Texture-based Image Retrieval in MPEG-7 Multimedia System

Mislav Grgic [1], Mohammed Ghanbari [2], Sonja Grgic [1]

[1] Department of Radiocommunications,
Faculty of Electrical Engineering and Computing,
University of Zagreb,
Address: Unska 3 / XII, 10000 Zagreb, CROATIA,
E-mails: mgrgic@ieee.org; sonja.grgic@fer.hr

[2] Department of Electronic Systems Engineering,
University of Essex,
Address: Wivenhoe Park, Colchester, Essex,
CO4 3SQ, UNITED KINGDOM,
E-mail: ghan@essex.ac.uk

Abstract

Texture analysis is important for many applications such as content-based image retrieval and scene analysis. In this paper texture-based image retrieval as one of the important parameters of rising MPEG-7 multimedia standard is described. Results for similarity texture retrieval using Brodatz textures are shown.

1. Introduction

The ability to effectively retrieve images or video according to their content is still an unsolved goal for multimedia applications. Therefore, this is currently an active research area, especially under MPEG-7 multimedia standard development. Content-based image retrieval (CBIR) is determined by the features used for the annotation of data. Its efficiency depends on the ability of extracted features to represent the data in such a way that a subjectively good response to different queries is achieved. In order to characterise images (to extract image features), many different approaches have been proposed, [1].

Similarity measures between textures are important for image understanding applications such as CBIR, texture segmentation and texture classification. In order to be useful, it is important that these similarity measures correspond to human texture perception. Existing similarity measures for textures can be divided into measures that are defined in the image domain and measures that are defined in the frequency domain, mostly by using Gabor filters. Gabor functions are commonly used in texture analysis, [2]. There is strong evidence that simple cells in the primary visual cortex can be modelled by Gabor functions tuned to detect different orientations and scales on a log-polar grid. When applied to images, these functions produce features, which are basis for many definitions of texture.

In this paper, Gabor functions are used in image retrieval system. Our work is particularly concentrated on texture, texture features extraction and similarity texture retrieval. Moreover, we introduce *a novel intersection method* for improved texture retrieval.

2. Emerging MPEG-7 multimedia standard

The growth of available digital audio-visual content imposed the need for development of efficient CBIR systems. So, MPEG group started a new work item, MPEG-7, formally called *Multimedia Content Description Interface*, which will define search features for both still image and video data, [3]. MPEG-7 is currently under development and the ISO standard is scheduled for September 2001. It is the only standard specifically aimed at representing multimedia content as the core of CBIR technology. MPEG-7 will specify a standard set of descriptors that can be used to describe various types of multimedia information. The objective of MPEG-7 is to provide an interoperable solution to extend the capabilities of today's proprietary solutions in identifying multimedia content.

Previous MPEG standards have concentrated on image compression (MPEG-1, MPEG-2) and ways of separately representing foreground objects and background (MPEG-4). These standards have had little impact on the information retrieval community, even those dealing with image data. The new MPEG-7 standard is the first to address the issue of multimedia content. It aims to set up a standard framework for describing all aspects of a multimedia item's content.

The potential benefits of the new standard are considerable. It should make the process of searching for a desired image a great deal easier, since future MPEG-7-based search engines will simply need to process values of defined standard parameters, rather than computing search features from scratch. For the same reasons, the standard will enormously enhance system interoperability, since all search engines will potentially be using compatible features. This will have a major impact on image searching on the Web, which will become a far more efficient process once a significant number of images enhanced with MPEG-7 metadata become available.

MPEG-7 visual description tools cover four basic visual features: colour, texture, shape and motion. Image texture has emerged as an important visual primitive for searching and browsing through large collections of similar looking patterns. An image can be considered as a mosaic of textures so that texture features associated with the regions can be used to index the image data. For instance, a user browsing an aerial image database may want to identify all parking lots in the image collection. A parking lot with cars parked at regular intervals is an excellent example of a textured pattern when viewed from a distance, such as in an airphoto. To support image retrieval or browsing, an effective representation of texture is required. For this representation, MPEG-7 uses texture descriptors. The computation of descriptors is based on filtering using scale and orientation selective kernels. Descriptors provide a quantitative description that can be used for accurate search and retrieval.

In our paper we used Gabor functions for texture representation. It provides a perceptual characterisation

0-7803-6490-2/01/$10.00

of texture, similar to a human characterisation. The computation of this descriptor proceeds as follows: first, the image is filtered with a bank of orientation and scale tuned filters (modelled using Gabor functions); from the filtered outputs, two dominant texture features are identified, which are used for similarity image retrieval. So, our approach is well suited to the MPEG-7 requirements for image retrieval in multimedia systems.

2. Gabor Wavelets

Spatial frequency analysis has provided excellent results for texture classification. Pre-processing of images by Gabor wavelets is chosen for its biological relevance and technical properties. The Gabor wavelets are of similar shape as the receptive fields (RFs) of simple cells in the primary visual cortex (V1). These RFs are restricted to small regions of space and they are highly structured, [4]. Pollen & Ronner, [5], examined the phase relation of adjacent cells in the visual cortex of cats. They concluded that the cells of a pair of adjacent cells (defined by similar frequencies, i.e. similar magnitude and similar direction) have certain symmetries. One has even and the other odd symmetry. This allows modelling both RFs of such a pair of cells by a complex-valued Gabor function.

The results of Jones' and Palmer's experiments, [4], suggest to model the shapes of the RFs by two-dimensional *Gabor filters*, a plane wave restricted by a Gaussian envelope function. A two dimensional Gabor function $g(x,y)$ can be written as

$$g(x,y)=\frac{1}{2\pi\sigma_x\sigma_y}\cdot e^{-\frac{1}{2}\left(\frac{x^2}{\sigma_x^2}+\frac{y^2}{\sigma_y^2}\right)}\cdot e^{i2\pi f_0 x} \qquad (1)$$

where σ_x and σ_y are the spreads of the Gaussian and f_0 is the spatial frequency of the harmonic wave. Using $g(x,y)$ it is possible to design Gabor filter dictionary, [2]. All filters are similar in the sense that they can be generated from one filter (called *mother wavelet* or *basis wavelet*) simply by translation, scaling and rotation. For this reason the set of filters can be seen as a set of wavelets.

With Gabor filters, input image can be decomposed into $s\times\theta$ filtered images, where s is the number of scales, and θ is the number of orientations. We have used $s=4$, $\theta=6$, resulting in 24 different filters which decompose the input image into 24 filtered images. These parameters are selected according to the properties of human visual system (4 decompositions) and according to practical experience (6 orientations). For every filtered image, two features are extracted (mean value and standard deviation). Hence, after Gabor decomposition, 48 features per input image are extracted.

3. Texture Similarity Measure

Using Brodatz photographic album [6], we have created a set of 112 textures (format 512×512 pixels). For every image, extraction of 48 features was made, Fig. 1. In this way, texture features database with 112×48 features is produced. The same texture features extraction process can be performed on the query texture image, producing 1×48 query features.

After creating the texture features database and extracting the query features, we can approach to the search algorithm. It is based on calculating the similarity measure between each set of 48 features per query image and the set of 48 features per each database image.

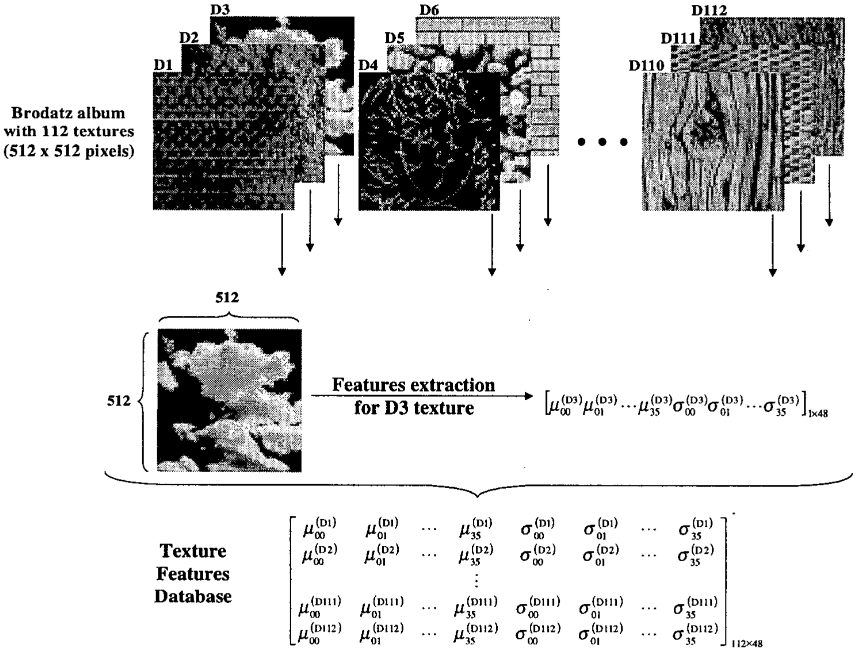


Fig. 1. Creating a texture features database (one feature vector per entire image)

The similarity measure (*sm*) is calculated according to formula for normalised Euclidean distance, [2]:

$$sm(i,j)=\left|\frac{\mu_{s\theta}^{(i)}-\mu_{s\theta}^{(j)}}{\sigma(\mu_{s\theta})}\right|+\left|\frac{\sigma_{s\theta}^{(i)}-\sigma_{s\theta}^{(j)}}{\sigma(\sigma_{s\theta})}\right| \qquad (2)$$

where $\mu_{s\theta}^{(i)}$ is a mean value of the image i on scale s and orientation θ, $\sigma_{s\theta}^{(i)}$ is a standard deviation of the image i on scale s and orientation θ, $\sigma(\mu_{s\theta})$ is a standard deviation of the mean values over the entire database, and $\sigma(\sigma_{s\theta})$ is a standard deviation of the standard deviations over the entire database. With this approach, 112 sets of different similarity measures are produced for each query image. Similarity measures are then sorted in increasing order. Due to the fact that every similarity measure is associated with appropriate *database* image number, we will use "retrieved image" or "result" formulation in the following text.

User-defined number of top matches (NTM) is used to extract NTM best results per each query image. This means that we are "cutting" *first* NTM results from each set of 112 sorted results and then these NTM results can be presented to the end-user.

4. Retrieval Results

An example of the similarity texture retrieval with NTM=12 is shown on Fig. 2. It is interesting to note that some of the incorrect matches actually look quite similar to the query texture (first image from the left is query image, and the rest are the matched images in the order of similarity from left to right column and from top to bottom row). Second interesting thing is that the first (most similar) result corresponds exactly to the query image. This fact means that our retrieval algorithm works good: if the query image is included in the database it must be retrieved as a most similar result (similarity measure must be equal to zero), because query texture features are identical to the texture features of the same image from the database.

5. Drawbacks and a New Retrieval Method

Considering achieved retrieval results, Fig. 2, it is obvious that the number of results is too big, especially using higher NTM. If we want to reduce the total number of retrieved images, naturally, it can be done by reducing the number of top matches per query image. But it would be also good if we can at the same time preserve or even improve the retrieval confidence. This reduction of the total number of retrieved images is necessary if we want to make functional image retrieval system that is acceptable for the end-user who can then display, view and scroll the images conveniently. Hence, we introduce *a novel intersection method* performed on the subimage level.

Each 512x512 image is divided into 4 non-overlapped subimages (format 256x256 pixels), producing a database of 448 subimages. The same image features extraction process is performed, but now, on each of the subimages, Fig. 3. Hence, a new texture features database is produced, with 448x48 texture features. The same procedure is made for the query image. It is split to 4 query subimages and for each subimage 48 texture features are extracted. For each query subimage retrieval process is performed separately. So, 4xNTM results are produced for one query image. Within the retrieved 4xNTM results, we are performing intersection according to this rule: *image will be considered as the intersection result if it appears 4 times in each of the 4 retrieved sets.*

An example of retrieved images for described intersection approach is shown on Fig. 4. It can be seen that intersection method reduces number of different retrieved images (from 12 to 4 in this example). On the other hand, selection of the results is much better than for the first approach that uses one features vector per entire image, because, subjectively, retrieval results are more accurate then in the first approach.

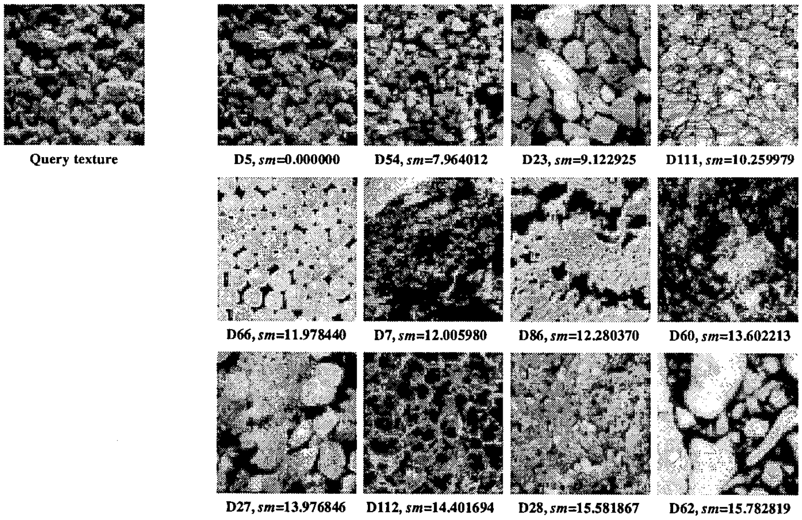

Query texture
D5, *sm*=0.000000 D54, *sm*=7.964012 D23, *sm*=9.122925 D111, *sm*=10.259979
D66, *sm*=11.978440 D7, *sm*=12.005980 D86, *sm*=12.280370 D60, *sm*=13.602213
D27, *sm*=13.976846 D112, *sm*=14.401694 D28, *sm*=15.581867 D62, *sm*=15.782819

Fig. 2. An example of texture-based image retrieval results; query texture: D5; NTM=12

6. Conclusion

We have presented an approach for texture-based image retrieval based on texture features extraction using Gabor filters. Considering achieved retrieval results, we showed that number of results is too big and results are not enough accurate. So, we introduced a new intersection method for significant improvement of texture image retrieval. The improvement is based on performing the intersection between retrieval results for 4 query subimages. Our approach has several qualities:

1. It reduces the total number of retrieval results;
2. It does not effect the number of top matches per query subimage (preserves the retrieval confidence);
3. Retrieval results are subjectively more accurate.

Our approach is well suited to the MPEG-7 requirements for texture-based image retrieval in multimedia systems.

Acknowledgement

Mislav Grgic would like to thank *The British Scholarship Trust, United Kingdom*, for their support in his research study.

References

[1] A.W.M. Smeulders, T. Gevers, M.L. Kersten, "Computer Vision and Image Search Engines", *Proceedings of the Workshop on Image and Video Content Based Retrieval*, 1998, Milano, Italy

[2] B.S. Manjunath, W.Y. Ma, "Texture Features for Browsing and Retrieval of Image Data", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 18(8), pp. 837-842, 1996

[3] ISO/IEC JTC1/SC29/WG11 N3445 (ed. Jose M. Martinez), "MPEG-7 Overview (version 3.0)", Geneva, Switzerland, May/June 2000

[4] J.P. Jones, L.A. Palmer, "An Evaluation of the Two-Dimensional Gabor Filter Model of Simple Receptive Fields in Cat Striate Cortex", *Journal of Neurophysiology*, 58(6), pp. 1233-1258, 1987

[5] D.A. Pollen, S.F. Ronner, "Phase Relationship between Adjacent Simple Cells in the Visual Cortex", *Science*, 212, pp. 1409-1411, 1981

[6] P. Brodatz, "Textures; A photographic album for artists & designers", *Dover*, New York, 1999 (reprint from 1966)

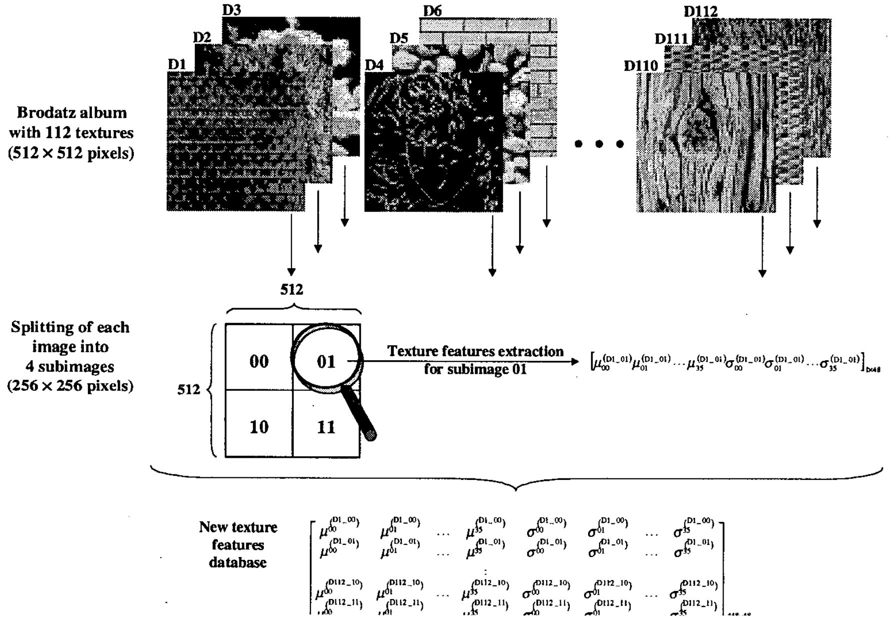
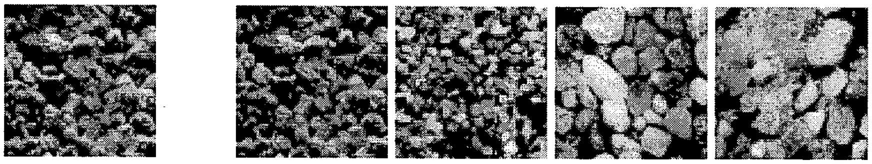


Fig. 3. Creating a new texture features database (4 feature vectors per image)



Query texture **D5** **D54** **D23** **D27**

Fig. 4. An example of texture-based image retrieval results by performing intersection method; query texture: D5; NTM=12 (per each subimage)