| Filter Bank | (*i, j*) | Mean of Power Spectrum of the Term Associated with $F_{ij}(z)$ in Eq. (10) | Estimated Flatness of the Power Spectrum of the Term Associated with $F_{ij}(z)$ in Eq. (10) |
|---|---|---|---|
| Haar Filter | (0, 0) | 1 | 0.52 |
| | (0, 1) | 0.090 | 0.69 |
| | (1, 0) | 0.090 | 0.69 |
| | (1, 1) | 0.63 | 0.57 |
| Johnston's 8-tap QMF Filter | (0, 0) | 1 | 0.67 |
| | (0, 1) | 0.0069 | 0.33 |
| | (1, 0) | 0.0069 | 0.33 |
| | (1, 1) | 0.70 | 0.55 |
| Mallat's QMF Filter in [8] | (0, 0) | 1 | 0.73 |
| | (0, 1) | 0.032 | 0.51 |
| | (1, 0) | 0.032 | 0.51 |
| | (1, 1) | 0.70 | 0.54 |
| 8-tap Optimal[1] PR-QMF Filter in Tab. 4.7 of [1] | (0, 0) | 1 | 0.52 |
| | (0, 1) | 0.0047 | 0.63 |
| | (1, 0) | 0.0047 | 0.63 |
| | (1, 1) | 0.33 | 0.59 |
| 8-tap Optimal[2] PR-QMF Filter in Tab. 4.9 of [1] | (0, 0) | 1 | 0.61 |
| | (0, 1) | 0.0022 | 0.51 |
| | (1, 0) | 0.0022 | 0.51 |
| | (1, 1) | 0.75 | 0.51 |
| Smith-Barnwell's 8-tap CQF Filter in [12] | (0, 0) | 1 | 0.61 |
| | (0, 1) | 0.0019 | 0.56 |
| | (1, 0) | 0.0019 | 0.56 |
| | (1, 1) | 0.75 | 0.51 |

**TABLE 4. Comparison of filter banks**

Note 1. The Optimality is based on energy compaction with zero mean high-pass filter.
Note 2. The Optimality is based on energy compaction with zero mean high-pass filter, and uncorrelated subband signals.

## Appendix B. Tables

| Image Name | Content | Size (WxH) | Speed Gain | (\|Δx\|, \|Δy\|) of peak |
|---|---|---|---|---|
| Airport | satellite image | 700x560 | 10.9 | (1, 0) |
| Baboon | close-up | 256x256 | 11.1 | (1, 0) |
| Earth | satellite image | 457x462 | 11.0 | (1, 1) |
| Hawaii | satellite image | 359x250 | 11.0 | (1, 0) |
| Lenna | head and shoulder | 256x240 | 11.1 | Fail |

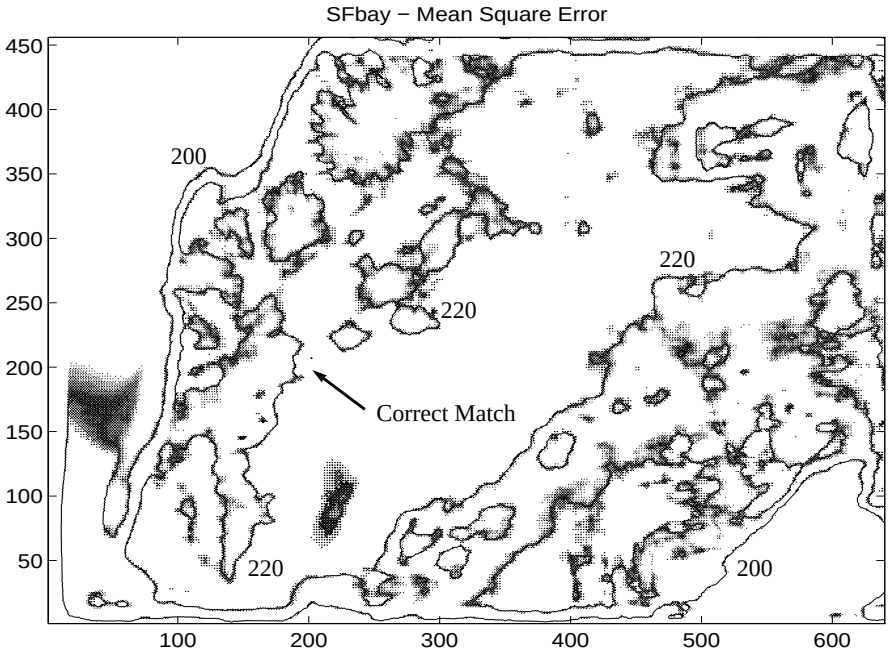
**TABLE 1. Test results of the speed and accuracy of the LL-LL subband correlation**

| Image Name | Size (WxH) | Template Size | (\|Δx\|, \|Δy\|) of peak |
|---|---|---|---|
| Earth | 457x462 | 16x16 | (1, 1) |
| | | 32x32 | (1, 0) |
| | | 64x64 | (0, 0) |
| Hawaii | 359x250 | 16x16 | (1, 0) |
| | | 32x32 | (2, 1) |
| | | 64x64 | (0, 0) |

**TABLE 2. Effect of template size on accuracy**

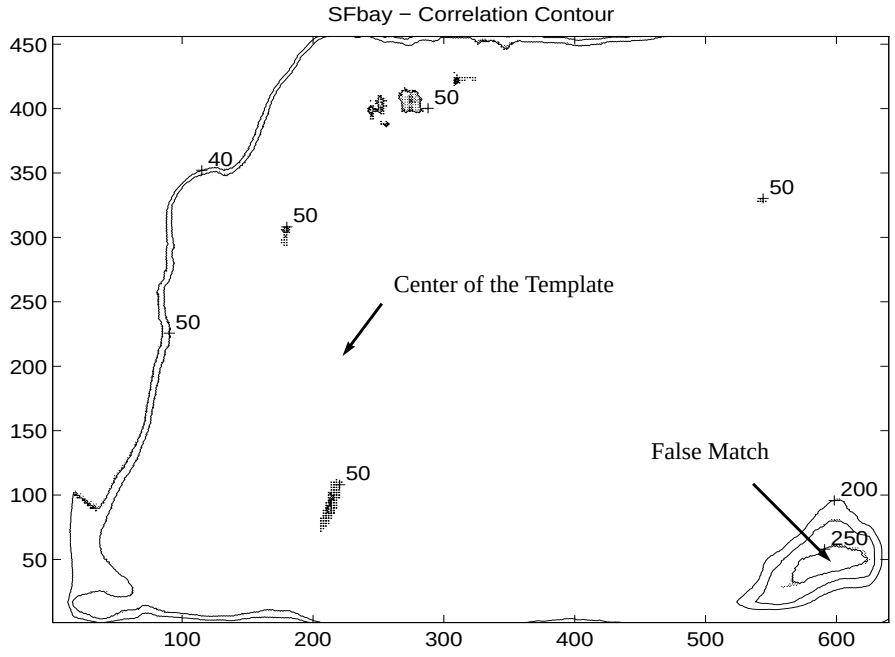
| Image Name | Size (WxH) | No. of Iterations | (\|Δx\|, \|Δy\|) of peak |
|---|---|---|---|
| Baboon | 256x256 | 1 | (0, 2) |
| | | 2 | (0, 4) |
| | | 3 | (0, 4) |
| Earth | 457x462 | 1 | (0, 0) |
| | | 2 | (0, 2) |
| | | 3 | Fail |
| Hawaii | 359x250 | 1 | (0, 0) |
| | | 2 | Fail |
| | | 3 | Fail |

**TABLE 3. Effect of the number of iterations in subband decomposition on accuracy**

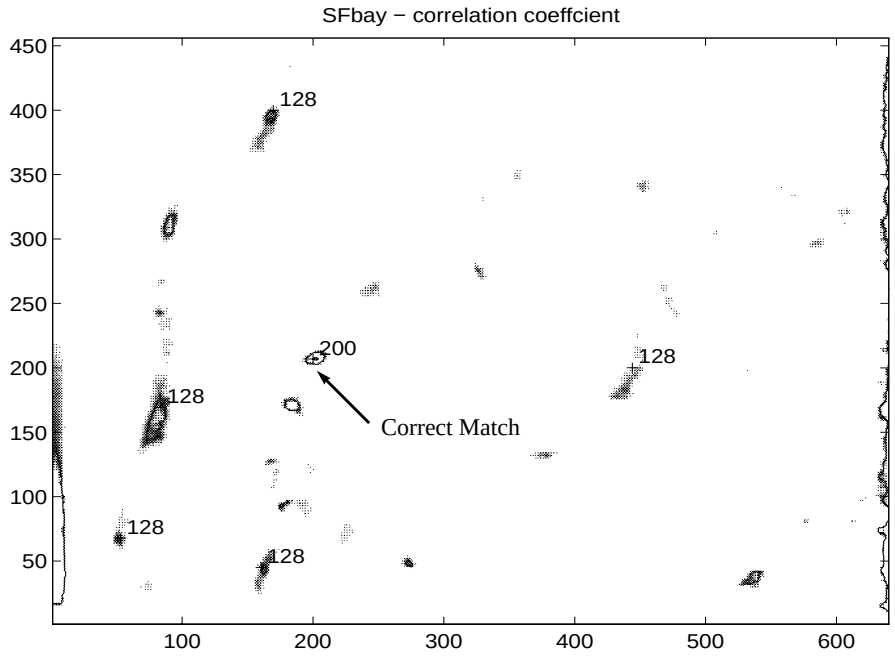



**(c) Mean square error of San Francisco Bay Image and the template in Fig. 10(b)**

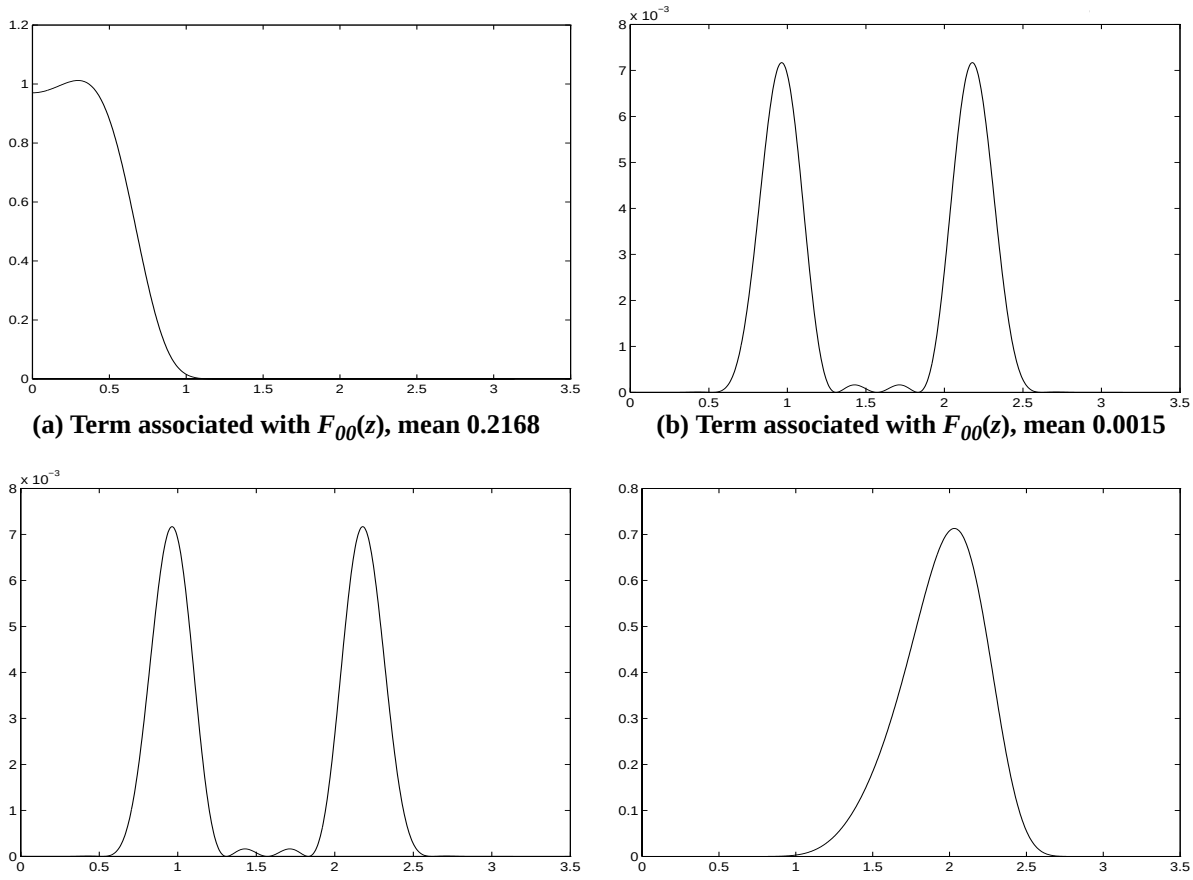
**FIGURE 11. Comparison of correlation, correlation coefficient, and mean square error methods**




**(a) Correlation of San Francisco Bay Image and the template in Fig. 10(b)**




**(b) Correlation coefficient of San Francisco Bay Image and the template in Fig. 10(b)**



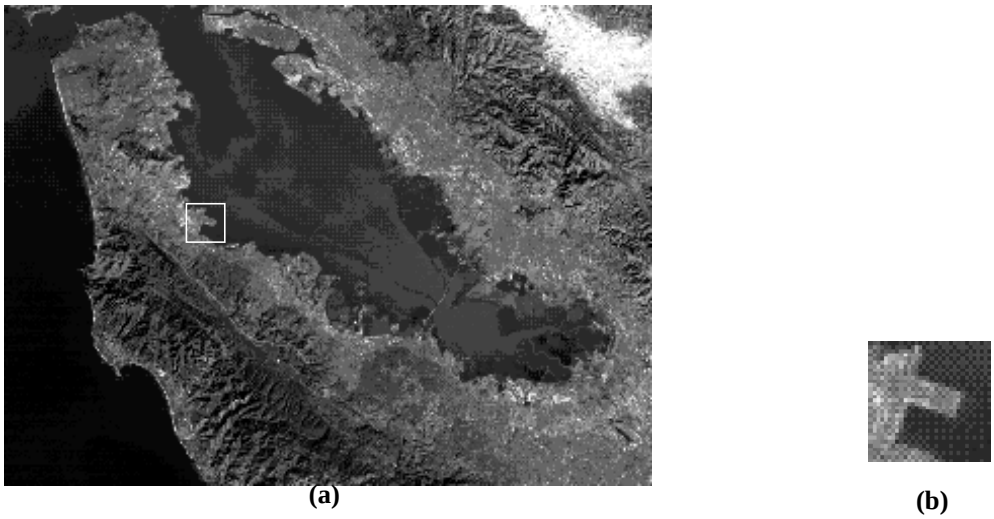

**(a) Term associated with $F_{00}(z)$, mean 0.2168**

**(b) Term associated with $F_{00}(z)$, mean 0.0015**

**(c) Term associated with $F_{10}(z)$, mean 0.0015**

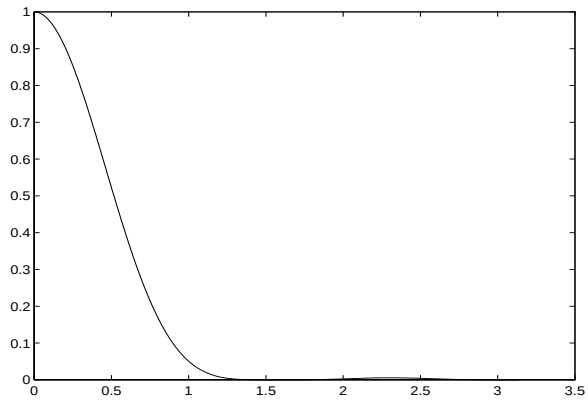
**(d) Term associated with $F_{11}(z)$, mean 0.1520**

**FIGURE 9. Power spectra of the four terms in Eq. (10), using Johnston's 8-tap QMF filters**

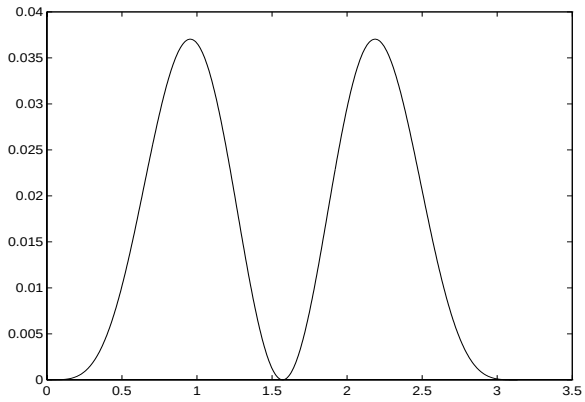


**(a)** **(b)**

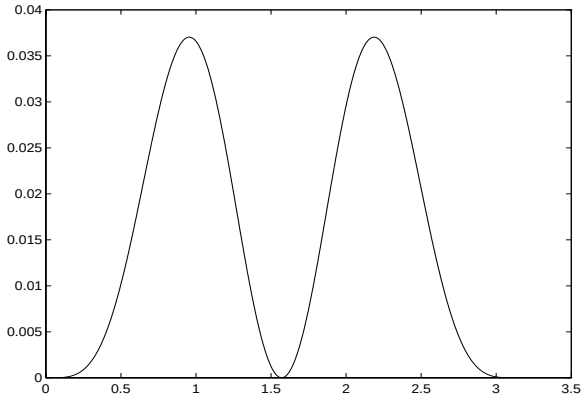
**FIGURE 10. (a) San Francisco Bay Image and (b) a 32x32 template cut from it**



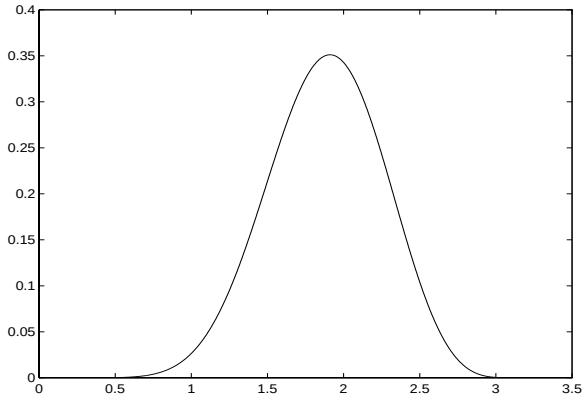

**(a) Term associated with $F_{00}(z)$, mean 0.1729**




**(b) Term associated with $F_{01}(z)$, mean 0.0156**



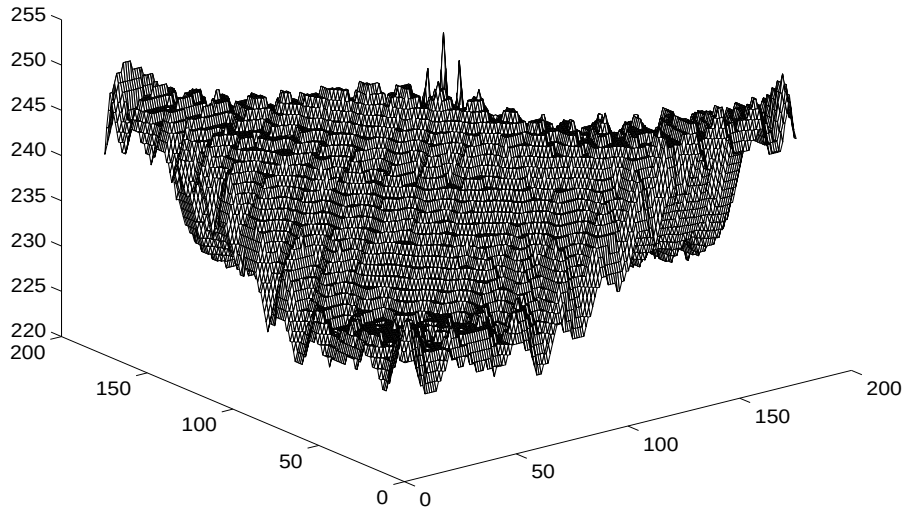

**(c) Term associated with $F_{10}(z)$, mean 0.0156**




**(d) Term associated with $F_{11}(z)$, mean 0.1094**

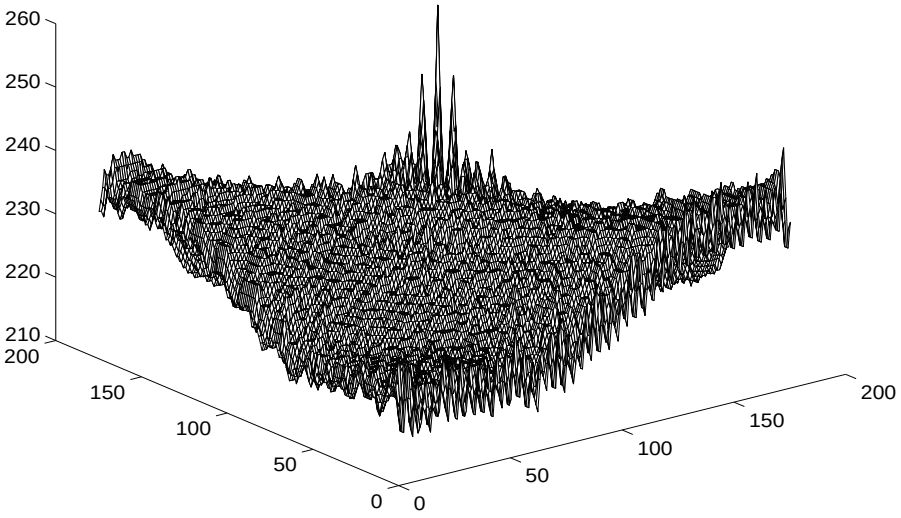
**FIGURE 8. Power spectra of the four terms in Eq. (10), using Haar filters**

Subband correlation of Star image and the template in Fig. 6(b), using only LL-LL term



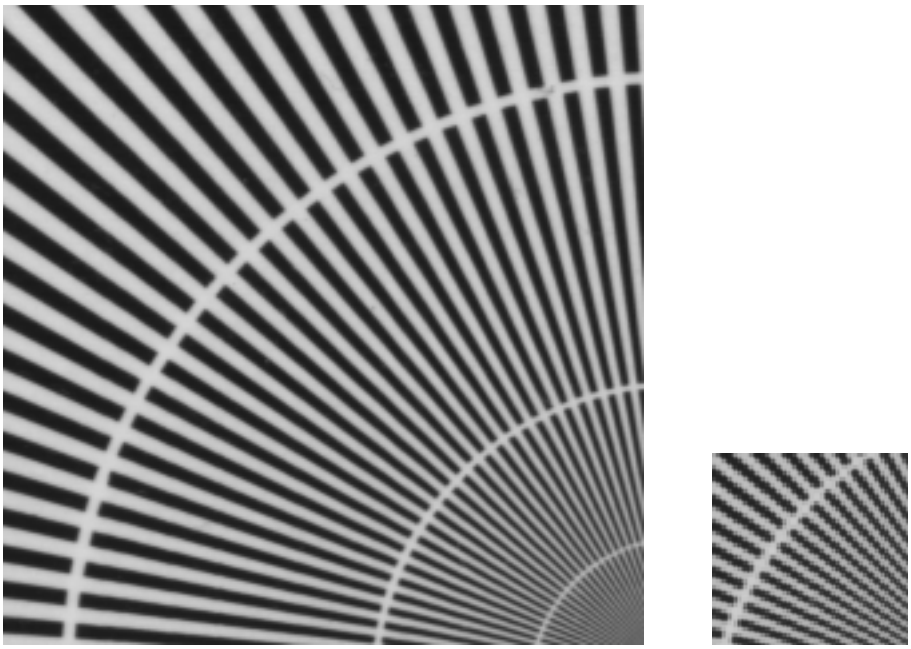

**(b) LL-LL term of subband correlation, two iterations of subband decomposition**

Subband correlation of Star image and the template in Fig. 6(b), using LL, LH, and HL terms in the second iteration of subband correlation




**(c) Adaptive subband correlation, two iterations of subband decomposition**

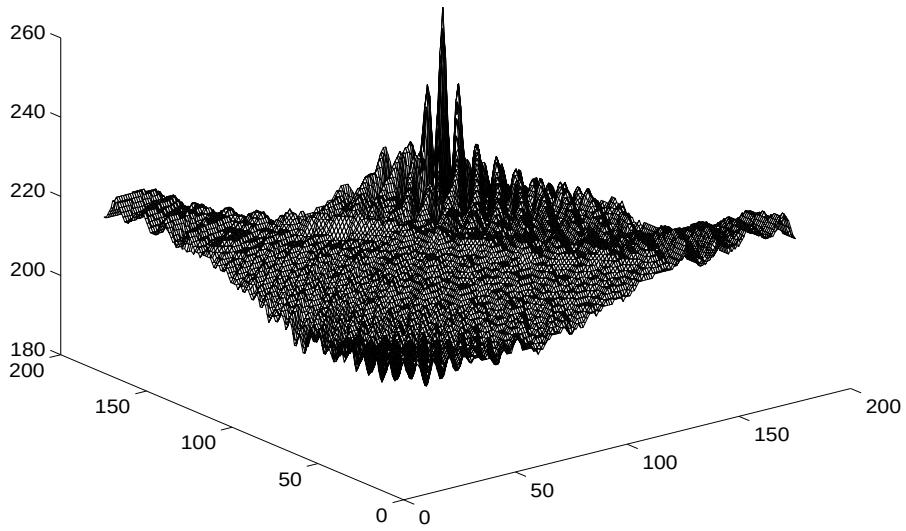
**FIGURE 7. Adaptive subband correlation in two-iteration case**



**(a)** **(b)**

**FIGURE 6. (a) Star image and (b) a 64x64 template cut from it**

Pixel domain correlation of Star image and the template in Fig. 6(b)




**(a) Pixel domain correlation, two iterations of subband decomposition**

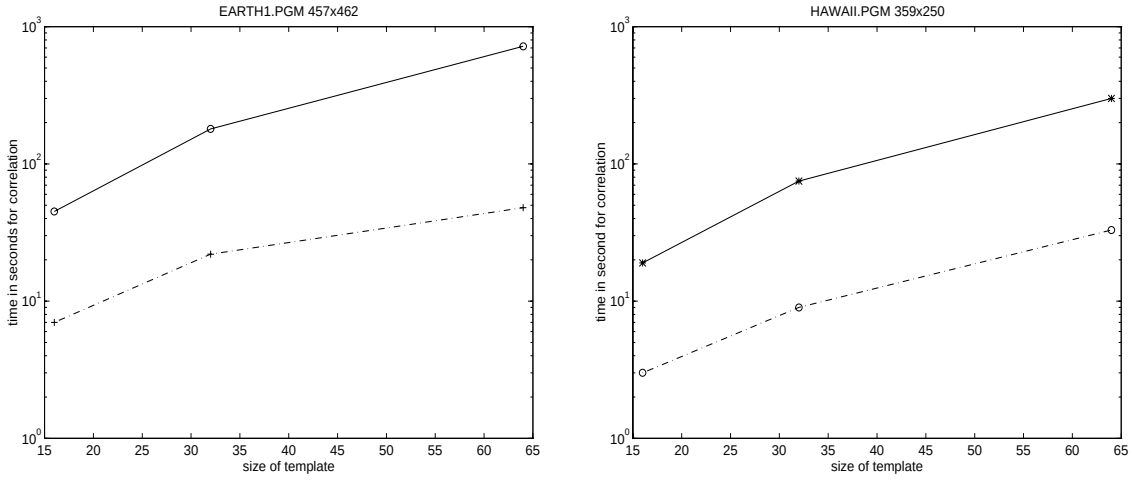



**FIGURE 4. Effect of template size on computation time**

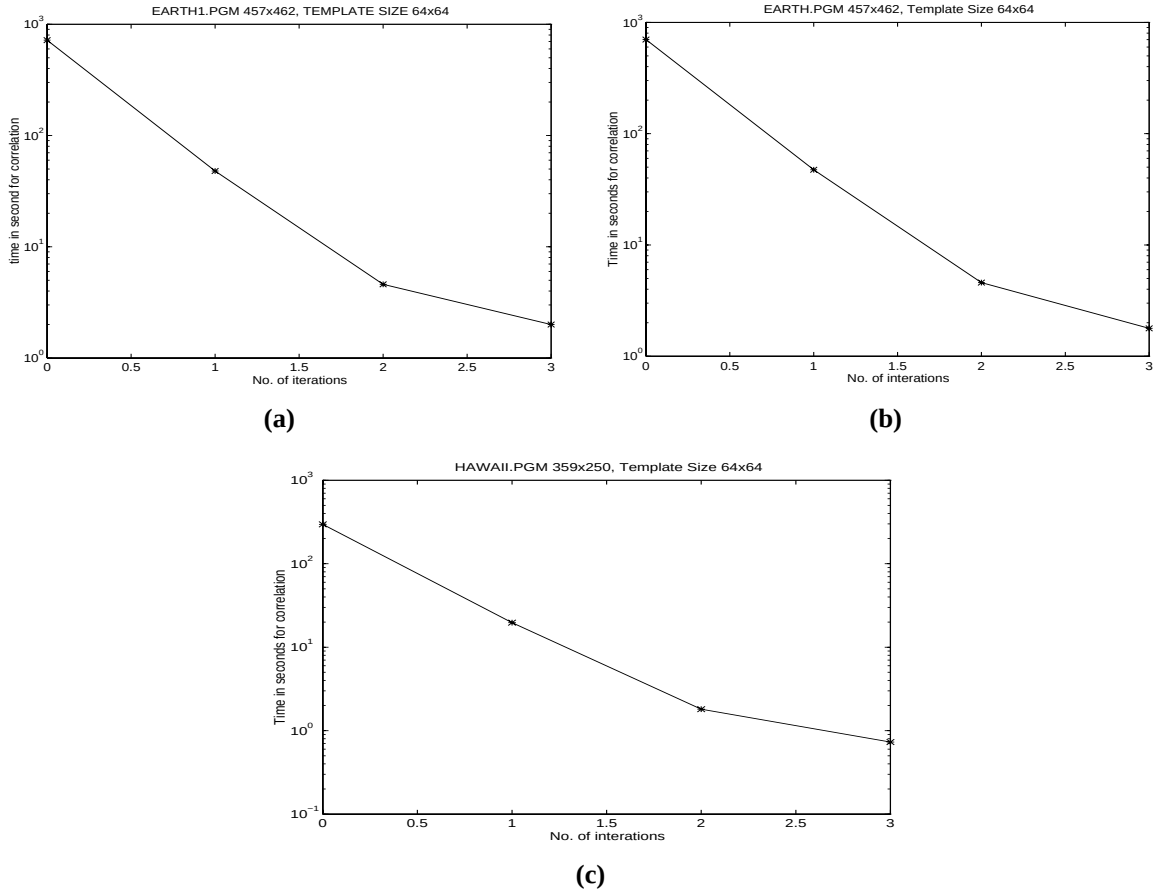
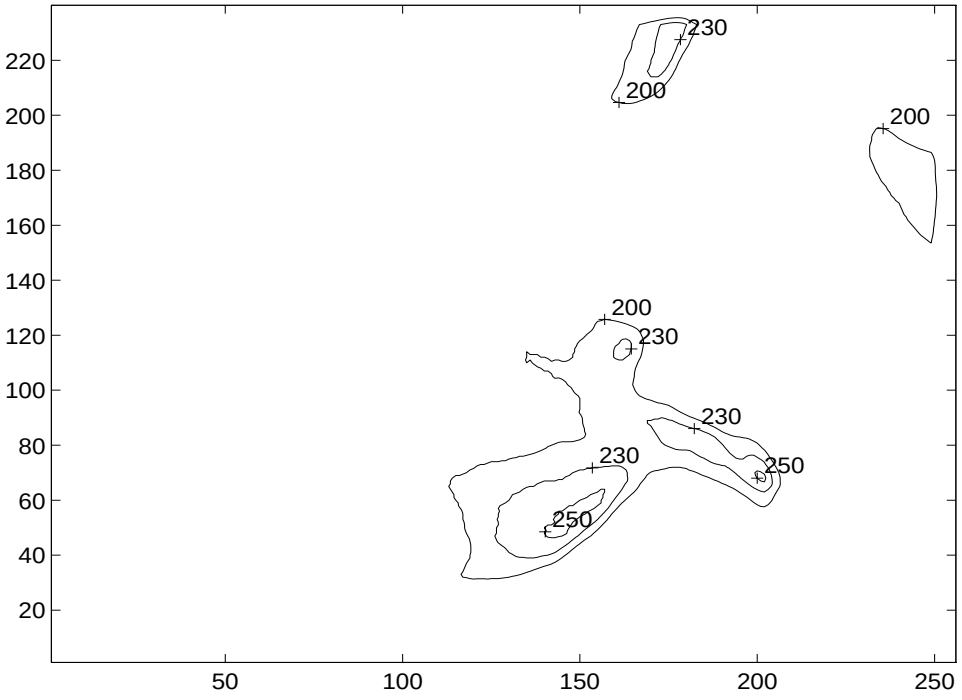
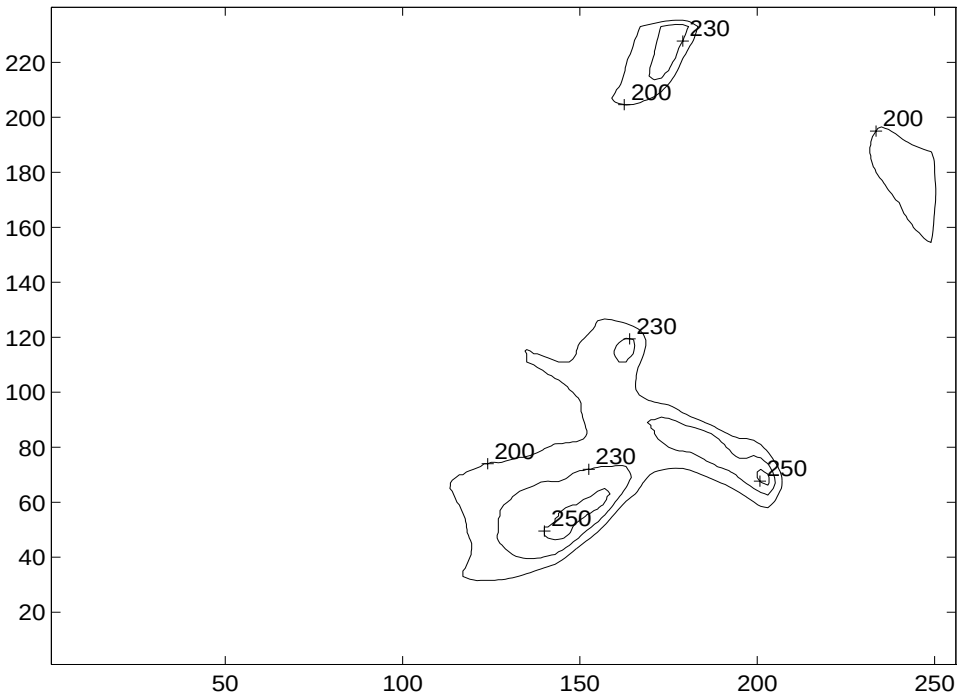



**FIGURE 5. Effect of number of iterations in subband decomposition on computation time**




**(a) Pixel domain correlation of Lenna image and the template in Fig. 2(b)**




**(b) LL-LL term of subband correlation of Lenna image and the template in Fig. 2(b)**

**FIGURE 3. Contour maps of correlation results**

## Appendix A. Figures

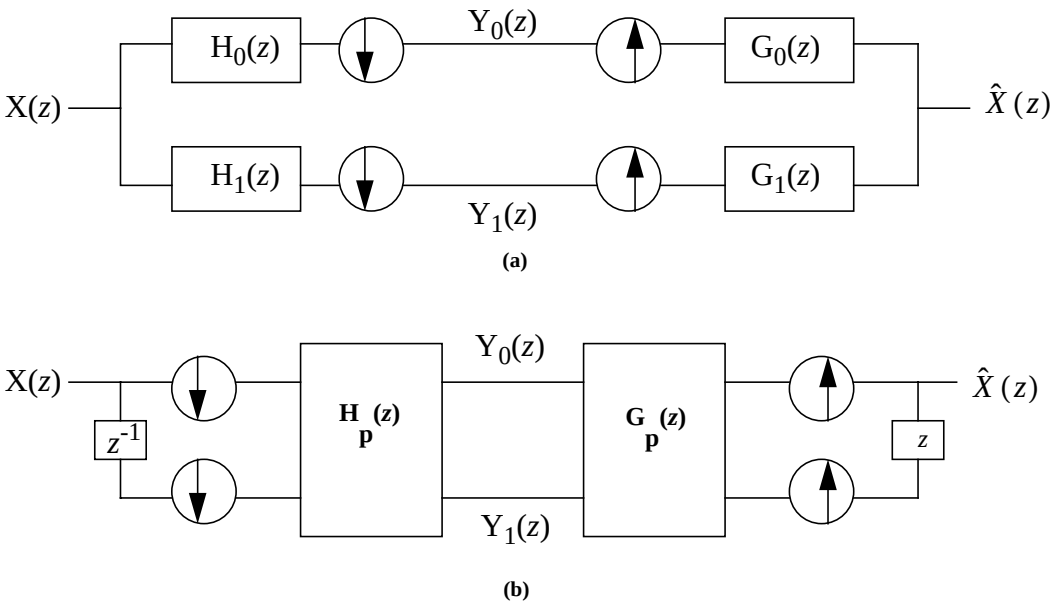



**FIGURE 1. (a) Two-channel filter bank and (b) its polyphase component equivalence**

**(a)** **(b)**

**FIGURE 2. (a) Lenna image and (b) a 16x16 template cut from it (magnified)**

[5] Faloutsos, C., Flickner, M., Niblack, W., Petkovic, D., Equitz, W., and Barber, R., "Efficient and Effective Query by Image Content," *IBM Research Journal*, No. 9453 (83074), August 3, 1993.

[6] Gonzalez, R.C. and Woods, R.E., *Digital Image Processing*, Addison-Wesley Publishing Co., 1992.

[7] Johnston, J.D., "A Filter Family for Use in Quadrature Mirror Filter Banks", *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 291-294, 1980.

[8] Mallat, S.G., "A Theory for Multiresolution Signal Decomposition: The Wavelet Representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 11, No. 7, pp. 674-693, July 1989.

[9] Papoulis, A., *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, Inc., 1984.

[10] Smith, J.R. and Chang, S.-F., "Single Color Extraction and Image Query," to appear in *Proceedings of the IEEE International Conference on Image Processing*, October 1995.

[11] Smith, J.R. and Chang, S.-F., "Texture Classification and Discrimination in Large Image Databases," *Proceedings of the IEEE International Conference on Image Processing*, Tx., November 1994.

[12] Smith, M.J.T. and Barnwell III, T.P., "Exact Reconstruction for Tree-Structured Subband Coders," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-34, No. 3, pp. 431-441, June 1986.

[13] Stone, H.S. and Li, C.S., "Image Matching by Means of Intensity and Texture Matching in the Fourier Domain," to appear in *Proceedings of SPIE Conference in Image and Video Databases*, San Jose, CA, January 1996.

[14] Swain, M. and Ballard, D., "Color Indexing," *International Journal of Computer Vision*, 7:1, pp. 11-32, 1991.

[15] Vaidyanathan, P.P., "Orthonormal and Biorthonormal Filter Banks as Convolvers, and Convolution Coding Gain," *IEEE Transactions on Signal Processing*, vol. 41, pp. 2210-2130, June 1993.

[16] Vetterli, M. and Kovacevic, J., *Wavelets and Subband Coding*, Prentice-Hall Inc., 1995.

[17] Woods, J.W. (Ed.), *Subband Image Coding*, Kluwer Academic Publishers, 1991.

[18] Woods, J.W. and O'Neil, S.D., "Subband Coding of Images," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. ASSP-34, No. 5, pp. 1278-1288, October 1986.

shown in Figure 11. The picture is 456x640 pixels and the template is 32x32 pixels. The template is cut from the original image and its center is at (206, 201) of the original image. We use the contour map of the correlation, correlation coefficient, and MSE to compare the effectiveness of the three methods.

The maximum value of the correlation and correlation coefficient are normalized to 1. The MSE is shown in the reverse intensity, with the minimal MSE mapped to maximum intensity. Then we look for the peak positions in these results to find the best match positions of the template. Because of the bright area in the lower right part of the image (mountain covered by snow), the correlation result gives the peak at a wrong position (Figure 11(a)). The MSE gives the minimal value at the correct match point, but there are too many candidate points with small MSEs in the whole image (Figure 11(c)). The correlation coefficient gives the correct match position and the peak value is much higher than those of the rest of the image (Figure 11(b)). In this example, the correlation coefficient method is the only one which gives the correct result with few false alarms.

## 6. Conclusion and Future Work

In this paper we focused on image correlation in subband domains and adaptive approaches of implementing it. First we proved the subband correlation theorem and showed that by using ideal half-band filters the cross terms of the inter-subband correlation are canceled. We further proposed adaptive methods of subband correlation to reduce computation while keeping accuracy. The effects of template size and number of iterations of subband decomposition were investigated. Then we analyzed examples of different filter banks and their effects on subband correlations. Finally we discussed other criteria of image matching which can be implemented based on correlation.

Correlation or correlation coefficients themselves may not necessarily represent visual content similarity. Invariances of size and rotation are not provided. Future work will address the issues of combining correlation with other image features derived directly from compressed domains and their efficient implementations.

## Acknowledgments

The authors would like to thank Dr. Harold Stone of NEC and Dr. Chung-Sheng Li of IBM for their helpful discussion in this project. This work was supported in part by NEC Research Institute, IBM under a 1995 Research Partnership (Faculty Development) Award, and the National Science Foundation under a CAREER award (IRI-9501266).

## References

[1] Akansu, A.N. and Haddad, R.A., *Multiresolution Signal Decomposition: Transforms, Subbands, and Wavelets*, Academic Press, Inc., 1992.

[2] Ballard, D.H. and Brown, C.M., *Computer Vision*, Prentice-Hall Inc., 1982.

[3] Campbell, J.D., "Edge Structure and the Representation of Pictures." *Ph.D. dissertation*, University of Missouri, Columbia, 1969.

[4] Chang, S.-F. and Smith, J.R., "Extracting Multi-Dimensional Signal Features for Content-Based Visual Query," *SPIE Symposium on Visual Communications and Signal Processing*, May 1995.

large area with very high intensity in the image, it is more likely to get high correlation values there, although the template may not be visually similar with that area. Other approaches based on image correlation are discussed in the following.

### 5.1 Mean Square Error

The mean square error (MSE) calculates the Euclidean distances between the template and different image regions. The smaller the error, the closer the template is to the image region. The computation of MSE can be decomposed into three components: the energy of the template, which is a constant; the energy of the image region overlapping with the template, which is slowly varying; and the correlation of the image and the template. With the correlation result prepared, the MSE can be easily computed. In ideal cases, MSE criterion is a reasonable method to find the best match. But in practice, the noise of pictures and different lighting or intensity scaling will greatly affect the result of this method. Furthermore, the image pattern we are going to find may not have the exact intensity level equivalent to the template. It is difficult for this method to find visually matched or related candidates. Features and patterns are more important than the absolute intensity in image matching. Hence, the MSE may not be a very good criterion without combining it with other matching criteria. In [13] average intensity is used as a second matching criterion besides MSE and the algorithm is computed in Fourier domain.

### 5.2 Correlation Coefficient

The correlation coefficient is defined as the correlation of two normalized random variables [9]. Consider one dimensional sequence $f$ with length $N$ and another one dimensional sequence $g$ with length $M$, their correlation coefficient is defined as (assume $M < N$)

$$\gamma(x) = \frac{\sum_n [f(n) - \bar{f}(n)]\,[g(n-x) - \bar{g}]}{\{\sum_n [f(n) - \bar{f}(n)]^2 \sum_n [g(n-x) - \bar{g}]^2\}^{1/2}} \tag{11}$$

where $\bar{g}$ is the average value of $g$(n) (computed only once), $\bar{f}$(n) is the average of $f$(n) in the interval coincident with the non-zero interval of $g$. It is clear that the numerator of Eq. (11) involves correlation of $f$ and $g$. Since the average intensity is subtracted and the residue normalized before doing the correlation, the absolute intensity does not affect the result of correlation here.

Computation of correlation coefficient includes similar components with MSE, i.e., energy of the template, energy of the image overlapped with the template window, and the correlation. It provides the advantage of lighting (average intensity) invariance at the cost of the higher computation complexity, when compared to correlation.

### 5.3 Comparisons

Given a remote sensing picture of San Francisco Bay and a template of the seashore, we use mean square error, correlation, and correlation coefficient to find the best match. The original image and template are shown in Figures 10(a) and 10(b), respectively. The results of the three methods are

found that Johnston's 8 tap QMF filters [7] is slightly better (within the difference of 2-3 pixels) than Haar filters.

The effects on speed is easier to predict. As shown in section 3.3 the computation can be divided into two parts: the correlation of subband images, whose complexity has little to do with filter banks; the filtering by $F_{ij}(z)$after the upsampling of the subband correlation, where the number of tap of filters determines the complexity. Hence it is clear that for large templates the effect of using various filter banks is small; while for small templates and the multiple iteration case the effect of using longer filters will be more prominent.

### 4.2 Comparisons of Various Filters

From Eq. (10) we know that it is desirable if most of the energy of the correlation signal falls into one of the four terms. Thus filter banks which are able to compact energy into one term or few terms in Eq. (10) are more favorable. As said in section 3.1 the filter bank will determine the spectrum of the filters $F_{ij}(z)$s as well as the spectra of the subband components. The spectrum of the original signal is in general not known and thus we first compare the power spectra of the four terms in Eq. (10). Figures 8 and 9 illustrate the cases for Haar filters and Johnston's 8-tap QMF filters.

From Figure 8 and Figure 9 we see that the energy of the cross-band terms is generally small. In the case of Haar filters the energy of the in-band terms are about ten times greater than the cross-band terms. In the QMF case the ratio is over one hundred which means the QMF filters have better energy compaction property. The high frequency in-band term is comparable with the low frequency in-band term. In practical situations since the image signals usually have much larger low frequency components, the first term will be dominant in most cases, as we described in section 3.

In Table 4 we give the numerical values of the mean of the power spectra in the four terms of Eq. (10) for various filter banks. Note that these mean values are normalized so that the mean of power spectrum of term associated with $F_{00}(z)$ is unity. Haar filter, Johnston's 8-tap QMF filter, Mallat's QMF filter presented in [8], 8-tap optimal perfect reconstruction QMF filter presented in Tables 4.7 and 4.9 of [1], and Smith-Barnwell's 8-tap CQF filter in [12] are compared. We can see from Table 4 that the CQF filter and Optimal filters have the tendency to dramatically suppress the energy of cross-band terms in Eq. (10) and thus seem to be more suitable for the subband correlation applications. The first Optimal filter can also suppress the energy in the high frequency in-band term in Eq. (10). That is because these filters are carefully designed for better energy compaction properties. The "flatness" of the power spectra in each of the terms in Eq. (10), which is estimated as the geometric mean divided by arithmetic mean, are also computed and compared for these filters. This flatness metric ranges from zero to one. We see that for most filters the power spectra have similar flatness (0.5-0.7) except in the case of Johnston's 8-tap QMF filters where cross-band terms are relatively narrow.

## 5. Other Criteria of Image Matching involving Correlation

The correlation of two images is a good way of describing the relation between their image contents. However, this method is affected by the absolute intensity of the images. When we have a

goes beyond two. Hence the reduction of computation grows more slowly as the number of iterations goes up.

As shown in Table 3, the peak positions deviate as the number of iterations goes up. In the Baboon case this deviation is small and the result is acceptable, while in the other cases the original peaks are lost if we use the LL-LL terms only after the iterations of subband decomposition. The difference between Baboon and the other two satellite images is that the latter have more high frequency components in the templates. For example, the 64x64 template of Hawaii case has the energy distribution in the four bands (LL, LH, HL, HH) as 53%, 21%, 15% and 11%. As the template is further decomposed and the LL band becomes smaller, the energy distribution tends to move toward high frequency bands and becomes more uniform, which is different from general large images.

These examples show that a reasonable number of iterations of subband decomposition may be two (or the size of the smallest LL image is larger than 16x16 pixels). If the template contains considerable high frequency energy, the LL-LL subband correlation alone will not give satisfying result.

### 3.5 Adaptive Method in Multiple Iteration Case

As shown in section 3.4, when we have more than one iteration of subband decomposition, it is not always reliable to use only the LL-LL subband correlation to approximate the original correlation. The adaptive method plays an important role here to get the correct result efficiently. An example is shown in Figure 7 with Haar filter bank used. We show the original correlation of image Star (Figure 6(a)) with a 64x64 template cut from it (Figure 6(b)) and the result using only the LL-LL subband correlations after two iterations of decomposition. The results are represented in Figures 7(a) and 7(b) in mesh graph with the height of the surface corresponding to the normalized correlation value. They differ significantly. Then we use the adaptive method in every step of iteration. In the first step the energy distributions are 86%, 6%, 6% an 1% for the image and 64%, 16%, 14% and 6% for the template. The low frequency components dominate so we just use the LL-LL subband correlation. As we go to the second iteration, we measure that the energy distributions change to 68%, 13%, 14%, 4% for the image and 45%, 19%, 17% and 19% for the template. Only the HH component of the image is low, so we take the in-band and cross-band correlations of LL, LH and HL components. The result of this adaptive method is shown in Figure 7(c). It can be seen that the result is much more closer to the original correlation result compared with the LL-LL only approach.

## 4. Variation of Filter Banks

### 4.1 Effects on Correlation Accuracy and Speed

As shown in section 2.3, if an ideal filter bank is used, only the in-band correlations are needed for the perfect reconstruction of the original correlation of the two images. For general filter banks this is not true. The properties of the filters will affect both the performance and complexity of the subband correlation. It is not easy to predict the effect of different filter banks on the accuracy of peak location. Based on several images we tested the accuracy of peak location of correlation. We

From Table 1 we see that given the same template size and filters, the speed gain of the method is almost fixed at 11. It is independent of the image size. The peak positions of the subband correlation are very close to the original ones except in the Lenna case where multiple correlation peaks exist. The LL-LL subband correlation gives good approximations in most of the examples with significant reduction of computation. In later sections we will show that other subbands than LL band only may need to be involved. For those cases, we will apply our energy-based adaptive matching method. We will also discuss the effects on the efficiency and accuracy of subband correlation by varying template size, number of iterations of subband decomposition, and filter type.

### 3.3 Speed and Accuracy vs. Size of Template

We still take the LL-LL subband correlation as the approximation of the original correlation but vary the template size. Haar filters are again used for subband decomposition. Templates are cut from the original images. Two satellite images are tested and the results are given in Figure 4 and Table 2.

As stated at the beginning of section 3, the computation of pixel domain correlation is in the order of $(NM)^2$, with $N$, $M$ the size of the image and template, respectively. As $M$ doubles, the computation goes up by four times. This is exactly the case in Figure 4(a). However, the computation of subband correlation does not go up as fast as the pixel correlation. Let n denote the number of taps of the filters $F_{ij}(z)$s in Eq. (10). For subband correlation, the computation will be in the order of $(MN/4)^2 + 2nN^2$ if we include computation associated with zero coefficients.* In our examples we use Haar filter bank where $n$ equals three. So the computation is in the order of $(M^2/16 + 6)N^2$. As M goes up, because the time spent for the filtering remains the same, the total computation time goes up more slowly than that of the pixel domain correlation. The above equation fits better with the example in Figure 4(a) than the example in Figure 4(b).

In the above examples, the size of the template does not significantly affect the accuracy of the peak location in subband correlation, as shown in Table 2.

### 3.4 Speed and Accuracy vs. Number of Iterations of Subband Decomposition

We use two images as examples to show the effects of the number of iterations of subband decomposition on the performance of subband correlation. As the LL band image is further decomposed, we still use only the LL-LL subband correlations in the multiple iterations to get an estimation of the pixel domain correlation. The sizes of the templates are 64x64 pixels. Haar filters are used for subband decomposition. Results are shown in Figure 5 and Table 3.

It can be seen from Figure 5 that after the second iteration of decomposition, the computation reduces greatly. However, the third iteration has much less effect on computation. Remember we estimate that the computation is in the order of $(MN/4)^2 + 2nN^2$ if computation associated with zero coefficients is included. As the number of iterations goes up by one, the first term of the equation reduces by 16 times, while the second term increases a bit because of the filtering of $F_{ij}(z)$ after upsampling. Thus the effect of the second term dominates as the number of iterations

* If we carefully omit the computation associated with zero coefficients, the complexity should be $(MN/4)^2+0.75nN^2$

nals themselves as well as the filters used. Here we consider the signals first. Effects of different filter banks will be discussed later.

Intuitively, the stronger a signal is, the more it will contribute to the correlation, because correlation involves the summation and multiplication of two signals. Therefore, it is reasonable to pick up the terms in Eq. (10) which involve the correlation of the subband components with greater energy and discard those trivial ones. Though it may not best represent the visual features of the image, such as texture and shape, this criterion is efficient and easy to implement, especially when the visual features are hard to extract or it is time-consuming to do so.

First we define the source images in the database as "*images*" and the query examples presented by the users as "*templates*". Then the energy based adaptive method for subband correlation can be summarized as follows. Given subband decomposed image and template, first compute the energy in each subband; then by applying a pre-determined threshold, discard the terms involving subband components with energy less than the percentage threshold; compute the necessary terms including upsampling and filtering and then sum them up to approximate the original correlation; finally detect the peak area of the correlation for possible matches.

When comparing the subband correlation results with original ones, we consider their speed and accuracy. Computer run time spent in the system are given for comparison of different approaches. All algorithms are written in C language and executed on an SGI Onyx workstation. The difference in pixels between the coordinate of the peak point in subband correlation and that in original correlation is given in absolute value ($|\Delta x|$, $|\Delta y|$) to measure the accuracy. If either $|\Delta x|$ or $|\Delta y|$ is greater than five pixels, we assume that the peak position is wrong in the result of subband correlation and the method fails in accuracy.

### 3.2 LL-LL Subband Correlation

It is known that for most natural images their energy concentrates in the low frequency bands. The high frequency components which correspond to the difference of the pixels usually have relatively small values and energy. Thus for many pictures the term involving LL-LL subband correlation will give a good approximation of the original correlation. Figure 3(a) shows the original correlation of the Lenna image (Figure 2(a)) with a 16x16 template (Figure 2(b)) cut from it. Figure 3(b) shows the corresponding subband correlation using only the LL-LL term. Both results are represented in contour maps. The height of each curve corresponds to the normalized correlation value. We can see that the peak positions and curves of the two contour maps matches well with each other. Hence the LL-LL subband correlation is a good approximation of the original one.

We tested five images including general images and satellite pictures for the speed and accuracy of the LL-LL subband correlation. In all cases we use a 16x16 template cut from the original image and Haar Filters for subband decomposition. The result is shown in Table 1. The size of the image is given in the form of width by height (pixels). All examples are gray-level images with 8 bits per pixel. "Speed Gain" is defined as the ratio of the computation time of pixel domain correlation over the computation time of subband correlation. "Fail" in the last column indicates that either $|\Delta x|$ or $|\Delta y|$ is greater than 5 pixels.

### 2.3 Ideal Filter Case

Eq. (10) has four terms including the weighted in-band and cross-band correlations of the subband components. However, if an ideal half-band lowpass filter and an ideal half-band highpass filter are used, only the in-band terms in Eq. (10) are needed. Consider the spectrum of the upsampled subband signals in the interval $(0, \pi)$. For ideal half-band brick-wall filters, $|Y_0(e^{j2\omega})|$ and $|Z_0(e^{j2\omega})|$ will be zero in the interval $(\pi/4, 3\pi/4)$; $|Y_1(e^{j2\omega})|$ and $|Z_1(e^{j2\omega})|$ will be zero in the intervals $(0, \pi/4)$ and $(3\pi/4, \pi)$ (Note that the period of Fourier spectrum of an upsampled signal is one half of that of the original signal). It is clear that the spectrum of the cross-band correlations in Eq. (10) will be zero in this case. Therefore only the in-band terms are needed for the computation of correlation if we use ideal filters.

### 2.4 Two Dimensional Signals and Wavelet Decomposition

Using separable two-dimensional filter banks, the subband correlation theorem can be easily extended to two-dimensional signals like images. In subband decomposition, the image is filtered in each dimension separately [18] and decomposed into four frequency bands, namely LL, LH, HL, and HH bands, where L stands for lowpass filter, and H for highpass filter. The first letter of L or H refers to row filtering, and the second one to column filtering. The correlation of the two images is then decomposed into the weighted sum of four in-band and twelve cross-band correlations. The filtering of the subband correlation by the weighting filter $F_{ij}(z)$ is also done in a separable manner, row filtering followed by column filtering, or vice versa.

Each subband component of a signal can be decomposed iteratively using the same filter bank and thus lead to wavelet or wavelet packet decomposition of the signal [8]. The subband correlation theorem can also be applied iteratively in these cases.

## 3. Adaptive Subband Correlation

Subband and wavelet decompositions preserve both partial spatial and frequency information. Correlation in the subband domain provides the possibilities of combining these spatial-frequency features for more efficient image matching or search. However, the straightforward implementation of Eq. (10) does not save computation. For example, suppose the image size is $N$x$N$ and the template size is $M$x$M$. In pixel domain correlation, the computation complexity is in the order of $(NM)^2$, while the computation for each of the subband correlation is $(NM/4)^2$. Adding up the sixteen correlations will result in the same computational complexity as the pixel domain correlation, plus the overhead of the subband filtering by $F_{ij}(z)$s. Thus we need even more computation. To solve this problem, we will show that it is not always necessary to compute the sixteen subband correlations to reconstruct the original correlation. By adaptively choosing the important terms in (10) we may reduce the computation of subband correlation significantly and keep the result accurate.

### 3.1 Energy Based Adaptation

From Eq. (10) we see that the correlation of one-dimensional signals equals the weighted sum of subband correlations. The contribution of each term to the original correlation depends on the sig-

From Eq. (7) we can easily see that if $G_p(z)H_p(z) = I$ then we will have perfect reconstruction of the original signal after the subband analysis and synthesis filters, i.e. $\hat{X}(z) = X(z)$. For orthogonal systems where $G_p(z) = \tilde{H}_p(z)$, the condition becomes $\tilde{H}_p(z)H_p(z) = I$.

We consider the correlation of two one-dimensional signals $x(n)$ and $w(n)$. The correlation of the signals can be represented in $z$-transform domain as $\tilde{W}(z)X(z)$. Suppose the subband decompositions of $x(n)$ and $w(n)$ are $\{y_0(n), y_1(n)\}$ and $\{z_0(n), z_1(n)\}$, respectively. Using above conditions with reference to Figure 1(a), we have:

$$\hat{X}(z) = \left[G_0(z)\ \ G_1(z)\right]\begin{bmatrix} Y_0(z^2) \\ Y_1(z^2)\end{bmatrix} = \left[\tilde{H}_0(z)\ \ \tilde{H}_1(z)\right]\begin{bmatrix} Y_0(z^2) \\ Y_1(z^2)\end{bmatrix} \tag{8}$$

and similarly, $$\hat{W}(z) = \left[\tilde{H}_0(z)\ \ \tilde{H}_1(z)\right]\begin{bmatrix} Z_0(z^2) \\ Z_1(z^2)\end{bmatrix} \tag{9}$$

so,

$$\tilde{W}(z)X(z) = \left[\tilde{Z}_0(z^2)\ \tilde{Z}_1(z^2)\right]\begin{bmatrix} H_0(z) \\ H_1(z)\end{bmatrix}\left[\tilde{H}_0(z)\ \ \tilde{H}_1(z)\right]\begin{bmatrix} Y_0(z^2) \\ Y_1(z^2)\end{bmatrix}$$

$$= \left[\tilde{Z}_0(z^2)\ \tilde{Z}_1(z^2)\right]\begin{bmatrix} H_0(z)\tilde{H}_0(z) & H_0(z)\tilde{H}_1(z) \\ H_1(z)\tilde{H}_0(z) & H_1(z)\tilde{H}_1(z)\end{bmatrix}\begin{bmatrix} Y_0(z^2) \\ Y_1(z^2)\end{bmatrix} = \sum_{i,j=0}^{1} F_{ij}(z)\tilde{Z}_i(z^2)Y_j(z^2) \tag{10}$$

$$\text{where} \quad F_{ij}(z) = H_i(z)\tilde{H}_j(z) \qquad \text{i, j = 0, 1} \tag{11}$$

Note that we assume reconstruction signals, $\hat{X}(z)$ and $\hat{W}(z)$, are close enough to the original signals, $X(z)$ and $W(z)$. This is true for filter banks with perfect reconstruction property. Eq. (10) shows that the correlation of two signals equals the weighted sum of their correlations in subbands. The original correlation is decomposed into four terms of subband correlation weighted by filters $F_{ij}(z)$ defined in Eq. (11). Hence the subband correlation theorem is proved. We define the terms associated with $F_{00}(z)$ and $F_{11}(z)$ as *in-band* terms and the other two as *cross-band* terms. As shown in Eq. (11), the weighting filter $F_{ij}(z)$ corresponds to the $z$-transform of the correlation of the impulse responses of analysis filters $H_i(z)$ and $H_j(z)$. Note that the $z^2$ in Eq. (10) corresponds to upsampling by two in time domain.

Related work on subband convolver can be found in [15], in which intra-subband convolutions (or correlations) are summed up to get a subsampled version of the original convolution (or correlation). Here we explore the representation of the complete correlation in terms of both intra-subband and cross-subband correlation, i. e., both in-band and cross-band terms in Eq. (10).

ing overhead. Furthermore, because the subband domain keeps partial spatial information of the images, we can find the locations of the image correlation peaks directly without inverse transform. It is also convenient to combine other features (like texture and multiscale edges) in the subband compressed domain with this method.

The size of template, the number of iterations of subband decomposition, and the type of filter bank will all affect the accuracy and speed of the proposed method. These effects are discussed with theoretical estimations and numerical results. Finally we discuss several image matching methods where the adaptive subband correlation scheme can be applied. Their relations with image correlation are explained and their levels of effectiveness are compared using an example.

## 2. Subband Correlation Theorem

We prove that the correlation of two discrete signals can be decomposed into the correlations of their subband components. Here we consider the case of one-dimensional signals and two-channel filter banks. The result can be generalized to two-dimensional signals and multi-channel filter banks easily.

### 2.1 Notations

In the following discussion we use $X(z)$ for the $z$-transform of a one-dimensional sequence $x(n)$; ~ (tilde) for transposition of the matrix, followed by conjugation of $z$-transform coefficients, and substitution of $z$ by $z^{-1}$ at last; $H_i(z)$ for analysis filters and $G_i(z)$ for synthesis filters; subscript 0 for lowpass filters and low frequency signals and subscript 1 the contrary; upward arrow for 1:2 upsampler and downward arrow for 2:1 subsampler (decimator); bold face characters for matrices.

### 2.2 Theorem

The subband decomposition and reconstruction of signals can be represented by the diagram in Figure 1(a), in which we assume lossless coding and ignore quantization. The analysis of such a subband system can be facilitated by using polyphase representation of the signals and the filters [16, 17]. Figure 1(b) shows the polyphase component equivalence of Figure 1(a). The relations between the two figures are:

Polyphase components of the signal

$$X_p(z) = \begin{bmatrix} X_0(z) \\ X_1(z) \end{bmatrix} \quad (1)$$

$$X(z) = X_0\left(z^2\right) + zX_1\left(z^2\right) \quad (2)$$

$$X_i(z) = \sum_n x(2n-i)\,z^{-n} \quad \text{(i=0, 1)} \quad (3)$$

Polyphase components of the filter

$$H_p(z) = \begin{bmatrix} H_{00}(z) & H_{01}(z) \\ H_{10}(z) & H_{11}(z) \end{bmatrix} \quad (4)$$

$$H_i(z) = H_{i0}\left(z^2\right) + z^{-1}H_{i1}\left(z^2\right) \quad \text{(i=0, 1)} \quad (5)$$

$$H_{ij}(z) = \sum_n h_i(2n+j)\,z^{-n} \quad \text{(i, j=0, 1)} \quad (6)$$

Output can be represented in terms of polyphase components as well

$$\hat{X}(z) = \begin{bmatrix} 1 & z \end{bmatrix} G_p\left(z^2\right) H_p\left(z^2\right) X_p\left(z^2\right) \quad (7)$$

# Adaptive Image Matching in the Subband Domain

Hualu Wang and Shih-Fu Chang
Department of Electrical Engineering & Center for Telecommunications Research
Columbia University, New York, NY 10027

**Abstract**

In this paper we discuss image matching by correlation in the subband domain with prospective applications. Theoretical proof is given to show that the correlation of two signals equals the weighted sum of the correlations of their decomposed subband signals. We propose an adaptive method to compute image correlation directly in the subband domain, which avoids decoding of the compressed data. Compared with pixel-domain correlation, this method reduces computation by more than ten times with satisfactory accuracy. We also compare the effects of template size, number of iterations of subband decomposition, and filter type on the speed and accuracy. Complexity estimations and test results are given. In addition, several techniques which involve image correlation are investigated for application in image matching.

## 1. Introduction

As the contents of databases evolve from text to multimedia information, the search and query methods of databases are also undergoing fundamental changes. Applying content-based search and visual query techniques [5], the objective is to efficiently retrieve visual information from large image and video databases with as little human involvement as possible.

Prominent visual features, such as texture, shape, and color, are usually used for query in image and video databases [4, 5, 10, 11, 14]. Users may specify a query by textual descriptions or, most likely, with the assistance of graphical user interface tools. In many other cases, they will give an image example to search for visually similar images or video scenes in the database. One way to achieve this is to extract features from the example and compare them with the visual features of the images in the database. This may become hard when the example lacks distinct visual features or accurate feature extraction is difficult. Another approach is to use the example itself as a template without extracting the features and search for similar images by template matching.

Image matching has been studied extensively and several popular methods are based on the correlation between the image and the template (i.e., the query example) [2, 6]. However, with two-dimensional convolution involved, these methods are computationally complex. Although FFT can be used to carry out the correlation in the Fourier Domain, there is no efficiency gain over the direct implementation when the size of the template is much smaller than the image, which is usually the case in image database searching [3]. Another disadvantage in using FFT is the inverse transform overhead at the end in order to find the correlation peaks, which correspond to possible matching positions.

This paper proposes an adaptive method which achieves image correlation in the subband domain. Compared with the conventional pixel-domain correlation, it reduces computational complexity by more than ten times. Based on the energy distribution over different bands, the adaptive method uses dominant subband correlation components to approximate the original correlation. This method can be applied directly in the subband compressed domain without decod-