# Image Search in a Visual Concept Feature Space with SOM-Based Clustering and Modified Inverted Indexing

Md Mahmudur Rahman
*Concordia University*
*Canada*

## 1. Introduction

The exponential growth of image data has created a compelling need for innovative tools for managing, retrieving, and visualizing images from large collection. The low storage cost of computer hardware, availability of digital devices, high bandwidth communication facilities and rapid growth of imaging in the World Wide Web has made all these possible. Many applications such as digital libraries, image search engines, medical decision support systems require effective and efficient image retrieval techniques to access the images based on their contents, commonly known as content-based image retrieval (CBIR). CBIR computes relevance of query and database images based on the visual similarity of low-level features (e.g., color, texture, shape, edge, etc.) derived entirely from the images Smeulders et al. (2000); Liua et al. (2007); Datta et al. (2008). Even after almost two decades of intensive research, the CBIR systems still lag behind the best text-based search engines of today, such as Google and Yahoo. The main problem here is the extent of mismatch between user's requirements as high-level concepts and the low-level representation of images; this is the well known *"semantic gap"* problem Smeulders et al. (2000).

In an effort to minimize the *"semantic gap"*, some recent approaches have used machine learning on locally computed image features in a *"bag of concepts"* based image representation scheme by treating them as visual concepts Liua et al. (2007). The models are applied to images by using a visual analogue of a word (e.g., *"bag of words"* ) in text documents by automatically extracting different predominant color or texture patches or semantic patches, such as, water, sand, sky, cloud, etc. in natural photographic images. This intermediary semantic level representation is introduced as a first step to deal with the *semantic gap* between low-level features and high-level concepts. Recent works have shown that local features represented by *"bags-of-words"* are suitable for scene classification showing impressive levels of performance Zhu et al. (2002); Lim (2002); Jing et al. (2004); Vogel & Schiele (2007); Shi et al. (2004); Rahman et al. (2009a). For example, a framework to generate automatically the visual terms ("keyblock") is proposed in Zhu et al. (2002) by applying a vector quantization or clustering technique. It represents images similar to the *"bags-of-words"* based representation in a correlation-enhanced feature space. For the reliable identification of image elements, the work in Lim (2002) manually identifies the visual patches ("visual keywords") from the sample images. n Jing et al. (2004), a compact and sparse representation of images is proposed based on the utilization of a region codebook generated by a clustering technique. A semantic

modeling approach is investigated in Vogel & Schiele (2007) for a small collection of images based on the binary classification of semantic patches of local image regions. A medical image retrieval framework is presented in Rahman et al. (2009b) that uses a visual concept-based feature space for which statistical models are built using a probabilistic multi-class support vector machine (SVM). The images are represented using concepts that comprise of color and texture patches from local image regions in a multi-dimensional feature space. It is demonstrated by experimental evaluation that approaches using intermediate semantic concepts are more appropriate to deal with the gap between low and high-level Boschet et al. (2007).

There exists a strong similarity between the keyword-based representation of documents in the vector space model (VSM) Yates & Neto (1999) and the majority of the concept-based image representation schemes as mentioned above. Besides the loss of all ordering structure, each concept is considered independent of all the other concepts in this model. However, this independent assumption might not hold in many cases as in general there exists correlated or co-occurring concepts in individual images as well as in a collection. For example, there is a higher probability of occurrence of a blue sky around the sun in an outdoor image. Whereas, a flower of yellow color has more probability to co-occur with green leaves in an image of a flower garden. In these examples, individual objects, such as sky, sun, flower, and leaf can be considered as visual concepts with their distinct color and texture patterns. Hence, there is indeed a need to exploit the correlation or co-occurrence patterns among the concepts to improve the effectiveness of the retrieval process. To overcome the limitation, we present a correlation-enhanced similarity matching and query expansion framework on the concept-based feature space. We explore the similarity matching technique based on the global correlation analysis of the concepts and query expansion based on a local neighborhood analysis of a SOM generated codebook by exploiting the topology preserving structure. 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. In this approach, a global similarity/correlation matrix or thesaurus is generated off-line and utilized in a quadratic form of distance measure to compare the query and database images. However, due to the quadratic nature, the distance 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 built on top of the codebook to reduce the search time and at the same time to ensure the retrieval effectiveness.

The organization of this chapter is as follows: In Section 2, the visual concept-based image representation approach is discussed. Section 3 presents the correlation-enhanced similarity matching approach based on the generation of several global matrices. In Section 4, we present the similarity matching approach in a modified inverted index. Exhaustive experiments and analysis of the results are presented in Sections 5 and 6. Finally, Section 7 provides our conclusions.

## 2. Visual concept-based image representation

By the term *"visual concept"*, we refer to the perceptually distinguishable color and/or texture patches that are identified locally in image regions. For example, in a heterogeneous collection of medical images, it is possible to identify specific local patches, such as homogeneous texture patterns in grey level radiological images, differential color and texture structures in microscopic pathology and dermoscopic images. The variation in these local patches can be effectively modeled by using unsupervised clustering or supervised classification techniques

Fukunaga (1990). There are three main steps to be considered before representing images in a visual concept-based feature space: the generation of a set of visual concepts from the local image regions; the construction of a codebook of prototype concepts analogous to a dictionary of keywords; and the encoding of the images with the concept indices of the codebook Rahman et al. (2009a).

**Definition 1** *A codebook $C = \{c_1, \cdots, c_j, \cdots, c_N\}$ is a set of prototype visual concepts where each $c_j$ is associated with a label $j$ and a vector $\mathbf{c}_j = [c_{j_1} \cdots c_{j_2} \cdots c_{j_d}]^{\mathrm{T}}$ of dimension d in a combined color and texture-based feature space.*

To generate the codebook, a reasonable training set of images needs to be selected either manually or in a random manner. Let $\mathcal{D}$ be an image database and let a subset $\hat{\mathcal{D}} = \{I_1, \cdots, I_j, \cdots, I_m\} \subset \mathcal{D}$ form a training set of images. After forming the training set, the next step is to segment the training images into regions and extract the low-level image features from each region as a representative of the initial visual concept vectors. Since, the automatic segmentation schemes usually offer only an unreliable object description, we use a fixed partitioning scheme. Let an image $I_j \in \hat{\mathcal{D}}$ be partitioned into an $r \times r$ grid of $l$ blocks as segmented regions to generate the region vectors $\{\mathbf{x}_{1_j}, \cdots, \mathbf{x}_{k_j}, \cdots, \mathbf{x}_{l_j}\}$ where each $\mathbf{x}_{k_j} \in \Re^d$ is a vector in a low-level feature space.

To represent each region as a feature vector $\mathbf{x}_i$, the mean and the standard deviation of each channel in the HSV (Hue, Saturation, and Value) color space as a 6-dimensional color feature vector and the second order moments (such as, energy, maximum probability, entropy, contrast, and inverse difference moment) as a 5-dimensional texture feature vector are extracted from a grey level co-occurrence matrix (GLCM) Haralick et al. (1973).

Finally the color and texture vectors are combined as a single region vector after re-scaling the feature attributes with zero mean and unit variance. There are in total $m$ training images. So, the partition scheme will generate $n = (l \times m)$ region vectors for all the training images and collectively we can refer to them as a set of vectors $\mathbf{X} = \{\mathbf{x}_1, \cdots, \mathbf{x}_i, \cdots \mathbf{x}_n\}$ where each $\mathbf{x}_i$ is a vector of dimension $d$. Since, the features from the blocks rather than individual pixels are used as vectors, some information on the spatial relationship among the neighboring pixels in the images are already retained. In general, there might be several similar regions in terms of the image features in an individual image as well as in different images in the same training set. Since our visual system should tolerate some small errors, if the difference between two regions is below a certain preset threshold, they are deemed to be the same. Hence, a subset of these representative vectors needs to be selected as a codebook of the visual concept prototype by applying a clustering algorithm, such as SOM Kohonen (1997).

## 2.1 Codebook generation by SOM

To generate a coodbook of prototype vectors (e.g., concept vectors) from the above features, we utilize the SOM-based clustering Kohonen (1997). The SOM is basically an unsupervised and competitive learning algorithm, which finds the optimal set of prototypes based on a grid of artificial neurons whose weights are adapted to match input vectors in a training set Kohonen (1997). It has been successfully utilized for indexing and browsing by projecting the low-level input features to the two-dimensional grid of the SOM map Laaksonen et al. (2002); Vesanto (2002); Yen & Zheng (2008). However, in this work it is utilized to generate a codebook of visual concepts based on a two-dimensional SOM map.

The basic structure of a SOM consists of two layers: an input layer and a competitive output layer as shown in Figure 1. The input layer consists of a set of input node vectors. The
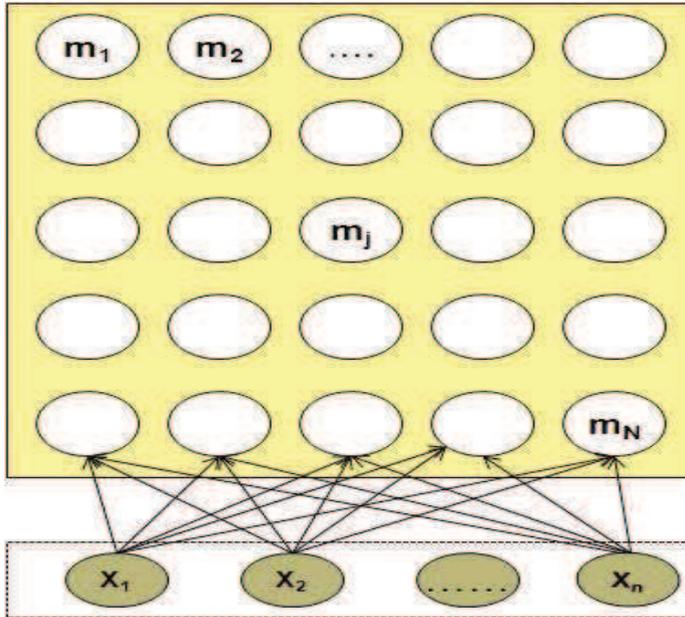
Fig. 1. Structure of the SOM

output map consists of a set of $N$ units organized into either a one-or two-dimensional lattice structure where each unit $m_j$ is associated with a weight vector $\mathbf{w}_j \in \Re^d$. During the training phase, the set of the input vectors is presented to the map a multiple number of times and the weight vectors stored in the map units are modified to match the distribution and topological ordering of the feature vector space. The first step of the learning process is to initialize the weight vectors of the output map. Then, for each input vector $\mathbf{x}_i \in \Re^d$, the distances between the $\mathbf{x}_i$ and weight vectors of all map units are calculated as

$$\|\mathbf{x}_i - \mathbf{w}_c\|^2 = \min_j \left\|\mathbf{x}_i - \mathbf{w}_j\right\|^2, \text{ for } j = 1, 2, \cdots, N \tag{1}$$

where $\|.\|^2$ is a distance measure in the Euclidean norm. The unit that has the smallest distance is called the *best-matching unit* (BMU) or the *winning node*. The next step is to update the weight vectors associated with the BMU, $m_c$ as

$$\mathbf{w}_j(t+1) = \mathbf{w}_j(t) + \alpha(t)\theta_{cj}(t)(\mathbf{x}_i(t) - \mathbf{w}_j(t)) \tag{2}$$

Here, $t$ is the current iteration, $\mathbf{w}_j(t)$ and $\mathbf{x}_i(t)$ are the weight vector and the target input vector respectively at the iteration $t$, and $\theta(t)$ and $\alpha(t)$ are the smooth neighborhood function and the time-dependent learning rate.

Due to the process of self-organization, the initially chosen $\mathbf{w}_j$ gradually attains new values such that the output space acquires appropriate topological ordering. After the learning phase, the map can be used as a codebook where the map units represent the prototype visual concepts and their associated weight vectors represent the prototype concept vectors. Hence, a weight vector $\mathbf{w}_j$ of unit $m_j$ resembles a visual concept vector $\mathbf{c}_j$ in the codebook $C$. In
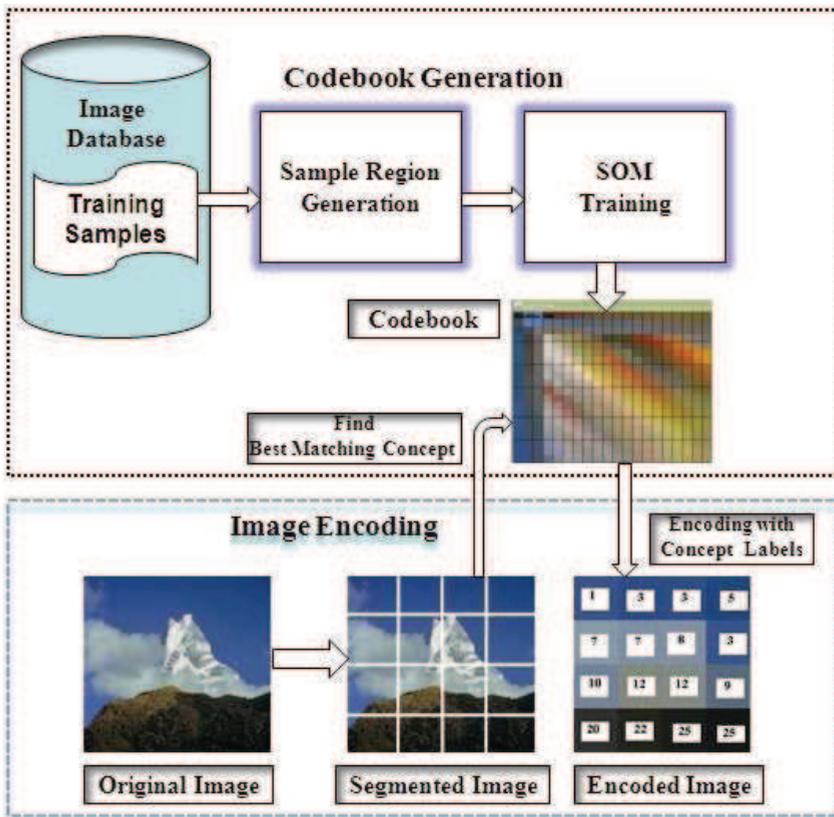
Fig. 2. Codebook generation and Image encoding process

general, the visual concept prototypes in the resulting codebook represent the most general structures extracted from all the training input vectors.

### 2.2 Image encoding and feature representation

The codebook can be effectively utilized as a simple image compression and representation scheme Zhu et al. (2002). To encode an image with the visual concept prototype labels or indices of the codebook, it is decomposed into an even grid-based $(r \times r)$ partition where similar low-level color and texture features are extracted from each region as is used as the training images. Let an image $I_j$ be partitioned into $l = (r \times r)$ blocks or regions to generate vectors $\{\mathbf{x}_{1_j}, \cdots, \mathbf{x}_{k_j}, \cdots, \mathbf{x}_{l_j}\}$ where each $\mathbf{x}_{k_j} \in \Re^d$. For each vector $\mathbf{x}_{k_j}$ in $I_j$ , the codebook is searched to find the best match concept prototype (e.g., BMU in the map)

$$c_k = \arg \min_{1 \le l \le N} \left\| \mathbf{x}_{k_j} - \mathbf{c}_l \right\|^2 \tag{3}$$

where $k$ denotes the label of $c_k$ and $\|.\|^2$ denotes the Euclidean distance between the region vectors of $I_j$ and the concept prototype vectors.

After this encoding process, each image is represented as a two-dimensional grid of concept prototype labels where the image blocks are linked to the corresponding best matching concept prototypes in the codebook. Figure 2 shows schematically the codebook generation and image encoding processes. The codebook generation is performed in the top portion of Figure 2 and the bottom portion shows how an example image is encoded with the indices (e.g., prototype concept labels) of the codebook. Based on this encoding scheme, an image $I_j$ can be represented as a concept vector

$$\mathbf{f}_j = [w_{1j} \cdots w_{ij} \cdots w_{Nj}]^{\mathrm{T}} \tag{4}$$

where each element $w_{ij}$ represents the normalized frequency of occurrences of the visual concept label of $c_i$ appearing in $I_j$.

## 3. Correlation-enhanced similarity matching

This section presents the similarity matching approach in the visual concept space by considering the correlations between the concepts in the collection. For the correlation analysis, we construct a global structure or thesauruses in the form of a correlation matrix where each element defines concept co-relationships. 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 Hafner et al. (1995) 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. Due to this property, it performs better compared to the Euclidean and histogram intersection-based distance measures for the color-based image retrieval Hafner et al. (1995). However, a similarity based on only the color feature does not always indicate semantic similarities between the images due to the *semantic gap* problem and does not imply any hidden correlation between feature attributes in a collection.

The visual concept-based feature representation is at a higher level then the simple pixel-based color feature representation due to the incorporation of both color and texture features in a region level. Since, the concept prototype vectors in the codebook are already represented in a feature space based on the color and texture features, we can use them directly to generate a *concept-concept* similarity matrix.

**Definition 2** *The concept-concept 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*

$$s_{u,v} = sim\,(\mathbf{c}_u, \mathbf{c}_v) = 1 - dis\,(\mathbf{c}_u, \mathbf{c}_v) \tag{5}$$

*where* $\mathbf{c}_u$ *and* $\mathbf{c}_v$ *are d-dimensional vector in a combined color and texture feature space and* $c_u, c_v \in C$ *where N is the size of the codebook C and dis* $(\mathbf{c}_u, \mathbf{c}_v)$ *denote the Euclidean distance between* $\mathbf{c}_u$ *and* $\mathbf{c}_v$.

Instead of using a matrix based on similarities in a color space, we can effectively utilize this global visual *concept-concept* similarity matrix $\mathbf{S}$ for the distance measure computation as

$$Dis_S(I_q, I_j) = \sqrt{(\mathbf{f}_q - \mathbf{f}_j)^T \mathbf{S}(\mathbf{f}_q - \mathbf{f}_j)} \tag{6}$$

However, the visual similarities between the concepts might not always imply semantic similarities or hidden correlations between the concepts as mentioned earlier. Hence, we
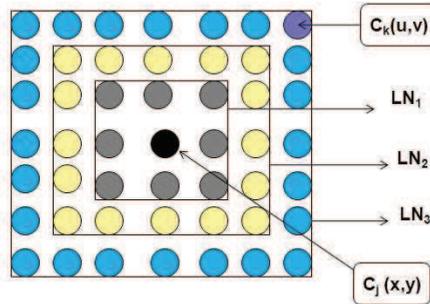
Fig. 3. Topological local neighborhoods

utilize another global matrix that is built by considering the co-occurrence of concepts inside images.

**Definition 3** *Let* $\mathbf{A} = [a_{uv}]$ *be a* $N \times N$ *matrix in which the rows and columns are associated with the concepts in the codebook C. Each entry* $a_{uv}$ *expresses a normalized correlation factor between concepts* $c_u$ *and* $c_v$ *as*

$$a_{uv} = n_{uv}/(n_u + n_v - n_{uv}) \tag{7}$$

*where* $n_u$ *be the number of images in* $S_l$ *that contain the concept* $c_u$, $n_v$ *be the number of images that contain the concept* $c_v$, *and* $n_{uv}$ *be the number of the images in the collection that contain both the concepts.*

The entry $a_{uv}$ measures the ratio between the number of images where both $c_u$ and $c_v$ appear and the total number of images in the collection where either $c_u$ or $c_v$ appear and its value ranges to $0 \leq a_{uv} \leq 1$. If $c_u$ and $c_v$ have many co-occurrences in images, the value of $a_{uv}$ increases and the images are considered to be more correlated. This matrix is termed as a *connection matrix* in Yasushi et al. (1991), which was successfully utilized in a fuzzy information retrieval approach. Finally we can easily replaces the above matrices with the distance matching function in (6) and perform the similarity search effectively.

## 4. Query expansion in a modified 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 Gaede & Gunther (1998). The inverted file is a very popular indexing technique for the vector space model in IR Yates & Neto (1999). 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 then ten thousands in Müller et al. (1999). Motivated by their success, we present an enhanced inverted index to reduce the search time that considers the similarities between the visual concept prototypes by exploiting the topology preserving property of the SOM generated codebook. Our goal is to decrease the response time where the codebook is acted as an inverted file to store

the mapping from concepts to images. In this index, for each visual concept 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 4** *Each visual concept prototype $c_j(x,y) \in C$ has a local $\gamma$-neighborhood $LN_\gamma(x,y)$ in a two-dimensional grid of codebook as depicted in Figure 3. We have*

$$LN_\gamma(x,y) = \{c_k(u,v) : \max\{|u-x|, |v-y|\}\}  \qquad (8)$$

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

In this approach, for each concept prototype $c_i \in I_q$ with a weight (e.g., *tf-idf* based weighting) $w_{iq}$, we expand it to other $\lfloor w_{iq} \times (|S_\gamma| - 1) \rfloor$ concept prototypes based on the topology preserving ordering in a codebook. Here, $S_\gamma$ contains all the concept prototypes including $c_i$ up to a local neighborhood level $LN_\gamma$. For example, Figure 3 shows the local neighborhood structure of a concept prototype in a two-dimensional codebook based on Definition 4. Here, each concept prototype is visualized as a circle on the grid and the black circle in the middle denotes a particular concept prototype $c_j(x,y)$. The concept prototype $c_k(u,v)$ is three neighborhood level (e.g., $LN_3$) apart from $c_j(x,y)$ based on Definition 4 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 3. As the value of $\gamma$ increases, the number of neighboring concept prototypes increases for $c_j$.

For the query expansion, the concepts other then $c_i$ are considered by subtracting it from $S_\gamma$. After the expansion, those images that appear in the list of expanded concepts are deemed
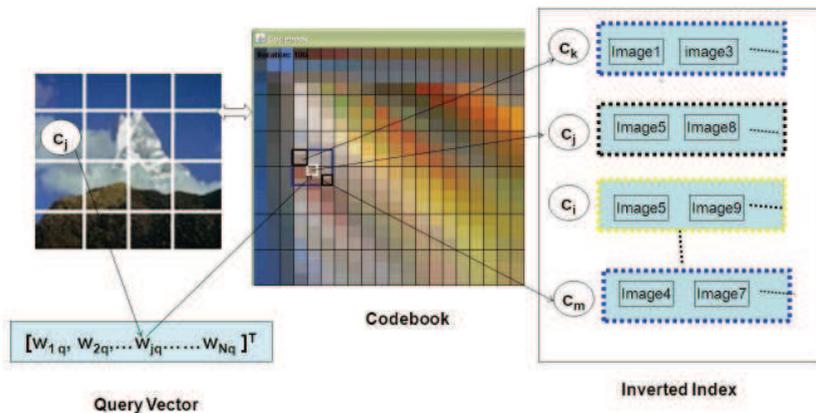


Fig. 4. Example process of Query Expansion in an Inverted File

---

**Algorithm 1** Similarity Matching in Modified Inverted File

---

1: Initially compute the global similarity matrix $\mathbf{S}$ off-line. Let, the feature vector of a query image $I_q$ be $\mathbf{f}_q = [w_{1q} \cdots w_{iq} \cdots w_{Nq}]^{\mathrm{T}}$ in a visual concept-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 concept 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 $LN_\gamma$ neighborhoods of $c_i$ to find related $|S_\gamma| - 1$ concept 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 concept 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 (6) between $I_q$ and $I_j$ based on the Matrix $\mathbf{S}$ or $\mathbf{A}$ .

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).

---

as candidates for further distance measure calculations while the other images are ignored. A larger $\gamma$ will lead to more expanded concepts, 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 $|S_\gamma| - 1$ concept 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 concepts are actually determined by both their closeness in the topology preserving codebook and their correlation or similarity obtained from the global matrices. Finally, the top $\lfloor w_{iq} \times (|S_\gamma| - 1) \rfloor$ concepts are selected as expanded concepts for $c_i$. Hence, a concept with more weight in a query vector will be expanded to the more closely related concepts and as a result will have more influence to retrieve candidate images. Therefore, the enhanced inverted index contains an entry for a concept 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 describes in Algorithm 1.

Figure 4 shows an example of the above processing steps. Here, for a particular concept $c_j$ with the associated weight in vector as $w_{jq}$ that is presented in the query image $I_q$, the corresponding location of the concept in the codebook is found out. 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 concept 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 4. In this way, the response time is reduced while the retrieval accuracy is still maintained.

## 5. Experiments

To evaluate the effectiveness and efficiency of the proposed concept-based image representation and similarity matching approaches, exhaustive experiments were performed on two different benchmark image collections under ImageCLEF [1]. The first collection is termed as the IAPR TC-12 benchmark that was created under the Technical Committee 12 (TC-12) of the International Association of Pattern Recognition (IAPR) Grubinger et al. (2006) and it has been used for the ad-hoc photographic retrieval task in ImageCLEF'07 Grubinger et al. (2007). This collection is publicly available for research purposes and currently contains around 20,000 photos taken from locations around the world that comprises a varying cross-section of still natural images. The domain of the images in this collection is very generic that covers a wide range of daily life situations. Unlike the commonly used COREL images, this collection is very general in content with many different images of similar visual content but varying the illumination, viewing angle and background. This making it more challenging for the successful application of image retrieval techniques.

The second collection contains more than 67,000 bio-medical images of different modalities from the RSNA journals Radiology and Radiographics under ImageCLEFmed'08 Müller et al. (2008). For each image, the text of the figure caption is supplied as free text. The contents of this collection represent a broad and significant body of medical knowledge, which make the retrieval more challenging. As the entire collection contains variety of imaging modalities, image sizes, and resolutions, it makes really difficult to perform similarity search based on the current CBIR techniques.

The training set used for the SOM-based learning consists of only around 1% images of each individual data set. We set the initial learning rate as $\alpha = 0.07$ and iteration number as 300 due to their better performances. Based on the retrieval results, we found the the optimal combination when the images are partitioned as 64 sub-images and a codebook size of 225 ($15 \times 15$) units is used. Hence, the images are indexed based on this configuration for both the collections for the experimental pupose.

For a quantitative evaluation of the retrieval results, we used *"query-by-example"* as the search method where the query images are specified by the topics that were developed by the CLEF organizers. Each topic is a short sentence or phrase describing the search topic with one to three "relevant" images. The query topics are equally subdivided into three categories: visual, mixed, and semantic Grubinger et al. (2007); Müller et al. (2008). A total of 60 topics were provided by the ImageCLEF'07 Grubinger et al. (2007) for the ad-hoc retrieval of general photographic images with a short description of each topic. Similarly, for the ad-hoc medical image retrieval in ImageCLEF'08, a total of 30 query topics were provided Müller et al. (2008) that were initially generated based on a log file of Pubmed [2].

Results for different methods are computed using the latest version of TREC-EVAL [3]

---

[1]http://imageclef.org/
[2]http://www.pubmed.gov
[3]http://trec.nist.gov/$trec-eval$/

| Method | MAP | GMAP | P(20) |
|---|---|---|---|
| EHD-Euclid | 0.0131 | 0.0028 | 0.0467 |
| CLD-Euclid | 0.0198 | 0.0040 | 0.0642 |
| Concept-Euclid | 0.0164 | 0.0028 | 0.0525 |
| Concept-Quad(A) | 0.0179 | 0.0037 | 0.0583 |
| Concept-Quad(S) | 0.0178 | 0.0035 | 0.0575 |
| Concept-Quad(A)-IF | 0.0176 | 0.0036 | 0.0581 |
| Concept-Quad(A)-IF($LN_2$) | 0.0178 | 0.0037 | 0.0583 |

Table 1. Retrieval Results (Photographic Collection)

software based on the relevant sets of all topics, which were crated by the CLEF organizers by considering top retrieval results of all submitted runs of the participating groups in ImageCLEF'08 Müller et al. (2008); Grubinger et al. (2007). Results were evaluated using un interpolated (arithmetic) Mean Average Precisions (MAP)to test effectiveness, Geometric Mean Average Precision (GMAP) to test robustness, and Precision at rank 20 (P20) because most online image retrieval engines like Google, Yahoo, and Altavista display 20 images by default.

## 6. Results

This section presents the experimental results of the retrieval approaches with and without using the correlation-enhanced similarity matching and inverted indexing schemes. The performances of the different search schemes are shown in Table 1 and Table 2 for the retrieval of the photographic and medical collections respectively based on the query image sets as discussed previously. The proposed correlation-enhanced similarity matching approach is compared (using different matrices) to the case when only Euclidean distance measure is used in the visual concept-based feature (e.g., method "Concept-Euclid") space. In addition, we consider the MPEG-7 based Edge Histogram Descriptor (EHD) and Color Layout Descriptor (CLD) Chang et al. (2001) (e.g., methods "EHD-Euclid" and "CLD-Euclid") and compare our search approach with these features based on the Euclidean distance measure. The EHD

| Method | MAP | GMAP | P(20) |
|---|---|---|---|
| EHD-Euclid | 0.0095 | 0.0004 | 0.0333 |
| CLD-Euclid | 0.0006 | 0.0001 | 0.0050 |
| Concept-Euclid | 0.0248 | 0.0009 | 0.0300 |
| Concept-Quad(A) | 0.0255 | 0.0014 | 0.0433 |
| Concept-Quad(S) | 0.0256 | 0.0012 | 0.0467 |
| Concept-Quad(A)-IF | 0.0251 | 0.0012 | 0.0430 |
| Concept-Quad(A)-IF($LN_2$) | 0.0254 | 0.0014 | 0.0433 |

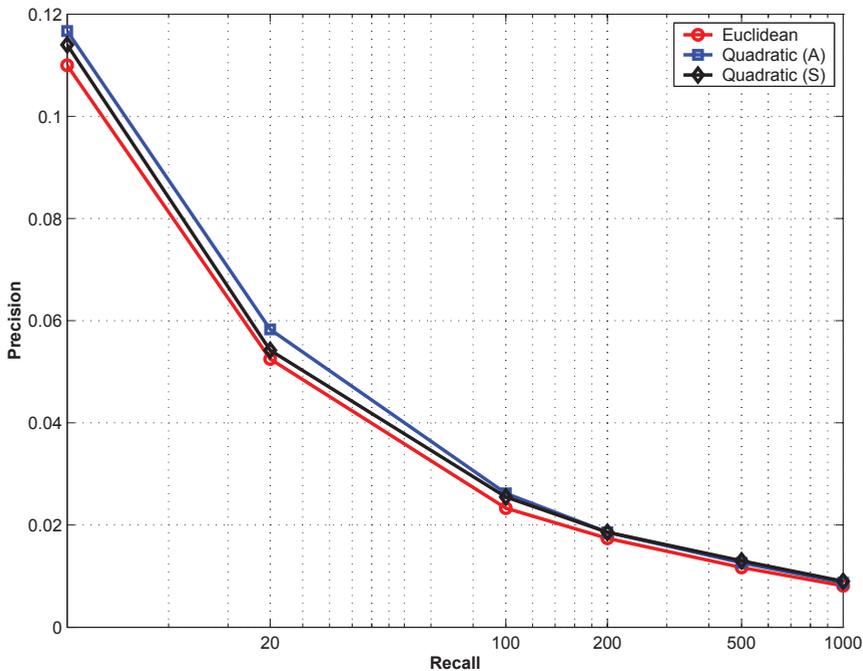Table 2. Retrieval Results (Medical Collection)

Fig. 5. Precisions at different ranks for the photographic collection

represents local edge distribution in an image by dividing the image into $4 \times 4$ sub-images and generating a histogram from the edges present in each of these sub-images. The CLD represents the spatial layout of the images in a very compact form. It is obtained by applying the discrete cosine transformation (DCT) on the 2-D array of local representative colors in the *YCbCr* color space where *Y* is the luma component and *Cb* and *Cr* are the blue and red chroma components Chang et al. (2001).

We can observe from Table 1 and Table 2 that the retrieval performances in general have improved on visual concept space for both the collections based on the different performance measures. Only the MPEG-7 based CLD feature performed better with a MAP score of 0.0198 in the photographic image collection due to the presence of many color images of natural scenery, whereas it performed worst in the medical collection due to the lack of presence of such images. On the other hand, the performance of the visual concept feature is quite consistent in both the collections due to the incorporation of both color and texture features for the codebook generation process based on the SOM learning. Overall, the retrieval results in both the collections in terms of different performance measures are quite comparable with the results of the participant of the previous ImageCLEF Müller et al. (2008); Grubinger et al. (2007). The low precision scores as obtained are due to the nature of the image collections and the complexity of the query topics.

In addition, we can observe the improvement in performances in terms of the MAP, GMAP, and BPREF scores, when searches are performed with different correlation-enhanced similarity matching functions based on using the global matrices: **S** (e.g., method "Concept-Quad(S)"), and matrix **A** (e.g., method "Concept-Quad(A)"), for both collections.
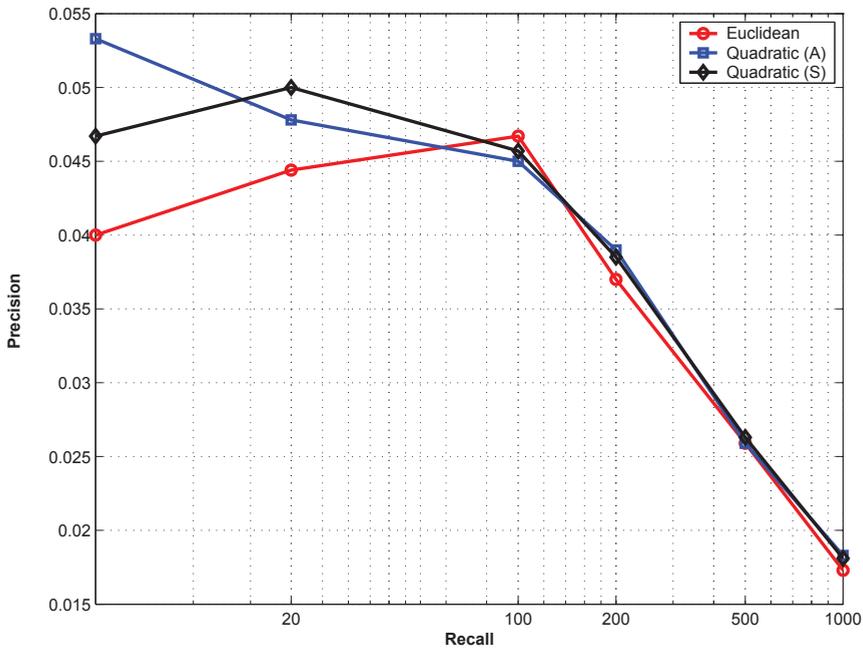
Fig. 6. Precisions at different ranks for the medical collection

Figure 5 and Figure 6 show the precision at different rank position (e.g., 5, 20, 100, 200, 500, and 1000) for the photographic and medical collections respectively. For a better visibility, the X-axis is represented by the logarithmic scale for both the figures. Although the precision curves are looking different in Figure 5 and Figure 6, we can conjecture one thing about the nature of the curves. It is that the precision in initial rank positions (up to 100) are comparatively better (especially in medical collection) for the quadratic distance measures when compared to the Euclidean distance measure in the concept space. This improvement in performance is important as users are usually interested only on the first few top retrieved images.

It is also noticeable that the performances of the retrieval result are decreased slightly for both collections when searches are performed in the inverted index with the quadratic distance measure based on the global connection matrix (e.g., method "Concept-Quad(A)-IF") and the performances are almost comparable to linear search when the matching are performed in the modified inverted index based on a local neighborhood of $LN_2$. The major gain in searching on a inverted index is that it takes less computational time compared to a linear search in the entire collections. Hence, to test the efficiency of the search schemes for the concept-based feature, we also compared the average retrieval time (in milliseconds) with and without the indexing scheme (in an Intel Pentium 4 processor with Windows XP as the operating system and 2 GB memory) for both the query sets.

From the results in Table 3, it is observed that the search with the quadratic distance measure with global matrix **A** in the inverted index of concept feature, is about two times faster as compared to the linear search for both the data sets. Although, it took longer time to perform search in the medical collection due to the presence of around three times more images

| DataSet | Euclid-Linear | Quad-Linear | Quad (A)-IF | Quad (A)-IF($LN_2$) |
|---------|---------------|-------------|-------------|---------------------|
| Photograph | 126 | 800 | 436 | 520 |
| Medical | 580 | 2470 | 1320 | 1570 |

Table 3. Average retrieval time (ms) for the query images

compared to the photographic collection. However, the percentage of improvement in terms of using an inverted index is almost same for both the collections. In addition, we found a trade-off in between the time and precision by performing the same search in the modified inverted index with a local neighborhood of $LN_2$. By observing closely in Table 1, Table 2, and Table 3, it is found that the MAP scores are increased slightly with the expense of a little longer search time when compared to the search without the modification of the inverted index. Hence, the quadratic distance matching in the modified inverted index with the query expansion has proved to be both effective and efficient.

## 7. Conclusions

We have investigated a correlation-enhanced similarity matching and a query expansion technique in CBIR domain, inspired by the ideas of the text retrieval domain in IR. The proposed technique exploit the similarities/correlations between the concepts based on a global analysis approach. Due to the nature of the image representation schemes in the concept-based feature spaces, there always exists enough correlations between the concepts. Hence, exploiting this property improved the retrieval effectiveness. For the feature representation, we limited our approaches by modeling only the intermediate level visual concepts. This limitation is obvious due to the current state of the object recognition techniques in the broad domain images. It would be more effective, if specific objects can be identified in large collections irrespective of their variations and occlusions. However, the main focus of our approach is to perform the retrieval that can exploit concept correlations at the global level. In future, when the object recognition techniques will be mature enough to a certain level, our approaches could be easily extendible to a higher level concept-based representation.

## 8. References

Smeulders, A.; Worring, M.; Santini, S.; Gupta, A. & Jain, R. (2000). Content-Based Image Retrieval at the End of the Early Years. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 12, 2000, 1349–1380

Liua, Y.; Zhang, D.; Lu, G.; & and Ma, W. Y. (2007). A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, Vol. 40, 2007, 262–282

Datta, R.; Joshi, D.; Li, J.; & Wang, J. Z. (2008). Image Retrieval: Ideas, Influences, and Trends of the New Age. *ACM Computing Surveys*, Vol. 40, No. 2, 2008, 1–60

Zhu, L.; Zhang, A.; Rao, A.; & Srihari, R. (2002). Theory of keyblock-based image retrieval. *ACM Transactions on Information Systems*, Vol. 20, No. 2, April 2002, 224–257, ISSN:1046-8188

Lim, J. H. (2001). Building visual vocabulary for image indexation and query formulation. *Pattern Analysis & Applications (Special Issue on Image Indexation)*, Vol. 4, No. 2-3, 125–139, DOI: 10.1007/PL00014574

Jing, F.; Li, M.; Zhang, H. J. & Zhang, B. (2004). An efficient and effective region-based image

retrieval framework. *IEEE Transaction on Image Processing*, Vol. 13, No. 5, May 2004, 699–709, PMID: 15376601

Vogel, J. & Schiele, B. (2007). Semantic Modeling of Natural Scenes for Content-Based Image Retrieval. *International Journal of Computer Vision* , Vol. 72, No. 2, April 2007, 133–157, ISSN:0920-5691

Shi, R.; Feng, H.; Chua, T. S. & Lee, C. H. (2004). An adaptive image content representation and segmentation approach to automatic image annotation. *Proceedings Lecture Notes in Computer Science*, Vol 3115, Springer 2004, ISBN 3-540-22539-0, Third International Conference on Image and Video Retrieval (CIVR 2004), Dublin, Ireland, July 21-23, 2004.

Rahman, M. M.; Bhattacharya, P. & Desai, B. C. (2009). A Unified Image Retrieval Framework on Local Visual and Semantic Concept-Based Feature Spaces. *Journal of Visual Communication and Image Representation*, Vol. 20, No. 7, October 2009, 450–462, ISSN:1047-3203

Rahman, M. M.; Antani, S. K.; & Thoma, G. R. (2009). A Medical Image Retrieval Framework in Correlation Enhanced Visual Concept Feature Space. *Proceedings of 22nd IEEE International Symposium on Computer-Based Medical Systems (CBMS)*, August 3-4, 2009, Albuquerque, New Mexico, USA.

Bosch, A.; Muoz, X.; & Mart, R. (2007). Which is the best way to organize/classify images by content?. *Image and Vision Computing*, Vol. 25, No. 6, June 2007, 778–791, doi:10.1016/j.imavis.2006.07.015

Fukunaga, K. (1990). *Introduction to Statistical Pattern Recognition*, 2nd Ed. Academic Press, ISBN:0-12-269851-7

Kohonen, T. (1997). *Self-Organizing Maps*, New York, Springer-Verlag, ISBN: 978-3-540-67921-9

Yates, R. B. & Neto, B. R. (1999). *Modern Information Retrieval*, 1st ed., Addison Wesley, ISBN-13: 978-0201398298

Yasushi, O.; Tetsuya, M. & Kiyohiko, K. (1991). A fuzzy document retrieval system using the keyword connection matrix and a learning method. *Fuzzy Sets and Systems*, Vol. 39, No. 2, 163–179

Haralick, R. M.; Shanmugam & Dinstein, I. (1973). Textural features for image classification. *IEEE Transactions on Systems, Man and Cybernetics*, Vol. SMC-3, No. 6, 610-621

Chang, S. F.; Sikora, T. & Puri, A. (2001). Overview of the mpeg-7 standard. *IEEE Transactions on Circuits and Systems for Video Technology*, Vol. 11, No. 6, June 2001, 688–695, ISSN: 1051-8215

Laaksonen, J.; Koskela, M. & Oja, E. (2002). PicSOM: Self-Organizing Image Retrieval With MPEG-7 Content Descriptors. *IEEE Transaction On Neural Networks*, Vol. 13, No, 4, Jul 2002, 841–853, ISSN: 1045-9227

Vesanto, J. (1999). SOM-Based Data Visualization Methods, *Intelligent Data Analysis*, Vol. 3, No. 2, August 1999, 111–126, doi:10.1016/S1088-467X(99)00013-X

Yen, G. G. & Zheng, W. (2005). Ranked Centroid Projection: A Data Visualization Approach With Self-Organizing Maps, *IEEE Transaction On Neural Networks*, Vol. 19, No. 2, Feb. 2008, 245–259, ISSN: 1045-9227

Gaede, V. & Gunther, O. (1998). Multidimensional Access Methods. *ACM Computing Surveys*, Vol. 30, No. 2, June 1998, 170–231, ISSN:0360-0300

Müller, H.; Squire, D. M.; Mueller, W. & Pun, T. (1999). Efficient access methods for content-based image retrieval with inverted files. *Proceedings of SPIE*, Vol. 3846,

461–472, Multimedia Storage and Archiving Systems IV, Sethuraman Panchanathan; Shih-Fu Chang; C.C. J. Kuo; Eds.

Hafner, J.; Sawhney, H. S.; Equitz, W.; Flickner, M. & Niblack, W. (1995). Efficient color histogram indexing for quadratic form distance functions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 17, No. 7, Jul 1995, 729-736, ISSN: 0162-8828

Grubinger, M.; Clough, P.; Müller, H. & Deselears, T. (2006). The IAPR-TC12 benchmark: A new evaluation resource for visual information systems. *Proceedings of International Workshop on Language Resources for Content-Based Image Retrieval*, 13–23, 2006.

Grubinger, M.; Clough, P.; Hanbury, A. & Müller, H. (2007). Overview of the ImageCLEF 2007 Photographic Retrieval Task. *Working Notes of the 2007 CLEF Workshop*, Sep., 2007, Budapest, Hungary.

Müller, H.; Cramer, J. K.; Kahn, C. E.; Hatt, W.; Bedrick, S. & Hersh, W. (2008). Overview of the ImageCLEFmed 2008 medical image retrieval task. *9th Workshop of the Cross-Language Evaluation Forum, CLEF 2008*, Aarhus, Denmark.
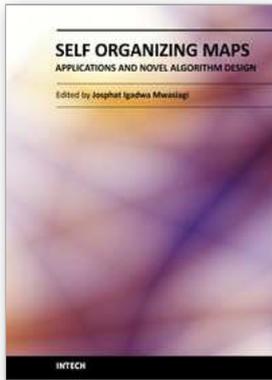
**Self Organizing Maps - Applications and Novel Algorithm Design**
Edited by Dr Josphat Igadwa Mwasiagi

Kohonen Self Organizing Maps (SOM) has found application in practical all fields, especially those which tend to handle high dimensional data. SOM can be used for the clustering of genes in the medical field, the study of multi-media and web based contents and in the transportation industry, just to name a few. Apart from the aforementioned areas this book also covers the study of complex data found in meteorological and remotely sensed images acquired using satellite sensing. Data management and envelopment analysis has also been covered. The application of SOM in mechanical and manufacturing engineering forms another important area of this book. The final section of this book, addresses the design and application of novel variants of SOM algorithms.

**How to reference**
In order to correctly reference this scholarly work, feel free to copy and paste the following:

Mahmudur Rahman (2011). Image Search in a Visual Concept Feature Space with SOM-Based Clustering and Modified Inverted Indexing, Self Organizing Maps - Applications and Novel Algorithm Design, Dr Josphat Igadwa Mwasiagi (Ed.), ISBN: 978-953-307-546-4, InTech, Available from:
http://www.intechopen.com/books/self-organizing-maps-applications-and-novel-algorithm-design/image-search-in-a-visual-concept-feature-space-with-som-based-clustering-and-modified-inverted-index

# INTECH
open science | open minds