

Ibrahim El-Henawy

Zagazig university, Faculty of Computer & Information
henawy2000@yahoo.com

Mohamed Eisa

Institute of Computer Science Albert-Ludwigs-Universität Freiburg
eisa@informatik.uni-freiburg.de

A. E. Elalfi

Mansoura university, Faculty of Specific Education
ael_alfi@yahoo.com

IMAGE RETRIEVAL USING LOCAL COLOUR AND TEXTURE FEATURES

Abstract

Colour histograms have proved to be successful in automatic image retrieval; however, their drawback is that all structural information is lost. So Siggelkow et al. extended the colour histogram approach by features that take into account the relations within a local pixel neighbourhood; they extracted features that are invariant with respect to translation and rotation by integrating nonlinear functions over the group of Euclidean motion. Gabor wavelets proved to be very useful texture analysis. In this paper we present an image retrieval method based on a nonlinear monomial kernel function and Gabor filters. Colour features are discovered by calculating the 3D colour histogram after applying the monomial kernel function onto the image. Texture features are found by calculating the mean and standard deviation of the Gabor filtered image. Experimental results are shown and discussed.

Key-words:- Image Retrieval, Knowledge Discovery, Data Mining, Colour Histograms, Intelligent Agent

1 Introduction

The recent advances in digital imaging and computing technology have resulted in a rapid accumulation of digital media in the personal computing and entertainment industry. In addition, large collections of such data already exist in many scientific application domains, such as that of medical imaging. Managing large collections of multimedia data requires the development of new tools and technologies however[4].

General methods for the construction of invariant features are explained in [1, 8]. We have concentrated on the invariant features for the group of translations and rotations and their theoretical invariance to global translation and rotation. These

features have proven to be robust as regards the independent motion of objects, different object constellations and articulated objects. Invariant features are developed so as to characterize images independently of absolute object positions. Thus, these features are suited for image retrieval [11].

Textural analysis has, along history and texture analysis algorithms, ranged from using random field models to multiresolution filtering techniques, such as wavelet transforming. This work has as its focus the multi-resolution representation as based on Gabor filters [3].

Basically, Gabor filters are a group of wavelets, with each wavelet capturing energy at a specific frequency and a specific direction. Expanding a signal using this basis provides a localized frequency description, therefore capturing local features/energies of the signal. Textural features can then be extracted from this group of energy distributions. The scale and orientation tuneable property of the Gabor filter makes it especially useful for texture analysis. Experimental evidence from human and mammalian vision supports the notation of a spatial-frequency analysis that maximizes the simultaneous localization of energy in both spatial and frequency domains [2, 7].

Texture and colour are two very important attributes in such image analysis. Most of the many different methods that are proposed for texture analysis are focused on gray level representation. In this paper we wish to combine the local colour features that can be extracted using nonlinear monomial kernel functions with the local texture features extracted by applying some of appropriate dilations and rotations of a Gabor function.

2 Construction of invariant features

Let $X = \{X(n_0, n_1)\}$, $0 \leq n_0 < N_0$, $0 \leq n_1 < N_1$ be a gray value image, with $X(n_0, n_1)$ representing the gray value at the pixel coordinate (n_0, n_1) [9,10]. G is denoted as a transformation group with elements $g \in G$ acting on the image.

For a group element g of the group of translations and rotations and an image X , the transformation image is denoted by gX and can be expressed as:

$$(gX)(n_0, n_1) = X(n'_0, n'_1) \quad (1)$$

Where

$$\begin{pmatrix} n'_0 \\ n'_1 \end{pmatrix} = \begin{pmatrix} \cos \varphi & -\sin \varphi \\ \sin \varphi & \cos \varphi \end{pmatrix} \begin{pmatrix} n_0 \\ n_1 \end{pmatrix} + \begin{pmatrix} t_0 \\ t_1 \end{pmatrix}$$

Let X be a given gray-value image and $f(X)$ is a kernel function; it is then possible to construct an invariant feature $F(X)$ by integrating $f(gX)$ over the translations and rotations group,

$$F(\mathbf{X}) = \frac{1}{|G|} \int_G f(g\mathbf{X}) dg \quad (2)$$

where

$$|G| = \int_G dg = 2\pi N_0 N_1$$

In practice, we can use the following discrete form

$$F(\mathbf{X}) = \frac{1}{|G|} \sum_G f(g\mathbf{X})$$

For the construction of rotation and translation invariant gray scale features we have to integrate over the group of rotation and translation. In this case the integral over the transformation group can be written as:

$$F(\mathbf{X}) = \frac{1}{2\pi N_0 N_1} \int_{t_0=0}^{N_0} \int_{t_1=0}^{N_1} \int_{\varphi=0}^{2\pi} f(g(t_0, t_1, \varphi)\mathbf{X}) d\varphi dt_1 dt_0 \quad (3)$$

In practice, the integrals are replaced by sums, with our choosing only integer translations and varying the angle in discrete steps while applying a bilinear interpolation for pixels that do not coincide with the image grid.

To provide some intuitive insights for equation (3) consider the following example: For the function $f(\mathbf{X}) = \mathbf{X}(0,0) \cdot \mathbf{X}(0,1)$ the group average is given by

$$F(\mathbf{X}) = \frac{1}{2\pi N_0 N_1} \int_{t_0=0}^{N_0} \int_{t_1=0}^{N_1} \int_{\varphi=0}^{2\pi} \mathbf{X}(t_0, t_1) \cdot \mathbf{X}(\sin \varphi + t_0, \cos \varphi + t_1) d\varphi dt_1 dt_0 \quad (4)$$

Consider the inner integral over φ . When varying φ between 0 and 2π the second coordinate given describes a circle with radius 1 around the center (t_0, t_1) . This means we have to take the product between the center value and points on the circle and average over all these products. The remaining integrals over t_0, t_1 mean that the center point has to be varied for the whole image. Therefore, the calculation of equation (3) can be generally reformulated in two steps: first for every pixel a local function is evaluated (integration over φ); then, all of the intermediate results from the local computations are summed up (integration over t_0, t_1). For more details of this technique, refer to [1, 11, 12].

3. Texture feature extraction

3.1 Gabor Functions and Wavelets

Gabor Elementary Functions are Gaussians modulated by complex sinusoids. In two dimensions they are represented by [2, 3, 5, 6].

$$G(x, y) = G_1(x, y) \exp(2\pi j W x) \quad (5)$$

Where

$$G_1(x, y) = \left(\frac{1}{2\pi\sigma_x\sigma_y} \right) \exp\left(-\frac{1}{2} \left(\frac{x^2}{\sigma_x^2} + \frac{y^2}{\sigma_y^2} \right) \right)$$

and W is the modulation frequency

The Fourier transform of G(x,y) is

$$H(u, v) = \exp\left(-\frac{1}{2} \left[\frac{(u-W)^2}{\sigma_u^2} + \frac{v^2}{\sigma_v^2} \right] \right) \quad (6)$$

$$\text{Where } \sigma_u = \frac{1}{2\pi\sigma_x} \text{ and } \sigma_v = \frac{1}{2\pi\sigma_y}$$

For a localized frequency analysis it is desirable to have a Gaussian envelope whose width adjusts itself with the frequency of the complex sinusoids. Let $G(x, y)$ be the mother Gabor wavelet; a filter set is then obtained by appropriate dilations and rotations of the mother wavelet using:

$$G_{mn}(x, y) = a^{-m} G(x', y') \quad (7)$$

$$a > 1, m, n = \text{integer}$$

$$x' = a^{-m} (x \cos \theta + y \sin \theta),$$

$$y' = a^{-m} (-x \sin \theta + y \cos \theta),$$

Where $\theta = \frac{n\pi}{N}$ and N is the total number of orientations. The scale factor a^{-m}

in equation (7) ensures that the energy is independent of scale.

$$E_{mn} = \int_{-\infty-\infty}^{\infty} \int_{-\infty-\infty}^{\infty} |G_{mn}(x, y)|^2 dx dy, \quad (8)$$

The non-orthogonality of Gabor wavelets implies redundant information for the filtered images, and the following strategy is used to reduce this redundancy.

Let U_l, U_h denote the lower and upper center frequencies of interest respectively. Let N be the number of orientations and M be the number of scales in the decomposition. The design strategy is to ensure that the half peak magnitude cross-

sections of the filter responses in the frequency spectrum touch each other. This results in the following formulas for computing the filter parameters σ_u and σ_v ,

$$a = \left(\frac{U_h}{U_l} \right)^{\frac{1}{M-1}}, \quad \sigma_u = \frac{(a-1)U_h}{(a+1)\sqrt{2\ln 2}}$$

$$\sigma_v = \tan\left(\frac{\pi}{2N}\right) \left[U_h - 2\ln 2 \left(\frac{\sigma_u^2}{U_h} \right) \right] \left[2\ln 2 - \frac{(2\ln 2)^2 \sigma_u^2}{U_h^2} \right]^{\frac{1}{2}} \quad (9)$$

Where $W = U_h$, $\theta = \frac{\pi}{N}$, $m = 0, 1, \dots, M - 1$

To eliminate the sensitivity of the filter response to absolute intensity values, the real (even) components of the 2-D Gabor filters are biased by adding a constant to give them zero mean. Each channel is formed by a pair of real Gabor filters. Let the output of each channel be given by

$$C_{ev}(x, y; U, \theta) = G_1(x, y) \cdot \cos(2\pi Ux') * X(x, y)$$

$$C_{odd}(x, y; U, \theta) = G_1(x, y) \cdot \sin(2\pi Ux') * X(x, y) \quad (10)$$

Where $G_1(x, y)$ is 2D Gaussian and $*$ denotes 2-D linear convolution. The channel output $C(x, y)$, is computed as

$$C(x, y; U, \theta) = \sqrt{C_{ev}^2 + C_{odd}^2} \quad (11)$$

Filters are implemented in a frequency domain for better computational efficiency. The mean value $M(U, \theta)$ of a channel is computed by

$$M(U, \theta) = \frac{1}{N_1 N_2} \sum C(x, y; U, \theta) \quad (12)$$

Where $N_1 N_2$ is the area of $C(x, y; U, \theta)$. This value depends on the filter center frequency U and orientation θ . The mean values provide powerful features for texture classification. These features are rotation dependent since $M(U, \theta_i) \neq M(U, \theta_j)$, for $i \neq j$. Since rotation of input image $X(x, y)$ corresponds to the translation of $M(U, \theta)$, DFT of $M(U, \theta)$ would be the rotation invariant feature [5]. The redundant data after DFT is removed.

3.2 Texture representation

After applying Gabor filters on the image with a different scale (m) and orientation (n), we obtain an array of magnitudes. These magnitudes represent the energy content at a different scale and orientation of the image. We are interested in images

or regions that have a homogeneous texture because the main purpose of texture-based retrieval is to find images or regions with similar textures.

The following mean M_{mn} and standard deviation S_{mn} of the magnitude of the transformed coefficient are used to represent the homogenous texture feature of the region:

$$M_{mn} = \frac{1}{N_1 N_2} \sum C_{mn}, S_{mn} = \frac{\sqrt{\sum (C_{mn} - M_{mn})^2}}{N_1 N_2} \quad (13)$$

$$m = 0, 1, \dots, M - 1; n = 0, 1, \dots, N - 1$$

A feature vector is constructed using M_{mn} and S_{mn} as feature components. In the experiments, we use five scales and six orientations, while the feature vector is given by:

$$F = (M_{00}, S_{00}, M_{01}, S_{01}, \dots, M_{45}, S_{45}) \quad (14)$$

4. Experiment results

We have conducted retrieval tests both on texture grey value images and real world colour images. The texture database used in the experiments consists of a set of 116 Brodatz different texture classes. The set includes different categories of textures in the Brodatz album such as regular textures, oriented textures, natural textures and bidirectional textures. Each of the 512×512 images is divided into 16 128×128 overlapping sub-images. The real world images consist of a set of nearly 2500 photographic images.

We used the monomial kernel function, $M(X) = \sqrt{X(1,0) \cdot X(0,3)}$ to extract the colour features from the three levels of RGB colour space, and the 3D $8 \times 8 \times 8$ colour histogram was applied to the intermediate image.

We used Gabor filters as a group of wavelets on the image with a different orientation ($K=6$) at different scales ($S=5$) and $U_l = 0.05$, $U_h = 0.4$, s and t range from 0 to 128, i.e., filter mask size is 128×128 , to obtain the texture features. The texture feature components are the first and second moments of energy that are extracted from each channel, as in (14).

A query image is any one of the images in the database. This query image is then processed to compute the texture feature vector and the colour $8 \times 8 \times 8$ histogram and is compared with the respective features of all the images in the database.

To compare the texture feature vectors of query image Q and a target image T in the database, we used

$$D(Q, T) = \sum_m \sum_n d_{mn}(Q, T)$$

where

$$d_{mn} = \sqrt{(M_{mn}^Q - M_{mn}^T)^2 + (S_{mn}^Q - S_{mn}^T)^2}$$

and to compare the colour histogram we used the well known statistical method χ^2 to determine whether two distributions differ:

$$\chi^2(Q, T) = \sum_i \frac{(Q_i - T_i)^2}{Q_i + T_i}$$

Let D_t be the result of the difference between the query image and a database based on texture feature vectors, and D_c is the result of the difference between the query image and a database based on a colour histogram. The total difference thus takes the form:

$$D_{tc} = \alpha \cdot D_t + \beta \cdot D_c, \text{ Where } \alpha + \beta = 1$$

Figure 1 gives a retrieval result from the texture image database. All the 20 similar textures in the database are retrieved in the first 20 images, and the remaining ones are also relevant. *Figure 2* shows a retrieval result from a real world image database. As can be seen, most of the retrieved images are images with similar textures and colours to that in the query image. It shows us that the proposed technique works well for retrieval of images with homogeneous colour and texture overall.

Figures 3 and 4 displays the comparison of the average recalls of the query images taken from texture image database and real world image database respectively, the recall for which is defined as:

$$\text{recall} = \frac{\text{number of relevant images retrieved}}{\text{number of relevant images}}$$

5. Conclusion

In this paper we have used the nonlinear monomial kernel function to capture the local structures of image content to find out the colour features. Such a technique does not require any segmentation and therefore works fully automatically [9, 12].

We have also used the Gabor function by applying some of appropriate dilations and rotations to base the texture features of the image on.

The extracted texture and colour features are then used to measure the similarities between the query image and the image database by weighting these feature types according to needs or according to image type. The retrieval results have been shown and examined.

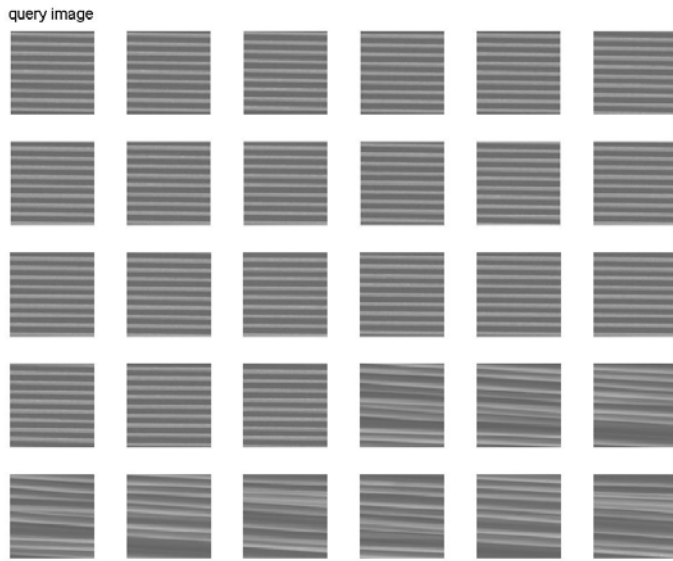


Figure 1: Image retrieval results from a texture image database



Figure 2: Image retrieval results from a real world image database

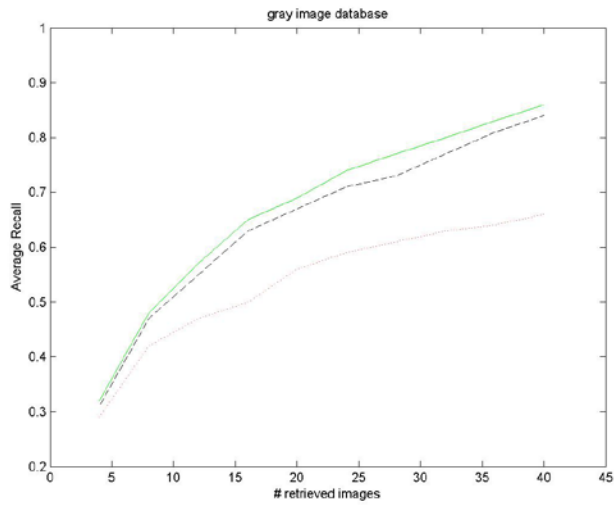


Figure 3: Comparison of the average recalls by using different features: a solid curve for the combination of colour and texture features, a dashed curve for texture feature and a dotted curve for the colour feature

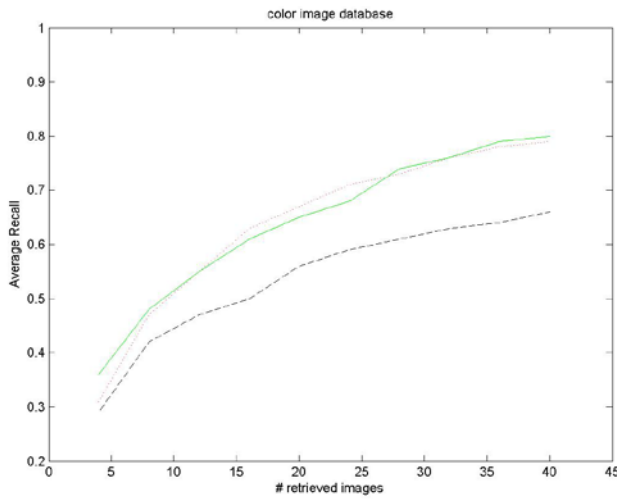


Figure 4: Comparison of the average recalls by using different features: a solid curve for the combination of colour and texture features, a dashed curve for texture feature and a dotted curve for the colour feature

References

- [1] H. Burkhardt and S. Siggelkow, "Invariant features in pattern recognition fundamentals and applications. In I.Pitas and C. Kotropoulos, editors, Non-linear Model Based Image/Video Processing and Analysis, pages 269–307. John Wiley & Sons, 2001.
- [2] George M. Haley and B. S. Manjunath, "Rotation invariant texture classification using a complete space-frequency model," *IEEE transactions on Image Processing*, vol. 8, no. 2, pp. 255–269, February 1999.
- [3] B. S. Manjunath and W. Y. Ma. "Texture features for browsing and retrieval for large image data" *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 18, NO. 8, August 1996, pp. 837–842.
- [4] B. S. Manjunath, P. Wu, S. Newsam, H. D. Shin, "A texture descriptor for browsing and similarity retrieval" *Signal Processing: Image Communication*, 2000.
- [5] R. Manthalkar, P. K. Biswas, "Rotation invariant texture classification using Gabor wavelets" *Asian Conference on Computer Vision*, pp. 493–498, Jan. 23–25, 2002, Melbourne Australia.
- [6] R. Manthalkar, P. K. Biswas, "Color Texture Segmentation using multichannel filtering" *Conference on Digital Image Computing Techniques and Application*, pp. 346–351, Jan 21–22, Melbourne, Australia, 2002.
- [7] R. Manthalkar, P. K. Biswas, B. N. Chatterji, "Rotation and scale invariant texture classification using Gabor wavelets" second international workshop on texture analysis and synthesis, 2002 Copenhagen.
- [8] H. Schulz-Mirbach, "Invariant features for gray scale images" In G. Sagerer, S. Posch, and F. Kummert, editors, 17. DAGM-Symposium, pages 1–14, Bielefeld, Germany, September 1995.
- [9] S. Siggelkow and H. Burkhardt, "Image retrieval based on colour and nonlinear texture invariants" In S. Marshall, N. Harvey, and D. Shah, editors, *Processings of the Noblesse Workshop on Non-Linear Model Based Image Analysis*, pages 217–224, Glasgow, United Kingdom, July 1998.
- [10] S. Siggelkow and H. Burkhardt, "Image retrieval based on local invariant features" In *IASTED International Conference on Signal and Image Processing (SIP)*, pages 369–373, Las Vegas, NV, October 1998.
- [11] S. Siggelkow, M. Schael, and H. Burkhardt "SIMBA- Search IMages By Appearance" In the B. Radig and S. Florczyk, editors, *Pattern Recognition, DAGM, LNCS 2191*, pages 9–16, München, Germany, September 2001.
- [12] Sven Siggelkow. "Feature Histograms for Content-based Image Retrieval" Ph.D thesis, Albert-Ludwigs-Universität Freiburg, Germany, 2002.