Histology Image Retrieval in Optimised Multi-Feature Spaces

Qianni Zhang and Ebroul Izquierdo, *Senior Member, IEEE,*

***Abstract*—Content based histology image retrieval systems have shown great potential in supporting decision making in clinical activities, teaching, and biological research. In content based image retrieval, feature combination plays a key role. It aims at enhancing the descriptive power of visual features corresponding to semantically meaningful queries. It is particularly valuable in histology image analysis where intelligent mechanisms are needed for interpreting varying tissue composition and architecture into histological concepts. This paper presents an approach to automatically combine heterogeneous visual features for histology image retrieval. The aim is to obtain the most representative fusion model for a particular keyword that is associated to multiple query images. The core of this approach is a *multi-objective learning* method, which aims to understand an optimal visual-semantic matching function by jointly considering the different preferences of the group of query images. The task is posed as an optimisation problem, and a multi-objective optimisation strategy is employed in order to handle potential contradictions in the query images associated to the same keyword. Experiments were performed on two different collections of histology images. The results show that it is possible to improve a system for content based histology image retrieval by using an appropriately defined multi-feature fusion model, which takes careful consideration of the structure and distribution of visual features.**

***Index Terms*—Content based image retrieval, histology image retrieval, feature fusion, multi-objective optimisation**

I. Introduction

In biology and medicine, histology is a fundamental tool that provides information on structure and composition of tissues at microscopic level. Nowadays, images of tissue slides are often digitized to document procedures and to support findings. These collections are often huge in size and thus hide a latent source of information that can be greatly exploited if suitable mechanisms are available for accessing the data [28]. Thus, a technology that can retrieve histological images according to given queries can potentially be a very useful tool for data archiving and analysis, teaching and training, assisting decision making in diagnosis and so on. When more complex systems are being considered, for instance, a system which provides suggestions on diagnosis based on existing histopathology databases, histological image retrieval can be employed as an essential component for initial search and indexing of content using carefully selected keywords, such as a basic tissue type.

Conventional medical image retrieval systems often rely on tags associated to images in the databases. However, text-based approaches are often limited in practice since tags may be both expensive and ambiguous. This is because generating manual annotation on images is extremely time-consuming, highly subjective and requires a good level of domain-related knowledge. Another kind of approaches exploit knowledge databases like Unified Medical Language Systems [22], but they rely on the availability of knowledge databases and their relativity to the application domain.

Alternatively, content-based image retrieval (CBIR) systems apply computer vision techniques to the image retrieval problem by analysing the actual contents of the image rather than semantic features such as keywords or tags. Recent research in CBIR indicates that such systems are capable of retrieving medical images according to its domain-specific image features [14]. Thus such systems form an important alternative and complement to traditional text-based retrieval systems. In image retrieval systems, given a query image, or a keyword query that is associated to a group of representative images, the goal is to retrieve from a reference library, the similar images whose semantic meaning is as close to the query as possible, regardless of its visual appearance. Although extraction algorithms for low-level image features are well-understood and able to capture subtle differences between colours, statistic and deterministic textures, global colour layouts, dominant colour distributions, etc., the link between such low-level primitives and high-level semantic concepts remains an open problem [31],[41]. This problem is referred to as 'the semantic gap'.

To alleviate the semantic gap problem, there have emerged many systems aiming at applying content-based approaches in more sophisticate ways in medical image retrieval [1], [2],[17], [19], [20], [37], [39]. Many diagnostic imaging modalities, such as X-ray, computed tomography (CT), magnetic resonance imaging (MRI) are currently available and routinely used to support clinical decision making, diagnosing, state of an illness-tracking, medical education and research, etc. Applications of content-based approaches in medical image retrieval have shown benefits in all these procedures. In addition to general image

Q. Zhang and E. Izquierdo are with School of Electronic Engineering and Computer Science, Queen Mary University of London, Mile End Road, London E1 4NS, UK. e-mail: {qianni.zhang, ebroul.izquierdo}@elec.qmul.ac.uk (see http://www.elec.qmul.ac.uk/mmv/people/qianniz.htm; http://www.elec.qmul.ac.uk/people/ebroul/index.htm).

retrieval systems applied to medical images, there are also many specialised image retrieval systems developed to enable retrieval of various specific kinds of medical images, such as breast cancer biopsy slides [40], positron emission tomographic functional images [5], ultrasound images [9], different types of pathology images [42], radiographic images [17], and histology images [8], [37]. More interestingly, the ImageCLEF medical image retrieval task targets modality classification of medical images and have attracted wide attention from the area in recent years [27].

Among these application domains, histology image retrieval has been an active research topic for modelling visual similarity measures and retrieving tissue slides in some semantic categories. In previously reported works on histology image classification and retrieval, accuracy is often limited due to the unreliable outputs from the feature metrics [14],[23]. This is because the systems often rely on single low-level features, which are not always capable of interpreting visual objects into varying and complicated semantic meanings. To tackle this problem, the combination of low-level features for semantic image retrieval has been widely considered in the literature. In [42], a system was proposed for retrieving histology images from prostate, liver and heart tissues, based on four different visual characteristics. The work in [38] described a system to index histology images of gastro-intestinal tract, by categorising image blocks into semantic classes based on local visual patterns. The approach proposed in [29] uses a boosting algorithm based on multiple distance measures computed on a fixed set of features to retrieve and classify breast histology slides. These works all intended achieving better performance in histology image understanding by employing heterogeneous visual features. However, the problems in feature combination are likely underestimated, when feature combination is done by simple concatenation of feature vectors.

Such approaches are called *early fusion*, and they share two main common problems. First, methods on concatenation of feature vectors can easily result in high dimensional feature space and thus suffer from "the curse of dimensionality" [10], [35]. Second but most importantly, different image features often have their own structures, distributions and metric spaces. Direct concatenation of feature vectors could result in meaningless similarity measures. For instance, feature histograms are usually compared using a similarity measure for probability distributions while feature vectors should be matched using Euclidean metrics. In addition, even if two different features are being compared with the same metric, their scale, domain and distribution may be completely different due to the intrinsic descriptor nature [16].

In order to avoid these problems, in this paper a *late-fusion* strategy is followed to combine low-level features for histology image retrieval. Similar approaches have been proposed in the literature, such as in [6], [15]. However, such approaches have one common disadvantage that they assume equal importance of features in fusion. In this way the performance of 'good' and 'bad' features is averaged and thus could not provide great improvements. Some other approaches assign weights to different features. For instance, genetic algorithm is used for finding the optimisation weights for features in image retrieval in [13]. This approach considers feature fusion independent of query, while it is not possible to expect the obtained fusion model to be optimal to all possible queries. In [7], multi-feature fusion was achieved using an optimal combination of multiple kernel functions. Kernel functions were fused using a weighted linear combination, whose weights were found by an optimization process that maximizes the correlation between low-level features, represented by kernel functions, and high-level semantic concepts. Another important direction is to rely on human interactions to obtain query specific weights [18]. Such approaches might improve the effectiveness of weighting factors, but they still require heavy workload and specific expertise from the user.

In this paper, we propose a histology image retrieval method aiming at retrieving images with relevant semantic meanings based on visual content. In each query process, the user first has in mind a semantic term that can be associated to a specific keyword, and then uses multiple query images that represent the semantic term for retrieving more images that are relevant to the query on the semantic level. In the proposed method, we use multiple images, namely, a *representative query group*, in the CBIR process for representing each semantic term. This is because, for a particular keyword, it is usually tricky to find a query image as an ideal visual representation. In [1], a multitiered CBIR system was proposed targeting microscopic images, enabling both multi-image query and slide-level image retrieval. In this work, the focus was to adopt multiple image queries in one retrieval process using the weighting terms. In comparison, we also consider using multiple image queries to represent each keyword, but the main focus of the proposed approach is to derive a fusion model for heterogeneous visual features.

In addition, for each keyword, a different representative query group is used and a unique feature fusion model is derived. This is because each semantic term is considered to have its own characteristic visual pattern and the feature fusion models should be keyword-specific. The proposed approach is able to automatically learn the relative importance of each feature space corresponding to the keyword from its associating representative query group. The learning of a suitable feature fusion model is posed as an optimisation problem. The optimisation is carried out using a Multi-Objective Learning (MOL) method, which involves a multi-objective optimisation (MOO) strategy [36]. The main advantage in the MOL method is that it is able to find a multi-feature model that can simultaneously encapsulate different aspects of the most representative visual patterns for each concept, without however assigning fixed relevance factors to each visual feature.

Without losing generality, several different texture features are considered in our experiments. Although, the

proposed method can be applied to more histological terms, the four fundamental tissue types in biology - *connective*, *epithelial*, *muscular* and *nervous* tissues - are considered in our experiments as the keywords for retrieval. It is demonstrated that, using the proposed approach, it is possible to retrieve histological images according to their semantic relevance by using properly combined visual features. The proposed approach has been tested in two collections of histology images. The first set contains over 20000 images, among which 2,828 images have manual labels of the four fundamental tissue types. The second smaller set contains around 442 histology images, and 130 of them have manual labels on those four tissue types.

The remainder of this paper is organised as follows: Section 2 presents the visual analysis steps in histology databases, including feature extraction, distance calculation and normalisation. Section 3 introduces the MOL method for feature fusion towards a multi-feature based retrieval in histology image databases. The evaluation procedure and experimental results are presented in Section 4 and Section 5 discusses some concluding remarks.

II. Visual feature extraction and analysis

The design and evaluation of an image retrieval system rely on properly defined visual features with suitable similarity matching metrics as well as correct normalisation functions. Without normalisation of feature spaces, comparison between different features becomes misleading and their combination is meaningless. Therefore in this section, we present our work on extracting several commonly used visual features, analysing their characteristics and deriving their normalisation functions.

A. Feature extraction

The main goal of the proposed research is to develop an approach for automatically combining low-level visual features for retrieval of histology images according to their fundamental tissue types. Therefore, the extraction and analysis of useful visual features in histology images is an essential step.

In the literature many well designed low-level visual features have been proposed, describing visual content from different perspectives including colour [26], texture [32], shape [34], local feature points [25], etc. It is worth noting that, the focus of this paper is on discussing how to obtain suitable fusion models of features. The proposed multi-feature combination approach is independent of selected features or their distance metrics. Evaluation of different features is out of the scope of this research.

Due to the nature of this specific dataset of histology images, texture features are suitable for analysing their visual patterns. In this research we selected several commonly used texture features together with architectural features due to their prominent characteristics for histology image analysis [4], [24]. Without losing generality, the eight features selected here to describe histology image contents are listed below. Unless specified otherwise, these features are extracted from 3x3 blocks of an image and then concatenated into one feature vector for that image.

(1) Gabor Textures (GT): Gabor filters possess outstanding ability of filtering in the spatial and frequency domain. The Gabor transform is a set of directional filters, thus it is shift invariant. To calculate a GT feature, a convolution with a Gaussian harmonic function is used, and 7 different frequencies, $freq = [1, 2, ..., 7]$, are considered to compute 7 descriptor values per block. As a result, the Gabor descriptor is composed of 63 descriptors.

(2) Tamura Textures (TT): Tamura proposed six texture features corresponding to human visual perception: *coarseness, contrast, directionality, line-likeness, regularity,* and *roughness*. From experiments testing the significance of these features with respect to human perception, it was concluded that the first three features are very important. In our experiments, two statistics are calculated for *contrast, directionality* and *coarseness*, providing 6 descriptors in each block. As a result, the Tamura descriptor has 54 descriptors.

(3) Zernike Moments (ZM): Zernike moments have many desirable properties, such as rotation invariance, robustness to noise, expression efficiency, fast computation and multi-level representation for describing the shapes of patterns. In this paper, the absolute values of the coefficients of the Zernike polynomial approximation are computed per block, providing 72 descriptors in each region. Then, the Zernike descriptor has 648 bins.

(4) SIFT-based dictionary (SIFT): SIFT feature is known for its ability in handling intensity, rotation, scale and affine variations. A histogram of SIFT converts each patch to 128-dimensional vector. Thus in this paper, each block in the process is represented by the rotation-invariant feature descriptor, using a histogram of 128 bins.

(5) DCT dictionary (DCT): DCT histograms are invariant to translation and rotation. In this paper, each block is represented by the coefficients of the Discrete Cosine Transform, applied to each channel of the RGB colour space. The 21 most significant coefficients per channel are preserved. In this way, the dictionary of patterns will have colour information as well.

(6) Gray-Level Co-Occurrence Matrix (GLCM): GLCM textures are obtained by a tabulation of how often different combinations of pixel brightness values occur in an image. When the co-occurrence matrix is formed using a set of offsets sweeping through 180 degrees (i.e. 0, 45, 90, and 135 degrees) at the same distance, it is able to achieve a degree of rotational invariance. A number of statistics can be derived from the co-occurrence matrix, which are calculated in the adjacency of one pixel in each of the four directions (horizontal, vertical, left and right diagonal). Four of them are considered - contrast, correlation, energy and homogeneity - to form a feature vector of four dimensions.

(7) MPEG-7 Edge Histogram (EH): EH describes the local edge distribution of an image. The descriptor is scale invariant and supports rotation invariant and rotation sensitive matching operations. It is obtained by first dividing

an image into 4x4 sub-images and then calculating the local-edge histogram bins. Edges in the 16 sub-images are categorised into five types - vertical, horizontal, diagonal 45 degrees, diagonal 135 degrees and non-directional - forming a histogram of 80 bins.

(8) MPEG-7 Homogeneous Textures (HT): A HT descriptor provides a quantitative representation using 62 numbers, including the image intensity average, standard deviation of the image pixels, energies of the 30 partitioned frequency channels based on the human visual system, and energy deviations of these 30 channels. The 30 partitioned channels ensure a scale and rotation-invariant description and matching of texture. To extract these numbers from an image, the image is first filtered with a bank of orientation and scale tuned filters using Gabor filters. The first and the second moments of the energy in the frequency domain , i.e., energies and energy deviations, in the corresponding channels are then used as the components of the texture descriptor.

B. Distance calculation and normalisation

In the next step, feature distances are calculated using a specifically defined distance metric of each feature space. The details of distance metrics used in this research are given in Section IV.

Then all the obtained distances are normalised. The goal of normalisation is to guarantee the appropriateness of comparing different measurements that differ in scale and domain, while preserving the underlying characteristics of the data. In this research, distance metrics computed from different image features are normalised based on the fitted probability density functions for their corresponding feature spaces.

Let d' be the distance value computed from a particular image feature. The statistical normalisation is computed as:

$$d = \frac{(d' - \mu)}{\sigma}, \tag{1}$$

where d is the normalised distance value and μ and σ are the mean and standard deviation of the underlying distance distribution.

The critical problem in deriving appropriate normalisation function is to precisely estimate the distance distribution of the feature space. The distribution of feature distances is highly dependent on the structure of the feature and image content. One can assume a normal distribution of the distances to estimate μ and σ without any further analysis. However, the true distribution of distances might be more precisely approximated using other Probability Distribution Functions (PDFs). Therefore in this research, six types of PDFs are considered, including Normal, Gamma, Laplace, Log-norm, Rayleigh and Exponential. These PDFs are denoted as $P_k, k = 1, 2, ..., 6$. Based on distance samples derived from the database, the parameters Θ_k of each possible P_k are estimated to determine whether the data is being drawn from the associated PDF or not.

TABLE I
PROBABILITY DISTRIBUTION FUNCTIONS IN THE SET OF POSSIBLE APPROXIMATIONS FOR EACH DISTANCE DISTRIBUTION.

Distribution	$\Theta =$	PDF	Mean	Standard deviation
Normal	(μ, σ)	$\frac{1}{\sqrt{2\pi\sigma^2}} exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$	μ	σ^2
Gamma	(k, θ)	$x^{k-1}\frac{exp(-x/\theta)}{\Gamma(k)\theta^k}$	$k\theta$	$k\theta^2$
Laplace	(μ, b)	$\frac{1}{2b} exp\left(-\frac{\lvert x-\mu\rvert}{b}\right)$	μ	$2b^2$
Log-norm	(μ, σ)	$\frac{1}{x\sqrt{2\pi\sigma^2}} exp\left(-\frac{(ln\, x-\mu)^2}{2\sigma^2}\right) e^{\mu}$	$\left(e^{\sigma^2}-1\right) e^{2\mu+\sigma^2}$	
Rayleigh	(σ)	$\frac{x}{\sigma^2} exp\left(\frac{-x^2}{2\sigma^2}\right)$	$\sigma\sqrt{\frac{\pi}{2}}$	$\frac{4-\pi}{2}\sigma^2$
Exponential	(λ)	$\lambda exp\,(-\lambda x)$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$

Finally, to select the best distribution approximation for the underlying data, the Kullback-Leibler (KL) divergence is evaluated between the histogram of actual distances and the estimated PDF. For two distributions P and Q, the KL divergence between them is estimated by:

$$D_{KL}(P||Q) = \int_{-\infty}^{\infty} p(x) log \frac{p(x)}{q(x)} dx, \tag{2}$$

where p and q denote the densities of P and Q. The procedure for distance normalisation can be summarised as the following steps:

1) For each $P_k, k = 1, 2, ..., 6$, estimate the parameters Θ_k using the data samples in distance matrix $\mathbf{D}$.
2) Build a histogram H of distances with m bins using the data samples in $\mathbf{D}$.
3) For each bin in the histogram H, calculate the value of P_k in $\tilde{P_k}$ to approximate the shape of H.
4) Calculate the KL-divergence between the histogram H and the approximation $\tilde{P_k}$.
5) Select P_k with the minimum value of KL divergence as the best fit PDF.

Table I shows the six PDFs considered in this work for approximation of distance distributions. Notice that for each PDF, the normalisation parameters μ and σ are calculated in different ways using the estimated parameters Θ, which are computed from the sample according to the corresponding rules.

For the remaining of this paper, when distance values are mentioned, they have been normalised following the steps described in Section II.

III. MULTI-FEATURE BASED HISTOLOGY IMAGE RETRIEVAL

The proposed approach to multi-feature based histology image retrieval relies on the MOL method, that is able to automatically learn a suitable multi-feature model from a representative group containing multiple query images as a visual representation for the keyword.

A. Representative query images

To define a suitable combination model for each keyword related to the histology database, multiple query images are employed in a query process. Due to the complex and varying visual appearance in different keywords, it is unrealistic to assume there exist a single image query giving an optimal representation of a semantic keyword. Therefore, in this paper, we use a group of query images, referred to as the *representative query group*, to approximate a suitable representation for one keyword. For a given keyword, let us use $R \subset \mathcal{G}$ to denote a representative group, where $\mathcal{G}$ is the complete image set. To improve the discriminative power of the low-level features, two kinds of representative query samples are considered. R^+ contains the most relevant samples for the corresponding keyword, referred to as *positive group*; R^- represents *negative group* in which the samples are irrelevant to the keyword of concern but may look similar to the positive query samples; $R = R^+ \cup R^-$. If new histology keywords are added or the database is populated with new images in new concepts, new representative query groups for the incoming concepts need to be generated.

Assume a total number of n feature spaces are considered. Having the representative group R ready, we can calculate the centroid $\bar{v}_j$ of representative group R in a given feature space $F_j, j = 1, 2, ..., n$. F_j could be any feature space described in Section II, or any other suitable visual features in the literature. $\bar{v}_j$ in feature space F_j is calculated by finding the sample in positive group R^+ with the minimal sum of distances to all other positive samples in R^+. Let $v_{i,j}$ and $v_{k,j}$ be the feature vectors extracted from positive samples r_i and r_k in feature space F_j, $i, k \in [1, |R^+|]$, and $d(v_{i,j}, v_{k,j})$ be the feature metric estimating the distance between these two images in feature space F_j. We note that $\bar{v}_j$ actually equals to the feature vector in F_j extracted from one of the positive samples in R^+. The centroid of representative group R in F_j can be defined as:

$$\bar{v}_j = argmin_{\{v_i, i \in [1,|R^+|]\}} \{ \sum_{k \in [1,|R^+|]} d(v_{i,j}, v_{k,j}) \}. \quad (3)$$

Taking $\bar{v_j}$ as an anchor, for a given image g_i in $\mathcal{G}$, where $i \in [1, |\mathcal{G}|]$, the distance from g_i to the centroid $\bar{v}_j$ of feature space F_j can be calculated as:

$$\bar{d}_{i,j} = d(\bar{v}_j, v_{i,j}). \quad (4)$$

For the set of n feature spaces $\{F_j | j = 1, 2, ..., n\}$, all the centroids across different feature spaces form a particular set of vectors $\bar{V} = \{\bar{v_1}, \bar{v_2}, ..., \bar{v_n}\}$, in which each $\bar{v_j}$ is the centroid vector of feature space F_j. In general, $\bar{V}$ is referred to as the *generalised centroid* of representative group R, since it does not necessarily attach to a positive query sample in R. Note that $\bar{V}$ is always calculated considering only positive samples. This is because the negative samples can be randomly scattered in the metric space and calculating the generalised centroid by taking them into account would be meaningless.

For a representative query group R with m image samples for a particular keyword, a distance matrix $\bar{M}$ of size $m \times n$ can thus be constructed.

$$\bar{M} = \begin{matrix} \bar{d}_{1,1} & \bar{d}_{1,2} & \cdots & \bar{d}_{1,n} \\ \bar{d}_{2,1} & \bar{d}_{2,2} & & \bar{d}_{2,n} \\ \vdots & & \ddots & \vdots \\ \bar{d}_{m,1} & \bar{d}_{m,2} & \cdots & \bar{d}_{m,n} \end{matrix} \quad (5)$$

Each element $\bar{d}_{i,j}$ (in row i and column j) in $\bar{M}$ is the distance from representive sample r_i to centroid $\bar{v}_j$ in feature space F_j. In this way, the keyword is represented by a distance matrix covering multiple feature spaces.

B. Optimisation of multi-feature model and histology image retrieval

The aim of MOL method is to define a suitable multi-feature model for the visual representation of a specific histological keyword. The core of this method is a learning process towards an optimal combination model by assigning each involved low-level feature space F_j a proper weight α_j. This can be achieved by optimising an objective function or a set of objective functions for variable α. Since several representative samples are used for a good visual representation of a keyword, the interest of each single query sample may conflict with others. Thus we construct an objective function for each query sample in R, and use a multi-objective optimisation strategy to find a solution that can achieve a common optimum for all these functions.

Based on the distance matrix $\bar{M}$ given in (5), a set of objective functions can be constructed for the optimisation of multi-feature model. Each objective function is formed as weighted linear combinations of feature-specific distances. Considering $\bar{M}$, a total number of m objective functions can be constructed:

$$D(A) = \left\{ \begin{array}{c} D_1 = \sum_{j=1}^{n} \alpha_j \bar{d}_{1,j}, \\ D_2 = \sum_{j=1}^{n} \alpha_j \bar{d}_{2,j}, \\ \vdots, \\ D_m = \sum_{j=1}^{n} \alpha_j \bar{d}_{n,j}. \end{array} \right\} \quad (6)$$

The MOL method seeks to learn from the representative group, a suitable set of weights $A = \{\alpha_1, \alpha_2, ..., \alpha_n\}$, subject to the constraint: $\sum_{j=1}^{n} \alpha_j = 1$. The problem of learning a multi-feature fusion model is now transformed to finding a solution that optimises each of these objective functions in (6).

Generally speaking, an optimum is usually defined as the maximum or minimum of some objective(s). The optimal solution $\hat{A}$ should lead to the maximal or minimal value of the objective function(s) in $D(A)$ among all possible scalar combination of A that satisfy the constraints. In the proposed MOL method, the optimum is regarded as the minimum of objective functions for positive samples in R^+ and maximum of objective functions for negative

samples in R^-. The problem of learning a multi-feature fusion model is now posed as to find a solution that optimises each of these, in some cases contradicting, objective functions. Observe that different representative query samples may display different visual characters but these differences need to be harmonised via the simultaneous optimisation of multiple objectives corresponding to different representative samples. This optimisation process for fusion model learning is achieved using the MOO strategy. The MOO strategy is able to find a general optimum across potentially conflicting objectives by taking the interest of each single objective into account. Thus it is widely used in real-life optimisation problems [36].

While various algorithms have been developed using MOO strategy, in this paper Pareto Archived Evolution Strategy (PAES) is adopted as the MOO algorithm to optimise the combination models for the following reasons. Studies have been carried out to compare PAES with other two well-known and respected multi-objective genetic algorithms - the Niched Pareto Generic Algorithm and the Nondominated Sorting Genetic Algorithm in [21]. Their results provided strong evidence that PAES performed consistently well on a range of multi-objective optimisation tasks. It was also shown that PAES required fewer comparison processes to perform selection and acceptance, and this claim was empirically evidenced by the timing of experiments. This algorithm usually generates a set of potential Pareto optimal solutions $\Phi = \{A_1, A_2, A_3, ...\}$. Thus, a second step is required to decide which one of these solutions is the most suitable or feasible. The meaning of optimum in our specific task can be described as to find the 'optimal' multi-feature distance in which all the points representing the positive samples in the target multi-feature space are closely gathered around the generalised centroid while the points for the negative samples are randomly scattered around the generalised centroid. Thus, a sensible selection criterion can be defined as to find the $\hat{A}$ that satisfies:

$$\hat{A} = argmin_{A\in\Phi} \frac{\sum_{j\in[1,|R^+|]} D_j(A)}{\sum_{i\in[1,|R^-|]} D_i(A)}. \quad (7)$$

For a particular tissue type, an optimal multi-feature distance fusion model of an image $g_i \in \mathcal{G}$ can be obtained:

$$\hat{D}(g_i) = \sum_{j=1}^{n} \hat{\alpha}_j \bar{d}_{i,j}, \quad (8)$$

where $\hat{\alpha}_j \in \hat{A}$. According to these multi-feature distances, histology images can be ranked and retrieved based on their multi-feature distances with respect to the query keyword.

C. Why multi-objective optimisation?

Visual descriptors are different in nature and may have conflicting interest when they are jointly used to represent a semantic term. In order to take into consideration the preferences of all different features, in this paper we consider estimating a set of weighting factors for the feature spaces according to their relevance to the representative query group. The core of the proposed method for deriving a concept-specific fusion model is the MOO strategy.

There are similar works in the literature using various optimisation algorithms for obtaining multi-feature fusion models, such as in [11], the downhill simplex method is used to optimise a single objective function. A voting kNN rule is used to derive an objective function for optimisation. However, in this case the objective function obtained is the result of joining the interests of different samples into a single one. The optimised model only reflects the preference of the overall objective, without considering to satisfy the preference of any individual sample.

Similarly in our case, to obtain suitable weighting factors for the fusion model, one might consider constructing an overall objective function out of (6) and simply optimise it. One possible approach can be to sum up all the functions in (6):

$$D'(A) = \sum_{i=1}^{m} D_i = \sum_{j=1}^{n} \alpha_j (\sum_{i=1}^{m} \bar{d}_{i,j}). \quad (9)$$

The task now is to find the set of weighting factors that minimize the objective function (9):

$$A = argmin_{A=\{\alpha_1,\alpha_2,...,\alpha_n\}} \left\{ \sum_{j=1}^{n} \alpha_j (\sum_{i=1}^{m} \bar{d}_{i,j}) \right\}. \quad (10)$$

For the particular representative group, the values of the term $\sum_{i=1}^{m} \bar{d}_{i,j}, j \in \{1, 2, ..., n\}$ are determined. Let $F_k, k \in \{1, 2, ..., n\}$ be the feature space in which the sum of distances from each representative sample to the centroid $\bar{v_k}$ is smaller than those in other feature spaces: $\sum_{i=1}^{m} \bar{d}_{i,k} \leq \sum_{i=1}^{m} \bar{d}_{i,j}$, where $k \in \{1, 2, ..., n\}, j \in \{1, 2, ..., n\} \setminus \{k\}$. The minimal value of (9) can then be achieved by a particular $\bar{A}$, in which each α has the following values:

$$\alpha_x = \begin{cases} 1, & x = k \\ 0, & x \neq k \end{cases} \quad x \in \{1, 2, ..., n\} \quad (11)$$

In this case, the resulting weight factors $\bar{A}$ indicate that all the credit is given to one of the feature space F_k and all other features are not taken into account at all. In the end the task becomes a selection of one 'best' feature rather than fusion of multiple features. For one particular representative group and the represented keyword, it might be possible to say F_k is the 'best'. However, it may not be the case when different testing images are present. In most cases, properly defined feature fusion model may produce better retrieval performance compared to using any of the single features.

In comparison to the single objective based optimisation schemes, the advantage of employing the proposed MOL method is that, each representative sample and its corresponding objective function are treated separately in an

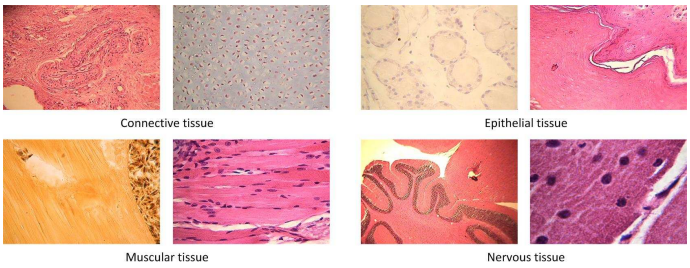


Fig. 1. Image samples of tissue types.

optimisation process. The interest of each representative sample is taken into account while the overall interest is being satisfied. MOO is used to find the solution that can achieve a balanced local optimum for each objective, without compromising the other objectives. The obtained results are regarded as the 'general optimum' for all objectives.

IV. Experiments

The first histology image set used for experiments in this research was the 'BiMed' database [3]. The experiment carried out on this dataset is referred to as 'experiment I' in the following text. This dataset contains about 20,000 histology image samples. The aim of this database was to support academic and research activities in biology by allowing medical students, doctors and researchers in biomedical field to access a wide variety of microscopy images in the four fundamental tissue types of living beings: *connective*, *epithelial*, *muscular* and *nervous*. These four tissue types are considered as keywords and in a query process, each of them is represented by a group of query images. A few examples of each tissue type in the database are shown in Figure 1, illustrating different biological structures and visual characters of the four types. Images in this collection have been acquired from tissue slides taken from different mice organs including brain, liver, heart, lung, kidney and skin among others. All samples were drawn from healthy specimens. Tissue slides were prepared using different staining methods including Hematoxylin and Eosin and Immunohisto chemical procedures. In addition, different zoom factors were used to acquire digital images according to the structure of interest. Most samples in this dataset were un-annotated, making these samples inaccessible using textual-based search methods. However, a portion of these images were annotated by expert biologists. This annotation contains information on some particular structures, organ, system and fundamental tissue. Thus, although experiments were performed on the whole dataset, we selected a subset of 2,828 image samples with full annotation on the four fundamental tissues, for evaluation of the proposed method. Five texture features were extracted from the images in the BiMed database and employed in experiment I. These features are GT, TT, ZM, SIFT and DCT.

The second dataset used for testing and validation purposes was the 'Blue Histology' images [33]. This set contains 442 histology images which were fully annotated, among which, 130 were labelled with the keywords of four fundamental tissue types: *connective*, *epithelial*, *muscular* and *nervous*. The experiment performed on this second set is called 'experiment II' in the following of the paper. Experiments were performed on all 442 images in this dataset, but the evaluation was conducted only against those 130 images with labels on the four tissue types. As explained before, the proposed approach is supposed to be independent of employed features. In order to prove this, in experiment II, a different feature bank was used, including GT, TT, GLCM, EH and HT. The aim of performing a second experiment with this dataset is to demonstrate the validity of the tested approach when database is different or a different feature bank is used.

A. Experiment I

1) Feature distance estimation: For experiment I, a total number of five texture features were extracted from the images in BiMed database. These features are GT, TT, ZM, SIFT and DCT.

Among the five features in experiment I, GT, TT and ZM are feature vectors. Each of these three feature vectors were computed per block in a 3×3 grid, leading to an image analysis in 9 different regions. Each feature vector was constructed by concatenating together the values computed in each block and preserving the spatial arrangement of the processed regions [30].

These feature vectors are evaluated using the Euclidean distance in the subsequent stages, which is computed as:

$$d_2(\mathbf{x}, \mathbf{y}) = \sqrt{\sum_{i=0}^{n} (x_i - y_i)^2}. \tag{12}$$

where $\mathbf{x}$ are $\mathbf{y}$ are two feature vectors in the one feature space, such as GT, TT or ZM, while x_i and y_i represent the feature numbers in the i_{th} dimension of that feature space.

The two histogram features, SIFT and DCT, were constructed using a bag-of-features approach, that may be considered as a texture analysis [12]. This strategy allows to estimate the presence of local patterns in images. First, a set of local patches or blocks are extracted from images and a local descriptor is computed for each of them. Then, a dictionary of patterns is constructed using a vector quantization algorithm to merge together patches with similar visual appearance. In our implementation, the k-means algorithm was used to cluster similar patches and to set cluster centroids as dictionary elements. Finally, a histogram is computed for each image, counting the occurrence of each element in the dictionary among the blocks extracted from the image. The most important parameters of this image representation are the selection of the local descriptor and the size of the dictionary. Two different strategies have been followed in this work, both using a dictionary size equal to 500 elements.

These feature histograms are evaluated using the Histogram Intersection measure:

TABLE II
NORMALISATION PARAMETERS

	PDF	Parameter I	Parameter II
DCT	Gamma	$k = 0.087$	$\sigma = 3124.743$
Gabor Textures	Log-norm	$\mu = -0.799$	$\sigma = 0.601$
SIFT	Normal	$\mu = 556.950$	$\sigma = 181.939$
Tamura Textures	Gamma	$k = 1.241$	$\sigma = 1712.942$
Zernike Moments	Log-norm	$\mu = -2.531$	$\sigma = 0.131$

$$d_{\cap}(\mathbf{x}, \mathbf{y}) = \sum_{i=0}^{n} min\{x_i, y_i\}, \tag{13}$$

where $\mathbf{x}$ and $\mathbf{y}$ are histograms and x_i and y_i are their corresponding i-th bins. This is a similarity measure instead of being a distance measure, i.e. the more similar two images are, the larger the score is.

2) Distance normalisation: As explained in Section II, an important step following calculating a distance is the normalisation of it. Original distances are estimated in each feature space using distance functions described above. Normalisation of features is a key step to guarantee the correctness of the derived multi-feature model. Since the distance distribution depends on the feature structure and image contents, the normalisation parameters are unknown *a priori*. The distribution of distances might be approximated using any PDF, whose parameters are estimated in different ways. To overcome this problem, we consider a set of six possible PDFs. Using the distance samples in the computed matrix, the parameters of each possible PDF are estimated in an attempt to match the true distribution of the distances. An approximation of the distribution is then calculated for each bin of the empirical histogram using the estimated parameters. To identify the best approximation to the distribution of distances, the PDF that minimises the KL-divergence score is selected.

Figure 2 shows the distribution of distances in each of the considered feature spaces, as well as the approximation with each of the PDFs considered in Table I. Using each PDF, a lowest possible KL-divergence score is calculated between the PDF approximation and the empirical histogram, as shown in Figure 2. For each feature space, the best fit PDF is the one with the lowest score among all considered PDFs. The best fit PDFs are indicated with a bold continuous curve, while the other PDFs are in dashed curves. The estimated parameters of the best fit PDF for each feature distribution are recorded for further normalisation purposes. For example, table II shows the estimated parameters for the best fit PDF for the feature distributions in experiment I.

3) Histology image retrieval: As mentioned before, in experiment I evaluation of retrieval performance was conducted against a subset containing 2,828 histology images with manual labels for the four tissue types. The subset contains 484 samples for *connective* tissue, 804 for *epithelial* tissue, 514 for *muscular* tissue and 1026 for *nervous* tissue. For evaluation, a 5-fold cross-validation scheme is used, in which the whole dataset is randomly divided into five equally sized groups. In each test, one of the five groups is used as the training set and the other four are used for testing. The positive representative group of each type of tissue contains 10 relevant samples that are randomly selected from the training set based on the ground-truth annotations. Using the retrieval results of each positive representative group, the corresponding negative representative group is selected as the first 10 retrieved non-relevant samples in the same training set. The performance measures presented include mean and standard deviation values across five folds in Average Precision (AP); R-Precision (R-prec), which is obtained at the point where precision and recall get the same value; and precision after the first 20 retrieved samples (Prec 20), as shown in Table III.

TABLE III
EXPERIMENT I: RETRIEVAL EVALUATION OF FOUR TISSUE TYPES USING MOL APPROACH ACROSS 5 FOLDS, MEAN AND STANDARD DEVIATION (SD) VALUES REPORTED.

Tissue types	AP		R-prec		Prec 20	
	Mean	SD	Mean	SD	Mean	SD
Connective	0.378	0.024	0.354	0.014	0.830	0.249
Epithelial	0.467	0.018	0.409	0.020	0.840	0.188
Muscular	0.338	0.024	0.303	0.045	0.760	0.167
Nervous	0.584	0.032	0.525	0.020	0.930	0.027

TABLE IV
EXPERIMENT I: RETRIEVAL EVALUATION OF THE PROPOSED RESULTS COMPARED TO SINGLE FEATURES AND LINEAR FUSION MODEL ACROSS 5 FOLDS, MEAN VALUES REPORTED.

Feature	mean AP	R-prec	Prec 20
GT	0.296	0.290	0.383
TT	0.272	0.255	0.313
ZM	0.273	0.267	0.343
SIFT	0.352	0.357	0.658
DCT	0.328	0.325	0.675
All features linear comb.	0.402	0.356	0.748
MOL feature comb.	**0.442**	**0.398**	**0.840**

As presented in Table III, among the four different tissue types, some results are better compared to the others. For instance, the muscular tissue results are relatively less accurate than the other three. There are probably two reasons for that. First, the number of muscular tissue images is less than the others. There are 484 connective, 804 epithelial, 514 muscular and 1026 nervous tissue samples in the evaluation dataset of experiment I. The task of retrieving images of a less popular query concept is usually more difficult than popular ones. Second, each different tissue type has its unique visual characteristics and patterns. Some of them may be trickier to recognise and differentiate from the others.

Table IV shows mean retrieval performance across four concepts, and using each of the five single features, and two different feature fusion models. All features linear comb. is the direct linear combination model of all the five features with the same importance weights for each feature space. This fusion model follows a direct linear combination approach. MOL feature comb. represents the results using the proposed MOL feature combination model. As it can be observed in Table IV, the proposed MOL method

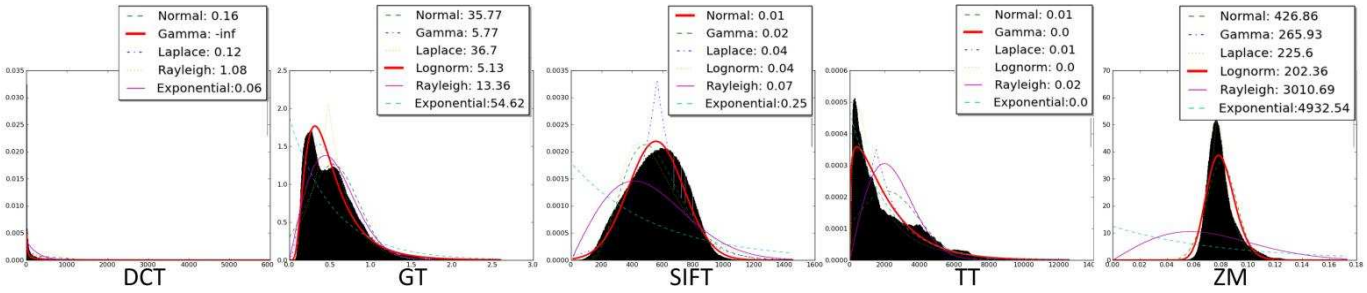


Fig. 2. Distribution of distances in each feature space. Plot labels show the KL-divergence scores for each PDF. The best fit PDF with minimum score is shown in a bold-continuous curve.

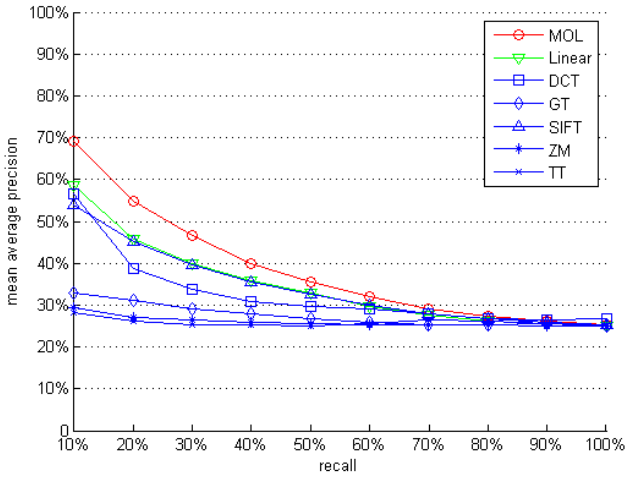


Fig. 3. Average precision-recall curves using MOL multi-feature model, linear multi-feature combination and each single features in experiment I.

performed the best out of the seven different retrieval methods.

To further analyse the results, a statistical significance test is performed based on binomial testing. Given the null hypothesis that the MOL based feature fusion model does not improve the retrieval average precision on top of the second best approach - the linear fusion model, the P-value is calculated to be 0.003. This means that we can confidently reject the null hypothesis and declare our approach has shown a statistically significant improvement in the experiment compared to the linear fusion model.

Figure 3 presents the mean average precision-recall curves corresponding to MOL model based multi-feature retrieval, linear fusion model for multi-feature retrieval and retrieval using each of the considered single features. This figure shows that MOL multi-feature retrieval has shown a clear advantage in the performance over the retrievals relying on direct linear combination or single features. Moreover, direct linear fusion multi-feature retrieval did not bring significant improvements in the retrieval performance based upon SIFT feature.

B. Experiment II

In experiment II, a similar experiment procedure to experiment I was performed. Among the 130 labelled images, 25 were labelled as *connective* tissue, 37 as *epithelial* tissue, 34 as *muscle* tissue, and 34 as *nervous* tissue. A different set of five features were extracted from this dataset, including three vector features: GT, GLCM and ZM, and two MPEG-7 texture features: EH and HT. As experiment II aims at demonstrating that the proposed approach for fusion model learning is independent of selected features, we replaced three features (ZM, SIFT and DCT) in experiment I with three other texture features - GLCM, EH and HT. The other two features, GT and TT which had relatively poor performance in experiment I, were kept and used again in experiment II. GLCM feature also uses the same Euclidean distance metric in (12). For the two MPEG-7 features, EH and HT, their recommended distance metrics in MPEG-7 standard were used.

After the feature extraction and distance calculation steps, a same distance normalisation process was performed following the same steps as described in experiment I. Here, the PDF parameters were estimated based on a combined set of all images in both two experiments, because the number of images in experiment II may not be big enough for acquiring appropriate estimations of parameters. Similar normalisation results were obtained but are not presented here due to the space limitation.

For each keyword, a representative group was randomly selected based on the ground-truth annotations, and the rest of image in the dataset were used for testing. For this experiment, a comparison of results are shown in Table V, based on five single features, the linear fusion model and the MOL based fusion model. Similarly, three evaluation scores for each tissue concept, MAP, R-prec and Prec 20, are presented.

An average precision-recall curve is presented for this set of experiment results in Figure 4. In Table V and Figure 4, a similar observation from experiment I can be obtained, that the proposed MOL feature fusion model out-performed the other models or single features. This observation conforms our assumption that the proposed feature fusion approach is independent of the set of testing features.

TABLE V
EXPERIMENT II: RETRIEVAL EVALUATION OF THE PROPOSED RESULTS COMPARED TO SINGLE FEATURES AND OTHER FEATURE FUSION MODELS

Feature	Evaluation	Connective	Epithelial	Muscular	Nervous	Average
GT	MAP	0.30	0.38	0.36	0.42	0.38
	R-prec	0.28	0.28	0.40	**0.40**	0.34
	Prec 20	0.30	0.20	0.35	**0.45**	0.33
TT	MAP	0.44	0.34	0.35	0.25	0.35
	R-prec	0.32	0.20	0.36	0.20	0.27
	Prec 20	0.35	0.20	0.45	0.15	0.29
GLCM	MAP	0.26	0.38	0.34	0.27	0.31
	R-prec	0.16	0.40	0.32	0.12	0.25
	Prec 20	0.15	0.35	0.35	0.15	0.25
EH	MAP	0.45	0.49	0.23	0.39	0.39
	R-prec	**0.44**	0.52	0.12	0.28	0.34
	Prec 20	**0.45**	0.55	0.15	0.35	0.38
HT	MAP	0.30	0.72	**0.49**	0.36	0.47
	R-prec	0.32	**0.84**	**0.48**	**0.40**	**0.51**
	Prec 20	0.35	0.80	**0.50**	0.40	0.51
All features linear comb.	MAP	0.41	0.68	0.35	0.38	0.45
	R-prec	0.40	0.72	0.28	0.32	0.43
	Prec 20	**0.45**	0.75	0.25	0.30	0.44
MOL feature comb.	MAP	**0.46**	**0.76**	0.45	**0.43**	**0.52**
	R-prec	0.40	**0.84**	0.40	**0.40**	**0.51**
	Prec 20	0.40	**0.95**	0.45	**0.45**	**0.56**

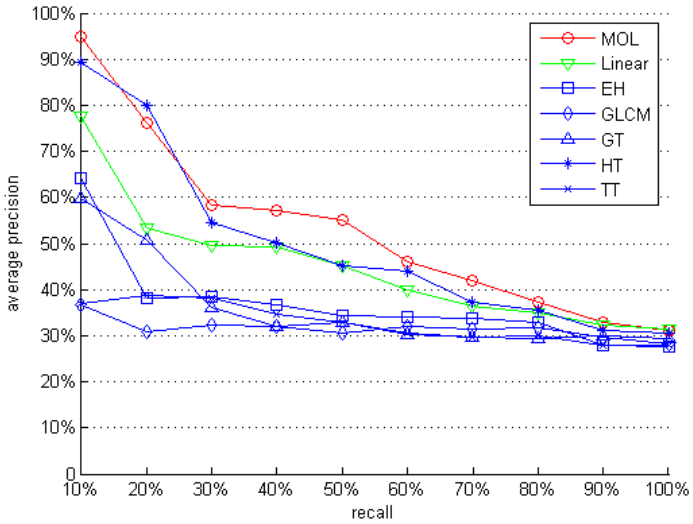


Fig. 4. Average precision-recall curves using MOL multi-feature model, linear multi-feature combination and each single features in experiment II.

V. CONCLUSIONS

This paper proposes a strategy for multi-feature based retrieval in histology image databases. The multi-feature fusion model is obtained using a MOL method, which automatically derives suitable model for feature combination based on multiple query images that are associated to the keyword in concern. The advantage of the proposed feature fusion approach is that it considers a fusion model for each keyword individually. Two different histology image datasets and different sets of low-level features were considered in the experiments. Experimental performance of the proposed approach, as well as a comparison to retrievals relying on single features and other similar fusion models, were presented and analysed. The evaluation of results showed that, in the used experimental set-ups, the proposed strategy was able to provide more precise retrieve results based on semantic keywords that are each represented by a set of query images.

ACKNOWLEDGMENT

The authors thank the BioIngenium Research Group at Universidad Nacional de Colombia and School of Anatomy and Human Biology at The University of Western Australia for sharing BiMed and Blue Histology image databases for free use in scientific research. The research that lead to this paper was partially supported by the European Commission under contract FP7-287704 CUBRIK.

REFERENCES

[1] H. Akakin and M. Gurcan, "Content-based microscopic image retrieval system for multi-image queries," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 16, no. 4, pp. 758 –769, july 2012.

[2] B. André, T. Vercauteren, A. M. Buchner, M. B. Wallace, and N. Ayache, "Learning semantic and visual similarity for endomicroscopy video retrieval," *IEEE Trans. Med. Imaging*, pp. 1276–1288, 2012.

[3] BioIngenium research group, Universidad Nacional de Colombia, http://www.informed.unal.edu.co:8084/BiMed/, Accessed in 2012.

[4] N. Bonnet, "Some trends in microscope image processing," *Micron*, vol. 35, no. 8, pp. 635–653, December 2004.

[5] W. Cai, D. Feng, and R. Fulton, "Content-based retrieval of dynamic PET functional images," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 4, no. 2, pp. 152 –158, june 2000.

[6] J. Caicedo and E. Izquierdo, "Combining low-level features for improved classification and retrieval of histology images," *Transactions on Mass-Data Analysis of Images and Signals*, vol. 2, no. 1, pp. 68–82, 2010.

[7] J. C. Caicedo, F. A. Gonzalez, and E. Romero, "Content-based histopathology image retrieval using a kernel-based semantic annotation framework," *Journal of Biomedical Informatics*, pp. 519–528, 2011.

[8] B. A. Canada, G. K. Thomas, K. C. Cheng, and J. Z. Wang, "SHIRAZ: an automated histology image annotation system for zebrafish phenomics," *Multimedia Tools Appl.*, vol. 51, no. 2, pp. 401–440, Jan. 2011.

[9] K. Chen, J. Lin, Y. Zou, and G. Yin, "Content-based medical ultrasound image retrieval using a hierarchical method," in *Image and Signal Processing, 2nd International Congress on*, oct. 2009, pp. 1 –4.

[10] W. Chen, P. Meer, B. Georgescu, W. He, L. A. Goodell, and D. J. Foran, "Image mining for investigative pathology using optimized feature extraction and data fusion," in *Computer Methods and Programs in Biomedicine*, 2005, pp. 59–72.

[11] D. Comaniciu, P. Meer, and D. J. Foran, "Image-guided decision support system for pathology," *Mach. Vision Appl.*, vol. 11, no. 4, pp. 213–224, Dec. 1999.

[12] G. Csurka, C. R. Dance, L. Fan, J. Willamowski, and C. Bray, "Visual categorization with bags of keypoints," in *Workshop on Statistical Learning in Computer Vision*, 2004.

[13] W. Cui and H. Shao, "Automatic feature weight assignment based on genetic algorithm for image retrieval," *Computer Engineering and Applications*, vol. 44, no. 2, pp. 106–108, 2008.

[14] R. Datta, D. Joshi, J. Li, and J. Z. Wang, "Image retrieval: Ideas, influences, and trends of the new age," *ACM Comput. Surv.*, vol. 40, no. 2, pp. 1–60, April 2008.

[15] K. Donald and A. Smeaton, "A comparison of score, rank and probability-based fusion methods for video shot retrieval," *Image and Video Retrieval*, pp. 61–70, 2005.

[16] P. Duygulu, K. Barnard, J. de Freitas, and D. Forsyth, "Object recognition as machine translation: Learning a lexicon for a fixed image vocabulary," in *Computer Vision – ECCV 2002*, ser. LNCS, 2006, vol. 2353, pp. 349–354.

[17] I. El-Naqa, Y. Yang, N. Galatsanos, R. Nishikawa, and M. Wernick, "A similarity learning approach to content-based image retrieval: application to digital mammography," *Medical Imaging, IEEE Transactions on*, vol. 23, no. 10, pp. 1233 –1244, 2004.

[18] A. Grigorova, F. De Natale, C. Dagli, and T. Huang, "Content-based image retrieval by feature adaptation and relevance feedback," *Multimedia, IEEE Transactions on*, vol. 9, no. 6, pp. 1183 –1192, oct. 2007.

[19] Y. Huang, J. Zhang, Y. Zhao, and D. Ma, "Medical image retrieval with query-dependent feature fusion based on one-class SVM," in *Computational Science and Engineering, 13th International Conference on*, dec. 2010, pp. 176 –183.

[20] J. Kim, W. Cai, D. Feng, and H. Wu, "A new way for multidimensional medical data management: Volume of interest (VOI)-based retrieval of medical images with visual and functional features," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 10, no. 3, pp. 598 –607, july 2006.

[21] J. Knowles and D. Corne, "Approximating the nondominated front using the Pareto archived evolution strategy," *Evolutionary computation*, vol. 8, no. 2, pp. 149–172, 2000.

[22] C. Lacoste, J.-H. Lim, J.-P. Chevallet, and D. Le, "Medical-image retrieval based on knowledge-assisted text and image indexing," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 17, no. 7, pp. 889 –900, july 2007.

[23] R. Lam, H. Ip, K. Cheung, L. Tang, and R. Hanka, "Similarity measures for histological image retrieval," in *Pattern Recognition, 2000. Proceedings. 15th International Conference on*, vol. 2, 2000, pp. 295 –298 vol.2.

[24] C. Loukas, "A survey on histological image analysis-based assessment of three major biological factors influencing radiotherapy: proliferation, hypoxia and vasculature," *Computer Methods and Programs in Biomedicine*, vol. 74, no. 3, pp. 183–199, 2004.

[25] K. Mikolajczyk and C. Schmid, "A performance evaluation of local descriptors," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, no. 10, pp. 1615 –1630, oct. 2005.

[26] A. Mojsilovic, "A computational model for color naming and describing color composition of images," *Image Processing, IEEE Transactions on*, vol. 14, no. 5, pp. 690 –699, may 2005.

[27] H. Müller, J. Kalpathy-Cramer, C. Kahn, W. Hatt, S. Bedrick, and W. Hersh, "Overview of the ImageCLEFmed 2008 medical image retrieval task," *Evaluating Systems for Multilingual and Multimodal Information Access*, pp. 512–522, 2009.

[28] H. Müller, N. Michoux, D. Bandon, and A. Geissbuhler, "A review of content-based image retrieval systems in medical applications–clinical benefits and future directions," *International Journal of Medical Informatics*, vol. 73, no. 1, pp. 1–23, February 2004.

[29] J. Naik, S. Doyle, A. Basavanhally, S. Ganesan, M. D. Feldman, J. E. Tomaszewski, and A. Madabhushi, "A boosted distance metric: application to content based image retrieval and classification of digitized histopathology," N. Karssemeijer and M. L. Giger, Eds., vol. 7260, no. 1. SPIE, 2009, p. 72603F.

[30] N. Orlov, L. Shamir, T. Macura, J. Johnston, D. M. Eckley, and I. G. Goldberg, "WND-CHARM: Multi-purpose image classification using compound image transforms," *Pattern Recognition Letters*, vol. 29, no. 11, pp. 1684–1693, August 2008.

[31] J. Philbin, M. Isard, J. Sivic, and A. Zisserman, "Descriptor learning for efficient retrieval," in *Computer vision, 11th European conference on*, 2010, pp. 677–691.

[32] C. Schmid and R. Mohr, "Local grayvalue invariants for image retrieval," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 19, no. 5, pp. 530 –535, may 1997.

[33] School of Anatomy and Human Biology - The University of Western Australia, http://www.lab.anhb.uwa.edu.au/mb140/Big/Big.htm, Accessed in 2012.

[34] A. W. M. Smeulders, M. Worring, S. Santini, A. Gupta, and R. Jain, "Content-based image retrieval at the end of the early years," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 12, pp. 1349–1380, December 2000.

[35] C. Snoek, M. Worring, and A. Smeulders, "Early versus late fusion in semantic video analysis," in *Proceedings of the 13th annual ACM International Conference on Multimedia*. ACM, 2005, pp. 399–402.

[36] R. Steuer and R. Steuer, *Multiple criteria optimization: Theory, computation, and application*. Wiley, 1986, vol. 233.

[37] A. Tabesh, M. Teverovskiy, H.-Y. Pang, V. Kumar, D. Verbel, A. Kotsianti, and O. Saidi, "Multifeature prostate cancer diagnosis and gleason grading of histological images," *Medical Imaging, IEEE Transactions on*, vol. 26, no. 10, pp. 1366 –1378, oct. 2007.

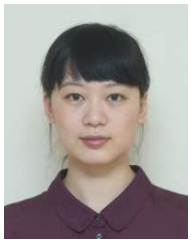
[38] H. L. Tang, R. Hanka, and H. H. S. Ip, "Histological image retrieval based on semantic content analysis," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 7, no. 1, pp. 26–36, 2003.

[39] D. Unay, A. Ekin, and R. Jasinschi, "Local structure-based region-of-interest retrieval in brain MR images," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 14, no. 4, pp. 897 –903, july 2010.

[40] L. Wei, Y. Yang, and R. M. Nishikawa, "Microcalcification classification assisted by content-based image retrieval for breast cancer diagnosis," *Pattern Recognition*, vol. 42, no. 6, pp. 1126 – 1132, 2009.

[41] K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.*, vol. 10, pp. 207–244, Jun. 2009.

[42] L. Zheng, A. W. Wetzel, J. Gilbertson, and M. J. Becich, "Design and analysis of a content-based pathology image retrieval system," *Information Technology in Biomedicine, IEEE Transactions on*, vol. 7, no. 4, pp. 249–255, 2003.



Qianni Zhang received her M.Sc. degree in 2004 and her PhD degree in 2007, both from Queen Mary University of London. She is now working as a postdoctoral researcher at the School of Electronic Engineering and Computer Science, Queen Mary University of London. Her research interests include medical image classification and understanding; content based multimedia retrieval, annotation and classification; 3D immersive environment applications.

Ebroul Izquierdo ,PhD, MSc, CEng, FIET, SMIEEE, MBMVA, is Chair of Multimedia and Computer Vision and head of the Multimedia and Vision Group in the school of Electronic Engineering and Computer Science at Queen Mary, University of London. For his thesis on the numerical approximation of algebraic-differential equations, he received the Dr. Rerum Naturalium (PhD) from the Humboldt University, Berlin, Germany. He has been a senior researcher at the Heinrich-Hertz Institute for Communication Technology (HHI), Berlin, Germany, and the Department of Electronic Systems Engineering of the University of Essex. Prof. Izquierdo is a Chartered Engineer, a Fellow member of The Institution of Engineering and Technology (IET), a senior member of the IEEE, a member of the British Machine Vision Association, past chairman of the IET professional network on Information Engineering, member of the Visual Signal Processing and Communication Technical Committee of the IEEE Circuits and Systems Society and member of the Multimedia Signal Processing technical committee of the IEEE. Prof. Izquierdo is or has been associated and guest editor of several relevant journals in the field including the IEEE Transactions on Circuits and Systems for Video Technology, the EURASIP Journal on Image and Video processing, the Elsevier journal Signal Processing: Image Communication, The EURASIP Journal on Applied Signal Processing, the IEE Proceedigs on Vision, Image & Signal Processing, the Journal of Multimedia Tools and Applications and the Journal of Multimedia. He has been member of the organizing committee of several conferences and workshops in the field and has chaired special sessions and workshops in ICIP, ICASSP and ISCAS. Prof. Izquierdo holds several patents in the area of multimedia signal processing and has published over 500 technical papers including chapters in books.