Proc. of 3th Int. Conf. on Visual Info. and Info. Systems, Amsterdam, The Netherlands, 1999

Invariant Image Retrieval using Wavelet Maxima Moment

Minh Do, Serge Ayer, and Martin Vetterli

Swiss Federal Institute of Technology, Lausanne (EPFL)
Laboratory for Audio-Visual Communications (LCAV)
CH-1015 Lausanne, Switzerland
{Minh.Do, Serge.Ayer, Martin.Vetterli}@epfl.ch

Abstract. Wavelets have been shown to be an effective analysis tool for image indexing due to the fact that spatial information and visual features of images could be well captured in just a few dominant wavelet coefficients. A serious problem with current wavelet-based techniques is in the handling of affine transformations in the query image. In this work, to cure the problem of translation variance with wavelet basis transform while keeping a compact representation, the wavelet transform modulus maxima is employed. To measure the similarity between wavelet maxima representations, which is required in the context of image retrieval systems, the difference of moments is used. As a result, each image is indexed by a vector in the wavelet maxima moment space. Those extracted features are shown to be robust in searching for objects independently of position, size, orientation and image background.

1 Introduction

Large and distributed collections of scientific, artistic, and commercial data comprising images, text, audio and video abound in our information-based society. To increase human productivity, however, there must be an effective and precise method for users to search, browse, and interact with these collections and do so in a timely manner.

As a result, image retrieval (IR) has been a fast growing research area lately. Image feature extraction is a crucial part for any such retrieval systems. Current methods for feature extraction suffer from two main problems: first, many methods do not retain any spatial information, and second, the problem of invariance with respect to standard transformations is still unsolved.

In this paper we propose a new wavelet-based indexing scheme that can handle variances of translation, scales and rotation of the query image. Results presented here are with the "query-by-example" approach but the method is also ready to be used in systems with hand-drawn sketch query. The paper is organized as follows. Section 2 discusses the motivation for our work. The proposed method is detailed in Sections 3 and 4. Simulation results are provided in Section 5, which is followed by the conclusion.

2 Motivation

A common ground in most of current IR systems is to exploit low-level features such as color, texture and shape, which can be extracted by a machine automatically. While semantic-level retrieval would be more desirable for users, given the current state of technology in image understanding, this is still very difficult to achieve. This is especially true when one has to deal with a heterogeneous and unpredictable image collection such as from the World Wide Web.

Early IR systems such as [2, 8] mainly relied on a *global* feature set extracted from images. For instance, color features are commonly represented by a global histogram. This provides a very simple and efficient representation of images for the retrieval purpose. However, the main drawback with this type of systems is that they have neglected *spatial* information. Especially, shape is often the most difficult feature to be indexed and yet it is likely the key feature in an image query.

More recent systems have addressed this problem. Spatial information is either expressed *explicitly* by the segmented image regions [9, 1, 6] or *implicitly* via dominant wavelet coefficients [4, 5, 12]. Wavelets have been shown to be a powerful and efficient mathematical tool to process visual information at multiple scales. The main advantage of wavelets is that they allow simultaneously good resolution in time and frequency. Therefore spatial information and visual features can be effectively represented by dominant wavelet coefficients. In addition, the wavelet decomposition provides a very good approximation of images and its underlying multiresolution mechanism allows the retrieval process to be done *progressively* over scales.

Most of the wavelet-based image retrieval systems so far employed traditional, i.e. orthogonal and maximally-decimated, wavelet transforms. These transforms have a serious problem that they can exhibit visual artifacts, mainly due to the lack of translation invariance. For instance, the wavelet coefficients of a translated function $f_\tau(t) = f(t - \tau)$ may be very different from the wavelet coefficients of $f(t)$. The differences can be drastic both within and between subbands. As a result, a simple wavelet-based image retrieval system would not be able to handle affine transformations of the query image. This problem was stated in previous works (eg. [4]), but to our knowledge, it still has not received proper treatment. On the other hand, the ability to retrieve images that contain interesting objects at different locations, scales and orientations, is often very desirable. It is our intent to address the invariance problem of wavelet-based image retrieval in this work.

3 Wavelet Maxima Transform

As mentioned above, the main drawback of wavelet bases in visual pattern recognition applications is their lack of translation invariance. An obvious remedy to this problem is to apply a non-subsampled wavelet transform which computes all the shifts [11]. However this creates a highly redundant representation and we have to deal with a large amount of redundant feature data.

To reduce the representation size in order to facilitate the retrieval process while maintaining translation invariance, an alternative approach is to use an adaptive sampling scheme. This can be achieved via the wavelet maxima transformation [7], where the sampling grid is automatically translated when the signal is translated.

For images, inspired by Canny's multiscale edge detector algorithm, the wavelet maxima points are defined as the points where the wavelet transform modulus is locally maximal along the direction of the gradient vector. Formally, define two wavelets that are partial derivatives of a two-dimensional smoothing function $\theta(x, y)$

$$\psi^1(x,y) = \frac{\partial\theta(x,y)}{\partial x} \text{ and } \psi^2(x,y) = \frac{\partial\theta(x,y)}{\partial y} \tag{1}$$

Let us denote the wavelets at dyadic scales $\{2^j\}_{j\in\mathcal{Z}}$ as

$$\psi^k_{2^j}(x,y) = \frac{1}{2^j}\psi^k\left(\frac{x}{2^j}, \frac{y}{2^j}\right) \qquad k = 1, 2 \tag{2}$$

Then the wavelet transform of $f(x, y)$ at a scale 2^j has the following two components

$$\begin{aligned} W^k f(2^j, u, v) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\psi^k_{2^j}(x-u, y-v)dxdy \\ &= \langle f(x,y), \psi^k_{2^j}(x-u, y-v)\rangle \qquad k = 1, 2 \end{aligned} \tag{3}$$

It can be shown [7] that the two components of the wavelet transform given in (3) are proportional to the coordinates of the gradient vector of $f(x, y)$ smoothed by $\theta_{2^j}(x, y)$. We therefore denote the wavelet transform modulus and its angle as:

$$Mf(2^j, u, v) = \sqrt{|W^1 f(2^j, u, v)|^2 + |W^2 f(2^j, u, v)|^2} \tag{4}$$

$$Af(2^j, u, v) = \arctan\left(\frac{W^2 f(2^j, u, v)}{W^1 f(2^j, u, v)}\right) \tag{5}$$

Definition 1 (Mallat et al. [7]). *Wavelet maxima at scale 2^j are defined as points (u_0, v_0) where $Mf(2^j, u, v)$ is locally maximum in the one-dimensional neighborhood of (u_0, v_0) along the angle direction given by $Af(2^j, u_0, v_0)$.*

If the smoothing function $\theta(x, y)$ is a separable product of cubic spline functions then the transform can be efficiently computed using a filter bank algorithm [7]. Figure 1 displays the wavelet maxima transform of an image at 3 scales.

The wavelet maxima transform has some useful properties for image retrieval applications. Apart from being compact and translation invariant, it has been shown to be very effective in characterization of images from multiscale edges (see Fig. 1). Therefore feature extraction based on the wavelet maxima transform captures well the edge-based and spatial layout information. Using wavelet maxima only, [7] can reconstruct an image which is visually identical to the original one. This reconstruction power of wavelet maxima indicates the significance of its representation. In addition, the "denoising" facility in the wavelet maxima domain can be exploited to achieve robustness in retrieving images which contain interesting objects against various image backgrounds.

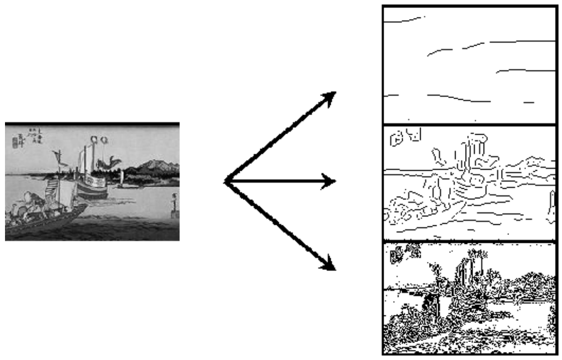

Fig. 1. Wavelet maxima decomposition. The right hand part shows the wavelet maxima points at scales 2^j where $j = 6, 3, 1$ from top to bottom, respectively (showing from coarse to detail resolutions)

4 Wavelet Maxima Moment

Given a compact and significant representation of images via wavelet maxima transform, the next step is to define a good similarity measurement using that representation. The result of wavelet maxima transform is multiple scale sets of points (visually located at the contours of the image) and their wavelet transform coefficients at those locations. Measuring the similarity directly in this domain is difficult and inefficient. Therefore we need to map this "scattered" representation into points in a multidimensional space so that the distances could be easily computed. Furthermore, we require this mapping to be invariant with respect to affine transforms.

For those reasons, we select the *moments* representation. Traditionally, moments have been widely used in pattern recognition applications to describe the geometrical shapes of different objects [3]. Difference of moments has also been successfully applied in measuring similarity between image color histograms [10]. For our case, care is needed since we use moments to represent wavelet maxima points which are dense along curves rather than regions (see the normalized moment equation (8)).

Definition 2. *Let us denote $\mathcal{M}^j$ is the set of all wavelet maxima points of a given image at the scale 2^j. We define the $(p+q)^{th}$-order moment of the wavelet* *maxima transform, or* wavelet maxima moment *for short, of the image as:*

$$m_{pq}^{j} = \sum_{(u,v)\in\mathcal{M}^{j}} u^{p} v^{q} Mf(2^{j},u,v), \qquad p,q = 0,1,2,\ldots \tag{6}$$

where $Mf(2^j,u,v)$ is defined in (4).

The reason for not including the angles $Af(2^j,u,v)$ in the moment computation is because they contain information about direction of gradient vectors in the image which is already captured in the locations of the wavelet maxima points. In the sequel the superscript j is used to denote scale index rather than power.

First, to obtain translation invariance, we centralize the wavelet maxima points to their center of mass $(\overline{u}^j,\overline{v}^j)$ where $\overline{u}^j = m_{10}^j/m_{00}^j$; $\overline{v}^j = m_{01}^j/m_{00}^j$. That is,

$$\mu_{pq}^{j} = \sum_{(u,v)\in\mathcal{M}^{j}} (u-\overline{u}^{j})^{p}(v-\overline{v}^{j})^{q} Mf(2^{j},u,v) \tag{7}$$

We furthermore normalize the moments by the number wavelet maxima points, $|\mathcal{M}^j|$, and their "spread", $(\mu_{20}^j + \mu_{02}^j)^{1/2}$, to make them invariant to the change of scale. The normalized center moments are defined as:

$$\eta_{pq}^{j} = \frac{\mu_{pq}^{j}/|\mathcal{M}^{j}|}{(\mu_{20}^{j}/|\mathcal{M}^{j}| + \mu_{02}^{j}/|\mathcal{M}^{j}|)^{(p+q)/2}} = \frac{\mu_{pq}^{j}}{(\mu_{20}^{j}+\mu_{02}^{j})^{(p+q)/2}|\mathcal{M}^{j}|^{1-(p+q)/2}} \tag{8}$$

Note that unlike computing moments for regions, in our case we can not use the first order moment μ_{00}^j for scale normalization. This is due to the fact that when the scale of an object reduces, for example, the number of wavelet maxima points may decreases because of *both* the reduction in size and also the lost of details in high frequencies.

Finally, to add in rotation invariance, we compute seven invariant moments up to the third order as derived in [3] for each scale, except invariants $\eta_{20}^j + \eta_{02}^j$ (which are always equal to 1 due to our scale normalization) are replaced by η_{00}^j. The current implementation of our system computes 4 levels of wavelet decomposition at scales 2^j, $1 \le j \le 4$, and 7 invariant moments ϕ_i^j, $1 \le i \le 7$, for each scale, thus giving a total of 28 real numbers as the signature for each indexed image.

For testing, we simply adapt the most commonly used similarity metric, namely the variance weighted Euclidean distance [2]. The weighting factors are the inverse variances for each vector component, computed over all the images in the database. The normalization brings all components in comparable range, so that they have approximately the same influence to the overall distance.

5 Simulation Results

In this section, we evaluate the performance of the proposed method in the *query-by-example* approach. Since we are particularly interested in the invariant aspect of extracted features, a test image database was synthetically generated. Figure 2 shows the object library which consists of twenty different foods in small size images 89 by 64 pixels. For each object, a class of 10 images was constructed by randomly rotating, scaling and pasting that object onto a randomly selected background. Scaling factor is a uniform random variable between 0.5 and 1. The position of pasted objects was randomly selected but such that the object would entirely fit inside the image. The backgrounds come from a set of 10 wooden texture images of size 128 by 128 pixels. The test database thus contains 200, 128x128 grey level images. Each image in the database was used as a query in order to retrieve the other 9 relevant ones.

Figure 3 shows an example of retrieval results. The query image is on the top left corner; all other images are ranked in the order of similarity with the query image from left to right, top to bottom. In this case, all relevant images are correctly ranked as the top matches following by images of very similar shape but are different in visual details.

The retrieval effectiveness evaluation is shown in Figure 4 in comparison with the ideal case. By considering different number of the top retrieval (horizontal axis), the average number of the images from the same similarity class is used to measure the performance (vertical axis). This result is superior in compared with [4] where the retrieval performance was reported to drop significantly, about *five* times, if the query was translated, scaled and/or rotated.

Fig. 2. The object library of 20 food images of size 89 x 64.

6 Conclusion

This paper has presented a wavelet-based image retrieval system that is robust in searching for objects independently of position, size, orientation and image background. The proposed feature extraction method is based on the marriage of the wavelet maxima transform and invariant moments. The important point is that neither a moment or a wavelet maxima method alone would lead to the good performance we have shown, as thus, the combination of the two is the key. This results in an extracted feature set that is *compact*, *invariant* to translation, scaling, rotation, and *significant* - especially for shape and spatial information.

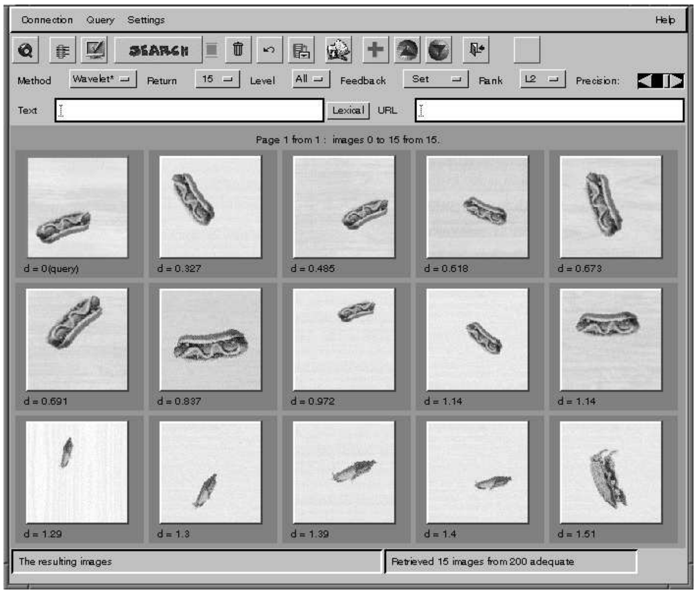

Fig. 3. Example of retrieval results from the synthetic image database.

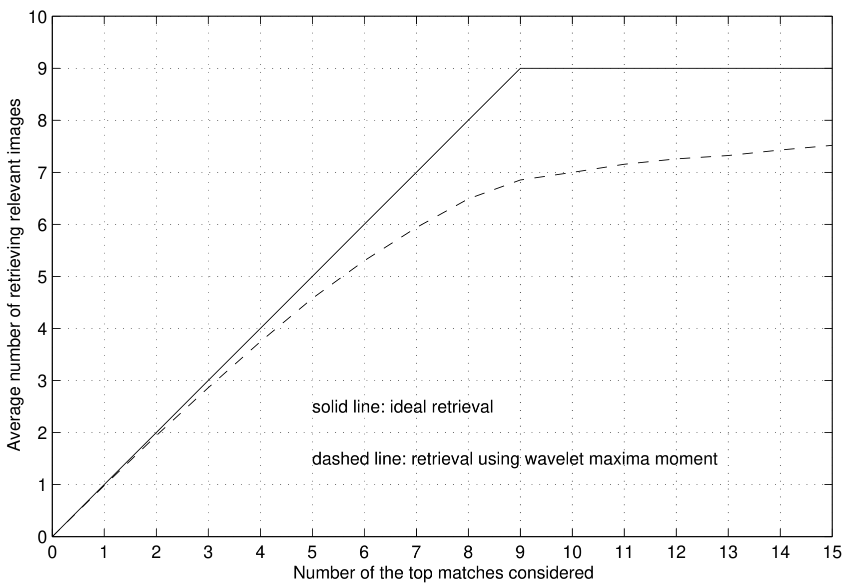

Fig. 4. Retrieval performance in comparison with the ideal case.

However, the presented retrieval system here is mainly based on configuration/shape related information. This is because of the moment computation puts emphasis on the positions of the wavelet maxima or edge points of the image. Extensions on extracting other types of image information from the wavelet maxima transform are being explored. In particular, color-based information can be efficiently extracted from the scaling coefficients which correspond to a low resolution version of the original image. Texture can be characterized by a set of energies computed from wavelet coefficients from each scale and orientation. To conclude, the main advantage of using wavelet transform in image retrieval application is that it provides a fast computation process to decompose image into meaningful descriptions.

Acknowledgments

The authors would like to thank Wen Liang Hwang , Stephane Mallat and Sifen Zhong for their Wave2 package and Zoran Pečenović for his user interface software.

References

1. C. Carson, S. Belongie, H. Greenspan, and J. Malik. Region-based image querying. In *IEEE Workshop on Content-based Access of Image and Video Libraries*, Puerto Rico, June 1997.
2. M. Flickner et al. Query by image and video content: The QBIC system. *Computer*, pages 23–32, September 1995.
3. M.-K. Hu. Visual pattern recognition by moment invariants. *IRE Trans. Info. Theory*, IT-8:179–187, 1962.
4. C.E. Jacobs, A. Finkelstein, and D.H. Salesin. Fast multiresolution image querying. In *Computer graphics proceeding of SIGGRAPH*, pages 278–280, Los Angeles, 1995.
5. K.-C. Liang and C.-C. Jay Kuo. Progressive image indexing and retrieval based on embedded wavelet coding. In *IEEE Int. Conf. on Image Proc.*, 1997.
6. W. Y. Ma and B. S. Manjunath. NETRA: A toolbox for navigating large image databases. In *IEEE International Conference on Image Processing*, 1997.
7. S. Mallat and S. Zhong. Characterization of signals from multiscale edges. *IEEE Trans. Pattern Anal. Machine Intell.*, 14:710–732, July 1992.
8. A. Pentland, R.W. Piccard, and S. Sclaroff. Photobook: Content-based manipulation of image databases. *International Journal of Computer Vision*, 18(3):233–254, 1996.
9. J.R. Smith and S.-F. Chang. VisualSEEk: a fully automated content-based image query system. In *Proc. The Fourth ACM International Multimedia Conference*, pages 87–98, November 1996.
10. M. Stricker and M. Orengo. Similarity of color images. In *Storage and Retrieval for Image and Video Databases III*, volume 2420 of *SPIE*, pages 381–392, 1995.
11. M. Vetterli and J. Kovacevic. *Wavelets and Subband Coding*. Prentice-Hall, Inc, 1995.
12. J. Z. Wang, G. Wiederhold, O. Firschein, and S. X. Wei. Wavelet-based image indexing techniques with partial sketch retrieval capability. In *Proceedings of 4th ADL Forum*, May 1997.