ЗБОРНИК РАДОВА
PROCEEDINGS

IX ТЕЛЕКОМУНИКАЦИОНИ ФОРУМ
IX TELECOMMUNICATIONS FORUM

ТЕЛФОР 2001
TELFOR 2001

Београд, 20., 21. и 22. новембар 2001.
Belgrade, 20, 21 and 22 November 2001

9th TELECOMMUNICATIONS FORUM, Belgrade, Yugoslavia, Nov. 20-22, 2001.

LOCATION OF OBJECTS IN A JPEG IMAGE WITH PROGRESSIVE WAVELET CORRELATION USING FOURIER METHODS

Stojanovic I.[1], Bogdanov M.[2]
[1]Customs Administration, Skopje, R. Macedonia
[2]Faculty of Electrical Engineering Skopje, R. Macedonia

I INTRODUCTION

A key tool that helped make the Internet universally useful is the text-search engine. The image-search engines available today are relatively crude. There are several techniques for image searching: descriptor-based search, pixel-based search and image understanding techniques. The fastest methods available today use descriptor-based search techniques. IBM QBIC (www.qbic.almaden.ibm.com) [1] is an example of this type of search engine. Images with higher information content, such as satellite images and medical images, are difficult to encapsulate with descriptors. Queries on images of this type require detailed analysis. Normalized correlation coefficients, an instance of pixel-based search techniques, measure the differences between images and patterns. They can be computed with progressive wavelet correlation using Fourier methods [2]. The images are mapped into the wavelet-frequency domain to take advantage of high-speed correlation.

The paper is organized as follows. Section II contains the brief description of progressive wavelet correlation using Fourier methods [2]. The results obtained by implementing progressive wavelet correlation using Fourier methods for location of objects in a JPEG image are given in Section III.

II PROGRESSIVE WAVELET CORRELATION USING FOURIER METHODS

In this section we summarize the technique described in [2]. Progressive wavelet correlation using Fourier methods is based on four theorems: the Wavelet-Correlation Theorem, the Fourier-Wavelet Correlation Theorem, the Fourier-Wavelet Subband Theorem and the Fourier-Wavelet Multiresolution Theorem. To simplify the discussion all data are assumed to be one-dimensional vectors.

The fundamental operation for searching is the circular correlation $\mathbf{x} \otimes \mathbf{y}$. The j^{th} entry of the circular correlation is defined as:

$$(\mathbf{x} \otimes \mathbf{y})_j = \sum_{i=0}^{N-1} x_{i+j \bmod N}\, y_i\,, \quad \text{for } j = 0, 1, \ldots, N-1 \qquad (1)$$

where **x** and **y** are column vectors of length N.

The matrix form is $\mathbf{x} \otimes \mathbf{y} = \mathbf{X}\mathbf{y}$, where **X** is left circulant matrix generated by **x**:

$$\mathbf{X} = \begin{bmatrix} x_0 & x_1 & \cdots & x_{N-1} \\ x_1 & x_2 & \cdots & x_0 \\ \cdots & \cdots & \cdots & \cdots \\ x_{N-1} & x_0 & \cdots & x_{N-2} \end{bmatrix} \qquad (2)$$

The notation $(\mathbf{P})_{i\downarrow R}$ denotes subsampling of **P** by taking components whose indices are equal to i modulo R. For example, if R = 4 and i = 2, the subsampling operation extracts the elements 2, 6, 10, 14, ... from **P**.

Wavelet-Correlation Theorem:

$$(\mathbf{x} \otimes \mathbf{y})_{0\downarrow R} = \sum_{k=0}^{R-1} \left((\mathbf{Hx})_{k\downarrow R}\right) \otimes \left((\mathbf{Hy})_{k\downarrow R}\right) \qquad (3)$$

where **H** is wavelet-packet transform. $\times$ is the Kronecker product of $\mathbf{I}_M$ and **W**, $\mathbf{H} = \mathbf{I}_M \times \mathbf{W}$, where $\mathbf{I}_M$ is MxM identity matrix and **W** is an RxR matrix with property $\mathbf{W}^T\mathbf{W} = \mathbf{I}_R$. The wavelet transform packet matrix **H** has a special structure. **H** is block diagonal with block size R. For instance, **W** can be 2x2 Haar matrix:

$$\mathbf{W} = \left(\frac{1}{\sqrt{2}}\right)\begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}. \qquad (4)$$

Fourier-Wavelet Correlation Theorem:

$$(\mathbf{x} \otimes \mathbf{y})_{0\downarrow R} = \mathbf{F}_M^{-1}\left(\sum_{k=0}^{R-1} \left(\mathbf{F}_M\left((\mathbf{Hx})_{k\downarrow R}\right)\right) * \left(\hat{\mathbf{F}}_M\left((\mathbf{Hy})_{k\downarrow R}\right)\right)\right) \qquad (5)$$

where $\mathbf{F}_M$ is the Fourier transform matrix of dimension M and $\hat{\mathbf{F}}_M$ is the complex conjugate of $\mathbf{F}_M$.

Fourier-Wavelet Subband Theorem:

$$\mathbf{F}_N\mathbf{x} = \left(\mathbf{T}_{N,M,R}\mathbf{H}^{-1}\right)\left(\mathbf{F}_{M,R}\mathbf{H}\mathbf{x}\right) \qquad (6)$$

where $N = MR$. The matrix $\mathbf{F}_{M,R}$ is an interlaced Fourier transform matrix with structure $\mathbf{F}_{M,R} = \mathbf{F}_M \times \mathbf{I}_R$, that is it has R interlaced copies of transform of size M. The matrix $\mathbf{T}_{N,M,R}$ is a Fourier update matrix that transforms $\mathbf{F}_{M,R}$ into $\mathbf{F}_N$: $\mathbf{F}_N = \mathbf{T}_{N,M,R}\mathbf{F}_{M,R}$.

Fourier-Wavelet Multiresolution Theorem:

$$\mathbf{F}_{N/R,R}\mathbf{H}_1\mathbf{x} = \left(\mathbf{T}_{N/R,N/R^2,R} \times \mathbf{I}_R\right)\mathbf{U}_{2,1}^{-1}\mathbf{F}_{N/R^2,R^2}\mathbf{H}_2\mathbf{x} \qquad (7)$$

$$\mathbf{F}_N\mathbf{x} = \mathbf{T}_{N,N/R,R}\mathbf{H}_1^{-1}\left(\mathbf{T}_{N/R,N/R^2,R} \times \mathbf{I}_R\right)\mathbf{U}_{2,1}^{-1}\mathbf{F}_{N/R^2,R^2}\mathbf{H}_2\mathbf{x} \qquad (8)$$

where $N = MR^2$. $\mathbf{H}_2$ is a coarse transform matrix that is block diagonal with block of size R^2 with the structure $\mathbf{H}_2 = \mathbf{I}_{N/R^2} \times (\mathbf{W}_1 \times \mathbf{W}_1)$ and operates on R^2 subbands, each of length N/R^2. $\mathbf{W}_1$ is an RxR wavelet filter matrix with property $\mathbf{W}_1^T\mathbf{W}_1 = \mathbf{I}_R$. $\mathbf{H}_1$ is a fine transform matrix that is block diagonal with block size N/R with structure $\mathbf{H}_1 = \mathbf{I}_{N/R} \times \mathbf{W}_1$. There is an update matrix $\mathbf{U}_{2,1}$ that refines $\mathbf{H}_1$ into $\mathbf{H}_2$, $\mathbf{H}_2 = \mathbf{U}_{2,1}\mathbf{H}_1$. The matrix $\mathbf{U}_{2,1}$ is block diagonal with block of size R^2 with the following structure:

$$\mathbf{U}_{2,1} = \mathbf{I}_{N/R^2} \times (\mathbf{W}_1 \times \mathbf{I}_R) \qquad (9)$$

III LOCATION OF OBJECTS IN A JPEG IMAGE

JPEG compression is based on the discrete-cosine transform (DCT) [3]. The matrix $\mathbf{C}_8$ is an 8x8 DCT matrix that is used to create transforms of 8x8 subimages in a JPEG representation of an image. The multiresolution search process relies on the factorization of the DCT matrix $\mathbf{C}_8 = \mathbf{V}_{8,4}\mathbf{V}_{4,2}\mathbf{V}_2$, where $\mathbf{V}_2$ and $\mathbf{V}_{4,2}$ are matrices built from Kronecker products of $\mathbf{W}$ and the identity matrix.

The matrix $\mathbf{V}_2 = \mathbf{I}_4 \times \mathbf{W}$ consists of 4 interlaced copies of $\mathbf{W}$ and is of size 8x8. The matrix $\mathbf{V}_{4,2}$ has a structure $\mathbf{V}_{4,2} = \mathbf{I}_2 \times (\mathbf{W} \times \mathbf{I}_2)$. If we write $\mathbf{C}_8 = \mathbf{V}_{8,4}\mathbf{V}_{4,2}\mathbf{V}_2$ where $\mathbf{V}_{8,4}$ is a matrix whose coefficients we want to compute, then

$$\mathbf{V}_{8,4} = \mathbf{C}_8\mathbf{V}_2^{-1}\mathbf{V}_{4,2}^{-1} \qquad (10)$$

obtain the last expression by multiplying both sides by $\mathbf{V}_2^{-1}\,\mathbf{V}_{4,2}^{-1}$. The matrix $\mathbf{V}_{8,4}$ satisfies equation $\mathbf{V}_{8,4} = \mathbf{V}(\mathbf{W} \times \mathbf{I}_4)$. The inverse of $\mathbf{V}$ is

$$\mathbf{V}^{1} = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & \gamma(2)\gamma(7) & 0 & \gamma(5)\gamma(6) & 0 & \gamma(3)\gamma(6) & 0 & \gamma(2)\gamma(1) \\ 0 & \gamma(6)\gamma(1) & 0 & \gamma(3)\gamma(2) & 0 & -\gamma(5)\gamma(2) & 0 & -\gamma(6)\gamma(7) \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & \gamma(2)\gamma(1) & 0 & -\gamma(3)\gamma(6) & 0 & \gamma(5)\gamma(6) & 0 & -\gamma(2)\gamma(7) \\ 0 & 0 & \gamma(6) & 0 & 0 & 0 & \gamma(2) & 0 \\ 0 & 0 & \gamma(2) & 0 & 0 & 0 & -\gamma(6) & 0 \\ 0 & -\gamma(6)\gamma(7) & 0 & \gamma(5)\gamma(2) & 0 & \gamma(3)\gamma(2) & 0 & -\gamma(6)\gamma(1) \end{bmatrix} \qquad (11)$$

where $\gamma(m) = \cos(2\pi m/32)$.

The matrix $\mathbf{H}$ is an NxN matrix with the structure $\mathbf{I}_M \times \mathbf{C}_8$ where $N = 8M$. It produces the JPEG transform of a vectors of length N. Let $\mathbf{x}$ be image stored as a JPEG transform of a vector $\mathbf{Hx}$ with an instance of a pattern $\mathbf{y}$ with JPEG transform $\mathbf{Hy}$. To locate object $\mathbf{y}$ in a JPEG image we search the image step by step. The algorithm consists of three incremental steps, each of which quadruples the number of correlation points. The process can be halted at any stage if the intermediate results indicate that the correlation will not result in a match.

The three incremental steps are:

1. *Coarse correlation* – Generate the Fourier transforms $\mathbf{F}_{M,8}\mathbf{Hx}$ and $\hat{\mathbf{F}}_{M,8}\mathbf{Hy}$. Multiply the transforms point by point and partition them into eight subbands of length M. Add these eights vectors and take the inverse Fourier transform of the sum. Every eighth point of the correlation is generated.
2. *Medium correlation* – Multiply $\mathbf{F}_{M,8}\mathbf{Hx}$ by $(\mathbf{T}_{2M,M,2} \times \mathbf{I}_4)(\mathbf{I}_M \times ((\mathbf{W} \times \mathbf{I}_4)\mathbf{V}^{-1}))$ and $\hat{\mathbf{F}}_{M,8}\mathbf{Hy}$ by $(\hat{\mathbf{T}}_{2M,M,2} \times \mathbf{I}_4)(\mathbf{I}_M \times ((\mathbf{W} \times \mathbf{I}_4)\mathbf{V}^{-1}))$. Multiply the resulting vectors point by point and partition them in four subband of length $2M$. Add the subbands and you will create a single vector of length $2M$. Taking the inverse Fourier transform of size $2M$ yields the correlation at indices that are multiples of 4 mod 8 of the full correlation.
3. *Fine correlation* – Multiply the $\mathbf{x}$ and $\mathbf{y}$ transform from the preceding step by $(\mathbf{T}_{4M,2M,2} \times \mathbf{I}_2)(\mathbf{I}_M \times \mathbf{V}_{4,2})$ and $(\hat{\mathbf{T}}_{4M,2M,2} \times \mathbf{I}_2)(\mathbf{I}_M \times \mathbf{V}_{4,2})$, respectively. Multiply the resulting vectors point by point and partition them in two subbands of length $4M$. Add the subbands and you will create a single vector of length $4M$. Take the inverse Fourier transform of size $4M$ to obtain the correlation at indices that are multiples of 2 mod 8 and 6 mod 8 of the full correlation.

4. *Full correlation* – Multiply the **x** and **y** transform from the last step by $\mathbf{T}_{8M,4M,2}(\mathbf{I}_M \times \mathbf{V}_2)$ and $\hat{\mathbf{T}}_{8M,4M,2}(\mathbf{I}_M \times \mathbf{V}_2)$, respectively. Multiply the resulting vectors point by point and take the inverse Fourier transform of size $8M$ to obtain the correlation at odd indices.

We investigate what happens in the two-dimensional case. The assumption is that the image size is N by N. In step 1, we have 64 subbands of length $N^2/64$. We perform one step of the inverse 2D **H** function, and one 2D step of the forward Fourier transform function. Fortunately, these steps are simple generalizations of the 1D functions. Specifically, if in 1D we compute **Hx**, where **H** is N by N and **x** is N by 1, then in 2D we compute **H2X** conjtrans (**H2**), where **X** is the 2D image to be transformed, **H2** is N by N, and each row of **H2** is equal to **H**. This is equivalent to applying **H** to each of the columns of **X** and **H** to each of the rows of **X**. The next step is to add the 64 subbands point by point to create a 2D array of size $N/8$ by $N/8$. If we take its inverse Fourier transform, we will obtain the correlations at points that lie on a grid that is coarser than the original pixel grid by a factor 8 in each dimension.

In step 2, we obtain 16 subbands of size $N^2/16$ by adding the 16 subbands point by point, and taking the Fourier inverse. We will obtain the correlation values on a grid that is coarser than the original grid by a factor of 4 in each dimension.

In step 3, we obtain 4 subbands of size $N^2/4$.

In step 4, we obtain the full resolution.

For instance, if x and y are 16 by 16 images, we will have the correlation at the points given by 1s below:

Step1

Coarse correlation
1000 0000 1000 0000
0000 0000 0000 0000
0000 0000 0000 0000
0000 0000 0000 0000
0000 0000 0000 0000
0000 0000 0000 0000
0000 0000 0000 0000
0000 0000 0000 0000
1000 0000 1000 0000
0000 0000 0000 0000
0000 0000 0000 0000
0000 0000 0000 0000
0000 0000 0000 0000
0000 0000 0000 0000
0000 0000 0000 0000
0000 0000 0000 0000

Step2

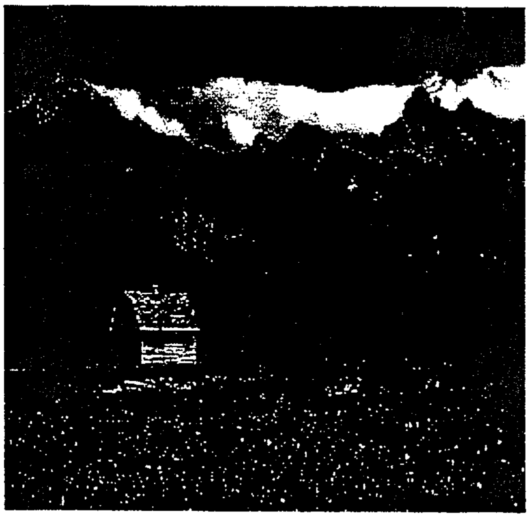
Medium correlation
1000 1000 1000 1000
0000 0000 0000 0000
0000 0000 0000 0000
0000 0000 0000 0000
1000 1000 1000 1000
0000 0000 0000 0000
0000 0000 0000 0000
0000 0000 0000 0000
1000 1000 1000 1000
0000 0000 0000 0000
0000 0000 0000 0000
0000 0000 0000 0000
1000 1000 1000 1000
0000 0000 0000 0000
0000 0000 0000 0000
0000 0000 0000 0000

Step3

Fine correlation
1010 1010 1010 1010
0000 0000 0000 0000
1010 1010 1010 1010
0000 0000 0000 0000
1010 1010 1010 1010
0000 0000 0000 0000
1010 1010 1010 1010
0000 0000 0000 0000
1010 1010 1010 1010
0000 0000 0000 0000
1010 1010 1010 1010
0000 0000 0000 0000
1010 1010 1010 1010
0000 0000 0000 0000
1010 1010 1010 1010
0000 0000 0000 0000

Formulas for calculating of normalized correlation coefficients for measuring differences between images and patterns are given in [2]. Normalized correlation coefficients can be computed from the correlations described above. The normalization is very important because it allows for a threshold to be set. Such a threshold is independent of the encoding of the images. The normalized correlation coefficient has a maximum absolute value of 1. Correlations that have absolute values above 0.9 are excellent, and almost always indicate a match found. Correlations of 0.7 are good matches. Correlations of 0.5 are usually fair or poor. Correlations of 0.3 or less are very poor. There is a tradeoff between the value of the threshold and the likelihood of finding a relevant match. Higher thresholds reduce the probability of finding something that is of interest, but they also reduce the probability of falsely matching something that is not of interest.

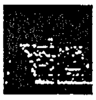
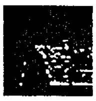
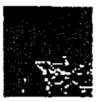
Location of objects in JPEG images was tested on a great number of pictures with MATLAB assistance. The following example shows the location of the object house in the picture house.jpg.



Pic.1. Picture house.jpg, which contains the object

The picture size is 280x280 pixels; the size of the object we aim to locate is 48x48 pixels. The object is positioned at $x_{min} = 57$ and $y_{min} = 153$. We search the image along the x-axis, and subsequently along the y-axis, for different thresholds assigned by the lim variable. The located objects for different thresholds values are given in the following pictures:

		
lim = 1	lim = 0.9	lim = 0.8
$x_{min} = 57$	$x_{min} = 49$	$x_{min} = 49$
$y_{min} = 153$	$y_{min} = 153$	$y_{min} = 145$

		
lim = 0.7	lim = 0.6	lim = 0.5
$x_{min} = 57$	$x_{min} = 49$	$x_{min} = 49$
$y_{min} = 137$	$y_{min} = 137$	$y_{min} = 129$

From the pictures we can conclude the following: decreasing the threshold value speeds up our algorithm, but too small threshold values may result in incorrect matches.

IV CONCLUSIONS

In this paper we presented our work that examines the technique of progressive wavelet correlation using Fourier methods for locating of objects in a JPEG image, which led us to the following conclusions.

This technique is not yet suitable for general practical commercial usage. The reason for that is the big number of operations per picture. In the following years, with increasing the processor speed, there should be a possibility for detail analysis at the rate of 1000 pictures per second. With this processing speed, it would be easy to construct a system, which is a combination of searching by description (descriptor-based search) and searching by pixels (pixel-based search). The descriptor can be used for isolating a certain part out of a big collection, which should be an object of a detail pixel-based search in the later phase.

References

[1] M. Flickner et al., "Query by image and video content: The QBIC system," *IEEE Comp.*, vol. 28, pp. 23-32, Sept. 1992.

[2] H. S. Stone, "Progressive Wavelet Correlation Using Fourier Methods," *IEEE Trans. Signal Processing*, vol. 47, pp. 97-107, Jan. 1999.

[3] G. K. Wallace, "The JPEG still-picture compression standard", *Commun. ACM*, vol. 34, no.4, pp. 30-44, Apr. 1991.

Abstract: In the paper we apply progressive wavelet correlation using Fourier methods to locate objects in a JPEG image. The image is searched step by step. The searching consists of three incremental steps, each of which quadruples the number of correlation points. The process can be halted at any stage if the intermediate results indicate that the correlation will not result in a match.

LOCATION OF OBJECTS IN A JPEG IMAGE WITH PROGRESSIVE WAVELET CORRELATION USING FOURIER METHODS,
Igor Stojanovic, Momcilo Bogdanov.