VisMed: A Visual Vocabulary Approach for Medical Image Indexing and Retrieval

Joo-Hwee Lim, Jean-Pierre Chevallet

VisMed: A Visual Vocabulary Approach for Medical Image Indexing and Retrieval

Joo-Hwee Lim[1] and Jean-Pierre Chevallet[2]

[1] Institute for Infocomm Research
21 Heng Mui Keng Terrace, Singapore 119613
joohwee@i2r.a-star.edu.sg
[2] Image Processing and Application Lab (IPAL)
French National Center for Scientific Research (CNRS)
21 Heng Mui Keng Terrace, Singapore 119613
Jean-Pierre.Chevallet@imag.fr

Abstract. Voluminous medical images are generated daily. They are critical assets for medical diagnosis, research, and teaching. To facilitate automatic indexing and retrieval of large medical image databases, we propose a structured framework for designing and learning vocabularies of meaningful medical terms associated with visual appearance from image samples. These VisMed terms span a new feature space to represent medical image contents. After a multi-scale detection process, a medical image is indexed as compact spatial distributions of VisMed terms. A flexible tiling (FlexiTile) matching scheme is proposed to compare the similarity between two medical images of arbitrary aspect ratios.
We evaluate the VisMed approach on the medical retrieval task of the ImageCLEF 2004 benchmark. Based on 2% of the 8725 CasImage collection, we cropped 1170 image regions to train and validate 40 VisMed terms using support vector machines. The Mean Average Precision (MAP) over 26 query topics is 0.4156, an improvement over all the automatic runs in ImageCLEF 2004.

1 Introduction

Medical images are an integral part in medical diagnosis, research, and teaching. Medical image analysis research has focused on image registration, measurement, and visualization. Although large amounts of medical images are produced in hospitals every day, there is relatively less research in medical content-based image retrieval (CBIR) [1]. Besides being valuable for medical research and training, medical CBIR systems also have a role to play in clinical diagnosis [2]. For instance, for less experienced radiologists, a common practice is to use a reference text to find images that are similar to the query image [3]. Hence, medical CBIR systems can assist doctors in diagnosis by retrieving images with known pathologies that are similar to a patient's image(s).

Among the limited research efforts of medical CBIR, classification or clustering driven feature selection and weighting has received much attention as

general visual cues often fail to be discriminative enough to deal with more subtle, domain-specific differences and more objective ground truth in the form of disease categories is usually available [3, 4].

In reality, pathology bearing regions tend to be highly localized [3]. Hence, local features such as those extracted from segmented dominant image regions approximated by best fitting ellipses have been proposed [5]. A hierarchical graph-based representation and matching scheme has been suggested to deal with multi-scale image decomposition and their spatial relationships [5]. However, it has been recognized that pathology bearing regions cannot be segmented out automatically for many medical domains [1]. As an alternative, a comprehensive set of 15 perceptual categories related to pathology bearing regions and their discriminative features are carefully designed and tuned for high-resolution CT lung images to achieve superior precision rates over a brute-force feature selection approach [1].

Hence, it is desirable to have a medical CBIR system that represents images in terms of semantic local features, that can be learned from examples (rather than handcrafted with a lot of expert input) and do not rely on robust region segmentation. In this paper, we propose a structured learning framework to build meaningful medical terms associated with visual appearance from image samples. These *VisMed* terms span a new feature space to represent medical image contents. After a segmentation-free multi-scale detection process, a medical image is indexed as compact spatial distributions of VisMed terms. A flexible tiling (FlexiTile) matching scheme is also proposed to compare the similarity between two medical images of arbitrary aspect ratios.

Indeed, the US National Cancer Institute has launched a cooperative effort known as the Lung Image Database Consortium (LIDC) to develop an image database that will serve as an international research resource for the development, training, and evaluation of computer-aided diagnostic (CAD) methods in the detection of lung nodules on CT scans. One of the key efforts is to create a visual nodule library with images of lesions that span the focal abnormality spectrum and the subset nodule spectrum [6]. All lesions have been characterized by a panel of experienced thoracic radiologists based on attributes that include shape, margin, internal structure, and subtlety. The library is intended to serve as a standard for the development of a practical radiologic definition of nodule as radiologists believe that "the expertise of the interpreter lies in a vast experience of seeing many thousands of radiologic patterns and sythesizing them into a coherent, organized, and searchable mental matrix of diagnostic meaning and pathologic features" [7].

In this paper, we evaluate the VisMed approach on ImageCLEF 2004 medical retrieval task. Based on 2% of the 8725 CasImage data, we cropped 1170 image regions to train and validate 40 VisMed terms using support vector machines [8]. The Mean Average Precision over 26 query topics is 0.4156, an increase over all the automatic runs in ImageCLEF 2004. We detail the VisMed framework and evaluation in the next two sections respectively.

2 VisMed: A Structured Learning Framework for Medical CBIR

In this paper, we aim to bridge the semantic gap between low-level visual features (e.g. texture, color) and high-level semantic terms (e.g. brain, lung, heart) in medical images for content-based indexing and retrieval. At the moment, we focus on visual semantics that can be directly extracted from image content (without the use of associated text) with computer vision techniques.

In order to manage large and complex set of visual entities (i.e. high content diversity) in the medical domain, we propose a structured learning framework to facilitate modular design and extraction of medical visual semantics, VisMed terms, in building content-based medical image retrieval systems.

VisMed terms are segmentation-free image regions that exhibit semantic meanings to medical practitioners and that can be learned statistically to span a new indexing space. They are detected in image content, reconciled across multiple resolutions, and aggregated spatially to form local semantic histograms. The resulting compact and abstract representation can support both similarity-based query and compositional visual query efficiently. In this paper, we only report evaluation results for similarity-based retrieval (i.e. query by image examples). For the unique compositional visual query method and its evaluation based on consumer images, please refer to another regular paper in the same proceeding [9]. We have also performed the semantic-based query based on the compositional visual query method for ImageCLEF 2005 queries and dataset. We will report this work elsewhere in the near future.

2.1 Learning of VisMed Terms

VisMed terms are typical semantic tokens with visual appearance in medical images (e.g. X-ray-lung, CT-head-brain, MRI-abdomen-liver, mouth-teeth). They are defined using image region instances cropped from sample images and modeled based on statistical learning.

In this paper, we have adopted color and texture features as well as support vector machines (SVMs) [8] for VisMed term representation and learning respectively though the framework is not dependent on a particular feature and classifier. The notion of using a visual vocabulary to represent and index image contents for more effective (i.e. semantic) query and retrieval has been proposed and applied to consumer images [10, 11].

To compute VisMed terms from training instances, we use SVMs on color and texture features for an image region and denote this feature vector as z. A SVM $\mathcal{S}_k$ is a detector for VisMed term k on z. The classification vector T for region z is computed via the softmax function [12] as

$$T_k(z) = \frac{\exp^{\mathcal{S}_k(z)}}{\sum_j \exp^{\mathcal{S}_j(z)}}. \quad (1)$$

That is, $T_k(z)$ corresponds to a VisMed entry in the 40-dimensional vector T adopted in this paper.

In our experiments, we use the YIQ color space over other color spaces (e.g. RGB, HSV, LUV) as it performed better in our experiments. For the texture feature, we adopted the Gabor coefficients which have been shown to provide excellent pattern retrieval results [13].

A feature vector z has two parts, namely, a color feature vector z^c and a texture feature vector z^t. We compute the mean and standard deviation of each YIQ color channel and the Gabor coefficients (5 scales, 6 orientations) respectively [11]. Hence the color feature vector z^c has 6 dimensions and the texture feature vector z^t has 60 dimensions. Zero-mean normalization [14] was applied to both the color and texture features. In our evaluation described below, we adopted RBF kernels with modified city-block distance between feature vectors y and z,

$$|y - z| = \frac{1}{2}\left(\frac{|y^c - z^c|}{N_c} + \frac{|y^t - z^t|}{N_t}\right) \tag{2}$$

where N_c and N_t are the numbers of dimensions of the color and texture feature vectors (i.e. 6 and 60) respectively. This just-in-time feature fusion within the kernel combines the contribution of color and texture features equally. It is simpler and more effective than other feature fusion methods that we have attempted.

2.2 Image Indexing Based on VisMed Terms

After learning, the VisMed terms are detected during image indexing from multi-scale block-based image patches without region segmentation to form semantic local histograms as described below.

Conceptually, the indexing is realized in a three-layer visual information processing architecture. The bottom layer denotes the pixel-feature maps computed for feature extraction. In our experiments, there are 3 color maps (i.e. YIQ channels) and 30 texture maps (i.e. Gabor coefficients of 5 scales and 6 orientations). From these maps, feature vectors z^c and z^t compatible with those adopted for VisMed term learning (Equation (2)) are extracted.

To detect VisMed terms with translation and scale invariance in an image to be indexed, the image is scanned with windows of different scales, similar to the strategy in view-based object detection [15, 16]. More precisely, given an image I with resolution $M \times N$, the middle layer, Reconciled Detection Map (RDM), has a lower resolution of $P \times Q, P \leq M, Q \leq N$. Each pixel (p, q) in RDM corresponds to a two-dimensional region of size $r_x \times r_y$ in I. We further allow tessellation displacements $d_x, d_y > 0$ in X, Y directions respectively such that adjacent pixels in RDM along X direction (along Y direction) have receptive fields in I which are displaced by d_x pixels along X direction (d_y pixels along Y direction) in I. At the end of scanning an image, each pixel (p, q) that covers a region z in the pixel-feature layer will consolidate the classification vector $T_k(z)$ (Equation (1)).

In our experiments, we progressively increase the window size $r_x \times r_y$ from 20×20 to 60×60 at a displacement (d_x, d_y) of $(10, 10)$ pixels, on a 240×360 size-normalized image. That is, after the detection step, we have 5 maps of detection

of dimensions 23×35 to 19×31, which are reconciled into a common RDM as explained below.

To reconcile the detection maps across different resolutions onto a common basis, we adopt the following principle: If the most confident classification of a region at resolution r is less than that of a larger region (at resolution $r+1$) that subsumes the region, then the classification output of the region should be replaced by those of the larger region at resolution $r+1$. For instance, if the detection of a face is more confident than that of a building at the nose region (assuming that both face and building (but not nose) are in the visual vocabulary designed for a particular application), then the entire region covered by the face, which subsumes the nose region, should be labeled as face.

To illustrate the point, suppose a region at resolution r is covered by 4 larger regions at resolution $r+1$ as shown in Figure 1. Let $\rho = max_k max_i T_i(z_k^{r+1})$ where k refers to one of the 4 larger regions in the case of the example shown in Figure 1. Then the principle of reconciliation says that if $max_i T_i(z^r) < \rho$, the classification vector $T_i(z^r) \ \forall i$ should be replaced by the classification vector $T_i(z_m^{r+1}) \ \forall i$ where $max_i T_i(z_m^{r+1}) = \rho$.

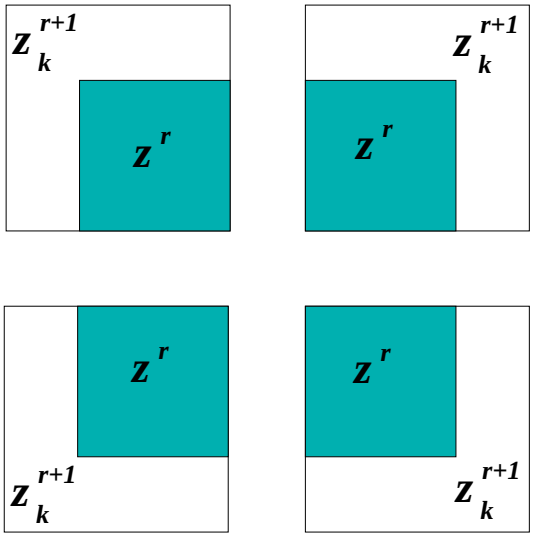

Fig. 1. Reconciling multi-scale VisMed detection maps

Using this principle, we compare detection maps of two consecutive resolutions at a time, in descending window sizes (i.e. from windows of 60×60 and 50×50 to windows of 30×30 and 20×20). After 4 cycles of reconciliation, the detection map that is based on the smallest scan window (20×20) would have consolidated the detection decisions obtained at other resolutions for further spatial aggregation.

The purpose of spatial aggregation is to summarize the reconciled detection outcome in a larger spatial region. Suppose a region Z comprises of n small equal regions with feature vectors $z_1, z_2, \cdots, z_n$ respectively. To account for the size of detected VisMed terms in the spatial area Z, the classification vectors of the reconciled detection map are aggregated as

$$T_k(Z) = \frac{1}{n} \sum_i T_k(z_i). \tag{3}$$

This is the top layer in our three-layer visual information processing architecture where a Spatial Aggregation Map (SAM) further tessellates over RDM with $A \times B, A \leq P, B \leq Q$ pixels. This form of spatial aggregation does not encode spatial relation explicity. But the design flexibility of s_x, s_y in SAM on RDM (the equivalent of r_x, r_y in RDM on I) allows us to specify the location and extent in the content to be focused and indexed. We can choose to ignore unimportant areas (e.g. margins) and emphasize certain areas with overlapping tessellation. We can even have different weights attached to the areas during similarity matching.

To facilitate spatial aggregation and matching of image with different aspect ratios ρ, we design 5 tiling templates for Eq. (3), namely 3×1, 3×2, 3×3, 2×3, and 1×3 grids resulting in 3, 6, 9, 6, and 3 $T_k(Z)$ vectors per image respectively. Since the tiling templates have aspect ratios of 3, 1.5, and 1, the decision thresholds to assign a template for an image are set to their mid-points (2.25 and 1.25) as $\rho > 2.25$, $1.25 < \rho \leq 2.25$, and $\rho \leq 1.25$ respectively based on $\rho = \frac{L}{S}$ where L and S refer to the longer and shorter sides of an image respectively. For more details on detection-based indexing, readers are referred to [11].

2.3 FlexiTile Matching

Given two images represented as different grid patterns, we propose a flexible tiling (FlexiTile) matching scheme to cover all possible matches. For instance, given a query image Q of 3×1 grid and an image Z of 3×3 grid, intuitively Q should be compared to each of the 3 columns in Z and the highest similarity will be treated as the final matching score. As another example, consider matching a 3×2 grid with 2×3 grid. The 4 possible tiling and matching choices are shown in Fig. 2.

The FlexiTile matching scheme is formalized as follows. Suppose a query image Q and a database image Z are represented as $M_1 \times N_1$ and $M_2 \times N_2$ grids respectively. The overlaping grid $M \times N$ where $M = \min(M_1, M_2)$ and $N = \min(N_1, N_2)$ is the maximal matching area. The similarity λ between Q and Z is the maximum matching among all possible $M \times N$ tilings,

$$\lambda(Q, Z) = \max_{m_1=1, n_1=1}^{m_1=u_1, n_1=v_1} \max_{m_2=1, n_2=1}^{m_2=u_2, n_2=v_2} \lambda(Q_{m_1,n_1}, Z_{m_2,n_2}), \tag{4}$$

where $u_1 = M_1 - M + 1, v_1 = N_1 - N + 1, u_2 = M_2 - M + 1, v_2 = N_2 - N + 1$ and the similarity for each tiling $\lambda(Q_{m_1,n_1}, Z_{m_2,n_2})$ is defined as the average

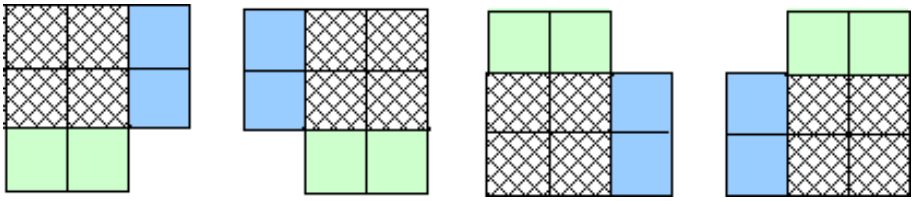

Fig. 2. Example to illustrate FlexiTile matching

similarity over $M \times N$ blocks as

$$\lambda(Q_{m_1,n_1}, Z_{m_2,n_2}) = \frac{\sum_i \sum_j \lambda_{ij}(Q_{m_1,n_1}, Z_{m_2,n_2})}{M \times N}, \tag{5}$$

and finally the similarity $\lambda_{ij}(Q_{m_1,n_1}, Z_{m_2,n_2})$ between two image blocks is computed based on L_1 distance measure (city block distance) as,

$$\lambda_{ij}(Q_{m_1,n_1}, Z_{m_2,n_2}) = 1 - \frac{1}{2} \sum_k |T_k(Q_{p_1,q_1}) - T_k(Z_{p_2,q_2})| \tag{6}$$

where $p_1 = m_1 + i, q_1 = n_1 + j, p_2 = m_2 + i, q_2 = n_2 + j$ and it is equivalent to color histogram intersection except that the bins have semantic interpretation as VisMed terms.

There is a trade-off between content symmetry and spatial specificity. If we want images of similar semantics with different spatial arrangement (e.g. mirror images) to be treated as similar, we can have larger tessellated block in SAM (i.e. the extreme case is a global histogram). However in applications such as medical images where there is usually very small variance in views and spatial locations are considered differentiating across images, local histograms will provide good sensitivity to spatial specificity. Furthermore, we can attach different weights to the blocks to emphasize the focus of attention (e.g. center) if necessary. In this paper, we report experimental results based on even weights as grid tessellation is used.

3 Experimental Evaluation

As part of the Cross Language Evaluation Forum (CLEF), the ImageCLEF 2004 track [17] that promotes cross language image retrieval has initiated a new medical retrieval task in 2004. The goal of the medical task is to find images that are similar with respect to modality (e.g. Computed Tomography (CT), Magnetic Resonance Imaging (MRI), X-ray etc), the shown anatomic region (e.g. lung, liver, head etc) and sometimes with respect to the radiologic protocol (e.g. T1/T2 for MRI (contrast agents alter selectively the image intensity of a particular anatomical or functional region)).

The dataset is called the CasImage database and it consists of 8725 anonymized medical images, e.g. scans, and X-rays from the University Hospitals of Geneva

(visit `www.casimage.com` for example images). Most images are associated with case notes, a written English or French description of a previous diagnosis for an illness the image identifies. The case notes reflect real clinical data in that it is incomplete and erroneous. The medical task requires that the first query step has to be visual (i.e. query by image example). Although identifying images referring to similar medical conditions is non-trivial and may require the use of visual content and additional semantic information in the case notes, we evaluate the VisMed approach on the medical task without using the case notes in this paper.

In the ImageCLEF 2004 medical retrieval task, 26 topics were selected with the help of a radiologist which represented the database well. Each topic is denoted by a query image (Fig. 3). An image pool was created for each topic by computing the union overlap of submissions and judged by three assessors to create several assessment sets. The task description of the 26 topics given to the assessors is listed in Table 1. The relevance set of images judged as either relevant or partially relevant by at least two assessors is used to evaluate retrieval performance in terms of uninterpolated mean average precision (MAP) computed across all topics using `trec_eval`. The sizes of the relevance sets for each topic are listed in the rightmost column in Table 1.

We evaluate the VisMed approach on the medical retrieval task of the ImageCLEF 2004 benchmark. We designed 40 VisMed terms that correspond to typical semantic regions in the CasImage database (Table 2). While the first 30 VisMed terms are defined on grey-level images, the last 10 VisMed terms are for the minority of color images. Note that "print-sketch" and "print-slide" refer to drawing and text in presentation slides respectively. With a uniform VisMed framework, dark background in the scan images (e.g. CT, MRI) and empty areas in drawing etc are simply modeled as dummy terms instead of using image preprocessing to detect them separately.

Based on 172 (approx. 2%) of the 8725 image collection, we cropped 1170 image regions with 20 to 40 positive samples for each VisMed term (Table 2). For a given VisMed term, the negative samples are the union of the positive samples of all the other 39 VisMed terms. The 172 images are selected to cover different appearances of the 40 VisMed terms. We ensure that they do not contain any of the 26 query images though Q06, Q15, Q16, and Q22 have a total of 7 duplications in the database.

The odd and even entries of the cropped regions are used as training and validation sets respectively (i.e. 585 each) to optimize the RBF kernel parameter of support vector machines. The best generalization performance with mean error 1.44% on the validation set was obtained with $C = 100, \alpha = 0.5$ [18]. Both the training and validation sets are then combined to form a larger training set to train a new set of 40 VisMed SVM detectors.

Both query and database images are indexed and matched using the framework as described in the previous section (Eq. (1) to (6)). However, to avoid spurious matching between very different grids (e.g. 3×1 and 1×3), we set the similarity to zero if the difference in a grid dimension between two image indexes

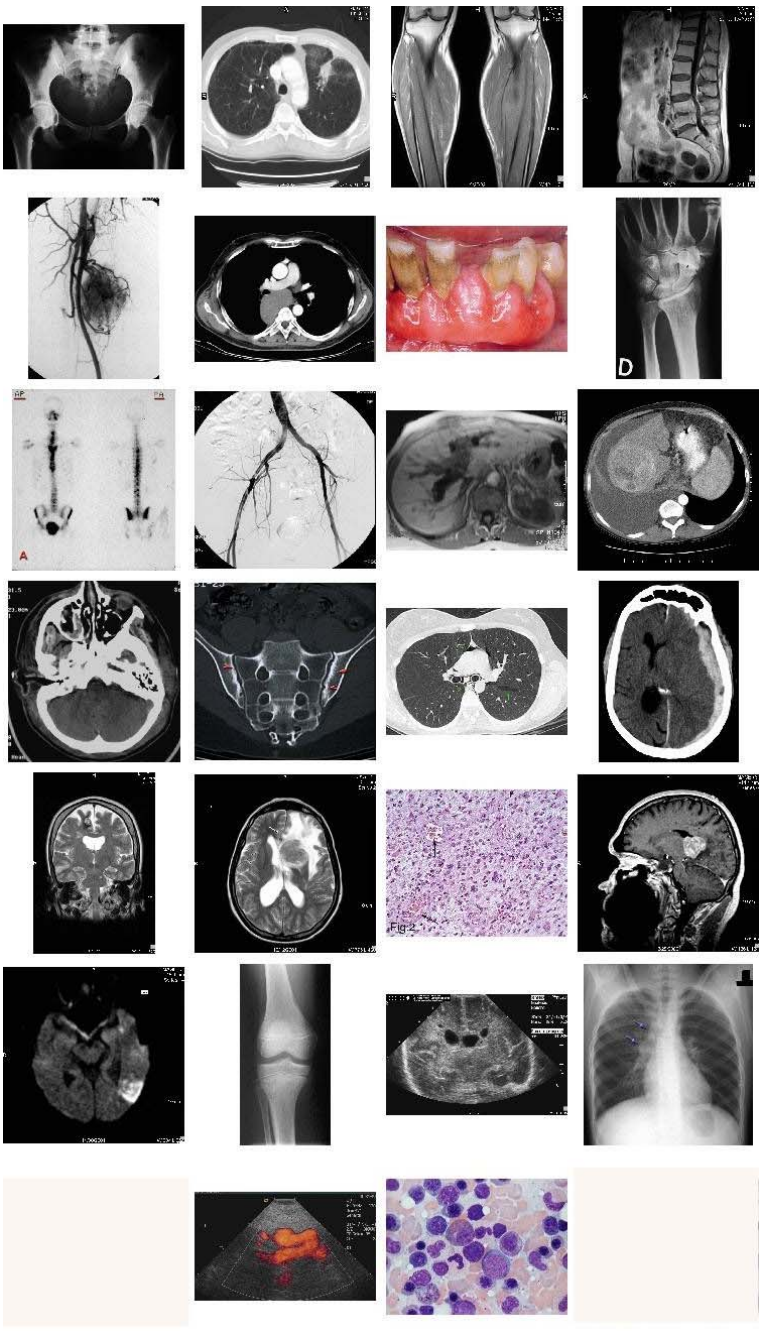

Fig. 3. Query images for the 26 topics

Table 1. Topic description for assessors and relevant set sizes

Query	Task Description	#
Q01	Frontal/Radiography/Pelvis	235
Q02	Axial/CT/Lung	320
Q03	Coronal/MRI/Legs	72
Q04	Sagittal/MRI/Abdomen & Spine	43
Q05	Arteriography/Contrast Agent	84
Q06	Abdominal CT Mediastin	252
Q07	Mouth photos showing Teeth	48
Q08	Radiography/Wrist	117
Q09	Szintigraphy/almost entire body	43
Q10	Arteriography/Contrast Agent	79
Q11	Axial/MRI/Liver	9
Q12	Abdominal CT/Liver	179
Q13	Axial/CT/Head with facial bones	95
Q14	Oblique cut/CT/Sacrum	11
Q15	Axial/CT/Lung	252
Q16	Horizontal/CT/Head, Cerebral	141
Q17	Coronal/MRI/Head/T2	31
Q18	Axial/MRI/Brain/T2	78
Q19	Histology/Cells/Color/Size	114
Q20	Sagittal/MRI/Head	27
Q21	Horizontal/MRI/Head/Diffusion	90
Q22	Frontal/Radiography/Knee Joint	171
Q23	Ultrasound/No colored parts	74
Q24	Frontal/Radiography/Thorax	409
Q25	Ultrasound/With colored parts	64
Q26	Hematology/Similar colors, size	53

Table 2. VisMed terms and numbers of region samples

VisMed Terms	#	VisMed Terms	#
arteriography-agent	40	xray-face	40
xray-neck-spine	30	xray-lung	40
xray-pelvis	40	xray-bone	40
xray-finger	30	xray-joint	40
xray-implant	30	ct-head-brain	20
ct-head-bones	20	ct-thorax-lung	30
ct-abdomen-mediastin	30	ct-abdomen-liver	30
ct-abdomen-intestine	20	ct-abdomen-sacrum	30
mri-head-brain	40	mri-head-bones	20
mri-head-face	30	mri-head-diffusion	30
mri-abdomen-spine	40	mri-abdomen-liver	30
mri-pelvis-tissue	40	mri-legs-coronal	40
ultrasound-grey	30	print-scintigraph	40
print-sketch	30	print-slide	20
print-blank	20	print-dark	20
pathology-pink	20	pathology-blue	20
pathology-purple-big	20	pathology-purple	40
pathology-brown	20	pathology-dark	20
ultrasound-color	20	mouth-teeth	20
mouth-tissue	20	mouth-lesion	30

is more than one. That is, two images are considered dissimilar if they exhibit very different aspect ratios. The MAP over 26 query topics is computed using the same `trec_eval` program used in ImageCLEF 2004.

Table 3 compares the MAPs of the automatic VisMed run with those of the top 5 automatic runs as reported in ImageCLEF 2004 [17] where the percentages of improvement are shown in brackets, "RF" stands for the use of pseudo relevance feedback and "Text" means the case notes were also utilized. The group "Buffalo", "imperial", and "aachen-inf" refer to State Univ. of New York (USA), Imperial College (UK), and Dept. Medical Informatics, RWTH, Aachen (Germany) respectively. All these systems used low-level visual features for image indexing and matching. Details are given in their working notes at `http://clef.isti.cnr.it/`. The best run in ImageCLEF 2004 was a manual run with a MAP value of 0.4214 by the University Hospitals of Geneva.

Table 3. Comparison of the VisMed approach with the top 5 automatic runs of ImageCLEF 2004

Group	Run ID	MAP (% up)	RF	Text
VisMed	vismed40	0.4156		
Buffalo	UBMedImTxt01	0.3488 (19.2)		X
imperial	ic_cl04_base	0.3450 (20.5)		
aachen-inf	i6-025501	0.3407 (22.0)		
aachen-inf	i6-qe02555010	0.3323 (25.1)	X	
Buffalo	UBMedImTxt02	0.3309 (25.6)		X

From Table 3, we conclude that the VisMed approach is very promising. It has attained a MAP of 0.4156, clearly an improvement over all the automatic runs in ImageCLEF 2004. the average precisions at top 10, 20, 30 and 100 retrieved images of the VisMed approach are 0.70, 0.65, 0.60, and 0.41 respectively (similar results for the runs of ImageCLEF 2004 are not available), which we consider reasonable for practical applications.

Representing a medical image as compact spatial distributions (i.e. regular grids of 40 dimensional vectors) of semantically meaningful terms, the VisMed approach also has the following advantages: enables efficient matching, provides explanation based on VisMed terms that are detected and matched, and supports new compositional queries expressed as spatial arrangement of VisMed terms [11].

4 Conclusion

Medical CBIR is an emerging and challenging research area. We have proposed a structured framework for designing image semantics from statistical learning. Using the ImageCLEF 2004 CasImage medical database and retrieval task, we

have demonstrated the effectiveness of our framework that is very promising when compared to the current automatic runs [17]. Indeed our adaptive framework is scalable to different image domains [11, 19] and embraces other design choices such as better visual features, learning algorithms, object detectors, spatial aggregation and matching schemes when they become available.

We reckon that a limitation of the current VisMed approach is the need to design the VisMed terms manually with labeled image patches as training samples. We have begun some work in a semi-supervised approach to discover meaningful visual vocabularies from minimally labeled image samples [20]. In the near future, we would also explore the integration with inter-class semantics [19] and other source of information such as text. Last but not least, we would also work with medical experts to design a more comprehensive set of VisMed terms to cover all the essential semantics in medical images.

Acknowledgments

We would like to thank Mun-Kew Leong and Changsheng Xu for their feedback on the paper. We also thank T. Joachims for making his SVM^{light} software available.

References

1. Shyu, C., Pavlopoulou, C., Kak, A., Brodley, C.: Using human perceptual categories for content-based retrieval from a medical image database. Computer Vision and Image Understanding **88** (2002) 119–151
2. Muller, H., Michoux, N., Bandon, D., Geissbuhler, A.: A review of content-based image retrieval systems in medical applications – clinical benefits and future directions. Intl. J. of Medical Informatics **73** (2004) 1–23
3. Dy, J., Brodley, C., Kak, A., Broderick, L., Aisen, A.: Unsupervised feature selection applied to content-based retrieval of lung images. IEEE Trans. on PAMI **25** (2003) 373–378
4. Liu, Y., et al.: Semantic based biomedical image indexing and retrieval. In Shapiro, L., Kriegel, H., Veltkamp, R., eds.: Trends and Advances in Content-Based Image and Video Retrieval. Springer (2004)
5. Lehmann, T., et al.: Content-based image retrieval in medical applications. Methods Inf Med **43** (2004) 354–361
6. Armato III, S., et al.: Lung image database consortium: developing a resource for the medical imaging research community. Radiology **232** (2004) 739–748
7. Wood, B.: Visual expertise. Radiology **211** (1999) 1–3
8. Vapnik, V.: Statistical Learning Theory. Wiley, New York (1998)
9. Lim, J., Jin, J., Luo, S.: A structured learning approach to semantic photo indexing and query. In: Proc. of AIRS 2005. (2005) (accepted)
10. Lim, J.: Building visual vocabulary for image indexation and query formulation. Pattern Analysis and Applications **4** (2001) 125–139
11. Lim, J., Jin, J.: A structured learning framework for content-based image indexing and visual query. Multimedia Systems Journal (2005) to appear.

12. Bishop, C.: Neural Networks for Pattern Recognition. Clarendon Press, Oxford (1995)
13. Manjunath, B., Ma, W.: Texture features for browsing and retrieval of image data. IEEE Trans. on PAMI **18** (1996) 837–842
14. Rui, Y., Huang, T., Mehrotra, S.: Content-based image retrieval with relevance feedback in mars. In: Proc. of IEEE ICIP. (1997) 815–818
15. Sung, K., Poggio, T.: Example-based learning for view-based human face detection. IEEE Trans. on PAMI **20** (1998) 39–51
16. Papageorgiou, P., Oren, M., Poggio, T.: A general framework for object detection. In: Proc. of ICCV. (1998) 555–562
17. Clough, P., Sanderson, M., Muller, H.: The clef cross language image retrieval track (imageclef) 2004. http://clef.isti.cnr.it/ (2004)
18. Joachims, T.: Making large-scale svm learning practical. In Scholkopf, B., Burges, C., Smola, A., eds.: Advances in Kernel Methods - Support Vector Learning. MIT-Press (1999) 169–184
19. Lim, J., Jin, J.: Combining intra-image and inter-class semantics for consumer image retrieval. Pattern Recognition **38** (2005) 847–864
20. Lim, J., Jin, J.: Discovering recurrent image semantics from class discrimination. EURASIP Journal of Applied Signal Processing (2005) to appear.