Improving performance of content based image retrieval system with color features

Aleš Hladnik[1], **Ante Poljičak**[2]

[1]Faculty of Natural Sciences and Engineering, University of Ljubljana, Slovenia

[2]Faculty of Graphic Arts, University of Zagreb, Croatia

Abstract

Content based image retrieval (CBIR) encompasses a variety of techniques with a goal to solve the problem of searching for digital images in a large database by their visual content. Applications where the retrieval of similar images plays a crucial role include personal photo and art collections, medical imaging, multimedia publications and video surveillance. Main objective of our study was to try to improve the performance of the query-by-example image retrieval system based on texture features – Gabor wavelet and wavelet transform – by augmenting it with color information about the images, in particular color histogram, color autocorrelogram and color moments. Wang image database comprising 1000 natural color images grouped into 10 categories with 100 images was used for testing individual algorithms. Each image in the database served as a query image and the retrieval performance was evaluated by means of the precision and recall. The number of retrieved images ranged from 10 to 80.

The best CBIR performance was obtained when implementing a combination of all 190 texture- and color features. Only slightly worse were the average precision and recall for the texture- and color histogram-based system. This result was somewhat surprising, since color histogram features provide no color spatial information. We observed a 23% increase in average precision when comparing the system containing a combination of texture- and all color features with the one consisting of exclusively texture descriptors when using Euclidean distance measure and 20 retrieved images. Addition of the color autocorrelogram features to the texture descriptors had virtually no effect on the performance, while only minor improvement was detected when adding first two color moments – the mean and the standard deviation. Similar to what was found in the previous studies with the same image database, average precision was very high in case of dinosaurs and flowers and very low with beach, food, monuments and mountains images.

Keywords: content based image retrieval, texture features, color features, image database

1. Introduction

Image retrieval [1] aims to address the problem of browsing, searching and retrieving images from a large database of still or video images. Applications where the retrieval of similar images plays a crucial role include personal photo and art collections, medical imaging, multimedia publications and video surveillance. Traditionally, two approaches have been adopted (Fig. 1). In content based image retrieval (CBIR) images are compared based on their low level descriptors (attributes), e.g. color, texture and/or shape. An alternative strategy is referred to as semantic – or textual – image retrieval and makes use of higher-level features or metadata such as keywords, tags, or textual descriptions associated with the image. Although this framework was widely used in the past and has been increasingly implemented in a combination with CBIR in recent years [2], we will restrict our discussion to the former approach.

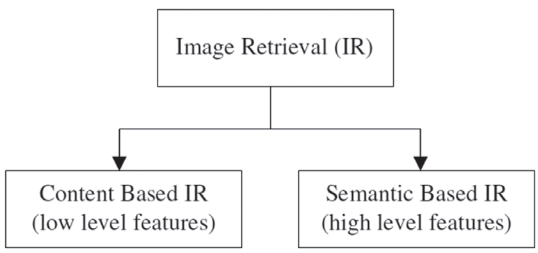


Fig. 1: Two basic image retrieval strategies.

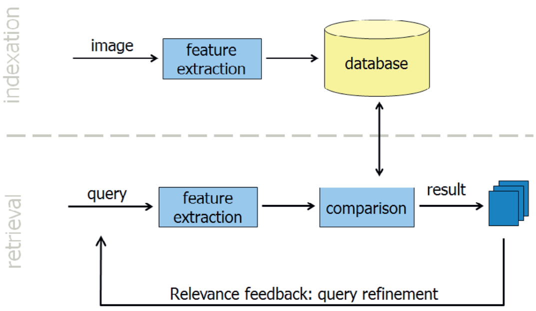


Fig. 2: Content based image retrieval workflow [3].

A general CBIR workflow is shown in Fig. 2. In the first – indexing – step, extraction of features for each and every image in the database is performed. The second step – searching or retrieval – consists of comparing the feature vector of the user-supplied image (so called query image) with those representing the other images and then, using a suitable similarity/distance measure, ranking images according to their relevance, or similarity, to the query image. In addition, in order to reduce the „semantic gap“ and to improve the obtained results, an option to provide a human feedback may be available, e.g. by the user‘s labelling the resulting images as „relevant“, „irrelevant“, or „neutral“ to the search query and then repeating the search with this new information.

1.1 Texture and color features

A typical CBIR system utilizes the elements of visual content of an image – descriptors – such as color, shape, texture, spatial location, salient points or any other information that can be derived from the image. The literature on the techniques, tools and algorithms developed during the last two decades is extremely rich [4], [5]. In the presented study we focused on a combination of texture and color features to generate an abstract representation of images – in a form of feature vectors – that will allow for their high discrimination and provide matches to the query image to be as relevant as possible.

Gabor filters/Gabor wavelets [6] and Wavelet transform [7] are two common spectral texture approaches adopted in a variety of modern CBIR systems. Gabor filter is a linear filter, similar to the Gaussian: a 2D Gabor filter is in the spatial domain a Gaussian kernel function modulated by a sinusoidal plane wave. It was shown that the visual cortex cells in mammalian brains can be modelled by Gabor functions [8], leading to an assumption that their application in image processing is similar to the perception of the human visual system. Gabor wavelets are a convenient expansion of the Gabor filters, where a set of Gabor filters with different frequencies and orientations is used in order to extract useful features from an image. In addition, applying wavelet tranform to an image and computing first few moments of the transform coefficients is another viable strategy how to obtain texture-related image features.

Furthermore, image description can be enhanced by including image-specific color information. Color histogram computation [9] is a simple technique for comparing color content of the query image to that of the database images. Color histogram contains occurrences of each color in a particular color space obtained counting all image pixels having that color. The method is known to be relatively insensitive to small changes in viewing conditions, but fails to incorporate spatial characteristics of the image colors. Before extracting color histogram, the image might be converted from RGB to a color space that is closer to a human perception, such as HSV, Lab or YCbCr. The number of the bins for each color component is usually decreased (quantization) from 255 to a lower number to reduce the algorithm computation time.

Color autocorrelogram [10] expresses how the spatial correlation of pairs of colors changes with distance and therefore, unlike color histogram, includes color spatial information.

Since the distribution of color in an image can be represented as a probability distribution, which in turn can be characterised by a number of unique moments [11], the idea of using first few color moments for image color description is logical. Usually the first – E_i, the average color in the image – and the second – σ_i, the standard deviation – color moments are computed:

$$E_i = \sum_{j=1}^{N} \frac{1}{N} p_{ij}$$

$$\sigma_i = \sqrt{\left(\left(\frac{1}{N} \right) \sum_{j=1}^{N} \left(p_{ij} - E_i \right)^2 \right)}$$

where N = number of pixels in the image, p_{ij} = value of the j-th pixel of the image at the i-th color channel.

To determine visual similarity between two images, the distance between feature representation of the query image and that of the image in the dataset is computed. If this distance is small, the images are considered similar. A number of distance measures (metrics) are available; some of the more frequently implemented are: Manhattan (= L1) Euclidean (= L2), Chebyshev and Cosine angle distance [12].

For evaluating retrieval performance of a particular CBIR algorithm, two parameters are predominantly used: the precision (P) and the recall (R). They are defined as follows [13]:

$$P = n/L \text{ and } R = n/M$$

where n is the number of relevant images retrieved, L is the number of retrieved images, and M is the number of all relevant images in the database. Other performance parameters include error rate (= number of irrelevant images retrieved/L), retrieval efficiency [3], etc.

2. Experimental

Our CBIR system was tested on a publicly available Wang database [14]. This is a collection of 1000 general-purpose images from the Corel Stock Photo Library that are saved in JPEG format of size 384x256 or 256x386 pixels. They belong to 10 distinct categories (classes) each of which is represented by 100 typical images. This simplifies the performance evaluation: given a query image, it is assumed that the user is searching for images from the same category, and therefore all images from the same category are considered relevant while the images from all other classes are considered irrelevant. In other words, M = 100.

Description of the texture and color features that were extracted from the database images is given in Table 1. The initial image retrieval system, and its feature vector, consisted of 48 + 40 = 88 texture features only (denoted as *Txtr* in the *Results and discussion* section). Our second dataset comprised a combination of 88 texture features and 32 color histogram features resulting in a 120-dimensional vector (*TxtrChist*). Third and fourth datasets were characterized by a combination of texture- and 64 color autocorrelogram features (*TxtrCauto*) and texture- and 6 color moments features (*TxtrCmom*), respectively. Feature scaling to the range (0-1) was performed in order to level the contributions of individual groups of textures and color features. Implementation of the algorithms was done in MATLAB based on the code available at MATLAB Central [15].

Table 1: Description of the texture and color features used in the study.

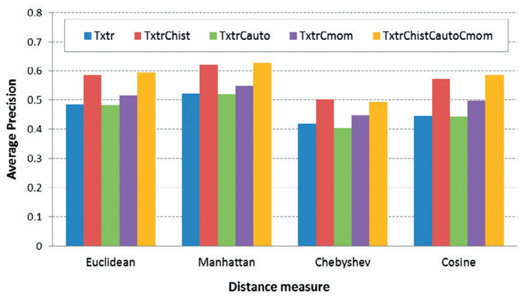
Features	Description	No. of features
Gabor wavelet	Application of Gabor wavelet filters spanning four scales (0.05, 0.1, 0.2 and 0.4) and six orientations ($\theta_0 = 0$, $\theta_{n+1} = \theta_n + 6/\pi$) to the image. Individual features are the mean and the standard deviation of the wavelet coefficients.	48
Wavelet moments	Wavelet transform with a 3-level decomposition is applied to the image. Individual features are the mean and the standard deviation of the transform coefficients.	40
Color histogram	RGB → HSV conversion followed by quantization of H, S and V components into 8, 2 and 2 bins, respectively, and computing histograms.	32
Color autocorrelogram	Quantization of the RGB image into 4 * 4 * 4 = 64 colors and computation of color autocorrelograms.	64
Color moments	Computation of the mean and the standard deviation – first two moments – from the R, G and B color components.	6

3. Results and discussion

In the experiment each image in the database was used as a query image. Each query returned top L results, i.e. images (L = 10, 20, …, 80), from the Wang database, according to an increased dissimilarity of each retrieved image to the query image. To quantify dissimilarity, several distance measures as explained in the *Texture and Color features* section were used. Fig. 3 shows the retrieved images using one particular query image/feature dataset/distance measure combination for L = 20. One can see that P = 11/20 = 0.55 and R = 11/100 = 0.11. Note that $P = R * 5$ when L = 20, $P = R * 4$ when L = 25, etc.

Fig. 4 demonstrates how the distance measure choice influences the performance of our five

CBIR systems for $L = 20$. Both P and R parameters are highest when using Manhattan distance, but trends are very similar with all four similarity metrics: combinations of texture- and all colour features (*TxtrChistCautoCmom*) as well as that of texture- and color histogram features (*TxtrChist*) evidently produce best results followed by *TxtrCmom* and finally *TxtrAuto and Txtr*. It is therefore clear that an addition of color attributes to an exclusively texture-based system has a beneficial effect on CBIR performance. Somewhat surprisingly, since they do not provide color spatial information, color histogram features are found to be almost as important as all three color-based attributes together. By implementing an appropriate feature selection technique that would eliminate irrelevant and/or redundant features though, performance of *TxtrChistCautoCmom* system could likely be improved [13].



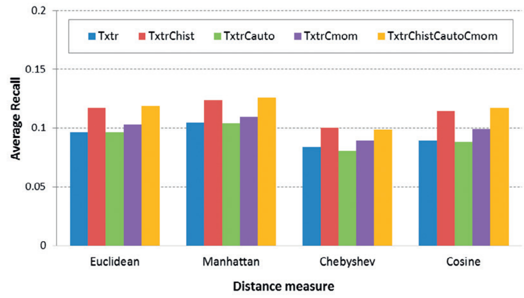
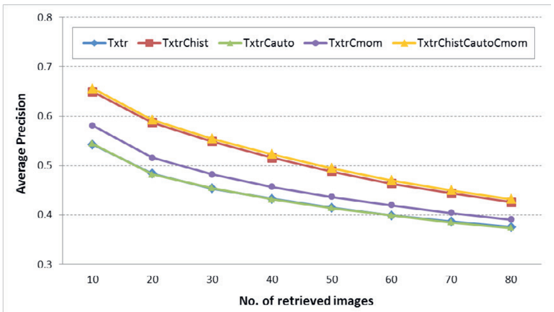


Fig. 4: Precision (left) and Recall (right) vs. distance measure for different feature datasets. L = 20.



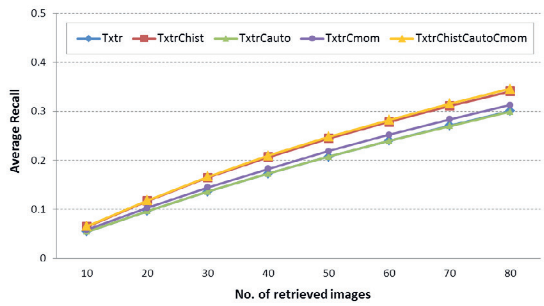


Fig. 5: Precision (left) and Recall (right) vs. number of retrieved images for different feature datasets. Distance measure = Euclidean distance.

Fig. 3: Example of a returned query. Feature dataset = TxtrCmom; distance measure = Euclidean distance; L = 20. Images irrelevant to the query image are marked with a red cross.

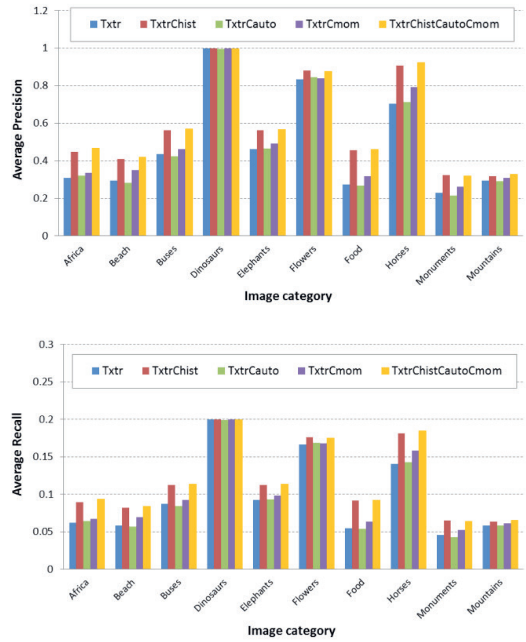


Fig. 6: Precision (left) and Recall (right) vs. image category for different feature datasets. Distance measure = Euclidean distance, L = 20.

Fig. 5 confirms the literature finding [4] that *P* and *R* follow an inverse relationship, i.e. that an increase in *L* is associated with a decrease in *P* and a simultaneous increase in *R* values regardless of the feature dataset being used.

Since human beings perceive big differences in visual content and appearance within individual Wang database images, it is interesting to see how successful are particular CBIR algorithms in finding matches similar to the query images, when these belong to different categories. Fig. 6 shows that categories such as dinosaurs or flowers, which contain relatively low within-class variability in visual attributes are characterized by high both *P* and *R* values. On the other hand, our algorithms had much bigger problems in retrieving similar images in case of colourful and diverse African scenery and people or texture- and color-rich themes (beach, food, monuments, mountains). For some categories, differences in performance among the five feature datasets were substantial (e.g. for food or horses), while for others they were very low (dinosaurs, flowers, mountains). This is probably related to the fact that some images have more pronounced texture features, while others are more sensitive to color features.

Finally, it should be noted that it is very difficult to compare our results to those obtained by other researchers in the field, since the applied methodologies differ widely – in terms of the number of retrieved images (*L*), the implemented distance measure, the way how individual features are combined into a feature vector, choice of performance metrics, etc. Nevertheless, the performance of our CBIR systems is in general comparable to the results of similar experiments described in the literature [13], [16].

Conclusion

Our study has shown that by combining both texture and color attributes of images, the performance of the CBIR system can be significantly improved. When comparing an exclusively texture-based system (*Txtr*) with the one containing all three color-related groups of features (*TxtrChistCautoCmom*), a 23% increase in average precision was found when using Euclidean distance measure and 20 retrieved images. Somewhat contrary to the expectations [17], very good results were obtained using color histogram features. Different weighting of individual groups of features in the feature vector might lead to an improved precision of algorithms that are based on color autocorrelogram and color moments.

In the future we intend to improve the performance of our CBIR system by incorporating high-level features that would bring semantics, i.e. image meaning and understanding, into the search. In addition, focusing on local – rather than global – image features (descriptors) such as SIFT [18] or SURF [19] might lead to a higher percentage of relevant querries returned by the system.

References

1. M.S. Lew, N. Sebe, C. Djeraba, R. Jain (2006): Content-based Multimedia Information Retrieval: State of the Art and Challenges. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 2(1), 1-19.
2. Y. Liu, D. Zhang, G. Lu, W.Y. Ma (2007): A survey of content-based image retrieval with high-level semantics. *Pattern Recognition*, 40, 262-282.
3. N. Vassilieva: Content Based Image Retrieval (CBIR). 2nd Russian Summer School in Information Retrieval. September 1-5, 2008, Taganrog, Russia. Web: http://romip.ru/russir2008/eng/program.html.
4. R. Datta, D. Joshi, J. Li, J.Z. Wang (2006): Image Retrieval: Ideas, Influences, and Trends of the New Age. *ACM Computer Surveys*, 40(2), 5:1-5:60.

5. T. Deselaers, D. Keysers, H. Ney (2008): Features for Image Retrieval: An Experimental Comparison. *Information Retrieval*, 11(2), 77-107.
6. W.Y. Ma, B. Manjunath (1997): NeTra: a toolbox for navigating large image databases. *Proceedings of the IEEE International Conference on Image Processing*. Oct 1997, Santa Barbara, CA, pp. 568-571.
7. J.Z. Wang, J. Li, G. Wiederhold (2001): SIMPLIcity: semantics-sensitive integrated matching for picture libraries, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 23(9), 947-963.
8. S. Marčelja (1980): Mathematical description of the responses of simple cortical cells. *Journal of the Optical Society of America*, 70(11), 1297-1300.
9. M. Swain, D. Ballard (1991): Color indexing. *International Journal of Computer Vision*, 7, 11–32.
10. J.Huang, S.R. Kumar, M. Mitra, W.-J. Zhu, R. Zabih (1997): Image Indexing Using Color Correlograms. *Proceedings of the IEEE Conference on Vision and Pattern Recognition*. June 1997, San Juan, pp. 762-768.
11. I.H. Sarker, S. Iqbal (2013): Content-based Image Retrieval Using Haar Wavelet Transform and Color Moment. *Smart Computing Review*, 3(3), 155-165.
12. H. Eidenberger (2011): *Fundamental Media Understanding*. atpress, Wien, ISBN: 978-3-842-37917-6; 230 p.
13. C.-H. Lin, R.-T. Chen, Y.-K. Chan (2009): A smart content-based image retrieval system based on color and texture feature. *Image and Vision Computing*, 27, 658-665.
14. J.Z. Wang's website: http://wang.ist.psu.edu/docs/related.shtml.
15. Chez on MatlabCentral: Content Based Image Retrieval. Web: http://www.mathworks.com/matlabcentral/fileexchange/42008-content-based-image-retrieval.
16. M. Singha, K. Hemachandran (2012): Content Based Image Retrieval using Color and Texture. *Signal & Image Processing: An International Journal*, 3(1), 39-57.
17. M. Stricker, M. Orengo (1995): Similarity of color images. *Proceedings of the SPIE Conference on Storage and Retrieval for Image and Video Databases III*. Feb. 1995, San Jose, CA.
18. D.G. Lowe (2004): Distinctive Image Features from Scale Invariant Features. *International Journal of Computer Vision*, 60(2), 91-110.
19. H. Bay, A. Ess, T. Tuytelaars, L.V. Gool (2008): SURF: Speeded Up Robust Features. *Computer Vision and Image Understanding*, 110(3), 346-359.