

Clinical Content Detection for Medical Image Retrieval

L. Chen¹, H. L. Tang¹ and I. Wells²

¹Department of Computing, University of Surrey, Guildford, Surrey, GU2 7XH, UK

²Department of Medical Physics, Royal Surrey County Hospital, Guildford, Surrey, GU2 7XX, UK

Abstract - Content-based image retrieval (CBIR) is the most widely used method for searching large-scale medical image collections; however this approach is not suitable for high-level applications as human experts are accustomed to manage medical images based on their clinical features rather than primitive features. Automatic detection of clinical features in a large-scale image database and realization of image retrieval by clinical content are still open issues. This paper presents a Markov random field (MRF) based model for clinical content detection. Multiple classifiers are applied to recognize a wide range of clinical features in a large-scale histological image database, and they are further combined to generate more reliable and robust estimation. Spatial contexts will cooperate with local estimations in the MRF based model to make a decision based on global consistency. The detected clinical features will provide a basis for image retrieval. Experiments have been carried out in a large-scale histological image database with promising results.

I. INTRODUCTION

Volumes of medical images are rapidly generated and how to effectively manage them has become a great challenge [1]. Traditional indexing and retrieval of images are based on patient names, identifiers, keywords and manual annotations, and these indexing methods vary with different data types, formats and approaches. Clinical experts find such database less useful when there is a need to retrieve clinical meaningful images. This applies equally to image resources available on the Internet and to image databases held by different departments in the same hospital. A promising solution is to index and retrieve images based on their individual content including both primitive and semantic properties.

CBIR has been widely used to retrieve medical images by matching visual features. This method is capable of managing large-scale image databases because of comparatively generic attributes of low-level feature

descriptors; however semantic/clinical features are desirable rather than low-level features in high-level applications [2]. This has prompted much research in the field of image retrieval by semantic content, although automatic semantic content recognition has not yet been achieved in practice. The broader the domain of the image data, the more difficult the semantics are to detect automatically [3]. The reason is that images in broad domains potentially have an unpredictable and complicated variability in their appearance even when they embody identical semantics. This great difficulty in deriving clinical features from primitive features has constrained many applications to examine a limited number of clinical features along a single organ or a narrow area of body.

The research presented in this paper will continue to investigate to examine a large-scale histological image database, which is obtained from the whole gastrointestinal tract including six organs (oesophagus, stomach, small intestine, large intestine, anus and appendix) [2]. This type of image is visually similar and usually differs only in small details, but such subtle differences may be of pathological significance [4]. This provides a great challenge for developing an automatic clinical content recognition mechanism, which can help retrieve images by clinical content. This paper proposes to apply multiple classifiers to cover complicated visual varieties associated with a wide range of clinical features in the large-scale histological image archive. These classifiers are combined with complementary information from each other, and provide reliable and robust local estimations. This cooperates with spatial contextual information in a MRF based model to jointly detect clinical features in a global view. Images are then retrieved based on these detected clinical features.

II. CLINICAL CONTENT DETECTION BASED ON A MRF BASED MODEL

Automatic detection of clinical features on an image depends on measurements of local regions and spatial contexts among these regions. An image is first segmented into small regions, which are assumed to represent clinical features from interest of histopathologists. Assume that N segmented regions, $R = \{R_1, R_2, \dots, R_N\}$, need to be recognized, and M potential clinical features, $\Omega = \{w_1, w_2, \dots, w_M\}$, are examined in a large-scale image archive. A set of random variables, $L = \{L_1, L_2, \dots, L_N\}$, is defined on these regions. The lowercase letter, $l = \{l_1, l_2, \dots, l_N\}$ denotes a realization/configuration of L , where $l_i \in \Omega$, $i = 1, 2, \dots, N$. $L = \{L_1, L_2, \dots, L_N\}$ is a set of Markov random variables if and only if $P(l) > 0$ and $P(l_i | l_j, j \neq i) = P(l_i | l_{n(R_i)})$, where $l_{n(R_i)} = \{l_i | i' \in n(R_i)\}$. $n(R_i)$ is a set of neighbours of region R_i .

With respect to Bayesian theorem, the task of detection of clinical features on an image becomes finding an optimal configuration l with the maximum joint probability, given observation d and domain knowledge pk . That is,

$$\text{assign } l \rightarrow L \text{ if } l = \arg \max_{l' \in \Psi} P(l' | d, pk) \quad (1)$$

where Ψ is the set of all possible configurations on L .

According to Markov-Gibbs equivalence,

$$P(l | d, pk) = Z^{-1} \times e^{-\frac{1}{T} U(l | d, pk)} \quad (2)$$

where $Z = \sum_l e^{-\frac{1}{T} U(l | d, pk)}$ is a normalizing constant called the partition function, T is a constant called the temperature which shall be assumed to be 1 unless specifically stated, and $U(l | d, pk)$ is the energy function.

Considering a four-neighbour system, the energy function is thus written as:

$$U(l | d, pk) = \sum_{i \in R} V_1(l_i | d, pk) + \sum_{i \in R} \sum_{j \in n(i)} V_2(l_i, l_j | d, pk) \quad (3)$$

where $V_i(\cdot)$ is called a clique function. The first part in (3) is called single-region clique function, and the second part is pair-region function. The energy function includes not only the information of single regions but also spatial contextual information among regions. A general principle for design of a clique function is that if the interpretation of regions (or a single region) tends to be consistent with the

measurements and spatial contexts, the clique function decreases, resulting in a decrease in the energy function, and vice versa [5].

A. Single-Region Clique Function by Combining Multiple Classifiers

The previous methods for designing single-region clique functions normally use heuristics or single detector. These methods however failed to deal with a large-scale data, which have more complicated visual appearances associated with a wide range of clinical features. This research proposes to apply multiple classifiers to handle these complicated visual and semantic varieties in a large-scale data, and generate a reliable and robust estimation on local regions by combining complementary information from individual classifiers.

Assume K statistically independent classifiers, denoted as e_1, e_2, \dots, e_K , to deal with a local region R_i . For an arbitrary classifier: $e_k(x) = j$, $k = 1, 2, \dots, K$, $j \in \Omega \cup \{w_{M+1}\}$, w_{M+1} represents rejected or unknown labels by a classifier, its confusion matrix records its detailed performance as below:

$$CM_k = \begin{pmatrix} n_{11}^k & n_{12}^k & \dots & n_{1M}^k & n_{1(M+1)}^k \\ n_{21}^k & n_{22}^k & \dots & n_{2M}^k & n_{2(M+1)}^k \\ \dots & \dots & \dots & \dots & \dots \\ n_{M1}^k & n_{M2}^k & \dots & n_{MM}^k & n_{M(M+1)}^k \end{pmatrix}$$

Each row i corresponds to class w_i , and each column j corresponds to $e_k(x) = w_j$. The element n_{ij}^k means that n_{ij}^k samples of category w_i are assigned to category w_j by $e_k(\bullet)$. The belief of the k^{th} classifier is calculated as [6]:

$$b_k(x \in w_i | e_k(x) = w_j, EN) = P(x \in w_i | e_k(x) = w_j, EN) = \frac{n_{ij}^k / n_{i \cdot}^k}{\sum_{t=1}^M (n_{ij}^k / n_{i \cdot}^k)} = \frac{n_{ij}^k}{\sum_{t=1}^M (n_{ij}^k \zeta_{it}^k)} \quad (4)$$

where $\zeta_{it}^k = \frac{n_{i \cdot}^k}{n_{i \cdot}^k}$, $i = 1, 2, \dots, M$ and $j = 1, 2, \dots, M + 1$. EN denote a normal classification environment.

By applying the averaging rule on multiple classifiers, the belief with a realization l_i on R_i is calculated as:

$$\begin{aligned}
b(l_i) &= b(L_i = l_i | e_1(i), e_2(i), \dots, e_k(i), pk) \\
&= P(L_i = l_i | e_1(i), e_2(i), \dots, e_k(i), pk) \quad (5) \\
&= \lambda \frac{1}{K} \sum_{k=1}^K b_k(L_i = l_i | e_k(i), pk)
\end{aligned}$$

where $\frac{1}{\lambda} = \sum_{l_i \in \Omega} \frac{1}{K} \sum_{k=1}^K b_k(l_i)$ to make sure that $\sum_{l_i \in \Omega} b(l_i) = 1$.

When the realization l_i is consistent with the domain knowledge and the statistical attributes on a local region R_i , the clique function should be as low as possible. The single-region clique function is thus defined as:

$$V_1(l_i | e_1(R_i), e_2(R_i), \dots, e_k(R_i), pk) = \delta_i (1 - \beta_i b(l_i)) \quad (6)$$

where $1 \geq \delta_i > 0$ and $1 \geq \beta_i > 0$ are the weights associated with the corresponding clique functions and they control the contributions of each clique function.

The belief calculation method is applicable to arbitrary classifiers, so the design of single-region functions has the generic capability of managing a large-scale image database. Meanwhile, complementary information from different classifiers makes parameter estimations on single regions more reliable and robust.

B. Pair-Region Clique Functions by a Knowledge Elicitation Subsystem

A knowledge elicitation subsystem [4] is used to learn spatial contextual information from the training data. The learning mechanism, through an intuitive interface, allows histopathologists provide sample knowledge which then is automatically modelled.

Given an arbitrary pair of regions $\{R_i, R_j\}$, its function is defined as:

$$V_2(l_i, l_j | D^{(2)}(R_i, R_j), pk) = \delta_{ij} T_{ij} \quad (7)$$

where $T_{ij}=1$ if the realization l_i and l_j are possibly next to each other; otherwise $T_{ij}=0$. δ_{ij} is weights for contributions from the clique function. $D^{(2)}(R_i, R_j)$ is the measurement of a pair of regions. Since the measurement between the regions R_i and R_j has no explicit statistical cues like boundary length and texture (or colour) contrast in the case of this research, Formula (7) is actually only based on spatial contextual knowledge acquired from the knowledge elicitation subsystem. If there are more complex constraints (for

example, a measurement of a combination of two regions) available for a pair of regions, their clique functions will become more and more complicated.

C. Optimization

A MRF based model has been constructed for the detection of clinical features in a large-scale histological image archive in terms of multiple processors and domain knowledge. Its optimal interpretation is a realization of a set of Markov random variables with the maximum probability. The *simulated annealing* (SA) algorithm is a popular method used for solving combinatorial optimization problems [5]. The SA algorithm is a stochastic iterative optimization procedure, however the SA needs a slow enough schedule to reach a global solution. In this research, a multi-population *steepest descent* algorithm [4] has been developed to search a global optimization in a quick way.

III. USING DETECTED CLINICAL FEATURES FOR IMAGE RETRIEVAL

Two similarity measurements [2] are designed to compare local neighbour pattern of semantic labels and clinical feature frequency distribution. Neighbourhood Similarity (NS) uses a matrix to record the co-occurrence frequencies of all clinical features on a four-node neighbour system. Each element $n(i, j)$ of the co-occurrence matrix records how many times label i is next to label j in an image. The similarity between a query image and an image in the database is calculated as follows [2]:

$$similarity = \sum_{j=1}^N \sum_{i=1}^N |n_q(i, j)s_q - n_r(i, j)s_r|$$

$$NS = \exp\left(-\frac{similarity}{W_N}\right)$$

where N is the number of the clinical features and W_N is the number of pair-regions in an image. s_q and s_r are the scaling factors to eliminate the influence of the trivial 'lumen' feature in the query and retrieved images, respectively.

$$s_q, s_r = \frac{w_t}{w_t - w_L}$$

where w_t and w_L are the total number of regions and the number of regions labelled with the 'lumen' feature respectively. Fig.1 illustrates an example of image retrieval by the NS similarity measurement. The first position on the

retrieved ranking list is reasonably the query image itself.

Clinical feature frequency distribution similarity (CFFDS) [2] counts the frequencies of the histological labels detected in an image. The CFFDS measurement is defined as:

$$\text{similarity} = \sum_{i=1}^N |F_q(i)s_q - F_r(i)s_r|$$

$$NS = \exp\left(-\frac{\text{similarity}}{F_N}\right)$$

where N is the number of the clinical categories, $F_q(i)$ and $F_r(i)$ are the frequencies of clinical feature i occurring in the query image and target image in the database, and F_N is the total number of regions in an image. s_q and s_r are defined same as above. Fig. 2 shows a retrieval example by applying the CFFDS measurement.

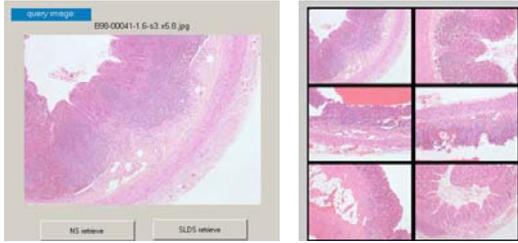


Fig. 1. Examples of NS similarity measurement

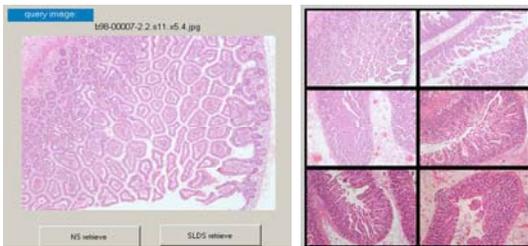


Fig. 2. Examples of CFFDS similarity measurement

The performance of the image retrieval is evaluated using the following method. The first N ($=6$) images on the ranking list are taken into the calculation as:

$$\text{measure_of_goodness} = \frac{1}{N} \sum_{i=1}^N C_i$$

$$\text{where } C_i = \begin{cases} N+1-i & \text{if } C_i \text{ is the correct answer} \\ 0 & \text{otherwise} \end{cases}$$

It should be noted that the original query image always appears at the first position and it is ignored from the ranking list during the calculation of *measure_of_goodness*. N should be 5 in this case. Points are awarded if the correct image appears in the first N positions. In an ideal situation, the first N positions are positive retrieved images and the

measure_of_goodness equals to 3. In a worst situation where the first N positions are negative samples, the *measure_of_goodness* should be 0. Experiments have been carried out to evaluate the *measure_of_goodness* in the large-scale histological image database. The average *measure_of_goodness* of the NS measurement is 2.36, and the average *measure_of_goodness* of the CFFDS measurement is 2.44. They are close to the ideal situation. The retrieval performance, to some extent, depends on the performance of automatic detection of clinical features.

IV. CONCLUSION

This paper presented a MRF based model for clinical content detection in a large-scale histological image database, with parameters estimated by combining multiple classifiers and spatial contexts acquired by a knowledge elicitation subsystem. Images are then retrieved based on the detected clinical content. Although the system has been evaluated in the gastrointestinal tract images, the proposed method has been designed in a generic way to suit for other types of image. We are currently planning to extend the approach to manage more medical image databases including mammography and prostate data.

REFERENCES

- [1] H. Müller, N. Michoux, et al., "A review of content-based image retrieval systems in medical applications – clinical benefits and future directions", *International Journal of Medical Informatics*, Vol. 73, pp. 1-23, 2004.
- [2] H. L. Tang, R. Hanka, et al., "Histological image retrieval based on semantic content analysis", *IEEE Transaction on Information Technology in BioMedicine*, 7(1), pp. 26-36, 2003.
- [3] A. Smeulders, M. Worring, et al., "Content-based image retrieval at the end of the early years", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(12), pp. 1349-1380, 2000.
- [4] L. Chen, *Semantic content recognition for a large-scale medical image archives*, PhD thesis, University of Surrey, UK. 2005.
- [5] J. W. Modestino and J. Zhang, "A Markov random field model-based approach to image interpretation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 14(6), pp. 606-615, 1992.
- [6] L. Chen and H. L. Tang, "Improved computation of beliefs based on confusion matrix for combining multiple classifiers", *IEE Electronics Letters*, 40(4), pp. 238-239. 2004.