Adaptive lifting for shape-based image retrieval

P.J. Oonincx[a,*], P.M. de Zeeuw[b]

[a] *Royal Netherlands Naval College, P.O. Box 10000 KI 013, 1780 CA Den Helder, The Netherlands*
[b] *CWI, P.O. Box 94079, 1090 GB Amsterdam, The Netherlands*

Abstract

We propose to use adaptive wavelet lifting for image retrieval systems that are based on shape detection and multiresolution structures of objects in a database against a background of texture. To measure the performance of our approach, feature vectors are computed based on moment invariants of detail coefficients produced by the adaptive lifting scheme and retrieval rates are obtained by measuring distances between these vectors. Retrieval rates are compared with the rates obtained when using non-adaptive wavelet filtering as a preprocessing step. A synthetic database is created for this simulation.
© 2003 Pattern Recognition Society. Published by Elsevier Ltd. All rights reserved.

Keywords: Image retrieval; Lifting scheme; Neville filters; Moment invariants; Adaptive scheme; Feature vectors; Affine transformations

1. Introduction

Content-based image retrieval (CBIR) is a widely used term to indicate the process of retrieving desired images from a large collection on the basis of features. The extraction process should be automatic (i.e. no human interference) and the features used for retrieval can be either primitive (color, shape, texture) or semantic (involving identity and meaning). In this paper, we confine ourselves to grayscale images of objects against a background of texture. This class of images occurs for example in various databases created for the combat of crime and holding tyre tracks and shoe sole impressions [1]. In this paper we restrict ourselves to the following problem. Given an image of an object (a so-called query) we want to identify all images in a database which contain the same object irrespective of translation, rotation or re-sizing of the object, lighting conditions and the background texture.

One of the most classical approaches to the problem of recognition of similar images is by the use of moment

invariants [2]. This method is based on calculating centralized moments in both the x- and y-direction of the image up to a certain order. Hu [2] has shown that certain homogeneous polynomials of these moments can be used as statistical quantities that attain the same values for images that are of the same class, i.e., that can be obtained by transforming one single original image (affine transforms and scaling). However, this method uses the fact that such images consists of a crisp object against a neutral background. If the background contains "information" (noise, environment in a picture) the background should be the same for all images in one class and should also be obtained from one background using the same transformations. In general this will not be the case. The kind of databases we consider in this paper consists of classes of different objects pasted on different background textures.

To deal with the problem of different backgrounds one may use some filtering process as a preprocessing step, e.g. the wavelet transform was exploited in Ref. [3] by means of the so-called wavelet modulus maxima method. The aim of this paper is not to build a new image retrieval system, but to introduce adaptive filtering by means of a lifting scheme as a new kind of preprocessing. To compare this method with existing non-adaptive methods we present simulation

---

* Corresponding author. Tel.: +31-223-657134; fax: +31-223-657414.

E-mail address: p.j.oonincx@kim.nl (P.J. Oonincx).

results for both the algorithm consisting of wavelet filtering and Hu's moment invariants and the algorithm built up by adaptive lifting and Hu's moment invariants.

The paper is organized as follows. In Sections 2, 3 and 4 we discuss the lifting scheme its adaptive version and its redundant version. In Section 5 the method of moment invariants is recapitulated and a new type of normalization is introduced. Furthermore, the mathematical consequences for the computation of moments of functions represented by wavelet (detail) coefficients are investigated. Section 6 discusses various aspects of the final retrieval algorithm. Simulation results of the algorithm for a synthetic database are presented in Section 7. Finally, some conclusions are drawn in Section 8.

2. The lifting scheme

The lifting scheme as introduced by Sweldens in 1997 (see Ref. [4]), is a method for constructing wavelet transforms that are not necessarily based on dilates and translates of one function. In fact the construction does not rely on the Fourier transform which makes it also suitable for functions on irregular grids. The transform also allows a fully in-place calculation, which means that no auxiliary memory is needed for the computations.

The idea of lifting is based on splitting a given set of data into two subsets. In the one-dimensional case this can mean that starting with a signal $x \in l^2(\mathbb{Z})$ the even and odd samples are collected into two new signals, i.e., $x = x_e + x_o$, where $x_e(n) = x(2n)$ and $x_o(n) = x(2n + 1)$, for all $n \in \mathbb{Z}$. The next step of the lifting scheme is to predict the value of $x_e(n)$ given the sequence x_o. This prediction uses a prediction operator P acting on x_o. The predicted value $(Px_o)(n)$ is subtracted from $x_e(n)$ yielding a "detail" signal d. An update of the odd samples x_o is needed to avoid alias sing problems. This update is performed by adding Ud to the sequence x_o, with U the update operator. The lifting procedure can also be seen as a 2-band filter bank. This idea has been depicted in Fig. 1. The inverse lifting scheme can immediately be found by undoing the prediction and update operators. In practice, this comes down in Fig. 1 to simply changing each $+$ into a $-$ and vice versa.

Compared to the traditional wavelet transform the sequence $d = x_e - P(x_o)$ can be regarded as detail coefficients of the signal x. The updated sequence $s = x_0 + U(d)$ can be
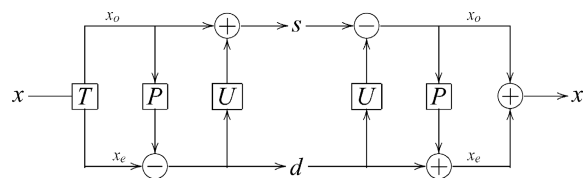
regarded as the approximation of x at a coarse scale. Using s again as input for the lifting scheme yields detail and approximation signals at lower resolution levels. We observe that every discrete wavelet transform can also be decomposed into a finite sequence of lifting steps [5].

Example 1. We take $P = I$ and $U = 1/2$. With these operators we get

$$s = x_e/2 + x_o/2,$$

$$d = x_e - x_o.$$

In this example the lifting transform corresponds to the classical Haar wavelet transform.

Example 2. For smoother approximation signals, we take for example

$$(Px)(n)$$
$$= [-x(n+1) + 9x(n) + 9x(n-1) - x(n-2)]/16,$$
$$(Ux)(n) = x(n+1)/4 + x(n)/4.$$

The lifting scheme can also be used for higher dimensional signals $x \in l^2(\mathbb{Z}^n)$. In our approach we use the so-called quincunx lattice to decompose image at several scaling levels. To explain this procedure we consider an image as a two-dimensional signal. An important example of the lifting scheme applied to such a signal is one that involves 2 channels ($M = 2$). We subdivide the lattice on which the signal has been defined into two sets on quincunx grids (see Fig. 2). This division is also called "checkerboard" or "red–black" division. The pixels on the red spots (○) are used to predict the samples on the black spots (●), while updating of the red spots is performed by using the detailed data on the black spots. An example of a lifting transform with second-order prediction and update filters is given by

$$(Px)(i, j)$$
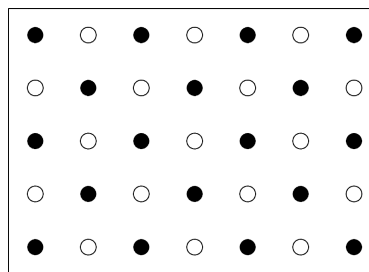$$= [x(i-1, j) + x(i, j-1) + x(i+1, j) + x(i, j+1)]/4,$$
$$i \bmod 2 \neq j \bmod 2,$$

Fig. 1. The lifting scheme: splitting, predicting, updating.

Fig. 2. A rectangular grid composed of two quincunx grids.

Table 1
Quincunx Neville filter coefficients

Order N	V_1	V_2	V_3	V_4	V_5	V_6	V_7
2	1/4	0	0	0	0	0	0
4	10/32	−1/32	0	0	0	0	0
6	$87/2^8$	$-27/2^9$	2^{-8}	$3/2^9$	0	0	0
8	$5825/2^{14}$	$-2235/2^{15}$	$625/2^{16}$	$425/2^{15}$	$-75/2^{16}$	$9/2^{16}$	$-5/2^{12}$

$(Ux)(i,j)$

$$= [x(i-1,j) + x(i,j-1) + x(i+1,j) + x(i,j+1)]/8,$$

$$i \bmod 2 = j \bmod 2.$$

The algorithm using the quincunx lattice is also known as the red–black wavelet transform by Uytterhoeven and Bultheel (see Ref. [6]). In general P can be written as

$$(Px)(i,j) = \sum_{(n,m)\in S_{\tilde{N}}} a_{\tilde{N}}(n,m)\, x(i+n, j+m),$$

$$i \bmod 2 \neq j \bmod 2 \qquad (1)$$

with $S_{\tilde{N}}$ a subset of $\{(n,m)\in \mathbb{Z}^2 \mid (n+m)\bmod 2 = 1\}$ and $a_{\tilde{N}}(s)$, $s\in S_{\tilde{N}}$, a set of coefficients in $\mathbb{R}$. In this case a general formula for U reads

$$(Ux)(i,j) = \sum_{(n,m)\in S_N} a_N(n,m)\, x(i+n, j+m)/2,$$

$$i \bmod 2 = j \bmod 2 \qquad (2)$$

with S_N depending on the degree of the used filter N. For several elements in S_N the coefficients $a_N(s)$ attain the same values. Therefore we take these elements together in subsets of S_N, i.e.,

$$V_1 = \{(\pm 1, 0), (0, \pm 1)\}, \qquad V_2 = \{(\pm 1, \pm 2), (\pm 2, \pm 1)\},$$

$$V_3 = \{(\pm 3, 0), (0, \pm 3)\}, \qquad V_4 = \{(\pm 2, \pm 3), (\pm 3, \pm 2)\},$$

$$V_5 = \{(\pm 1, \pm 4), (\pm 4, \pm 1)\}, \qquad V_6 = \{(\pm 5, 0), (0, \pm 5)\},$$

$$V_7 = \{(\pm 3, \pm 4), (\pm 4, \pm 3)\}. \qquad (3)$$

Table 1 indicates the values of all $a_N(s)$, $s\in V_k$, for different values of N (2–8) when using quincunx Neville filters (see Ref. [7]), which are the filters we use in our approach. We observe that $S_8 = V_1 + \cdots + V_7$ and so a 44 taps filter is used as prediction/update if the required filter order is 8.

We observe that the quincunx lattice yields a non-separable 2D-wavelet transform, which is also symmetric in both horizontal and vertical direction. Furthermore, we only need one prediction and one update operator for this 2D-lifting scheme, which reduces the number of computations.

3. Adaptive lifting

When using the lifting scheme or a classical wavelet approach, the prediction and update filters or wavelet/scaling functions are chosen in a fixed fashion. Generally they can be chosen in such way that a signal is approximated with very high accuracy using only a limited number of coefficients. Discontinuities mostly give rise to large detail coefficients which is undesirable for applications like compression. For our purpose large detail coefficients near edges in an images are desirable, since they can be identified with the shape of objects we want to detect. However, they are undesirable if such large coefficients are related to the background of the image. This situation occurs if a small filter is used on a background of texture that contains irregularities locally. In this case a large smoothing filter gives rise to small coefficients for the background.

These considerations lead to the idea of using different prediction filters for different parts of the signal. The signal itself should indicate (for example by means of local behavior information) whether a high or low order prediction filter should be used. Such an approach is commonly referred to as an adaptive approach. Many of these adaptive approaches have been described already thoroughly in the literature (e.g. Refs. [8–12]). In this paper, we follow the approach proposed by Baraniuk et al. [13], called the space-adaptive approach. This approach follows the scheme as shown in Fig. 3.

After splitting all pixels of a given image I into two complementary groups I_r and I_b (red/black), the pixels in I_r are used to predict the values in I_b. This is done by means of a prediction filter acting on I_r, i.e., $P(I_r)$. In the adaptive lifting case this prediction filter depends on local information of the image pixels I_r. Choices for P may vary from high
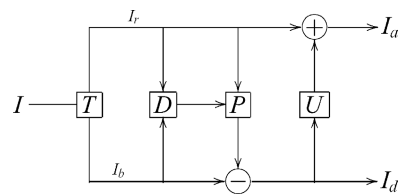

Fig. 3. Generating coefficients via adaptive lifting.

to low order filters, depending on the regularity of the image locally. For the update operator, we choose the update filter that corresponds to the prediction filter with lowest order from all possible to be chosen P. The above leads us to use a second-order Neville filter for the update step and an Nth order Neville filter for the prediction step, where $N \in \{2, 4, 6, 8\}$.

3.1. Relative local variance

We propose a measure on which the decision operator in the 2D adaptive lifting scheme can be based on, namely the *relative local variance* (RLV) of an image. This RLV of an image I is given by

$$\text{rlv}[I](i,j) = \sum_{k=i-T}^{i+T} \sum_{l=j-T}^{j+T} (I(k,l) - \overline{\mu_{i,j}})^2 / \text{var}(I) \tag{4}$$

with

$$\overline{\mu_{i,j}} = \sum_{k=i-T}^{i+T} \sum_{l=j-T}^{j+T} I(k,l) / (2T+1)^2. \tag{5}$$

For the window size we take $T = 5$, since with this choice all $I(k,l)$ that are used for the prediction of $I(i,j)$ contribute to the RLV for (i,j), even for the 8th order Neville filter.

For all pixels (i,j) to be predicted, we first compute $\text{rlv}[I](i,j)$. Then quantizing the values of the RLV yields a decisionmap indicating which prediction filter should be used at which positions. Values above the highest quantizing level induce a 2nd order Neville filter, while values below the lowest quantizing levels induce an 8th order Neville filter. For the quantizing levels we take multiples of the mean of the RLV. Test results have shown that $[\mu(\text{rlv})\ 1.5\,\mu(\text{rlv})\ 2\mu(\text{rlv})]$ are quantizing levels that yield a good performance in our application. In Fig. 4 we have depicted an image (left) and its decision map based on the RLV (right).

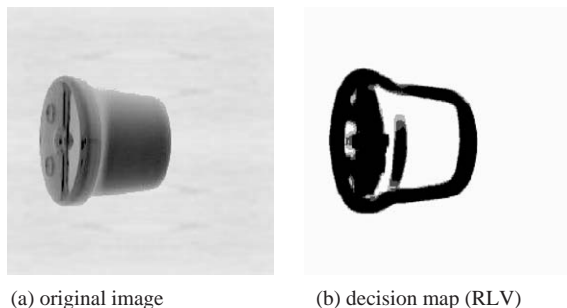(a) original image   (b) decision map (RLV)

Fig. 4. An object on a wooden background and its RLV (decision map): white=8th order; and black = 2nd order.

4. Redundant lifting

Although both traditional wavelet analysis and the lifting scheme yield detail and approximation coefficients that are localised in scale and space, they are not rotation/reflection invariant. Moreover, these coefficients are not even translation invariant.

For the classical wavelet transform a solution for translation invariance is given by the redundant wavelet transform [14], which is a non-decimated wavelet (at all scales) transform. This means that one gets rid of the decimation step. As a consequence, the data in all subbands have the same size as the size as the input data of the transform. Furthermore, at each scaling level, we have to use zero padding to the filters in order to keep the multiresolution analysis consistent. Not only more memory is used by the redundant transform, unfortunately also the computing complexity of the fast transform increases.

Whether the described redundant transform is also invariant under reflections and rotations depends strongly on the filters (wavelets) themselves. Symmetry of the filters is necessary to guarantee certain rotation and reflection invariances. This is a condition that is not satisfied by many well-known wavelet filters.

The redundant wavelet transform can also be translated into a redundant lifting scheme. In one dimension this works out as follows. Instead of partitioning a signal $x \in l^2(\mathbb{Z})$ into x_e and x_o we copy x to both x_e and x_o. The next step of the lifting scheme is to predict x_e by

$$x_e = x_e - P^{(j)} x_o. \tag{6}$$

The prediction filter $P^{(j)}$ is the same filter as used for the non-redundant case, however now it depends on the resolution level, since at each level zero padding is applied to P. This holds also for the update filters $U^{(j)}$. So, the update step reads

$$x_o = x_o + U^{(j)} x_e. \tag{7}$$

We observe that in our approach Neville filters on a quincunx lattice are used. Due to their symmetry properties (see Table 1), the redundant scheme does not only guarantee translation invariance, but also invariance under rotations over multiples of $\pi/2$ and reflections in the horizontal, vertical and diagonal axis is assured. Invariance under other rotations and reflections cannot be guaranteed by any prediction and update filter pair, since the quincunx lattice is not invariant under these transformations.

5. Moment invariants

At the outset of this section we give a brief introduction into the theory of statistical invariants for imaging purposes, based on centralized moments. Traditionally, these features have been widely used in pattern recognition applications to recognize the geometrical shapes of different objects [2].

Here, we will compute invariants with respect to the detail coefficients as produced by the wavelet lifting schemes of Sections 2–4. We use invariants based on moments of the coefficients up to third order. We show how to construct a feature vector from the obtained wavelet coefficients at several scales. It is followed by proposals for normalization of the moments to keep them in comparable range.

5.1. Introduction and recapitulation

We regard an image as a density distribution function $f \in S(\mathbb{R}^2)$, the Schwartz class. In order to obtain translation invariant statistics of such f we use central moments of f for our features. The $(p+q)$th order central moment $\mu_{pq}(f)$ of f is given by

$$\mu_{pq}(f) = \int_{\mathbb{R}} \int_{\mathbb{R}} (x - x_c)^p (y - y_c)^q$$
$$\times f(x, y) \, \mathrm{d}(x - x_c) \, \mathrm{d}(y - y_c) \tag{8}$$

with the center of mass

$$x_c = \frac{\int_{\mathbb{R}} \int_{\mathbb{R}} x f(x, y) \, \mathrm{d}x \, \mathrm{d}y}{\int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y) \, \mathrm{d}x \, \mathrm{d}y} \quad \text{and}$$

$$y_c = \frac{\int_{\mathbb{R}} \int_{\mathbb{R}} y f(x, y) \, \mathrm{d}x \, \mathrm{d}y}{\int_{\mathbb{R}} \int_{\mathbb{R}} f(x, y) \, \mathrm{d}x \, \mathrm{d}y}. \tag{9}$$

Computing the centers of mass x_c' and y_c' of $g(x, y) = f(x - a, y - b)$ yields

$$x_c' = x_c + a \quad \text{and} \quad y_c' = y_c + b.$$

Combining this with Eq. (8) shows $\mu_{pq}(g) = \mu_{pq}(f)$, i.e., the central moments are translation invariant.

We also require that the features should be invariant under orthogonal transformations (rotations). For deriving these features we follow Ref. [2]. Here polynomials with variables μ_{pq} were constructed in such way that there values are invariant under rotations and most of them also under reflections. The first two invariant polynomials that were derived are only built by second-order moments. These two invariants read

$$I_1 = \mu_{20} + \mu_{02} \quad \text{and} \quad I_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2.$$

These two polynomials are invariant under both rotations and reflections.

In the same way as for the second-order moments, Hu also derived three polynomials, all invariant under both rotations and reflections, that are only built up by third-order moments. These four invariants are

$$I_3 = p_1^2 + p_2^2, \quad I_4 = p_3^2 + p_4^2 \quad \text{and}$$

$$I_5 = 2(p_1 p_3^3 - 3p_1 p_3 p_4^2 + 3p_1 p_3^2 p_4 - p_1 p_4^3)$$

with

$$p_1 = \mu_{30} - 3\mu_{12}, \quad p_2 = \mu_{03} - 3\mu_{21}, \quad p_3 = \mu_{30} + \mu_{12},$$

$$p_4 = \mu_{03} + \mu_{21}.$$

A fourth invariant polynomial built by only third-order moments was given by

$$I_6 = 2(p_1 p_4^3 + 3p_2 p_3 p_4^2 - 3p_1 p_3^2 p_4 - p_2 p_3^3).$$

This polynomial is also invariant under rotations but changes sign after reflection. Since we want to include reflections as well in our set of invariant transformations we will use $|I_6|$ instead of I_6 in our approach. From now on, we will identify $|I_6|$ with I_6.

Finally, by combining second- and third-order moments a seventh polynomial has been constructed by Hu. This polynomial is again invariant under both rotations and reflections. It reads

$$I_7 = 2p_5 p_3^2 - 2p_5 p_4^2 + 8\mu_{11} p_3 p_4$$

with $p_5 = \mu_{20} - \mu_{02}$ and p_3, p_4 as above. Altogether seven invariant measures have been created in this way. Furthermore, we observe that all possible linear combinations of these invariants are also invariant under proper orthogonal transformations and translations. Therefore we can call these seven invariants also invariant generators.

At the end of this section we elaborate briefly on the computation of the moments. Using the values of the image pixels (or lifting coefficients) we construct an interpolating function based on piecewise constant approximation. The piecewise constant basisfunctions have their support on squares center around the pixels. Furthermore, the image is normalized in such way that the size of the shortest side is equal to 1. The size of the longest side is scaled accordingly. Observe that though the above interpolating function is not in the Schwartz class, it has compact support, is measurable and can be integrated. So we can now perform the integration in Eqs. (8) and (9) numerically and thereby compute the moments.

5.2. Normalizations

Here we increase the number of circumstances under which images should be considered invariant. Firstly a change in the dimensions of the object should still lead to the conclusion that the object remains the same. Invariance under this type of scaling, the similitude invariance, can be obtained by normalizing the moments μ_{pq}. Secondly, differences in luminosity also affect images of the same object. We will allow for a moderate variance with varying luminosity.

5.2.1. Similitude invariance

Uniform dilations (by a scalar $\alpha > 0$) of the whole image or objects in an image against a neutral background will result in new central moments given by [2]

$$\mu_{pq}' = \alpha^{p+q+2} \mu_{pq}. \tag{10}$$

It follows in particular that $\mu'_{00} = \alpha^2 \mu_{00}$, and also $\mu'_{20} + \mu'_{02} = \alpha^4(\mu_{20} + \mu_{02})$. Combining this result with Eq. (10) yields

$$\frac{\mu'_{pq}}{(\mu'_{00})^{(p+q+2)/2}} = \frac{\mu_{pq}}{\mu_{00}^{(p+q+2)/2}} \quad \text{and}$$

$$\frac{\mu'_{pq}}{(\mu'_{20} + \mu'_{02})^{(p+q+2)/4}} = \frac{\mu_{pq}}{(\mu_{20} + \mu_{02})^{(p+q+2)/4}},$$

respectively. As we recall that both μ_{00} and $\mu_{20} + \mu_{02}$ are invariants w.r.t. rotation and reflection this demonstrates how to normalize the moments to achieve invariance under dilation. The first choice leads to the following new set of invariant generators

$$I'_1 = I_1/\mu_{00}^2, \quad I'_2 = I_2/\mu_{00}^4, \quad I'_3 = I_3/\mu_{00}^5, \quad I'_4 = I_4/\mu_{00}^5,$$

$$I'_5 = I_5/\mu_{00}^{10}, \quad I'_6 = I_6/\mu_{00}^7, \quad I'_7 = I_7/\mu_{00}^{10}. \tag{11}$$

The second choice leads to a different but similar result. It may be more suitable (as a starting point) in case the density distribution is given by wavelet detail coefficients (Section 6.3).

5.2.2. The homogeneity condition

We consider the simple model that for a change in luminosity of an object the distribution function f is mapped onto a different f' by means of an affine transformation. For images this means that the grayvalue of each pixel is multiplied by a scalar $\lambda > 0$ and shifted along a distance b (the offset). A scalar multiplication of the distribution function f does not affect the center of mass (9). From Eq. (8) it follows directly that

$$\mu_{pq}(\lambda f) = \lambda \mu_{pq}(f) \quad \text{for all } \lambda \neq 0, \tag{12}$$

however the vector $I \equiv (I_1 \ I_2 \ I_3 \ I_4 \ I_5 \ I_6 \ I_7)^{\mathrm{T}}$, consisting of all seven presented invariants, changes unevenly in its entries, due to the fact that all I_k are polynomials of different degree. To overcome this inhomogeneous change in the feature vector I due to scaling of the object we introduce the homogeneity condition, meaning that we demand a homogeneous change in the elements of a feature vector in the case of scaling. We observe that neither the vector I nor the normalized vector I' derived from Eq. (11), satisfies the homogeneity condition, as multiplication of f by positive λ leads to the new feature vector

$$I = (\lambda I_1 \ \lambda^2 I_2 \ \lambda^2 I_3 \ \lambda^2 I_4 \ \lambda^4 I_5 \ \lambda^3 I_6 \ \lambda^4 I_7)^{\mathrm{T}}$$

and the new normalized vector

$$I' = (\lambda^{-1} I'_1 \ \lambda^{-2} I'_2 \ \lambda^{-3} I'_3 \ \lambda^{-3} I'_4 \ \lambda^{-6} I'_5 \ \lambda^{-4} I'_6 \ \lambda^{-6} I'_7)^{\mathrm{T}}.$$

The following operator

$$R_p(u) = \text{sign}(u)|u|^{1/p} \quad \text{with } p \in \mathbb{N} \text{ and } u \in \mathbb{R} \tag{13}$$

when applied to an invariant I_k produces again an invariant. It is a "legal" operation that invariants can be subjected to, i.e., neither their invariance properties nor their discriminative power are lost. The feature vectors

$$\tilde{I} = (I_1 \ R_2(I_2) \ R_2(I_3) \ R_2(I_4) \ R_4(I_5) \ R_3(I_6) \ R_4(I_7))^{\mathrm{T}} \tag{14}$$

and

$$\tilde{I}' = (I'_1 \ R_2(I'_2) \ R_3(I'_3) \ R_3(I'_4) \ R_6(I'_5) \ R_4(I'_6) \ R_6(I'_7))^{\mathrm{T}}, \tag{15}$$

now both satisfy the homogeneity condition as can be easily verified. By numerical experiments it is shown that also hereby the vector elements remain in comparable range. At this point we might consider to introduce a distance measure between feature vectors F and F' which vanishes altogether if $F = \lambda F'$, $\lambda \in \mathbb{R}$, however this is a matter for future investigation.

5.2.3. Offset in the affine transform

The offset b in the affine luminosity model is not a matter of concern when using our wavelet lifting approach. The detail coefficients produced by the lifting scheme are invariant to the offset b in the affine transform of our luminosity model. To show this we construct a new image $y \in l^2(\mathbb{Z}^2)$, given an image $x \in l^2(\mathbb{Z}^2)$, by

$$y(i,j) = x(i,j) + b,$$

for all $i, j \in \mathbb{Z}$ and for a given $b \in \mathbb{R}$. According to Eq. (1) the detail coefficients of y are given by

$$\begin{aligned}
d_y(i,j) &= y(i,j) - (Py)(i,j) \\
&= x(i,j) + b - \sum_{(n,m)\in S_{\tilde{N}}} a_{\tilde{N}}(n,m) \\
&\quad \times (x(i+n,j+m) + b) \\
&= x(i,j) + b - b - \sum_{(n,m)\in S_{\tilde{N}}} a_{\tilde{N}}(n,m)x(i+n,j+m) \\
&= d_x(i,j),
\end{aligned}$$

if and only if

$$\sum_{(n,m)\in S_{\tilde{N}}} a_{\tilde{N}}(n,m) = 1$$

for the given prediction filter. This necessary and sufficient condition is satisfied for the Neville filters we use in our approach, independent of the filter order.

5.2.4. The number of coefficients

To create feature vectors from a given image, using a wavelet lifting approach, we first compute the detail coefficients d_1 that appear in the lifting scheme for this image. The coarse scale approximation data s_1 is used again in the lifting scheme to obtain d_2 and s_2. After K recursive lifting steps we can use the detail coefficients $d_1, \ldots, d_K$ for computing feature vectors, as follows. After each step we select a relatively small set of coefficients, which we expect to correspond with the object of interest and not with the background. The set of these coefficients is given by W_j. All other detail coefficients $d_j(n,m)$, $(n,m) \notin W_j$ are put to zero. Of this newly constructed "detail image" d_j the moments μ^j_{pq} are computed. However, before calculating the invariants at this scale j we have to take into account that

the number of coefficients L_j may not be a constant as a function of j. Therefore we choose to use moments v_{pq} that are normalized to the number of coefficients as well, i.e., $v_{pq}^j = \mu_{pq}^j / L_j$. The criterions to select the coefficients may be based on thresholding, or, which is more sophisticated and reliable, on edge detection (see Section 6.1).

After K lifting steps we have at our disposal $\tilde{I}'^{(j)}$, $j = 1, \ldots, K$. Combining them into one $7K$-dimensional vectors yields the feature vector. We take $K = 4$, a common choice with other wavelet approaches [3,15].

6. The retrieval algorithm

6.1. Reducing the set of coefficients

As discussed in the previous chapter, our approach is based on the computation of statistics of detail coefficients. We also mentioned that, in order to get a sparse representation of the crisp object, we only use locally dominant coefficients, related to the object, for computing the moment invariants. One way to do this is to compute for each detail coefficient an approximate value for its gradient. This gradient based information is used to select the coefficients that are likely to be related to the object an not to the background. This is done by means of a threshold value.

So, for a given image I the detail coefficients we first compute the matrices consisting of detail coefficients after each lifting step, i.e., $d_1, \ldots, d_K$. At each level j the small set of coefficients W_j that are used to compute the moment invariants are given by the following rule

$$d_j(n,m) \in W_j \quad \Leftrightarrow \quad |\nabla d_j(n,m)|$$

$$\geqslant \sqrt{\sum_{n=1}^{N} \sum_{m=1}^{M} |\nabla d_j(n,m)|^2 / NM} \quad (16)$$

with N and M, respectively, the number of rows and columns in the matrix d_j. The gradients are approximated numerically. The threshold value in the right-hand side of Eq. (16) is known as the root mean-square (RMS).

6.2. Preprocessing

When using our algorithm we have to deal with the fact that in practice most affine transformations do not map an image on one (Cartesian) lattice onto an image defined on the same lattice. Generally, an interpolation filter is used to get a new image, which is defined on the original lattice. It can be shown that if the difference between the image f, that was not defined on the original lattice, and its interpolated version g defined on the lattice, is small, then also the difference in the moments μ_{pq} will be small. So, in a certain sense, the moments are continuous in the image functions.

Applying a high pass filter (prediction filter) on both the interpolated and the original image yields some other results.

Although the detail coefficients of both images will only slightly differ from each other, they will induce large differences in the central moments μ_{pq}. Simulation results have shown that they are highly influenced by this phenomenon.

As a remedy we propose to use a blurring filter on all images in the database at each resolution level. In our setup a (5×5) 2D cubic spline filter has been used, but other smoothing filters (Gaussian) gave a good performance as well.

6.3. Similitude invariance revisited

In Section 5.2.1 we have already seen that uniform dilations (by $\alpha > 0$) result in a multiplication of Hu's moments by $\mu'_{pq} = \alpha^{p+q+2} \mu_{pq}$. However, if the object in an image can just be represented by an orbit on a neutral background, e.g., a circle, then a uniform dilation yields $\mu'_{pq} = \alpha^{p+q+1} \mu_{pq}$. Assuming that the selected sparse set of detail coefficients also induce orbits instead of regions, this change in moments also holds for the moments computed from these coefficients. To increase stability in the lifting approach all moments were also divided by L_j, the number of selected coefficients at level j, yielding new moments v_{pq}^j. For these moments we have

$$v_{pq}^j = \alpha^{p+q} v_{pq}^j.$$

Following the computations in Section 5.2.1 invariance of the wavelet moments v_{pq}^j is achieved by dividing them by $(v_{20}^j + v_{02}^j)^{(p+q)/2}$, yielding at level j the new moments

$$\xi_{pq}^j = \frac{v_{pq}^j}{(v_{20}^j + v_{02}^j)^{(p+q)/2}} = \frac{\mu_{pq}^j L_j^{(p+q)/2 - 1}}{(\mu_{20}^j + \mu_{02}^j)^{(p+q)/2}}.$$

We observe that invariant I_1 is now always equal to 1 due to this normalization. Therefore we replace this invariant by $I_1 = \mu_{00}$.

6.4. Outline and recapitulation of the retrieval algorithm

Given the many aspects that have been discussed before, we give an outline of the retrieval algorithm. This algorithm is needed to measure the performance of the adaptive lifting approach versus non-adaptive filtering approaches. Firstly, we need to build a database with images of interest. In the database each image is accompanied by its feature vector which has to be derived. So-called query images are images which (usually) do not reside in the database and are compared to the elements of the database. Also from a query image a feature vector has to be derived to make comparison possible. Below follows a recap for the construction of a feature vector.

From image to feature vector.

(1) Application of a smoothing filter to the image (preprocessing, see Section 6.2).
(2) Computation of the RLV at each pixel (see Section 3).
(3) For each pixel a prediction stencil is chosen. The choice depends on the relative local variance (step 2) by means of a decision operator.

Fig. 5. Object library of eight images of size 128×128.

(4) Computation of the detail coefficients via lifting. It involves adaptive filtering (see Section 3).
(5) Selection of relevant detail coefficients.
(6) Computation of moments up to third order w.r.t. the coefficients resulting from step 4 (see Section 5.1).
(7) Composition of a feature vector based on moment invariants.
(8) Normalization of the feature vector by application of the homogeneity condition (see Section 5.2.2).

(9) Application of the smoothing filter of step 1 to the approximation coefficients.
(10) Repetition of steps 2–9 up to K levels.
(11) Composition of one feature vector from the K feature vectors resulting from step 8.

Once all feature vectors have been constructed distances between two images are computed by a weighted Euclidean distance, with one over the variance of each single feature, measured upon the whole database, as weighting coefficients. The distance between two images is identified with the distance between the corresponding feature vectors. For a given query image we find the image in the database which is at the smallest distance. This concludes the retrieval algorithm.

7. Results

7.1. A synthetic database

To give a "proof of principle" of our proposed adaptive lifting approach, we constructed a synthetic database of 64
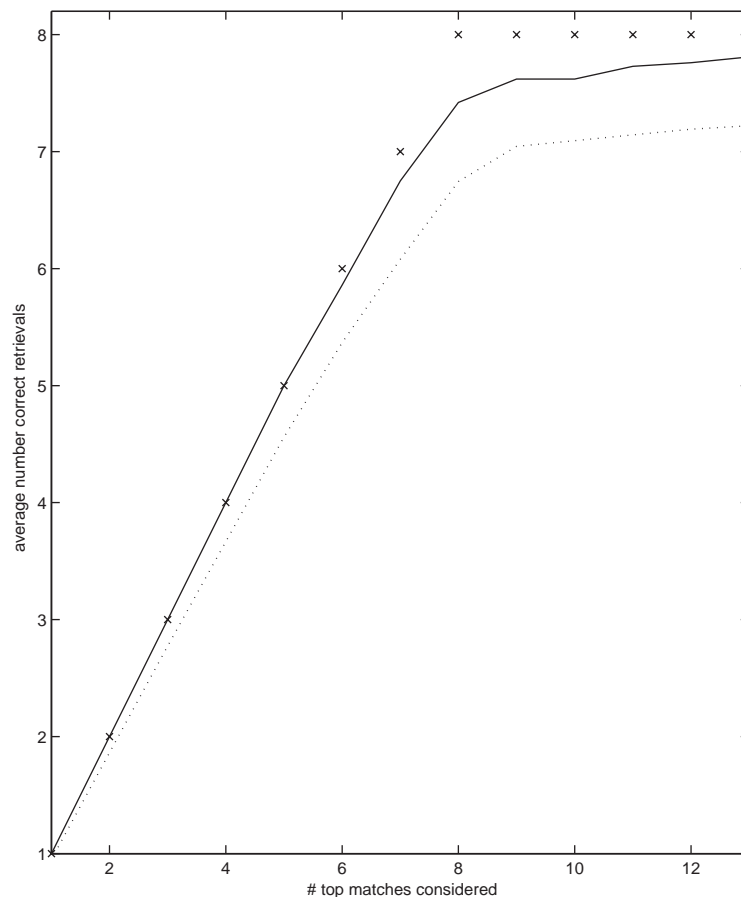

Fig. 6. Retrieval performance of adapted (solid) and non-adapted (dotted) approaches.

images. These images can be divided into eight classes, each one consisting of images of one of the objects depicted in Fig. 5, but subjected to translation over various distances, reflection over various angles, rotation and the similitude transform (zooming in). Moreover, the images are pasted on an arbitrarily chosen wooden background texture (256×256). For the simulation each image was used as a query to retrieve the other seven relevant ones.

7.2. The effectiveness of adaptivity

For several types of filters we performed a simulation with our synthetic databases consisting of eight classes of eight images (of the same object). All 64 images were used successively as input query image. So for each image 63 distances were calculated. In Fig. 6 the retrieval rates have been depicted for the ideal case (crosses), our adaptive approach (solid line) and the best possible non-adaptive approach using Neville filters (dotted line). These rates are given by the mean number of retrieved images in the class of the query image versus the number of considered top results (smallest distances). Ideally both number are equal to each other until the number of top results equals the number of remaining images in one class (7).

From Fig. 6 it is obvious that retrieval rates were improved by using adaptive lifting filters instead of classical non-adaptive wavelet filters. Moreover, the dotted line is related to the best possible retrieval rates using Neville filters of fixed order. It turned out that 8th order Neville filters gave slightly better results than the other ones we used. The values of the solid line increase the values of the dotted line by 5–10%, which can be seen as the quantitative improvement of using adaptive lifting techniques in this simulation.

Besides, we expect that the followed approach offers also the opportunity to deal with database of objects photographed against many other types of background textures. However more simulation with other databases are needed to confirm this conjecture.

8. Conclusions

In this paper, we described how adaptive lifting can be used for certain types of content based image retrieval. Test results show that the adaptive approach performs better than non-adaptive approaches both in a qualitative and a quantitative manner. The combination of adaptive lifting coefficients and moment invariants yields an improved retrieval system based on shape related information. Presumably the adaptive scheme can also be used in combination with other methods to extract other types of features in an image.

Furthermore, we have seen that the classical method of computing moment based invariants encounters some serious difficulties when applying them on images, that have been filtered by a high pass filter (detail coefficients). Both the computations of the moments themselves as well as certain invariants have to deal with these problems.

References

[1] W. Ashley, What shoe was that? The use of computerised image database to assist in identification, Forensic Sci. Int. 82 (1996) 7–20.

[2] M. Hu, Visual pattern recognition by moment invariants, IRE Trans. Inf. Theory IT-8 (1962) 179–187.

[3] M. Do, S. Ayer, M. Vetterli, Invariant image retrieval using wavelet maxima moment, Proceedings of the Visual '99, Amsterdam, The Netherlands, 1999, pp. 451–458.

[4] W. Sweldens, The lifting scheme: a construction of second generation wavelets, SIAM J. Math. Anal. 29 (2) (1997) 511–546.

[5] I. Daubechies, W. Sweldens, Factoring wavelet transforms into lifting steps, J. Fourier Anal. Appl. 4 (3) (1998) 267–345.

[6] G. Uytterhoeven, A. Bultheel, The red–black wavelet transform, TW Report 271, Department of Computer Science, Katholieke Universiteit Leuven, Leuven, Belgium, 1997.

[7] J. Kovacevic, W. Sweldens, Wavelet families of increasing order in arbitrary dimensions, IEEE Trans. Image Process. 9 (3) (2000) 480–496.

[8] E. Candes, D. Donoho, Curvelets, multiresolution representation, and scaling laws, Proc. SPIE 4119, Wavelet Applications in Signal and Image Processing VIII, San Diego, USA, 2000, pp. 1–12.

[9] T. Chan, H. Zhou, Adaptive ENO-wavelet transforms for discontinuous functions, Technical Report 21, Computational and Applied Mathematics UCLA, Los Angeles, USA, 1999.

[10] D. Donoho, Wedgelets: nearly minimax estimation of edges, Technical Report, Statistics Department, Stanford University, USA, 1997.

[11] S. Mallat, Z. Zhang, Matching pursuits with time frequency dictionaries, IEEE Trans. Signal Process. 41 (12) (1993) 3397–3415.

[12] W. Trappe, K. Liu, Adaptivity in the lifting scheme, Proceedings of the 33rd Conference on Information Sciences and Systems, Baltimore, USA, 1999, pp. 950–955.

[13] R. Baraniuk, R. Claypoole, G. Davis, W. Sweldens, Nonlinear wavelet transforms for image coding via lifting, IEEE Trans. Image Process., to appear.

[14] G. Nason, B. Silverman, The stationary wavelet transform and some statistical applications, in: Antoniadis, Oppenheim (Eds.), Wavelets and Statistics, Springer, New York, 1995, pp. 281–299.

[15] B. Manjunath, W. Ma, Texture features for browsing and retrieval of image data, IEEE Trans. Pattern Anal. Mach. Intell. 18 (8) (1996) 837–842.

About the Author—PATRICK OONINCX received his M.S. degree (with honors) in mathematics at Eindhoven University in 1995 and Ph.D. degree in mathematics at University of Amsterdam in 2000. His Ph.D. thesis was also published as a monograph "Mathematical Signal Analysis: Wavelets, Wigner distribution and a seismic problem". From 2000 to 2002 he worked as a post-doctoral researcher on multiresolution image processing at the National Research Institute for Mathematics and Computer Science (CWI) in Amsterdam. Currently he is working as Assistant Professor in Mathematics and Signal Processing at the Royal Netherlands Naval College in Den Helder, the Netherlands. His research interests are wavelet analysis, time-frequency signal representations and multiresolution imaging.

About the Author—PAUL M. DE ZEEUW studied mathematics and computer science at the University of Leiden. Since 1979 he works at the National Research Institute for Mathematics and Computer Science (CWI) in Amsterdam where he authored and co-authored many papers on multigrid methods. His current research interests include applications of multiresolution methods. From 1997 till 2002 he was the secretary of the Dutch-Flemish Numerical Analysis Society.