1999

Combining Wavelets and Hotelling Transforms in Image Query

Sol Neeman Ph.D.
Johnson & Wales University - Providence, sneeman@jwu.edu

Combining Wavelets and Hotelling transforms in Image Query

Sol Neeman
Johnson and Wales University

Introduction

In the "content based" image query, also referred to as "query by example", "similarity retrieval" or "sketch retrieval", the query image is provided by the user either as a sketch of the object, as the output of a scanner or a video camera. Some of the difficulties associated with content based image query are described in [1], e.g. significant differences between the "query image" and the "target image", artifacts and poor resolution of the query image make a straightforward comparison of images using L^1 and L^2 metrics not effective.

In [2] a new strategy is suggested based on wavelet decomposition of the query image and the database images combined with a metric which is designed to be insensitive to small differences in the query process. This approach is found to be fast and overcomes the above mentioned problems.

Wavelet coefficients of an image may very strongly when the image is displaced or rotated (unlike color histogram of an image which is invariant under displacement and rotation). Although the metric suggested in [2] is more robust to these errors when compared to the L^1 and L^2 metric (but worse when compared to the metric based on color histograms), still the error is significant.

In this paper we provide some experimental data on the sensitivity of the wavelet coefficients to displacement and rotation in the context of the standard characters and suggest an integration of the Hotelling transform to improve on this sensitivity. We also provide some experimental data on the distribution of the largest wavelet coefficients at different levels of the wavelet decomposition and discuss some questions relevant to the approach of using a set of the largest wavelet coefficients for image query.

Overview of this paper

Section 2 describes the content based image query using wavelet decomposition as described in [1] and [2], including the L^q metric used to compare the wavelet coefficients of the query image and the target image. In section 3 the Hotelling transform is described including its algebra. Section 4 describes some experimental results using MATLAB regarding the sensitivity of the L^q metric to rotation and displacement of an image. Section 5 will describe the use of the Hotelling transform to improve the image query process. In section 6 we discuss some questions regarding the set of the largest wavelet coefficients as representing an image: its size, its distribution at different levels of the wavelet decomposition and the L^q metric used to compare two such sets.

2. Content based image query using wavelet decomposition

Some of the previous approaches to content based image query, include the use of certain properties of the images or a combination of these properties. For example, the user may specify a color combination (color histogram), a texture, geometrical features(e.g. edges, shapes or major axis orientation) or a rough sketch of the image. The multiresolution approach suggested in [2] appears to have a success rate at least as good as that of other systems that work from a simple user sketch. This approach has some advantages when compared to the previous ones. First, It decouples the resolution of the query image and the target image and hence the query image can be specified at any resolution. Second, the performance (running time) is independent of the resolution of the database images(only the set containing the largest wavelet coefficients is used in the comparison regardless of the resolution of the images) and third, it provides a simple algorithm and a compact code in the implementation.

As described in [2], a wavelet decomposition is applied to the database images(color images, of size 128x128) , using Haar wavelets (simple to compute) and standard decomposition. Then, all but the 60 (experimental result) largest (in magnitude) coefficients are truncated for each color channel . These remaining 60 coefficients are then quantized to two levels only, +1 and -1 for positive and negative coefficients respectively. The metric developed (denoted by L^q), compares only the indices of these 60 largest, quantized coefficients of the query image with those of the data base images, and the scores of this comparisons are used to pull the 20 best matches from the database. The truncation and quantization are significant in this approach, since they reduce the search time and the storage requirements, but more importantly they improve the performance of the L^q metric used in comparing the query image to the target. This is because the truncation and quantization make the L^q metric tolerant to small differences between the query image and the target while giving more weight to the significant features that are closely matched. When compared to the L^1 and L^2 metrics (which take in account all the 128^2 wavelet coefficients of the images), and the L^c metric, the metric that uses color histograms, the L^q metric has a better success rate.

As to the robustness of the multiresolution approach to distortion due to image rotation and displacement of the query image, when compared to the L^1 and L^2 metrics, the L^q metric has a better success rate, but when compared to the

L^c metric, its performance is worse (since color histograms are not sensitive to rotation and translation).

3. The Hotelling Transform and the principal axis of an image

The Hotelling transform is a linear transformation of a set of n dimensional vectors that decorrelates the n coordinates. When applied to an 2-dimensional image, the transformed image will be aligned along its principal axis.

The Hotelling transform can be used to improve the robustness to distortion of the image due to rotation and displacement. The improvement though, would increase the running time of the image query.

This transformation is a combination of displacement and rotation of the object. The centroid of the image(the average of the x and y coordinates of all pixels) is shifted to the origin and the image is rotated by an angle that minimizes its moment of inertia[4]. Geometrically the transformed image will be oriented in the direction in which it seems to be the most 'elongated'.

In image recognition this transformation is helpful, since the identity of the object is not known and aligning the image with its principal axis can help remove the effects of translation and rotation in the analysis.

First we describe the Hotelling transform. Given a set k of n-dimensional column vectors: $X_1, X_2, ..., X_k$,the covariance matrix of the vector population is given by:

$$C_x = E[(X - m_x)(X - m_x)^T]$$

where m_x is the mean vector of the population.

C_x is a nxn real symmetric matrix whose c_{ij} entry equals the covariance between the i^{th} and the j^{th} coordinates of the vector population. If $c_{ij} = 0$, then the i^{th} and the j^{th} coordinates are decorrelated. When $c_{ij} > 0$ or $c_{ij} < 0$ then the there is a positive or a negative correlation between the i^{th} and the j^{th} coordinates, respectively.

The Hotelling transform maps the given vector population, X, into a vector population Y (that consists of k, n-dimensional vectors) such that $m_y = 0$ and C_y , the covariance of the new vector population is a nxn diagonal matrix and therefore the i^{th} and the j^{th} coordinates of the new vector population are decorrelated for all $i \neq j$. To see the relevance of the Hotelling transform to aligning 2-dimensional image with its principal axis, consider an image composed of mxn pixels . If the (x, y) coordinates of each pixel is presented as a 2-dimensional vector, the image then is represented by a set of mxn, 2-dimensional, vectors. When the Hotelling transform is applied to this vector population, the mean vector of the new vector population, $m_y = 0,$ and its covariance will be diagonal, which means that the x and y coordinates are decorrelated.

Geometrically this would mean that the centroid of the original image was shifted to the origin and the image was rotated so that its principle axis is aligned with the X-axis.

Algebra of the Hotelling transform

Let X be a set of k, n-dimensional column vectors: $X_1, X_2, ..., X_k$

The mean vector of the population, m_x , is given by:

$$m_x = \frac{1}{K}\sum_{i=1}^{K} X_i$$

and the covariance matrix of the vector population ,X, is:

$$\begin{aligned}
C_x &= E[(X - m_x)(X - m_x)^T] = \\
&= E(XX^T) - E(Xm_x^T) - E(m_x X^T) + E(m_x m_x^T) = \\
&= E(XX^T) - m_x m_x^T - m_x m_x^T + m_x m_x^T = \\
&= \frac{1}{K}\sum_{i=1}^{K} X_i X_i^T - m_x m_x^T
\end{aligned}$$

Now, C_x is a nxn real and symmetric matrix (because it is the sum of the products of vectors by their transpose and the entries of the vectors, which are the pixel coordinates, are real values); therefore it has n real eigen values and n real orthogonal eigen vectors[5].

Let $\lambda_1, \lambda_2, ..., \lambda_n$ be the n eigen values with $\lambda_1 \geq \lambda_2 \geq ... \geq \lambda_n$ and $V_1, V_2, ..., V_n$ be the corresponding set of orthonormal eigen vectors, where V_1corresponds to λ_1, the largest eigen value.

Define the matrix A whose rows are the n eigen vectors $V_1, V_2, ..., V_n$:

$$A = \begin{bmatrix} V_1 - - - - > \\ V_2 - - - - > \\ . \\ . \\ V_n - - - - > \end{bmatrix}$$

The Hotelling transform is then the linear transformation given by:

$$Y_i = A(X_i - m_x) \qquad i = 1, 2, ..., k$$

The mean of the vector population Y, is 0, since

$$\begin{aligned} m_y &= \frac{1}{K}\sum_{i=1}^{K} Y_i = \\ &= \frac{1}{K}\sum_{i=1}^{K} A(X_i - m_x) = \\ &= A\frac{1}{K}\sum_{i=1}^{K}(X_i - m_x) = 0 \end{aligned}$$

and the covariance matrix of the vector population Y is diagonal:

$$\begin{aligned} C_y &= E[(Y - m_y)(Y - m_y)^T] = \\ &= \frac{1}{K}\sum_{i=1}^{K} Y_i Y_i^T - m_y m_y^T = \\ &= \frac{1}{K}\sum_{i=1}^{K} Y_i Y_i^T = \\ &= \frac{1}{K}\sum_{i=1}^{K}[A(X_i - m_x)][A(X_i - m_x)]^T = \\ &= \frac{1}{K}\sum_{i=1}^{K} A(X_i - m_x)(X_i - m_x)^T A^T = \\ &= A[\frac{1}{K}\sum_{i=1}^{K}(X_i - m_x)(X_i - m_x)^T]A^T = \\ &= AC_xA^T = \end{aligned}$$

$$= \begin{bmatrix} V_1 \longrightarrow \\ V_2 \longrightarrow \\ \cdot \\ \cdot \\ V_n \longrightarrow \end{bmatrix} C_x \begin{bmatrix} V_1 & V_2 & \cdot\cdot & V_n \\ \downarrow & \downarrow & \downarrow & \downarrow \\ & & & \end{bmatrix} =$$

$$= \begin{bmatrix} V_1 \longrightarrow \\ V_2 \longrightarrow \\ \cdot \\ \cdot \\ V_n \longrightarrow \end{bmatrix} \begin{bmatrix} \lambda_1 V_1 & \lambda_2 V_2 & \cdot\cdot & \lambda_n V_n \\ \downarrow & \downarrow & \downarrow & \downarrow \\ & & & \end{bmatrix} =$$

$$= \begin{bmatrix} \lambda_1 & & \bigcirc \\ & \lambda_2 & \\ & & \cdot & \\ \bigcirc & & & \lambda_n \end{bmatrix}$$

(where the last equality is because the eigen vectors are orthonormal).

4. Sensitivity of the Wavelet Transform to rotation and displacement of an image

Next we present some experimental data on the sensitivity of the wavelet transform to rotation and displacement in the context of the standard characters.

Let an image and its rotated version be represented by the matrices X and X_R respectively and the wavelet transform be represented by the matrix W. Since the wavelet basis is an orthonormal basis, it preserves the L^2 norm [8] and so the relative error when comparing X and X_R , i.e.$\frac{\|X-X_R\|}{\|X\|}$, equals the relative error when comparing their wavelet coefficients, i.e. $\frac{\|WX-WX_R\|}{\|WX\|}$.

But when the wavelet coefficients are truncated and quantized , the relative error in comparing only a limited set of the coefficients(largest in magnitude) depends on the distribution of the coefficient values. The wavelet transform will contain many coefficients whose magnitude is small for images with high degree of regularity. Following [2], we compare the 60 largest coefficients and present the error as the number of mismatches.

The experimental results were produced using MATLAB Wavelet Toolbox and the Image Processing Toolbox. The images contain monochromatic bitmap alphabet characters, standard upper case New Courier, size 72, created using PAINT-BRUSH. Each character was first aligned so that its centroid was at the center of a 228x228 image.

The results given are relevant in the context of the standard alphabet characters and they may vary for different types of database images, hence for a specific application this experiment should be duplicated to get a better representation of the errors.

The first test provides the data on the sensitivity of the largest 60 wavelet coefficients to rotation of the image. To avoid the error resulting from the mismatch of the edges of the rotated image when compared to the original, the central (128x128) portion (which includes the character) of the 228x228 image was rotated and the largest 60 wavelets coefficients were quantized to the values ± 1 and compared to the largest 60 wavelet coefficients of the center portion of the unrotated character.

We use the same range and resolution as in [2] to present the errors associated with rotation and translation of the image.

Fig. 1 shows the relative error in % (i.e., number of mismatches *100%/60) for some of the standard alphabet characters. Results are presented for angle of rotation in the range 0 to 45 degrees in steps of 5 degrees. The relative error is closely linear with respect to the angle of rotation. Note that the relative error for the characters 'C' and 'D' which have radial symmetry, is less than the error for the characters 'B', 'A' and 'Z'. The errors for the latter characters are significantly high and are in the range of 40% to 50% for a rotation of 10 degrees.

In the second test we checked the effect of displacement on the largest wavelet coefficients. The character was displaced by a certain percentage of the width of the character(0 to 50%, in steps of 5%) and the 60 largest wavelet coefficients were compared to those of the original image. Fig. 2 shows the results for the same standard alphabet characters. While the relative error for the different characters is closely the same, note that the error in general is very significant due to displacement for all 5 characters. For example, a displacement of the image by 0.1 of the width of the image results in a relative error in the range of 75%-85% . The error due to displacement will be removed by the Hotelling transform since the transformed image will be aligned with its centroid at the origin.

Fig. 3 and 4 provide some results on the error due to rotation and translation for a few different wavelet bases. The 'haar' wavelet is more sensitive to rotation and displacement when compared to 'db2', 'db4' and 'coif' wavelets. This is because the latter wavelets have a wider support and use a larger neighborhood resulting in stronger averaging effect which make them less sensitive to rotation and displacement. On the other hand, this means that they would have worse performance in discriminating between different images (compared to 'haar').

5. Combining the Hotelling transform in the query process

As was mentioned before, the error due to displacement will be eliminated by the Hotelling transform since the image centroid will be aligned with the origin. The problem is not simple when it comes to the error due to rotation. The angle of orientation of an image, depends on the geometry of the image. Thus some images may have a principle axis which may be sensitive to slight variations in the geometry of the image and others may not. For example, objects with a geometry that strongly resembles a rectangular would have a principal axis not sensitive to small distortions, while a geometry that resembles a square or circle would be sensitive to small distortions. Because of that, the Hotelling transform cannot be integrated in a straightforward manner and rather than improving the query, it may make it worse, since these objects may be rotated by an unpredictable and undesirable angle.

To overcome this problem, once the query image is given and its principal axis computed by the Hotelling transform, the angle of rotation that corresponds to the principal axis can be checked. If the value is within some permissible window(which will depend on the application) the transformed image(i.e. the centered and rotated image according to its principal axis) can be used as the query image. Otherwise, we form a set of images, (which we call an image query set) that will include the original query (with its centroid aligned at the origin) and a few rotated (both clockwise and anti clockwise) versions of the image query. This set then will be used in the query process. Rather than comparing a single query image, each of the images in the query set will be compared to the targets in the database, and the score can be a weighted sum of these comparisons(or alternatively, the score can be that of the best match among the rotated versions).

As to what is the permissible angle and how many images the query set should include, this would depend on the application, the type of images, the desirable search speed and the level of error since there is a trade off between the query time and the success rate. Data that provide the relative error as a function of rotation can help in determining the width of the permissible window and the number of images in the image query set.

In the implementation of the Hotelling transform, the eigen vector corresponding to the largest eigen value of the covariance matrix provides the angle that the image has to be rotated so that its principal axis would be aligned with the X-axis.

Table 1 provides some experimental results illustrating the improvement that can be achieved by using the Hotelling transform to align an image before its comparison. The Hotelling transform was implemented in MATLAB using the 'Haar' wavelet, and the non-standard decomposition. The table shows first the error in comparing a hand sketched query character with its standard counterpart and the error in comparing the transformed same hand sketched character and its counterpart standard character. In the implementation of the Hotelling transform, once the image is centered, the rotation angle provided by the Hotelling transform was checked. If the angle was less than 10 degrees, the image would be rotated otherwise it will only be shifted so that its centroid is at the origin. Results are presented for a few different alphabets with a distinct principal axis.

Table1 : The relative error (%) between the largest 60 coeff's of a hand sketched query character and its counterpart standard character with and without the Hotelling transform

	Ralative Error(%)	
character	H.T. applied to query char./	Not applied
A	45	91
I	85	98
J	85	98
M	60	81
S	86	86

6. Optimal set of the largest wavelet coefficients and their distribution at different levels of the wavelet decomposition

Next we present some questions relevant to the approach that uses the set of the largest wavelet coefficients for image query.

The first question is how reliable can a set of largest wavelet coefficients be as a discriminating tool in image query. Clearly the answer depends on the type of images used in the data base. Can certain features in the images comprising the data base(e.g. edges, contour lines, the gradient of an image) can be captured by the set of the largest wavelet coefficients so that they can be used as a discriminating tool in image query? As an experimental approach one may take a representative sample of the data base, then compute a matrix which provides the error in all pairwise comparisons in this set, then examine the data to see if there is a significant difference between the error in comparing images that closely resemble each other and the error in comparing images that differ significantly. As an example, table 2 provides such data for some of the standard alphabet characters using the 60 largest coefficients.

table 2 : The relative error(%) in comparing (pairwise) some of the alphabet characters

	A	B	C	D	E	F	G	X	Y	Z
A	0	65	68	63	68	73	57	58	83	73
B		0	68	48	58	78	58	60	73	42
C			0	50	73	72	52	75	70	73
D				0	65	73	65	68	75	65
E					0	77	68	62	80	67
F						0	80	75	68	82
G							0	75	77	65
X								0	72	47
Y										78
Z										

The second question relates to the optimal size of this set (of the largest coefficients). As was mentioned taking only a 'small' set enables to develop a metric which is insensitive to small differences but gives more weight to the significant features. On the other hand if the set is too small the metric may not be able to distinguish between images that differ in 'nonsignificant' details. It seems that the answer to this question depends strongly on the application and the data base (in [2] for example, the query process will result in the 20 images that best match the query image rather than a single best match).

The distribution of the largest set of wavelet coefficients at different levels of the wavelet decomposition

The third question relates to the distribution of the indices of the largest coefficients at different levels of the wavelet decomposition. Recalling that at each level of the wavelet decomposition a filter bank composed of a low pass filter and a high pass filter are applied to the data followed by the decimation of the results, the high pass filter at the first level of the decomposition will extract the highest frequency components of the data while the low pass filter will average the data. The second level of the decomposition will remove the next level of high frequency components from the output of the previous low pass filter and so forth, so that for an image with high level of regularity one would expect that the largest wavelet coefficients will appear in the later levels of the decomposition. For an image with random entries one would expect the inverse to happen, i.e. the first levels of the decomposition will contain the largest coefficients.

In an experiment this distribution was calculated for images of size 128x128 using 7 levels of nonstandard 'haar' decomposition.

The results are summarized in table 3. They include the standard letters A, B and Z (bitmap images), the standard indexed color images 'woman' and 'wbarb' (from MATLAB) and an indexed color image composed of random entries. The data refers to the distribution of the largest 60 wavelet coefficients.

Table 3 : *The distribution of the largest 60 coefficients at different levels of the wavelet decomposition for various images.*

	Decomposition Level						
	1	2	3	4	5	6	7
$'A'$	0	10	15	18	4	12	1
$'B'$	0	15	20	8	4	12	1
$'Z'$	0	22	20	5	3	9	1
random_entries	35	22	1	0	1	0	1
$'woman'$	0	0	5	24	19	8	4
$'wbarb'$	0	0	12	17	20	7	4

Note that for the image with random entries, almost all the 60 largest coefficients are at the first and second levels of the decomposition. In contrast, for all the other images, none of the coefficients is at the first level of decomposition and for the color indexed images 'woman' and 'wbarb' none is at the first or the second levels of the decomposition. In the standard wavelet decomposition, 3/4 of the total 128^2 coefficients are at level 1 and 15/16 of the coefficients are at levels 1 and 2. This means that for an application with a similar distribution, the search for the largest 60 coefficients can be limited only in the set of coefficients at levels 2 through 7 which is only 1/4 of the total coefficients or even in the set of coefficients at levels 3 through 7 which comprises only 1/16 of the total coefficients, thus increasing the speed of the query algorithm.

The L^q Metric:

In [2], the relative error is based on computing the overlap (or the common elements) of the indices of the 60 largest coefficients in the query and target images. If instead we allow a tolerance on the value of the indices when we compare the common elements, can this improve the metric?. Would this increase the gap between the error in comparing images that closely resemble each other and the error in comparing images which differ significantly? more research needs to be done to see if allowing a tolerance on the value of the indices can improve the metric.

Acknowledgment

I would like to thank Professor Balla Ravikumar of University of Rhode Island, Computer Science Department for his helpful comments, his advise, review efforts and encouragement during the work on this research. His guidance and endless

patience were crucial in completing this project.

References

[1] [1] E. Stollnitz, T. Derose and D. Salesin, *Wavelets for Computer Graphics*, Morgan Kaufmann Publishers, San Fransisco, 1986. pages 43-57.

[2] [2] C.E. Jacobs, A. Finkelstein and D. Salesin, *Fast Multiresolution Image Querying,in Proceedings of SIGGRAPH '95.*

[3] [3] Rafael C. Gonzalez, *Digital Image Processing*, Addision Wesley, 1993, pp 148-151.

[4] [4] J. Ritter and G. Wilson, *Handbook of Computer Vision Algorithms in Image Algebra*, CRC Press, 1996, pp 274-276.

[5] [5] B. Friedman, *Principles and Techniques of Applied Mathematics,* Dover Publications.pages 98-99.

[6] [6] MATLAB, *Reference Guide*, The Mathworks Inc.

[7] [7] MATLAB *Wavelet Toolbox*, The Mathworks Inc.

[8] [8] Gilbert Strang and Truong Nguyen, *Wavelets and Filter Banks*, Wellesley Cambridge. pages 26-27.

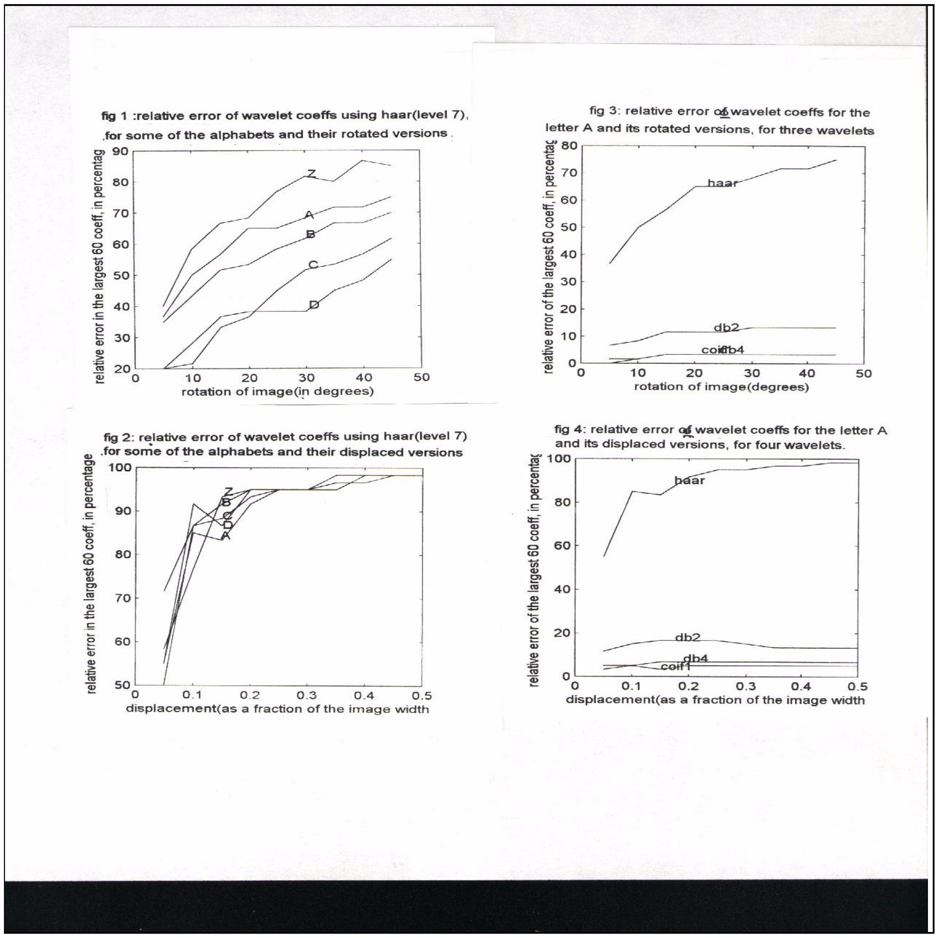
fig 1 :relative error of wavelet coeffs using haar(level 7),
.for some of the alphabets and their rotated versions
relative error in the largest 60 coeff, in percentag
rotation of image(in degrees)
fig 3: relative error of wavelet coeffs for the
letter A and its rotated versions, for three wavelets
relative error of the largest 60 coeff, in percentag
rotation of image(degrees)
haar
db2
fig 2: relative error of wavelet coeffs using haar(level 7)
.for some of the alphabets and their displaced versions
relative error in the largest 60 coeff, in percentage
displacement(as a fraction of the image width
fig 4: relative error of wavelet coeffs for the letter A
and its displaced versions, for four wavelets.
relative error of the largest 60 coeff, in percentag
haar
db2
db4
coif1
displacement(as a fraction of the image width

Figure 0.1: