A Multiple Perspective Spectral Approach to Object Detection

Robert J. Bonneau

Air Force Research Lab
Radar Signal Processing Branch
Rome, NY 13441

Abstract

Many applications for detection of objects such as video analysis require that candidate objects be observed over a range of perspectives in 3 dimensional space. As a result we must have a robust model and detection process for these objects in order to accurately detect them through a range of geometric transformations. In order to keep our detection process computationally efficient, we use a compact multiresolution model to represent the range of geometric transformations possible in the object to be detected. Additionally, we form an integrated likelihood ratio detection statistic to optimize the detection performance over the entire space of targets being examined. To demonstrate the performance of this algorithm we apply our results to a compressed video sequence and show the improvement of our integrated three dimensional model as a function of model order.

1. Wavelet Markov Data Model

First we shall focus on wavelet filter bank transform structures. The Mallat filter bank structure [7] shown in Figure 1 is the standard wavelet decomposition of most compression applications. The G and H filters are the high and lowpass filters respectively and each one is applied along the x and y axis alternately to extract the HH, HL, LH, and LL frequency band decompositions of the signal.

Recently there have been many studies that have shown the optimally of the wavelet transform domain for detection [4]. In order to work in this domain we must first represent our signal and noise process with an appropriate data model. Such a model is known as the wavelet Markov random field model

To represent this Markov random field [7,5] we define a given node in the quad tree structure as s, its children nodes as $s\alpha_{NW}, s\alpha_{NE}, s\alpha_{SE}, s\alpha_{SW}$ and its parent node as $s\bar{\gamma}$ where $\bar{\gamma}$ shifts the wavelet coefficients from parent $s\bar{\gamma}$ to child s as is shown in Figure 3. A Kth order model defined on the multiresolution structure is defined in either 1 or 2 dimensions with $t \in \{1, 2, \ldots, K(T+1)\}$.

We can define a similar Markov structure based on a DCT transform [9] as well. After labelling the 64 DCT coefficients as in Figure 2 we identify the parent children relationships between DCT coefficients as follows. The parent of coefficient i is [i/4] for $0 < i < 64$ while the set of four children associated with coefficient j is {4j,4j+1, 4j+2, 4j+3 for $0 < j < 16$. The DC coefficient 0 is the root of the DCT coefficient tree which has only three children: coefficients 1, 2, and 3.

Now, defining a MRF on a $2^N \times 2^N$ lattice, a state at the mth level represents the values of the MRF at

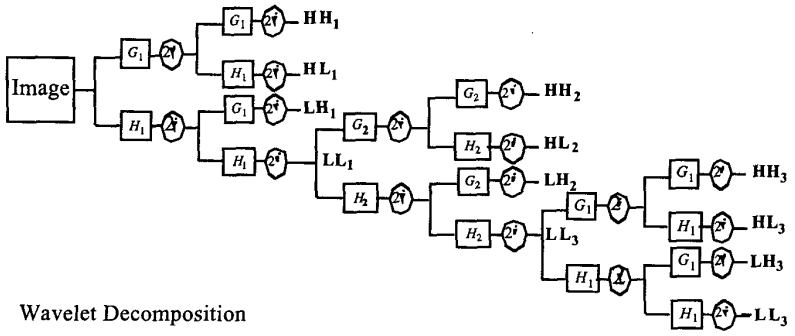

Figure 1 Wavelet Decomposition

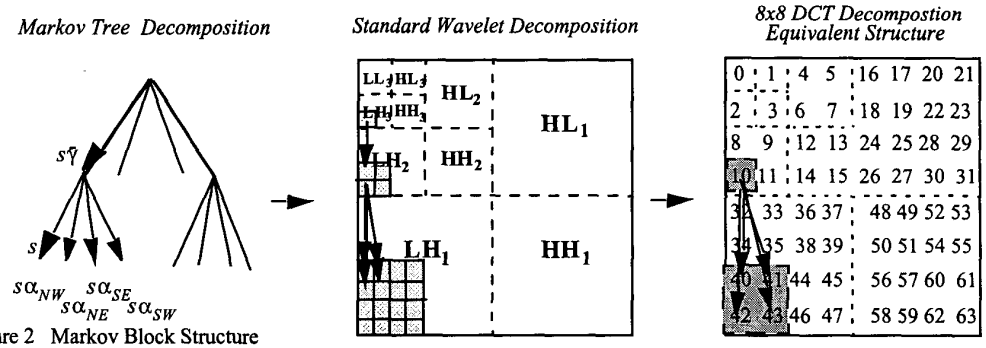

Figure 2 Markov Block Structure

$16(2^{N-m}-1)$ points. This set of points is denoted as Γ_s and it is the union of 4 mutually exclusive subsets. In general we can divide Γ_s into four set sets of $4(2^{N-m(s)}-1)$ points in a similar fashion, and we denote these subsets as $\Gamma_{s,i}, i \in \{NW,NE,SE,SW\}$. Now if we have the random variable Z representing the current state of any Γ_s at any stage of the tree then we insert our local scale iterative relationship, the basic probabilistic Markov relationship is defined as

$$p_{Z_t, t\in\Gamma_{s\alpha_i}|Z_T, T\in\Gamma_s}(Z_t, t\in\Gamma_{s\alpha}|Z_T, T\in\Gamma_s) = p_{Z_t, t\in\Gamma_{s\alpha_i}|Z_T\, T\in\Gamma_{s,i}}(Z_t, t\in\Gamma_{s\alpha}|Z_T\, T\in\Gamma_{s,i}) \quad (7)$$

3. Auto-regressive Structure

Once we have defined the Markov structure from the wavelet or DCT transform we next take the individual coefficient elements and represent them using an autoregressive set of equations as is shown in equation 17. Our target represented by the polynomial coefficients [4] A(s) added to a Gaussian noise component w(s) represented by the B(s) coefficients.

$$x(s) = A(s)x(s\tilde{\gamma}) + B(s)w(s) \quad (8)$$

In the image context we can represent the elements of the image Markov structure in terms of the recursive scale structure shown in Figure 3 with where the superscripts R represent the scale and its associated coefficients.

$$I(s) = a_1(s)I(s\tilde{\gamma}) + \ldots + a_R(s)I(s\tilde{\gamma}^R) + w(s), a_i(s) \in \Re \quad (9)$$

Representing equation 18 in matrix form we have equations 10 and 11.

$$x(s) = [I(s) \quad I(s\tilde{\gamma}) \quad \ldots \quad I(s\tilde{\gamma}^{R-1})]^T \quad (10)$$

$$x(s) = \begin{bmatrix} a_{1,m(s)} & a_{1,m(s)} & \cdots & a_{1,m(s)} & a_{1,m(s)} \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ : & & : & & : \\ 0 & \ldots & 0 & 1 & 0 \end{bmatrix} x(s\tilde{\gamma}) + \begin{bmatrix} 1 \\ 0 \\ 0 \\ : \\ 0 \end{bmatrix} w(s) \quad (11)$$

We can characterize a given signal or texture in our image by solving for the autoregressive coefficients a(s) equation 20 first for our representative texture or object. Then using these coefficients we attempt to use them to predict the target signal in a given input signal x(s). The residual between the target signal and the input signal is then w(s). This model assumes that our target signal is uncorrelated with the input signal.

4. 3-Dimensional Model

We now extend the traditional 2-dimensional spatial Wavelet Markov Model to a 3 dimensional perspective model. To do this we form individual models for each potential perspective that a user would have for a given object. The model from each perspective k is then denoted

$$x_k(s) = [I_k(s) \quad I_k(s\tilde{\gamma}) \quad \ldots \quad I_k(s\tilde{\gamma}^{R-1})]^T \quad (10)$$

This process is shown in Figure 3. The entire model for the collection of all perspectives is then denoted

$$\tilde{X} = [x_1, x_2, \ldots, x_k, \ldots, x_n] \quad (11)$$

Authors [6,8] have denoted such a collection of perspectives as in terms of manifolds. However, such

manifold representations generally deal with rigid bodies observed over all 3-d geometric transformations. Our model is not this rigid but can deal with varying model orders according to the adaptive criteria of the detection process. Additionally, we can handle objects which deform such as humans moving through perspective changes. Because we are using a multiresolution structure for the elements of our model we are able to have compact representations of many perspectives without and excessive amount of computation as in the eigenvalue case.

This model may be utilized in several ways. For instance we may have multiple cameras looking at the same object simultaneously and trying to use their combined information to detect and register the object. Thus each perspective is then defined as an element of the vector $\tilde{X}$. Another example is in the case of one camera observing the object move past it through multiple geometric configurations. Even though the objects shape changes in two dimensions it may still be effectively detected and identified using the three dimensional vector of position information.

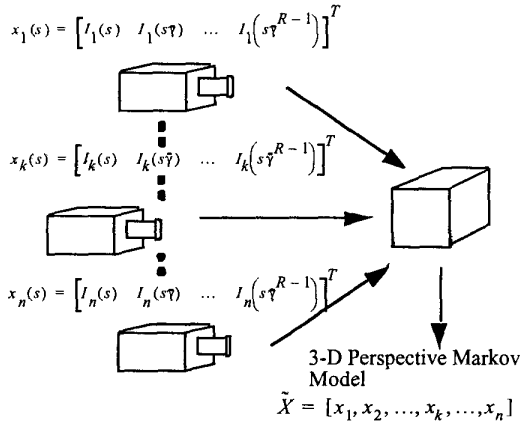

Figure 3 Multiple Perspective 3 Dimensional Markov Structures

5. Transform Domain Detection Statistic

If we assume that our conditional probability density of transition between two successive scales has a white noise signal vector $W_k(s)$ as is described below

$$P_{x(s)|x(s\bar{\gamma})}(X(s)|X(s\bar{\gamma})) = p_{w(s)}(W_k(s)) \qquad (12)$$

we can then express the likelihood function that an image object is H_f vs. H_g as is shown in equation 13.

$$l = \sum_s \log[p_{w(s)|H_g}(w(s)|H_g)] - \sum_s \log[p_{w(s)|H_f}(w(s)|H_f)] \qquad (13)$$

This residual vector W(s) can then be expressed as in equation 14 as

$$W_k(s) = I_k(s) - [(a_{1,m(s)}(s)I_k(s\bar{\gamma}) + \dots + a_{R,m(s)}(s)I_k(s\bar{\gamma}^R))] \qquad (14)$$

In our search algorithm, a given texture or object in an image is described [5] by attempting to predict one or more coefficients in an object with the representative coefficients a(s). The coefficients that minimize this difference are described by 15a as

$$a_m = \arg \min_{a_m \in \Re^R} \left\{ \sum_{\{s|m(s)=m\}} [I_k(s) - a_{1,m}I_k(s\bar{\gamma}) - \dots - I_k(s\bar{\gamma}^R)]^2 \right. \qquad (15a)$$

If we use this minimum distance between multiple scales we have equation 15b as a vector of a coefficients

$$a_m = [a_{1,m} \quad a_{2,m} \quad \dots \quad a_{R,m}]^T \qquad (15b)$$

We now compute a test statistic based on the residual between predicted coefficient and any given image coefficient as is shown in equation 16 as $w(s_c)$. μ_k is the average of the expected residual for the object model and p_c is the standard deviation of those coefficients used to normalize the statistic $\varsigma(s_c)$

$$\varsigma_k(s) = \frac{w(s) - \mu_k}{\sqrt{p_k}} \qquad (16)$$

$$c_2(s) = min[((\zeta_1(s), \zeta_2(s)), \dots, \zeta_k(s)), \dots, \zeta_n(s)] \qquad (17)$$

This normalization process makes the decision threshold T to decide between a H_f vs. H_g a uniform quantity independent of image noise properties. Our minimum value over the sequence of detection results for our model represents the best result over the entire sequence. We can assess the detection performance of this metric by computing probability of detection P_D and probability of false alarm P_{FA} using know targets with equations 18 and 19

$$\hat{P}_D = \frac{\text{\# of detected pixels | target}}{\text{\# of target pixels}} = \frac{n_{tgt}}{N_{tgt}} \qquad (18)$$

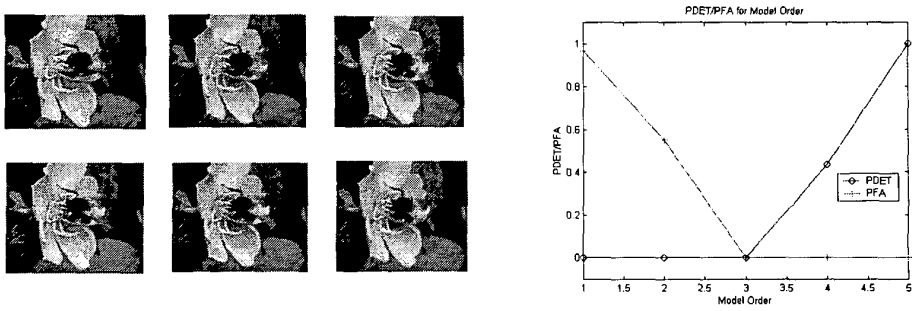

Figure 4 Bee Image Sequence With Prob of Detection and Probability of False Alarm With Increasing Model Order

$$\hat{P}_{FA} = \frac{\#\text{ of detected pixels} \mid \text{clutter}}{\#\text{ of target pixels}} = \frac{n_{fa}}{N_{tgt}} \quad (19)$$

6. Results

Our first application of our multiple perspective model is to show the improved detection performance of a 3 dimensional object from one video sensor moving through a sequence of poses. We build our multi-perspective model against the bee image sequence found in Figure 4 with the objective of detecting the bee throughout the entire video of approximately 100 frames using the Daubchies 8 basis set. Note that model order one corresponds to conventional detection methods. Our preliminary results in Figure 4. show that our detection performance improves while our probability of false alarm dramatically drops as our model order increases.

7. Conclusion

Our multiple perspective approach for object detection allows us to reduce the complexities of 3 dimensional objects in 2 dimensional scenes for improved detection performance. The multi-resolution Markov model and associated test statistic allow us much more flexibility with low overhead for many applications. Video registration for stereo matching, 3-d object recognition, and compressed domain video analysis are among the many uses of this approach.

References

[1] R. Bonneau, "The multiresolution transform and its application to image coding", SPIE/IEEE Conference on Visual Communication and Image Processing: Wavelets and Fractals, Orlando, March 1996.

[2] G. Davis, "Adaptive Self-Quantization of Wavelet Subtrees: A Wavelet-Based Theory of Fractal Image Compression", SPIE Conference on Mathematical Imaging: Wavelet Applications in Signal and Image Processing, San Diego, July 1995.

[3] I. Daubechies, "Ten lectures on wavelets," CBMS-NSF Series Appl. Math. SIAM, 1991.

[4] Fosgate C.H.., Krim H., Irving W.W., Karl W., Willsky A.S., "Multiscale Segmentation and Anomaly Enhancement of SAR Imagery", IEEE Transactions on Image Processing, Vol 6 No 1, January 1997.

[5] Luettgen, M., Karl, W., Willsky, A., Tenny, R., "Multiscale Representations of Markov Random Fields", IEEE Transactions on Signal Processing, Vol 41, No 12, pp. 3377-3396, December 1993.

[6] Nayar, S.K., Nene, S.A., Murase, H., "Real-Time 100-Object Recognition System, Proceedins of ARPA Understanding Workshop, San Francisco, February 1996.

[7] S. Mallat, "Multiresolution Approximations and Wavelet Orthonormal Bases of L2R", Transactions of the American Mathematicall Society, Volume 315, Number 1, September 1989.

[8] Murase, H., Nayar, S.K, "Image Spotting of 3D Objects Using Parametric Eigenspace Representation", Proceedings of 9th Scandanavian Conference on Image, Analysis pp. 325-332, June 1995.

[9] Ramchadndran, K, Orchard T., "An Investigation of Wavelet-Based Image Coding Using an Entropy -Constrained Quantization Framework", IEEE Transactions on Signal Processing, Vol 46, No 2, pp342-353, February 1998.

[10] J. Shapiro, "Embedded Image Coding Using Zerotrees of Wavelet Coefficients", IEEE Transactions on Signal Processing, Vol 41, No 12, December 1991.