

Color Patterns for Pictorial Content Description

Daniela Stan
Intelligent Information Engineering Laboratory
Dept. of Computer Science & Engineering
Oakland University
Rochester, Michigan 48309-4478
1-248-370-2137
dstan@oakland.edu

Ishwar K. Sethi
Intelligent Information Engineering Laboratory
Dept. of Computer Science & Engineering
Oakland University
Rochester, Michigan 48309-4478
1-248-370-2200
isethi@oakland.edu

ABSTRACT

In this paper, we propose a new type of image feature, which consists of *patterns of colors and intensities* that capture the latent associations among images and primitive features in such a way that the noise and redundancy are minimized. Incorporating our feature model into a Content-based Image Retrieval (CBIR) system moves the research in image retrieval beyond simple matching of images based on their primitive features and creates a ground for learning image semantics from visual content. A system developed using our proposed feature model, will have the capability of learning associations between not only semantic concepts and images, but also between semantic concepts and patterns. We evaluated the performance of our system based on the retrieval accuracy and on the perceptual similarity order among retrieved images. When compared to standard image retrieval methods, our preliminary results show that, even if the feature space was reduced to a significantly lower dimensional space, the accuracy and perceptual similarity for our system remain the same or better depending on the category of images.

Keywords

Latent semantic indexing, content-based image retrieval, clustering, annotation, color patterns

1 INTRODUCTION

In the last ten years, the multimedia super-highway has expanded exponentially, bringing vast repository of information to the desktop in a few mouse clicks; therefore, there is an ever-growing demand for tools to locate information by content with greater accuracy and efficiency. In particular, methods for CBIR have drawn the most attention as many of the underlying techniques

can be easily applied to other multimedia artifacts with some suitable modifications. A CBIR system can be viewed as two main components: feature extraction and the search for similar images in a feature space. The approach that we are presenting in this paper falls within the first component of a CBIR system. Using Latent Semantic Indexing (LSI) technique from Information Retrieval, we propose a new type of image feature that is meant to capture the hidden associations among visual feature elements within an image and across the image database.

Several attempts were made to exploit the benefits of LSI in the CBIR field. In 1998, M. La Cascia et al. [3] proposed the LSI technique to extract semantic content in the form of keywords from the text surrounding the images on the web (alt-tags, title, near text). Then, the LSI textual representation for an HTML document was associated with each visual representation of the images contained therein. The visual representation was derived using color histogram and dominant orientation histogram and Principal Component Analysis (PCA) technique was used to reduce the dimension of the visual features. Their experimental results show that the retrieval accuracy increases when both visual and textual information are used. In 1999, T. Westerveld et al. [13] proposed an integrated framework, in which first the terms and the visual content are combined and then LSI is applied on the unified vector to capture the associations between images and text. The difference between PCA and Singular value Decomposition (the underlying statistical technique on which LSI is based) is that PCA is applicable on multiple observations while SVD is applicable on single observations [6].

The two approaches assume that some textual information is provided with the images. However, sometimes there is little to no textual information available for an image database. Therefore, there is a need for CBIR systems that exploits the correlations present in the visual content of images, extracted from the raw pixel data.

We propose a new type of image feature that will allow retrieval and semi-automatic across-image annotation based on groups of image data rather on mere pixels. The idea of grouping image data for semi-automatic across-image labeling was also used by T.P. Minka and R.W. Picard in [4]. They used a clustering approach to form within-image groups of image regions and, since clustering deals with only one type of entity (feature or image), the clustering approach had to be applied one more time on the within-image groups in order to derive the across-image groupings. Our approach utilizes LSI's property of dealing with

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SAC 2002, Madrid, Spain.

© 2002 ACM 1-58113-445-2/02/03...\$5.00.

different types of entities at the same time and thus, both low-level feature elements and images are considered together as the input. The output is a new feature space in which features and images are expressed as points, and the coordinates of each point derive the importance of the corresponding feature element in forming the new patterns. If there are any groups of feature element points or image points in this space, they will produce *within or across contexts* for the image database, respectively. Therefore, in addition to the dimension reduction property of LSI previously used by W. Niblack et al. in [5] to achieve scalable retrieval performance, we exploit the LSI property of creating clusters of co-occurring features having the goal of generating image patterns, the analog of concepts derived from co-occurring terms in text retrieval. Fig.1 shows an example of a cluster of color bins and images obtained applying LSI on the color histograms used as images' visual content representations.



Figure 1: Example of a region in the LSI space used for image content interpretation. The small rectangles represent the color bins modeling the real color patterns present in the images from the corresponding region. The bins depicted produce a within context while the images produce an across context.

Since the images belonging to the cluster have the same semantic meaning, the textual characterization, 'sunset', can also be associated with the color pattern modeled by the cluster's bins. This association shows how the proposed approach can be used as an important step towards transforming the low-level features to meaningful high-level features.

As a result, instead of looking at an image as a meaningless pixel matrix and at a database as a sequence of images that make no sense when placed near each other, we create contexts in which meaningful perceptual impressions can be formed about the content and the similarity among images. In our approach, the *within-image groupings will not serve for segmentation as in [4], but rather for recognition of patterns of color and intensities in the image.* In this paper, we emphasize on the image pattern extraction and its application to image retrieval and annotation will be presented in more details in a future paper.

2 LOW-LEVEL FEATURE EXTRACTION

We implement and describe our system using only color information, but any other feature (visual or text -based) can be easily incorporated in our system since the information contained

in LSI is given by the notion of occurrence, and not by numeric abstract values [2]. We choose global histograms to represent the color information. The advantages of using global histograms are their robustness to translation and rotation about the viewing axis, slow modification with change in viewpoint and scale, and occlusion [11]. To capture the human perception into the representation, we choose the HSV space to represent the color information; this space correlates well with the human perception and is commonly used by artists to represent color information present in images [1]. After the color space transformation, three histograms (for hue, saturation, and value) are calculated and thus, every image is encoded by a m -dimensional feature vector ($m = 3 \times M$, M stands for the number of bins of every histogram). An analysis of the importance of the bins within a feature vector representation shows that not all bins are equal for describing image content. To capture this fact, an information-theoretic weighting scheme can be applied [2].

It is worth mentioning here that LSI extracts the latent relationships among different bins from their relative usage without considering their order or spatial information. This motivated us to use a global histogram representation instead of a local one; however, a global histogram loses the spatial information and thus, the discrimination power of the histogram is saturated in the context of very large image databases. To keep up performance, additional low-level features (such as texture, shape, or directed edges) can be incorporated to bring more information about images from the database. The explosion of dimensions of the histogram, produced by the incorporation of additional features, is overcome by the LSI property of considering the quality of the information along each dimension; LSI reduces the high dimensional low-level feature space such that noise and redundancy are eliminated.

3 PATTERN EXTRACTION

3.1 Singular Value Decomposition

Singular Value Decomposition (SVD), the statistical method on which LSI is based, is performed on a matrix W_0 , whose rows stand for bins and columns for images, and each entry represents the weight of a given bin in a given image. By definition, the SVD of W_0 is a factorization of the form [12]:

$$W_0 = T_0 \times \Sigma_0 \times D_0', \quad (1)$$

where T_0 , D_0 are two $m \times m$ and $n \times n$ (n is the number of images from database) orthonormal matrices respectively. Σ_0 is a $m \times n$ diagonal matrix, $\Sigma_0 = \text{diag}(\sigma_1, \sigma_2, \dots, \sigma_p)$, $p = \min(m, n)$, with the diagonal elements having the property that $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$ and being the singular values of W_0 .

The main idea behind SVD is that proper choice of T_0 and D_0 makes most of σ_i zero; that is, most of the important information gets concentrated in a few dimensions. Let k , $k \leq p$, be the number of the first dimensions that contain this information; the remaining smaller singular values are set to zero. Since zeros were introduced into Σ_0 , the representation can be simplified by deleting the rows and columns of Σ_0 to obtain a reduced

diagonal matrix Σ , and then deleting the corresponding columns of T_0 and D_0' to obtain T and D' , respectively. This results in a reduced model:

$$W = T \times \Sigma \times D', \quad (2)$$

which gives the rank- k model with the best possible least-squares-fit to W_0 [12].

3.2 Color Patterns

Let us consider the interpretation and the advantages of LSI for our image information extraction purposes. It is important for LSI technique that the derived W matrix does not reconstruct the original bin by image matrix W_0 exactly. It rearranges the color bin-space to reflect the major color association patterns in the data and reduces the noise and redundancy present in the image data. Since the columns $t^j, j = 1 \dots k$, of T form a basis for the space spanned by W 's columns [12], they can be considered the axis of the rearranged space. *First axis, t^1 , reflects the first major pattern, named $pattern_1$, present in the image database: t_1^1 shows the contribution of bin_1 in $pattern_1$, t_2^1 shows the contribution of bin_2 in $pattern_1$ and so on up to the last t_m^1 that shows the contribution of bin_m in the first pattern. Second axis, t^2 , reflects the second major pattern in the data, named $pattern_2$, and so on up to axis t^k corresponding to $pattern_k$. Therefore, LSI allows the replacement of individual bins (used initially as descriptors of images) by 'independent patterns' that can be specified by any one of several bins or combination of thereof. A decreasing order of the elements of t^j will reveal the most important bins in forming the $pattern_j, j = 1 \dots k$.*

Let us consider the new image representations in the pattern space. Equation (2) can also be written as:

$$[W^1 W^2 \dots W^n] = (T \times \Sigma) \times (D'^1 D'^2 \dots D'^n), \quad (3)$$

where W^j stands for the weighted bin representation of $Imag^j, j = 1 \dots n$ and $(D')^j$ is the j^{th} column of matrix D' . By definition [12], the product between a matrix and a vector is a combination of matrix's columns. Therefore, every $W^j, j = 1 \dots n$, is a linear combination of the columns of $T \times \Sigma$:

$$W^j = \sum_{l=1}^k d_{jl} (\sigma_l t^l). \quad (4)$$

Equation (4) gives the new representation of $Imag^j, j = 1 \dots n$, in which $d_{j1}, d_{j2}, \dots, d_{jk}$ indicate the strength of association of $Imag^j$ with the discovered $pattern_1, pattern_2 \dots pattern_k$, respectively; when images are represented as points, $d_{j1}, d_{j2}, \dots, d_{jk}$ are the coordinates of the point

representing $Imag^j, j = 1 \dots n$ in the pattern space whose axes t^l are rescaled by $\sigma_l, l = 1 \dots k$.

An example of a color pattern composition is presented in Fig.2 and the images in which the corresponding pattern is mostly present are shown in Fig.3.



Figure 2: The first eleven color bins of a color pattern obtained using LSI approach. The most important bin is the left most one and the importance decreases as you go from left to right.

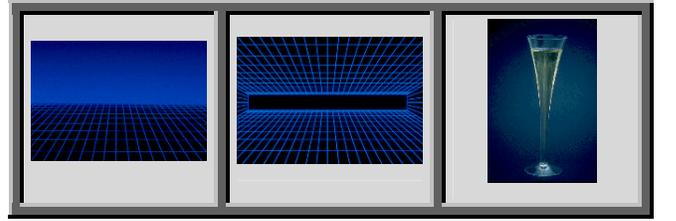


Figure 3: The first three images in which the color pattern from Fig.2 is mostly present.

We would like to point out that the dimension k of the new space represents the *embedded dimension* [7] of the visual data in the original feature space. This dimension should be large enough to fit all the real structure in the data and thus yield good retrieval performance, but small enough to exclude the unimportant details. Experimental results (Section 5) show that the embedded dimension is significantly smaller than the original dimension, which exactly correlates with White and Jain's affirmations [14]. The lower dimension of the pattern space allows overcoming of some problems introduced by the high-dimensionality of the color histograms; even with severe quantization, the histogram feature space can occupy over 100 dimensions [8] which complicates the calculation of the distance functions and requires new indexing structures for image retrieval [7].

Furthermore, our proposed model does not only reduce the dimension of the feature space, but also reveals the colors that although being different, are similarly used across the image database. To better understand this capability of our model, we will use a series of direct comparisons between LSI for text retrieval and LSI for image retrieval. The application of LSI in text retrieval allows words synonymous with each other to be near each other in the LSI space; a group of synonym words defines a concept that later is used to perform retrieval by concepts instead of term matching. For example, terms as 'physician', 'doctor', 'surgeon' (spelled different, but with same meaning) will form a concept and the retrieval results for a query based on the term 'physician' will also contain documents in which 'doctor' or 'surgeon' appears even if the 'physician' word is not present in those documents. We use the same idea in our image model (Table 1). *Bins representing different colors but when similarly used across the image database can be considered as synonymous*

with each other and the pattern that they define is the analogous of a concept from text retrieval.

Table 1: Summary of the analogy between image and text models.

	Text Example	Image Example
Same root/color used in different contexts	Doctor, Doctorate	
Different root/color used in same contexts	Doctor, Physician	

4 IMAGE ANNOTATION AND RETRIEVAL IN THE CONTEXT

Summarizing our model, we obtained a common ground for images and features, specifically the lower k -dimensional pattern space. In this space, both images and bins are uniformly represented as points, and if there are any groups of bins and images in this space, they will be considered as *contexts* for the image database. We use clustering approach [9,10] to perform the process of grouping bin and image points in the pattern space. The group of bin points placed near each other will form a *within-image context* and the points representing images belonging to the same group will form an *across image context*. If these images have the same semantic interpretation, *that semantic meaning can also be associated with the color pattern whose most important bins form the corresponding within-image context*. Thinking about patterns as a new type of image feature situated between the low-level features and the high-level ones (keywords, annotations etc), our model provides information about which low-level feature elements are most important, how they should be combined for a given pattern and which patterns are most important for a given annotation. This makes the proposed CBIR system to be an important step towards semantic retrieval.

A CBIR system having the pattern extraction component incorporated allows mainly two types of image database searches:

- *Search by image example*: the image query is represented in the pattern space and the process of searching for similar images is based on pattern similarity instead of individual bin similarity. We refer to this retrieval as *image retrieval in the context*: even if an image is not indexed by the query low-level features, if it shares some bins with another image that is relevant to the given query, it is likely to be retrieved. This kind of retrieval moves the CBIR system beyond simple matching of images based on primitive features.

- *Search by pattern example*: the most important patterns in the database are presented to the user. The user selects a pattern of interest and then the system provides the images in which the pattern is mostly present. Furthermore, the user chooses an image of interest and the system retrieves the most similar images.

The search by pattern example is very important when a user does not look for a particular image, but for a pattern of colors that will be present in the image; for instance, a web designer who searches an image database to find a background image to use for a web page. He does not have in mind any conceptualization of the image that he is looking for, what he knows is that the 'background should have some blue in it'. Our system can prompt the designer to identify which of the different patterns most closely resembles the wanted result that in turn helps the designer refine his search.

5 EXPERIMENTAL RESULTS

In this section, we present the preliminary results obtained in the evaluation of our approach. We implemented our system using a general-purpose database of 1000 RGB images. First, we transformed the images into HSV color space and then, we calculated separately three 256-bin global histograms (for hue, saturation, and value) for each image. The *svd* code from MATLAB provided us with the set of singular values already arranged in sorted order. This is very convenient to us, because when we vary the number k of singular values to obtain our reduced pattern representation, we want to pick the largest singular values such that the noise and redundancy are eliminated and the best retrieval performance is obtained.

In evaluating the retrieval performance of our approach, we focus only on the local topology of the pattern space and not on the used similarity metric. As a ground truth, for each query, a set of most similar images is assigned together with a relative ranked order of their similarity.

We present the results for a query image (left image from Fig. 4) belonging to a 'landscape' image category from the database. We define an image as being 'landscape', if it contains sky, water, and land. The image database contains 98 images of this category and the second and third images from Fig.4 represent the ground truth in measuring the retrieval accuracy for the chosen query image, q .



Figure 4: The most similar two images with the query image



Figure 5: Images retrieved using histogram intersection on global histograms



Figure 6: Images retrieved when cosine measure is used and k=20

Comparisons between the results of histogram intersection in the original 768-dimensional space (Fig.5) and the results of cosine measure in the pattern space (Fig.6) show that, even if the feature space was reduced to only 20 features, the retrieval accuracy is the same as for the original space. Moreover, if the users are asked to evaluate the performance, the retrieval results from Fig.6 are more perceptual appealing than those from Fig.5 because the images considered as the ground truth appear closer to the query image in Fig.6 than those in Fig.5 do. That is, using color patterns, the images considered as ground truth appear as the first and second best matches (not considering the query image itself), while using global histograms, same images appear as the third and sixth best matches. This observation is very important in the context of image retrieval where we are interested in displaying only the top few retrieval results. For example, if the system had been designed to display, let us say, only the first two retrieved images, no images from the ground truth would have been displayed to the user in the case of global histogram.

In order to quantify the perception issue discussed above, let $GT(q)$ be the ground truth ordered with respect to human perception and $R(q) = \{N_1, \dots, N_p\}$ be the set of the nearest neighbors ordered with respect to a defined similarity measure (cosine or histogram intersection). The parameter $\sigma(q)$ in formula (5) measures how close the most relevant images with the query image are situated in the list $R(q)$ of the nearest neighbors and a lower value of the parameter indicates a better retrieval result:

$$\sigma(q) = \sum_{s_i \in R(q) \cap GT(q)} \text{rank}(N_i, R(q)) \quad (5)$$

The $\text{rank}(\cdot, \cdot)$ notation gives the position of an element in a list and is defined as $\text{rank}(N, R(q)) = i$, whenever $N = N_i$. Moreover, if $\sigma(q)$ is the same in pattern space and original space, an additional objective measure $\Delta(q)$ defined in formula (6) can be considered to quantify the relative positions of the most relevant images and a lower value of $\Delta(q)$ indicating a better retrieval:

$$\Delta(q) = \sum_{i=1}^{L(q)} \left| i - \text{rank}(N_i, GT'(q)) \right| \quad (6)$$

where $L(q)$ is the cardinality of $GT(q) \cap R(q)$ and $GT'(q)$ is the set $GT(q) \cap R(q)$ ordered with respect to human perception (a subset of the ground truth).

Fig.7 gives the normalized values of $\sigma(q)$ and $\Delta(q)$ for landscape queries, for which the precision and recall are the same, and therefore, we calculated sigma ($\sigma(q)$) and delta ($\Delta(q)$) in order to evaluate the overall accuracy for every space*.

* To see the actual queries and retrieved results, visit our home page at <http://iilab-secs.secs.oakland.edu>.

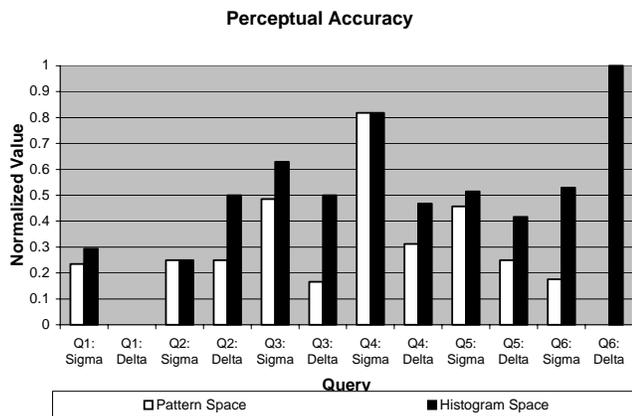


Figure 7: Perceptual accuracy evaluation as calculated with formula 5 and 6.

As an example, considering again the query from Fig.4 (Q_6 in Fig.7), we calculate $\sigma(q)$ and $\Delta(q)$ for the pattern space and the original space. The value of $\sigma(q)$ in the pattern space is 3, while in the original space is 9, verifying the better perceptual retrieval results when considering the pattern space. In the event of needing further evaluation, the value of $\Delta(q)$ will be equal to 0 and 2 for the pattern space and original space, respectively.

6 CONCLUSIONS AND FUTURE WORK

In this paper, we proposed a new type of image feature, patterns of colors and intensities, that incorporated into a CBIR system moves the research in image retrieval beyond simple matching of images based on their primitive features and creates a ground for learning image semantics from visual content. We evaluated the performance of our system based on the retrieval accuracy and on the perceptual similarity order among retrieved images. When compared to standard image retrieval methods, our preliminary results show that, even if the feature space was reduced to a lower dimensional space, the accuracy and perceptual similarity for our system remain the same or better depending on the category of images. Since the used similarity metrics did not take into account the human perception, we believe that the better perceptual similarity of images came from the different nature of the pattern space.

A complete characterization of the time performance goes beyond this paper, but we make a few remarks here. The use of a hierarchy of clusters as an efficient indexing tool and a branch-and-bound algorithm for fast calculation of the nearest neighbors will both speed-up the retrieval process. With respect to the time of calculating the patterns, we will pursue future work in applying faster algorithms for Latent Semantic Indexing. We will also implement our system using other low-level feature in addition to those based on color information and experiment the model on larger image databases.

7 REFERENCES

- [1] Castleman K.R. Digital Image Processing. Prentice Hall, Englewood Cliffs, New Jersey, 1996.
- [2] Deerwester S., Dumais S., Furnas G. et al. Indexing by latent semantic analysis. Journal of American Society for Information Science, vol. 41, 1990, 391-407.

- [3] La Cascia M., Sethi S., and Sclaroff S. Combining textual and visual cues for content-based image retrieval on the world wide web. IEEE Workshop on Content-based Access of Image and Video Libraries, 1998.
- [4] Minka T.P. and Picard R.W. Interactive learning using a "society of models". M.I.T Media Laboratory Perceptual Computing Section. Technical Report, nr. 349, 1996.
- [5] Niblack W., Barber R., Equitz W. et al. The QBIC project: querying images by content using color, texture, shape. Proc. SPIE: Storage and Retrieval for Images and Video Databases vol. 1908, 1993, 173-187.
- [6] Petrou M. and Bosdogianni P. Image Processing: The fundamentals, John Wiley & Sons Ltd, 1999
- [7] Rui Y., Huang T.S., and Chang S.F. Image Retrieval: past, present, and future. Journal of Visual Communication and Image Representation, 10, 1999, 1-23
- [8] Smith J.R. and Chang S.F. Automated image retrieval using color and texture. IEEE Transactions on Pattern Analysis and Machine Intelligence, 1996.
- [9] Stan D. and Sethi I.K. Mapping low-level image features to semantic concepts. Proceedings of SPIE: Storage and Retrieval for Media Databases, 2001, 172-179.
- [10] Stan D. and Sethi I.K. Image retrieval using a hierarchy of clusters. Proceedings of Second International Conference on Advances in Pattern Recognition, 2001, 377-386.
- [11] Swain M. J. and Ballard D. H. Color indexing. International Journal of Computer Vision, 7(1), 1991, 11-32.
- [12] Trefethen L. and Bau D. Numerical linear algebra. SIAM, 1997, 28-29.
- [13] Westerveld T., Hiemstra D., and F. de Jong. Extracting bimodal representations for language-based image retrieval. Multimedia '99, Proceedings of the Eurographics Workshop, 1999, 33-42.
- [14] White D.W. and Jain R. Similarity indexing: algorithms and performance. Proceedings SPIE Storage and Retrieval for Image and Video Databases, 1996.

BIBLIOGRAPHY

Daniela Stan received her B.S. degree in Mathematics from University of Bucharest and M.A. degree in Computer Science from Wayne State University. She is currently a Ph.D. candidate in Computer Science and Engineering at Oakland University. Her research interests include content-based and semantic-based image retrieval, digital libraries, pattern recognition, and data mining and knowledge discovery.

Ishwar K. Sethi is currently Professor and Chair of Computer Science and Engineering at Oakland University. His research interests include pattern recognition, data mining, and multimedia information processing and indexing. Professor Sethi serves on the editorial boards of several international journals including *IEEE Trans. Multimedia*. He is a fellow of IEEE.