

Biomedical CBIR using “Bag of Keypoints” in a Modified Inverted Index

Md Mahmudur Rahman, Sameer K. Antani, George R. Thoma
U.S. National Library of Medicine, National Institutes of Health,
Bethesda, MD, USA

Abstract

This paper presents a “bag of keypoints” based medical image retrieval approach to cope with a large variety of visually different instances under the same category or modality. Keypoint similarities in the codebook are computed using a quadratic similarity measure. The codebook is implemented using a topology preserving SOM map which represents images as sparse feature vectors and an inverted index is created on top of this to facilitate efficient retrieval. In addition, to increase the retrieval effectiveness, query expansion is performed by exploiting the similarities between the keypoints based on analyzing the local neighborhood structure of the SOM generated codebook. The search is thus query-specific and restricted to a sub-space spanned only by the original and expanded keypoints of the query images. A systematic evaluation retrieval results on a biomedical image collection of 5000 biomedical images of different modalities, body parts, and orientations shows a halving in computation time (efficiency) and 10% to 15% improvement in precision at each recall level (effectiveness) when compared to individual color, texture, edge-related features.

1 Introduction

In a heterogeneous medical collection with multiple modalities, such as ImageCLEFmed benchmarks¹, images are often captured with different views, imaging and lighting conditions, similar to the real world photographic images. Distinct body parts that belong to the same modality frequently present great variations in their appearance due to changes in pose, scale, illumination conditions and imaging techniques applied. Ideally, the representation of such images must be flexible enough to cope with a large variety of visually different instances under the same category or modality, yet keeping the discriminative power between images of different modalities.

Recent advances in computer vision and pattern recognition techniques have given rise to extract such robust and invariant features from images, commonly termed as affine region detectors [1]. The regions simply refers to a set of pixels or interest points which are invariant to affine transformations, as well as occlusion, lighting and intra-class variations. This differs from classical segmentation since the region boundaries do not have to correspond to changes in image appearance such as color or texture. Often a large number, perhaps hundreds or thousands, of possibly overlapping regions are obtained. A vector descriptor, such as scale invariant feature transform (SIFT) [2] is then associated with each region, computed from the intensity pattern within the region. This descriptor is chosen to be invariant to viewpoint changes and, to some extent, illumination changes, and to discriminate between the regions. The calculated features are clustered or vector quantized (features of interest points are converted into visual words or keypoints) and images are represented by a bag of these quantized features (e.g., bag of keypoints) so that images are searchable in a similar manner with “bag of words” in text retrieval [3].

The idea of clustering invariant descriptors of image patches and represent images with “bag of keypoints” has already been applied to the problem of texture classification and recently for generic visual categorization with promising results [4, 5]. For example, the work described in [5] presents a computationally efficient approach which has shown good results for objects and scenes categorization. Besides, being a very generic method, it is able to deal with a great variety of objects and scenes. Motivated by this, we present a correlation-enhanced “bag of keypoints” based biomedical image retrieval approach. In this approach, the SIFT features are extracted at first from the interest points and then vector quantized by the Self-Organizing Map (SOM)-based clustering to build a visual vocabulary of keypoints. By mapping the interest points extracted from one image to the words in the visual vocabulary, their occurrences are counted and the resulting histogram is called the “bag-of-keypoints” for that image. The similarities/correlations between the keypoints are analyzed

¹<http://ir.ohsu.edu/image/>

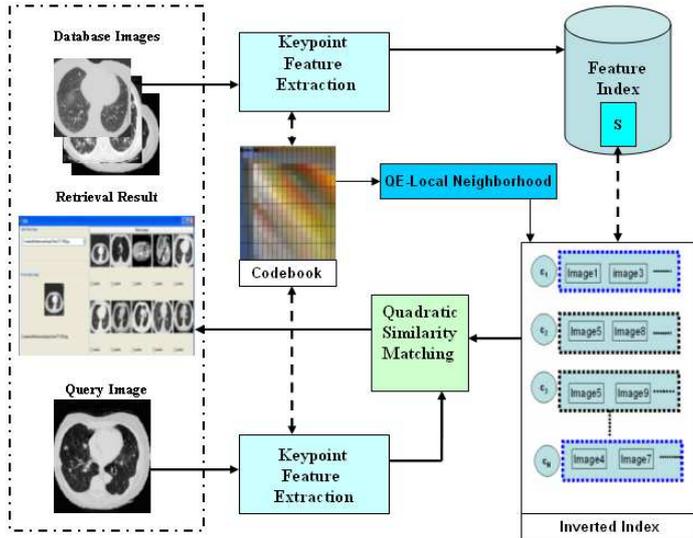


Figure 1. Block diagram of the retrieval framework

in the collection as a whole to construct a global similarity thesaurus that is finally utilized in a distance measure function to compare query and target images in a database. However, due to the quadratic nature, the similarity measure is computationally intensive. To overcome this, only a subset of the images of the entire collection is compared based on a local neighborhood analysis in an inverted index. The codebook or topology preserving SOM map is utilized to represent images as sparse feature vectors and an inverted index is created on top of this to facilitate efficient retrieval. The block diagram of the proposed image retrieval framework is shown in Fig. 1.

The organization of the paper is as follows: In Section 2, the “Bag of Keypoints”-based image representation approach is discussed. Section 3 presents a correlation-enhanced similarity matching approach and Section 4 presents the searching approach in an inverted index. The experiments and analysis of the results are presented in Sections 5 and 6. Finally, Section 7 provides our conclusions.

2 “Bag of Keypoints” based Image Representation

A major component of this retrieval framework is the detection of interest points in scale-space, and then determine an elliptical region for each point. Interest points are those points in the image that possess a great amount of information in terms of local signal changes [1]. In this study, the Harris-affine detector is used as interest point detection methods [6]. In this case, scale-selection is based on the

Laplacian, and the shape of the elliptical region is determined with the second moment matrix of the intensity gradient.

A vector descriptor which is invariant to viewpoint changes and to some extent, illumination changes is then associated with each interest point, computed from the intensity pattern within the point. We use a local descriptor developed by Lowe [2] based on the Scale-Invariant Feature Transform (SIFT), which transforms the image information in a set of scale-invariant coordinates, related to the local features. SIFT descriptors are multi-image representations of an image neighborhood. They are Gaussian derivatives computed at 8 orientation planes over a 4×4 grid of spatial locations, giving a 128-dimension vector. Recently in a study [1] several affine region detectors have been compared for matching and it was found that the SIFT descriptors perform best. SIFT descriptor with affine covariant regions gives region description vectors, which are invariant to affine transformations of the image. A large number of possibly overlapping regions are obtained with the Harris detector. Hence, a subset of the representative region vectors is then selected as a codebook of keypoints by applying a SOM-based clustering algorithm [7].

For each SIFT vector of interest point in an image, the codebook is searched to find the best match keypoint based on a distance measure (generally Euclidean). Based on the encoding scheme, an image I_j can be represented as a vector of keypoints as

$$\mathbf{f}_j^{\text{KV}} = [w_{1j} \cdots w_{ij} \cdots w_{Nj}]^T \quad (1)$$

where each element w_{ij} represents the normalized frequency of occurrences of the keypoints c_i appearing in I_j . This feature representation captures only a coarse distribution of the keypoints that is analogous to the distribution of quantized color in a global color histogram.

3 Quadratic Similarity Matching

This section presents the similarity matching approach in the keypoint feature space by considering the similarities between the keypoints in the codebook. For the correlation analysis, we construct a global structure or thesaurus in the form of a similarity matrix where each element defines the keypoint similarities in an Euclidean space. Finally, this global matrix is utilized in a Quadratic form of distance measure to compare a query and database images.

The quadratic distance measure is first implemented in the QBIC [8] system for the color histogram-based matching. It overcomes the shortcomings of the L-norm distance functions by comparing not only the same bins but multiple bins between color histograms. The keypoint-based feature representation is at a higher level than the simple pixel-based color feature representation due its invariant nature.

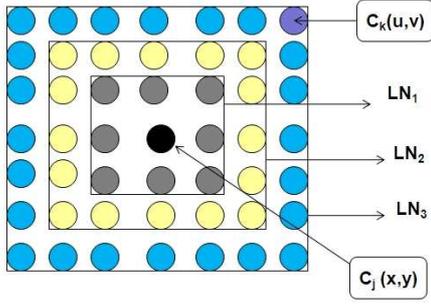


Figure 2. Topological local neighborhoods

Since, the keypoint vectors in the codebook are already represented in a feature space based on the SIFT feature [2], we can use them directly to generate a *keypoint-keypoint* similarity matrix.

Definition 1 The keypoint-keypoint similarity matrix $\mathbf{S}_{N \times N} = [s_{u,v}]$ is built through the computation of each element $s_{u,v}$ as the Euclidean similarity values between two vectors \mathbf{c}_u and \mathbf{c}_v of concept prototypes c_u and c_v as

$$s_{u,v} = \text{sim}(\mathbf{c}_u, \mathbf{c}_v) = \frac{1}{1 + \text{dis}(\mathbf{c}_u, \mathbf{c}_v)} \quad (2)$$

where \mathbf{c}_u and \mathbf{c}_v are 128-dimensional SIFT feature vector and $c_u, c_v \in C$ where N is the size of the codebook C and $\text{dis}(\mathbf{c}_u, \mathbf{c}_v)$ denote the Euclidean distance between \mathbf{c}_u and \mathbf{c}_v .

Finally, the global matrix \mathbf{S} is utilized in the quadratic form of distance measure to compare a query and database images as

$$\text{Dis}(I_q, I_j) = \sqrt{(\mathbf{f}_q - \mathbf{f}_j)^T \mathbf{S} (\mathbf{f}_q - \mathbf{f}_j)} \quad (3)$$

Here, \mathbf{f}_q and \mathbf{f}_j are the feature vector for the query image I_q and a target image I_j respectively.

4 Searching in Inverted Index

The distance measure described in Section 3, computes the cross correlations/similarities between the concepts, hence it requires longer computational time as compared to the L -norm (e.g., Euclidean) or cosine based distance measures. One solution is to compare only a subset of images from the entire collection. In large database applications, the indexing or pre-filtering techniques are essential to avoid exhaustive search in the entire collection. The inverted file is a very popular indexing technique for the vector space model in IR [3]. An inverted file contains an entry for every possible terms and each term contains a list of the documents if the documents have at least one occurrence

of that particular term. In CBIR domain, an inverted index is used in a suitable sparse set of color and texture feature space of dimension more than ten thousands in [9]. Motivated by their success, we present an enhanced inverted index to reduce the search time that considers the similarities between the keypoint prototypes by exploiting the topology preserving property of the SOM generated codebook. Our goal is to decrease the response time where the codebook is used as an inverted file to store the mapping from concepts to images. In this index, for each keypoint prototype in a codebook, a list of pointers or references to images that have at least one region map to this concept is stored in a list. Hence, an image in the collection is a candidate for further distance measure calculations if it contains at least one region that corresponds to a concept c_i in a query image.

Now, to consider the similarity factor between the concepts, the simple lookup strategy in inverted index is modified slightly.

Definition 2 Each keypoint prototype $c_j(x, y) \in C$ has a local γ -neighborhood $LN_\gamma(x, y)$ in a two-dimensional grid of codebook as depicted in Figure 2. We have

$$LN_\gamma(x, y) = \{c_k(u, v) : \max\{|u - x|, |v - y|\}\} \quad (4)$$

Here, the coordinates (x, y) and (u, v) denote the row and column-wise position of any two keypoint prototypes c_j and c_k respectively where $x, u \in \{1, \dots, P\}$ and $y, v \in \{1, \dots, P\}$ for a codebook of size $N = P \times P$ units. The value of γ can be from 1 up to a maximum of $P - 1$.

In this approach, for each keypoint prototype $c_i \in I_q$ with a weight (e.g., *tf-idf* based weighting) w_{iq} , we expand it to other $\lfloor w_{iq} \times (|S_\gamma| - 1) \rfloor$ keypoint prototypes based on the topology preserving ordering in a codebook. Here, S_γ contains all the keypoint prototypes including c_i up to a local neighborhood level LN_γ . For example, Figure 2 shows the local neighborhood structure of a keypoint prototype in a two-dimensional codebook based on Definition 2. Here, each keypoint prototype is visualized as a circle on the grid and the black circle in the middle denotes a particular keypoint prototype $c_j(x, y)$. The keypoint prototype $c_k(u, v)$ is three neighborhood level (e.g., LN_3) apart from $c_j(x, y)$ based on Definition 2 as the maximum distance between them (coordinate-wise) either in horizontal or vertical direction is three. Basically, all the gray circles within the square are positioned in the LN_1 neighborhood, the gray and yellow circles are positioned up to LN_2 and gray, yellow and blue circles in combine are positioned up to LN_3 neighborhoods of c_j as shown in the Figure 2. As the value of γ increases, the number of neighboring keypoint prototypes increases for c_j .

For the query expansion, the keypoints other than c_i are considered by subtracting it from S_γ . After the expansion, those images that appear in the list of expanded keypoints

are deemed as candidates for further similarity matching while the other images are ignored. A larger γ will lead to more expanded keypoints, which means that more images need to be compared with the query. This might lead to more accurate retrieval results in a trade off of the larger computational time. After finding the $|S_\gamma| - 1$ keypoint prototypes, they are ranked based on their similarity values with c_i by looking up the corresponding entry in the matrix \mathbf{S} . This way the relationship between two keypoints are actually determined by both their closeness in the topology preserving codebook and their similarity obtained from the matrix \mathbf{S} . Finally, the top $\lfloor w_{iq} \times (|S_\gamma| - 1) \rfloor$ keypoints are selected as expanded keypoints for c_i . Hence, a keypoint with more weight in a query vector will be expanded to the more closely related keypoints and as a result will have more influence to retrieve candidate images. Therefore, the enhanced inverted index contains an entry for a keypoint that consists of a list of images as well as images from closely related concepts based on the local neighborhood property. The steps of the above process are described in Algorithm 1. Figure 3 shows an example of the above processing steps. Here, for a particular keypoint c_j with the associated weight in vector as w_{jq} that is presented in the query image I_q , the corresponding location of the keypoint in the codebook is determined. Suppose, based on the LN_1 neighborhood of the above algorithm, only two concepts c_k and c_m are further selected for expansion. After finding the expanded keypoint prototypes, the images in their inverted lists are merged with the original set of images and considered for further distance measure calculation for ranked-based retrieval. Therefore, in addition to considering all the images in the inverted list of c_j (images under black dotted rectangle), we also need to consider the images in the list of c_k and c_m (under the blue dotted rectangle) as candidate images. Due to the space limitations, all the actual links are not shown in Figure 3. In this way, the response time is reduced while the retrieval accuracy is still maintained.

5 Experiments

The image collection for experiment comprises of 5000 bio-medical images of 32 manually assigned disjoint global categories, which is a subset of a larger collection of six different data sets used for medical image retrieval task in ImageCLEFmed 2007 [10]. In this collection, images are classified into three levels. In the first level, images are categorized according to the imaging modalities (e.g., X-ray, CT, MRI, etc.). At the next level, images at each of the modalities is classified according to the examined body parts (e.g., head, chest, etc.) and finally images are further classified by orientation (e.g., frontal, sagittal, etc.) or distinct visual observation (e.g. CT liver images with large blood vessels). The disjoint categories are selected only from the leaf nodes

Algorithm 1 Similarity Matching in Modified Inverted File

- 1: Initially compute the global similarity matrix \mathbf{S} offline. Let, the feature vector of a query image I_q be $\mathbf{f}_q = [w_{1q} \cdots w_{iq} \cdots w_{Nq}]^T$ in a keypoint-based feature space. Initialize the list of candidate image as $L = \phi$.
- 2: **for** $i = 1$ to N **do**
- 3: **if** $w_{iq} > 0$ (i.e., $c_i \in I_q$) **then**
- 4: Locate the corresponding keypoint prototype c_i in the two-dimensional codebook C .
- 5: Read the corresponding list L_{c_i} of images from the inverted file and add it to L as $L \leftarrow L \cup L_{c_i}$.
- 6: Consider up to LN_γ neighborhoods of c_i to find related $|S_\gamma| - 1$ keypoint prototypes.
- 7: For each $c_j \in S_\gamma - \{c_i\}$, determine its ranking based on the similarity values by looking up corresponding entry s_{ij} in matrix \mathbf{S} .
- 8: Consider the top $k = \lfloor w_{iq} \times (|S_\gamma| - 1) \rfloor$ ranked keypoint prototypes in set S^k for further expansion.
- 9: **for each** $c_k \in S^k$ **do**
- 10: Read the corresponding list as $L(c_k)$ and add to L as $L \leftarrow L \cup L_{c_k}$ after removing the duplicates.
- 11: **end for**
- 12: **end if**
- 13: **end for**
- 14: **for each** $I_j \in L$ **do**
- 15: Apply the distance matching functions of Equation (3) between I_q and I_j based on the Matrix \mathbf{S} .
- 16: **end for**
- 17: Finally, return the top K images by sorting the distance measure values in ascending order (e.g., a value of 0 indicates closest match).

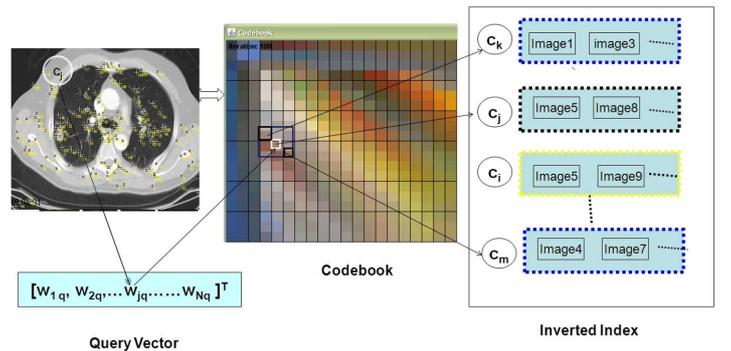


Figure 3. Example process of Query Expansion in an Inverted File

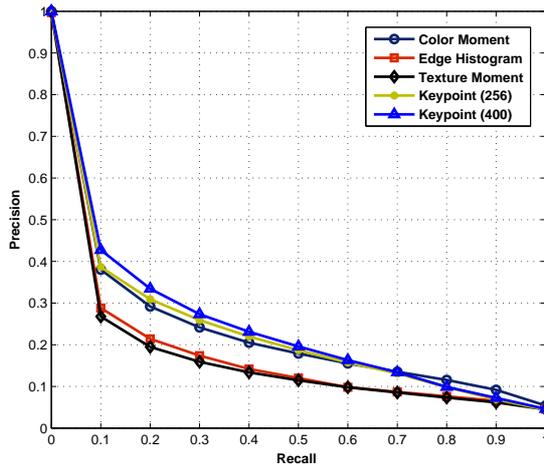


Figure 4. PR-graphs of different feature spaces.

(grey in color) to create the ground-truth data set.

To build the codebook based on the SOM clustering, a training set of images is selected beforehand for the learning process. The training set used for this purpose consists of 10% images of the entire data set (5000 images) resulting in a total of 500 images. For a quantitative evaluation of the retrieval results, we selected all the images in the collection as query images and used *query-by-example (QBE)* as the search method. A retrieved image is considered a match if it belongs to the same category as the query image out of the 32 disjoint categories at the global level.

6 Results

Fig. 4 shows the precision-recall (PR) curves of the keypoints-based image representation with two different codebook sizes (e.g., 256 (16×16) and 400 (20×20) units) by performing the Euclidean similarity matching. The performances were also compared to three low-level color, texture, and edge related features to judge the actual improvement in performances of the proposed methods. The reason of choosing these three low-level feature descriptors is that they present different aspects of medical images. For color feature, the first (mean), second (standard deviation) and third (skewness) central moments of each color channel in the RGB color space are calculated to represent images as a 9-dimensional feature vector. The texture feature is extracted from the gray level co-occurrence matrix (GLCM). A GLCM is defined as a sample of the joint probability density of the gray levels of two pixels separated by a given displacement and angle [11]. We obtained four GLCM for four

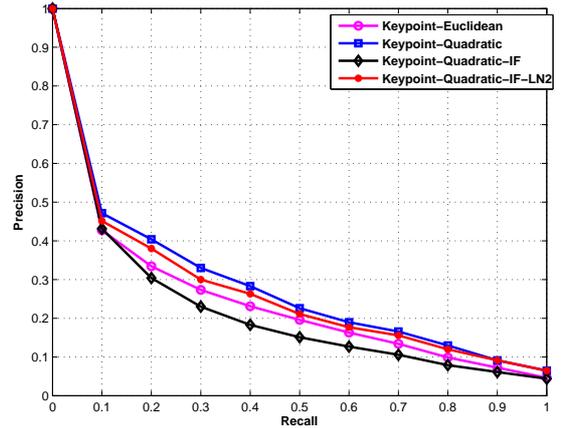


Figure 5. PR-graphs of different searching criteria.

different orientations (horizontal 0° , vertical 90° , and two diagonals 45° and 135°). Higher order features, such as energy, maximum probability, entropy, contrast and inverse difference moment are measured based on each GLCM to form a 5-dimensional feature vector and finally obtained a 20-dimensional feature vector by concatenating the feature vector for each GLCM. Finally, to represent the shape feature, a histogram of edge direction is constructed. The edge information contained in the images is processed and generated by using the Canny edge detection (with $\sigma = 1$, Gaussian masks of size = 9, low threshold = 1, and high threshold = 255) algorithm [12]. The corresponding edge directions are quantized into 72 bins of 5° each. Scale invariance is achieved by normalizing this histograms with respect to the number of edge points in the image. By analyzing the Fig. 4, we can observe that the performance of the proposed keypoints-based feature representation is better when compared to the low-level features in term of precision at each recall level. The better performances are expected as the keypoints-based features are more localized in nature and invariant to viewpoint and illumination changes.

Fig. 5 shows the PR-curves of the keypoints-based image representation (codebook size of 400) by performing the Euclidean (e.g., “Keypoint-Euclidean”) similarity matching and the Quadratic similarity matching (e.g., “Keypoint-Quadratic”). From Fig. 5, we can also observe that, the Quadratic similarity matching approach performed much better when compared to the Euclidean similarity matching. Although, we observe a decrease in performance when the search is performed in the inverted index based on the associated keypoints in each image (e.g., “Keypoint-Quadratic-IF”). This is due to the fact that there might be some quantization and encoding errors which oc-

curred during the codebook generation and image representation steps. However, when we exploited the local neighborhood structure of the codebook by considering up to two levels and performed search in the inverted index with the expanded keypoints (e.g., “Keypoint-Quadratic-IF-LN2”), performance is increased compared to the Euclidean-based matching or without using the modified indexing as shown in Fig. 5. Overall, the improved result indicate that the correlations among the keypoints are not negligible and can be exploited effectively in the similarity matching function as well as in the inverted index.

The major gain in searching on a inverted index is that it takes less computational time compared to a linear search in the entire collection. Hence, to test the efficiency of the search schemes for the keypoint-based feature, we also compared the average retrieval time with and without the indexing scheme. The search time is significantly reduced (nearly half) with the use of inverted index for both Euclidean and Quadratic similarity matching functions compared to the linear searching. In addition, the search time in the modified inverted index is slightly more compared to using without any modification. However, with this slight increase in time, we achieved better precision at each recall level as shown in Fig. 5, justifying its use. The Quadratic similarity matching in the modified inverted index has proved to be both effective and efficient.

7 Conclusions

We have investigated the “bag of keypoints” based image retrieval approach in medical domain inspired by the ideas of the text retrieval. Due to the nature of the keypoint-based image representation scheme, there always exists enough correlations between the keypoints in medical images. Hence, exploiting this property in the similarity matching and the inverted indexing schemes improved the retrieval effectiveness and efficiency as shown in the experimental section. In future, when the object recognition techniques will be mature enough, our approaches would be easily extendible to a higher level concept-based image representation and retrieval approaches.

Acknowledgment

This research is supported by the Intramural Research Program of the National Institutes of Health (NIH), National Library of Medicine (NLM), and Lister Hill National Center for Biomedical Communications (LHNCBC). We would like to thank the ImageCLEFmed [10] organizers for making the database available for the experiments.

References

- [1] K. Mikolajczyk, T. Tuytelaars, C. Schmid, A. Zisserman, J. Matas, F. Schaffalitzky, T. Kadir, and L. Van Gool, “A Comparison of Affine Region Detectors”, *International Journal of Computer Vision*, vol. 65, pp. 43–72, 2005.
- [2] D. G. Lowe, “Distinctive image features from scale-invariant keypoints”, *International Journal of Computer Vision*, vol. 60 (2), pp. 91–110, 2004.
- [3] R. B. Yates, and B. R. Neto, *Modern Information Retrieval*, 1st ed., Addison Wesley, 1999.
- [4] S. Lazebnik, C. Schmid, and J. Ponce, “Sparse texture representation using affine-invariant neighborhoods”, *Proc. International Conference on Computer Vision & Pattern Recognition*, pp. 319–324, 2003.
- [5] G. Csurka, C. Dance, J. Willamowski, L. Fan, and C. Bray, “Visual categorization with bags of keypoints,” *Proc. Workshop on Statistical Learning in Computer Vision*, pp. 1–22, 2004.
- [6] K. Mikolajczyk and C. Schmid, “An affine invariant interest point detector”, *Proc. of European Conference on Computer Vision*, pp. 128–142, 2002.
- [7] T. Kohonen, *Self-Organizing Maps*, New York, Springer-Verlag, 1997.
- [8] J. Hafner, H. S. Sawhney, W. Equitz, M. Flickner, and W. Niblack, “Efficient color histogram indexing for quadratic form distance functions,” *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 17, no. 7, pp. 729–736, 1995.
- [9] H. Müller, D. M. Squire, W. Mueller, and T. Pun, “Efficient access methods for content-based image retrieval with inverted files,” *Proc. SPIE*, vol. 3846, pp. 461–472, 1999.
- [10] H. Müller, T. Deselaers, E. Kim, C. Kalpathy, D. Jayashree, M. Thomas, P. Clough, and W. Hersh, “Overview of the ImageCLEFmed 2007 Medical Retrieval and Annotation Tasks”, *8th Workshop of the Cross-Language Evaluation Forum (CLEF 2007)*, Proc. of LNCS, 5152, 2008.
- [11] R. M. Haralick, Shanmugam, and I. Dinstein, “Textural features for image classification,” *IEEE Trans. Syst. Man Cybernetics*, vol. 3, pp. 610–621, 1973.
- [12] J. Canny, “A computational approach to edge detection”, *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 8, pp. 679–698, 1986.