considered, offering important medical information as the relationships of a sick region with another. Although we present the results achieved in endoscopic images analysis, our methods can be used to analyze other types of medical images. The prototype system was applied to real datasets and the results show high accuracy. Ill. 2, bibl. 7 (in English; summaries in English, Russian and Lithuanian).

С. Ц. Удристойу, А. Л. Ион. Медицинская диагностика с использованием шаблонных семантических правил // Электроника и электротехника. – Каунас: Технология, 2010. – № 2(98). – С. 75–78.

Разработанные методы основаны на шаблонах правил и поддерживают медицинскую диагностику используя изображении. Они имеют важные характеристики, которые отличают их от других методов автоматизированного проектирования: процесс полностью автоматизирован, с возможностью определения большого числа диагноз; методы могут быть применены к любой медицинской области, потому что визуальные особенности, семантические показатели остаются неизменными и семантические правила генерируются путем изучения помеченых примеров изображений; выбор набора визуальных характеристик основан на точность их поиска; оценена пространственная информация в регионах, предлагая важную медицинскую информацию по отношению больных областей. Хотя мы представили результаты, достигнутые используя эндоскопичесий анализ изображений, наши методы могут быть использованы для анализа других типов медицинских изображений. Экспериментальная система была применена на реальных данных и результаты показывают высокую точность. Ил. 2, библ. 7 (на английском языке; рефераты на английском, русском и литовском яз.).

S. C. Udristoiu, A. L. Ion. Medicininės diagnozės nustatymas naudojant semantines šablono taisykles // Elektronika ir elektrotechnika. – Kaunas: Technologija, 2010. – No. 2(98). – P. 75–78.

Sukurti metodai remiasi taisyklių šablonais ir padeda priimti medicininę diagnozę pagal nuotraukose matomą vaizdą. Nuo kitų CAD metodų jie skiriasi tokiomis svarbiomis charakteristikomis: procesas yra visiškai automatiškas, galima apibrėžti didelį skaičių diagnozių, metodus galima taikyti bet kuriai medicinos sričiai, kadangi vizualiniai aspektai ir semantiniai indikatoriai išlieka nepakitę, o semantinės taisyklės kuriamos sistemą apmokant naudojant suklasifikuotas pavyzdines nuotraukas; vizualinių charakteristikų parinkimas remiasi jų nustatymo tikslumu; analizuojama erdvinė sričių informacija, pateikiama svarbi medicininė informacija apie vieno ligos pažeisto regiono sąsają su kitu. Nors pateikti rezultatai gauti naudojant endoskopinių nuotraukų analizę, metodus galima taikyti analizuojant ir kitokio pobūdžio medicininius vaizdus. Sistemos prototipas buvo išbandytas su realių duomenų rinkiniais ir rezultatai rodo, jog tikslumas yra didelis. Il. 2, bibl. 7 (anglų kalba; santraukos anglų, rusų ir lietuvių k.).