REPORT DOCUMENTATION PAGE

Form Approved
OMB NO. 0704-0188

1. AGENCY USE ONLY (*Leave blank*)

2. REPORT DATE
17 January 1997

3. REPORT TYPE AND DATES COVERED
Technical (June 1996 - November 1996)

4. TITLE AND SUBTITLE
Use of Wavelets to Extract Structural Features from Images

5. FUNDING NUMBERS
Contract:
DAAH04-95-C-0003

6. AUTHOR(S)
Steven Schwartz and James Meany

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
McDonnell Douglas Aerospace
P.O. Box 516, Mailcode S270-1115
St. Louis, MO 63166

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSORING / MONITORING AGENCY REPORT NUMBER
ARO 33584.6-MA

11. SUPPLEMENTARY NOTES
This work was performed on a contract under the ARPA University-Industry ATR Initiative.

12a.. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for public release; distribution unlimited.

12b. DISTRIBUTION CODE
N/A

13. ABSTRACT
Summarizes work performed by McDonnell Douglas Aerospace on a contract under the ARPA University-Industry ATR Initiative. The focus is on a preliminary investigation into the use of wavelets to extract local feature points to be used for matching 2-D and 3-D structure in multiple images. The approach is to detect dominant local discontinuities (i.e., point features, corners, curvature discontinuities) as a function of wavelet coefficients.

14. SUBJECT TERMS
Multiresolution Analysis, Wavelets, Automatic Target Recognition, Feature Extraction

15. NUMBER OF PAGES
11

16. PRICE CODE
N/A

18. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	18. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

USE OF WAVELETS TO EXTRACT STRUCTURAL FEATURES FROM IMAGES

TYPE OF REPORT: TECHNICAL

STEVEN SCHWARTZ AND JAMES MEANY

17 JANUARY 1997

U. S. ARMY RESEARCH OFFICE

CONTRACT / GRANT NUMBER: DAAH04-95-C-0003

MCDONNELL DOUGLAS AEROSPACE
P.O. BOX 516, MAILCODE S270-1115
ST. LOUIS, MO 63166

TABLE OF CONTENTS

LIST OF FIGURES

1. Introduction

This report summarizes work performed by McDonnell Douglas Aerospace on a contract under the ARPA University-Industry ATR Initiative. The focus is on a preliminary investigation into the use of wavelet coefficients for the extraction of local feature points ("tie points") to be used for matching 2-D and 3-D structure in multiple images. The ultimate objective of this effort is to develop techniques to automatically estimate 3-D scene structure from sensor image sequences or multi-look sensor imagery. The preliminary effort described here focuses on the extraction of local structural features computed as a function of wavelet coefficients and the matching of feature points between images from FLIR image sequences.

2. Approach

The general approach to be investigated is depicted in Figure 1. The wavelet transforms of a pair of images are computed. Features which measure local discontinuity are computed as a function of the wavelet coefficients. The largest or dominant feature points are isolated to provide candidate local features to be matched between the images. Ultimately, 2-D or 3-D scene structure is to be inferred based on the correspondence of match points.

The research areas in this approach are: (1) determining appropriate linear or nonlinear functions of wavelet coefficients for extraction of local feature points, (2) developing methods for identifying correspondence between feature points, and (3) developing methods for inferring 2-D and 3-D structure based on matched feature points. This paper focuses on preliminary investigations into the first two problems.

3. Experiment - Using Wavelet Features to Match Features in FLIR Image Sequences

Figure 2 depicts preliminary experiments investigating the extraction of wavelet-based feature points and simple local search methods for determining correspondence between feature points in successive images of forward-looking infrared (FLIR) video. First, a biorthogonal wavelet transform is applied to a pair of successive images from the FLIR sequence. Second, local feature values are thresholded to leave only dominant feature points. Next, centroids of feature groups are identified to eliminate spurious feature points in a cluster. Finally, a local search is performed to find correspondence between match points in the pair of images.

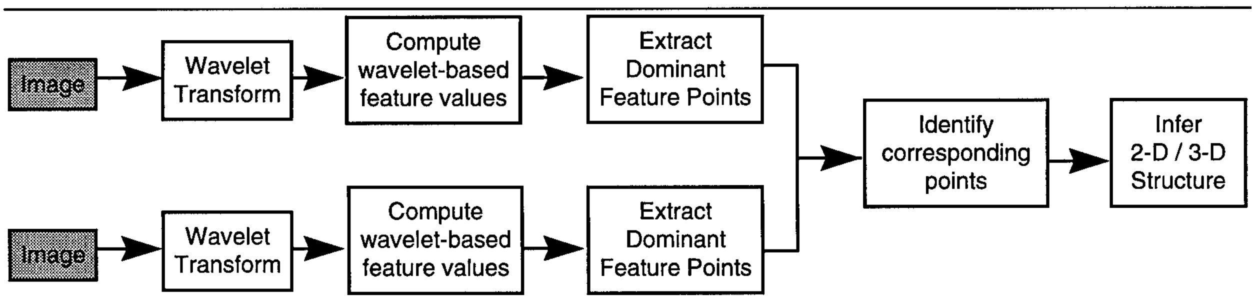


Figure 1. General Approach for Use of Wavelets to Extract Structural Features from Images

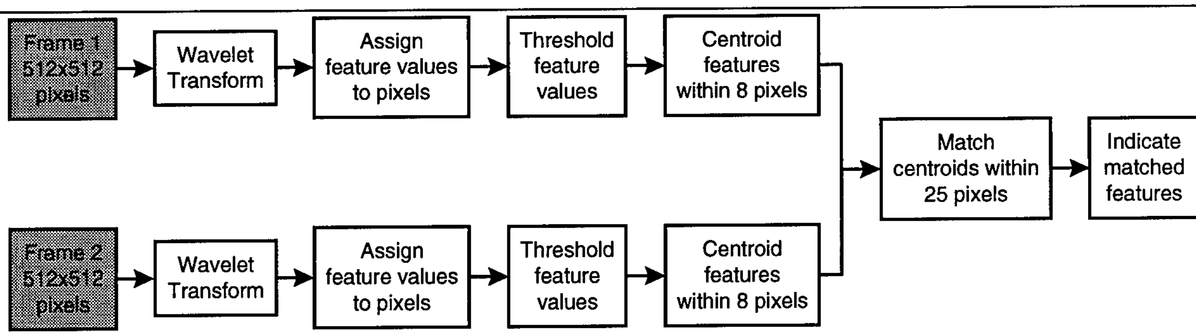


Figure 2. Approach for Wavelet Feature Extraction Experiments

3.1 Extraction of Wavelet Features

Wavelets have proven to be well-suited for the extraction of localized image features such as edges and texture features.[1-3] Here our objective is to extract point features (such as corners and curvature discontinuities) as a function of the wavelet coefficients. These point features are then used as "tie points" for correlating scene structure between images.

Four resolution levels of the wavelet transform of each image are computed using biorthogonal filters with 5 and 3 taps.[4] The coefficients of these filters are -1/8, +2/8, +6/8, +2/8, and -1/8 for the scaling (lowpass) filter; and -1/4, +2/4, and -1/4 for the wavelet (highpass) filter.

One of the drawbacks of the wavelet transform for image and signal analysis applications is that it lacks invariance to translation. It is clearly impossible for a subsampled transform to be invariant to translation (unless the translation occurs at a multiple of the subsampling factor). However, if the sampling rate is sufficiently high, it is possible to achieve a weaker form of translation invariance, in the sense that the energy in a particular band can be preserved as the signal or image is translated.[5] Thus, we sample the wavelet coefficients at twice the standard critically sampled rate (in each dimension). For an M x N image, this results in subbands having dimensions of M x N at the finest resolution level, M/2 x N/2 at the second finest level, M/4 x N/4 at the third level, and M/8 x N/8 at the coarsest level. The oversampled transform ends up with 2M x 2N transform coefficients, including the approximation coefficients at the coarsest scale. Figure 3 depicts the structure of this oversampled wavelet decomposition, with the subband dimensions indicated. A notation which we will use for the coefficient variables in shown in the respective subband blocks, with a denoting an approximation coefficient and r denoting a detail (residual) coefficient. The subscripts v, d, and h represent the vertical, diagonal, and horizontal orientation preferences respectively. The superscripts denote the resolution levels.

While the individual wavelet subbands are sensitive to certain oriented features (i.e., vertical edges, horizontal edges, diagonal edges), we want to extract not edges, but localized point features such as corners, which should produce a response at multiple resolutions and in multiple orientation bands. To do this, we compute at each pixel location a single feature value, which is a function of the wavelet coefficients (from all three orientations and from each resolution level) associated with that pixel location. Coefficients are associated to pixel locations by defining a square centered on the support of the corresponding basis function. (This association has some

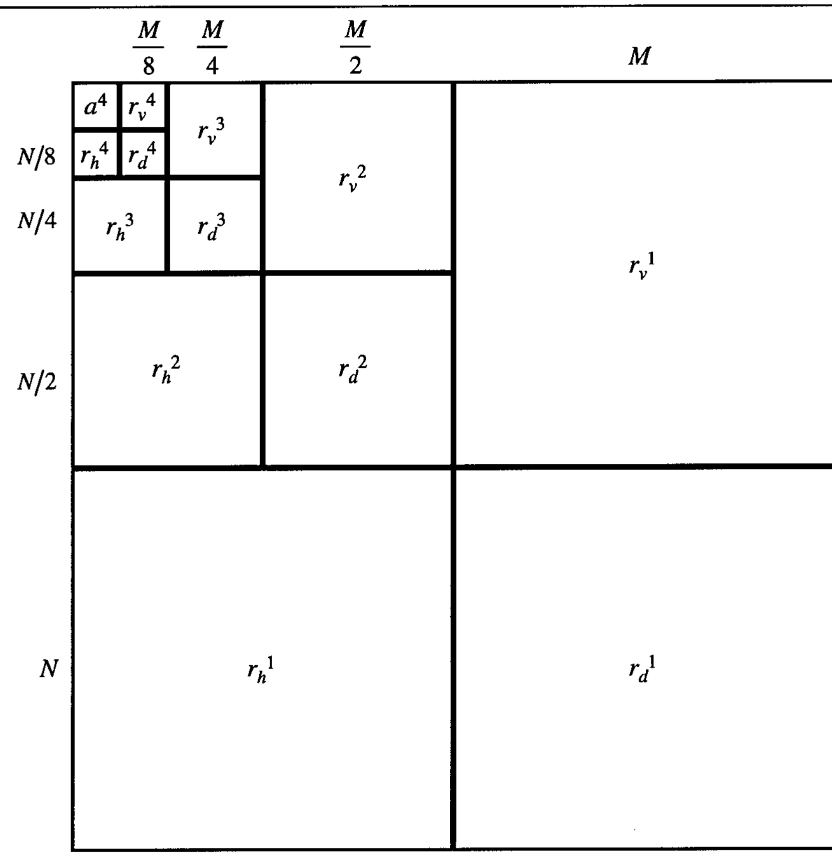


Figure 3. Structure of Oversampled Wavelet Decomposition

drawbacks which we will discuss below.) We currently use a linear combination of the absolute values of the wavelet coefficients. This is given by:

$$feature = \sum_{i=1}^{4} \left(|d_v^i| + |2d_d^i| + |d_h^i| \right)$$

Among several linear combinations that were evaluated, this choice seemed to provide the best performance for use with the point correspondence process described below. Doubling the weight of the diagonal channels improved the selection of point features over linear features. We found no advantage to weighting of resolution levels. We believe that the feature values should ultimately be a nonlinear function of the coefficient absolute values, so that a high contrast edge, which is localized in only one dimension and which produces a strong response in one orientation, can be suppressed in favor of corners and points which are localized in both dimensions and which produce a strong response in multiple orientation bands.

Figures 4, 5, and 6 depict the feature extraction and matching process for three image pairs from FLIR image sequences of urban clutter, a hospital complex, and a power plant, respectively. Figures 4a, 5a, and 6a show the original image pairs. The white border line near the perimeter of the left image in each pair indicates the limits of the region within which feature values are computed. At points outside this border, the features are not computed due to boundary effects of the wavelet filters.

Figures 4b, 5b, and 6b display the feature values for the image pairs. For the purposes of display, the feature values are normalized to a scale with 256 possible values, with black indicating the minimum feature value and white indicating the maximum feature value. Visual inspection of these feature images reveals an undesirable blocky character. This is attributable to the fact that even though the coarser resolution levels 2, 3, and 4 are oversampled in comparison to a standard wavelet decomposition, they are still subsampled with respect to the original pixel resolution. The association of these subsampled coarse resolution coefficients to locations at full pixel resolution leads to blocking of the feature values. In order to locate point features precisely, it would be preferable to eliminate this blocky characteristic in the feature values. Some potential strategies for accomplishing this are discussed in Section 4.

An adaptive threshold is applied to the feature values to obtain a set of about 20 candidate dominant (large) feature points. The dominant feature points often occur in clusters of adjacent pixel locations. In such cases, the centroid of the cluster is computed, and is used as a single feature point. For the purposes of our experiment, an 8 pixel radius was used to establish such clusters. Figures 4c, 5c, and 6c depict the positions of the dominant feature points, as indicated by squares centered on the points. Visual inspection of the results will show that the feature locations correspond to point features and corners within the images.

3.2 Determining Correspondence Between Feature Points

- Because the image pairs used in this experiment were from successive frames of FLIR image sequences, the frame-to-frame motion was strongly constrained. As a simple method to establish correspondence between feature points, we searched for matched feature pairs occurring within 25 pixels of each other in the pair of FLIR images. The resulting matched feature pairs are depicted in Figures 4d, 5d, and 6d. While this matching process is admittedly simple, visual inspection shows that in most cases, the matched points identified corresponding structure in the image pairs.

4. Conclusions and Future Work

The results of these preliminary experiments indicate that wavelet-based features offer potential value for extraction of feature points for matching structure in multi-look images and image sequences. Although the contract under which this work was performed has ended, we intend to extend this effort under Independent Research and Development (IRAD) funding. Future work is indicated in several areas:

- The blocky character of the feature image should be eliminated. The strategy of oversampling the wavelet transform by a factor of two preserves the energy in each band, but still does not match the pixel resolution. What we would like is to generate values at full resolution for each subband. Possible strategies are:

  - An obvious strategy is to fully sample the transform, but this is computationally costly.

  - Another strategy is to compute values at all pixel locations using bilinear interpolation from the nearest coefficients.

  - The strategy which we prefer is to invert each oversampled subband back into the pixel domain. This will produce an appropriately band-limited signal with full-resolution.

- The computation of feature values should be improved. As discussed above, a linear combination of coefficient absolute values will produce large feature values along high contrast edges, resulting in spurious point features which do not in fact correspond to any localized discontinuity. For the detection of point features, the feature value computation should include a nonlinearity to ensure that there is a significant response in multiple orientation bands.

- The search for correspondence must be able to handle significant motion between feature points because the image pairs will not always come from successive frames of an image sequence, and also because the estimation of 3-D structure by stereopsis techniques can require significant disparity between match points.

- The search for correspondence should be computationally efficient and should include considerations of consistency when sets of points may be the vertices of a rigid 2-D or 3-D structure. Consider, for example, that when the transformation between a pair of images consists of translation, scaling, and rotation within the image plane, then angles and log-distances from a given point feature to other point features provide an invariant signature which may be used to register the images.[6] We will consider possible extensions of this property to find pairings of feature points which are consistent with the motion of the vertices of a larger structure. Sets of points which exhibit motion which is consistent in some fashion will provide a potential basis for inferring 2-D scene structure or for inferring 2-D facets within 3-D scene structure.

- These efforts will be coordinated with future efforts by Summus, Ltd. on the use of partial differential equations (PDEs) for the extraction and minimal representation of 2-D and 3-D image features.

References

1. B. Deng and B. Jawerth and G. Peters and W. Sweldens, "Wavelet probing for compression based segmentation", *Mathematical Imaging: Wavelet Applications in Signal and Image Processing*, A. Laine (ed.), pp. 266-276, SPIE, 1993.

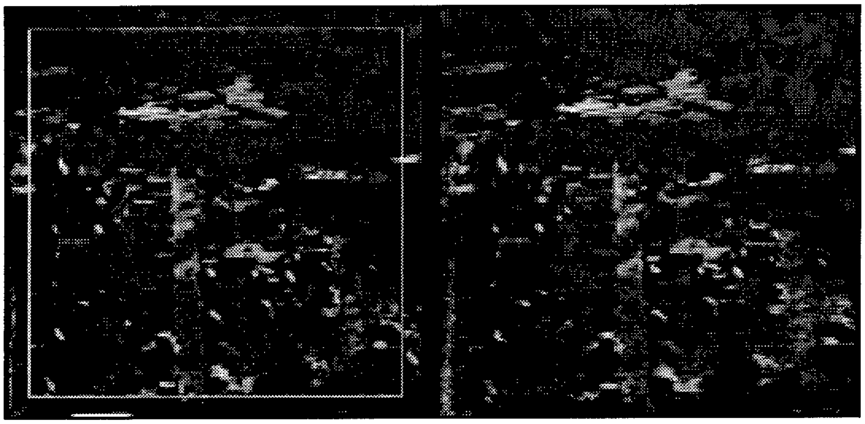
2. A. Laine and J. Fan, "An adaptive approach for texture segmentation by multi-channel wavelet frames", Technical Report 25, Univ. of Florida, 1993.

3. B. S. Manjunath, R. Chellappa, and C. Malsburg, "A feature based approach to face recognition," Proc. IEEE Conf. Comput. Vision and Pattern Recognition, Champaign, Illinois, pp. 373-378, 1992.

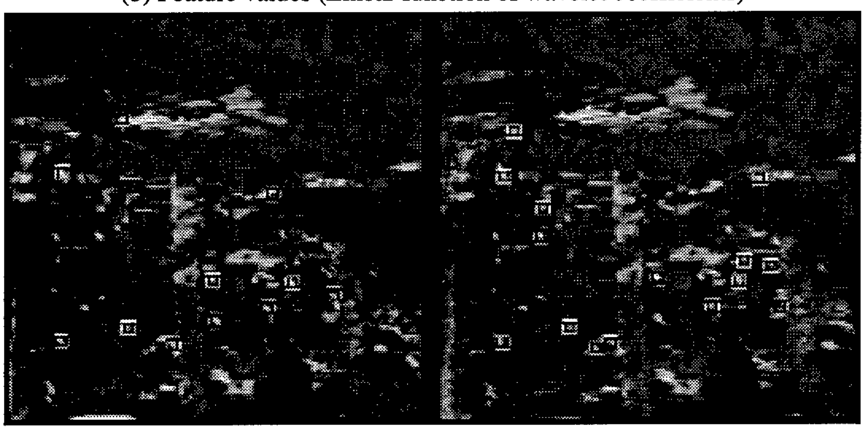
4. A. Cohen, "Biorthogonal Wavelets", *Wavelets – A Tutorial in Theory and Applications*, C.K. Chui (ed.), Academic Press, 1992.

5. E. P. Simoncelli, W. T. Freeman, E. H. Adelson, and D. J. Heeger, "Shiftable Multiscale Transforms," *IEEE Transactions on Information Theory*, vol. 38, pp. 587-607, 1992.

6. P. Fiore, "Image Registration Using Both Distance and Angle Information," Proc. IEEE International Conf. on Image Processing, Austin, Texas, 1995.


(a) Sensed FLIR images (from sequence)


(b) Feature values (Linear function of wavelet coefficients)


(c) Extracted feature point locations (indicated by squares)

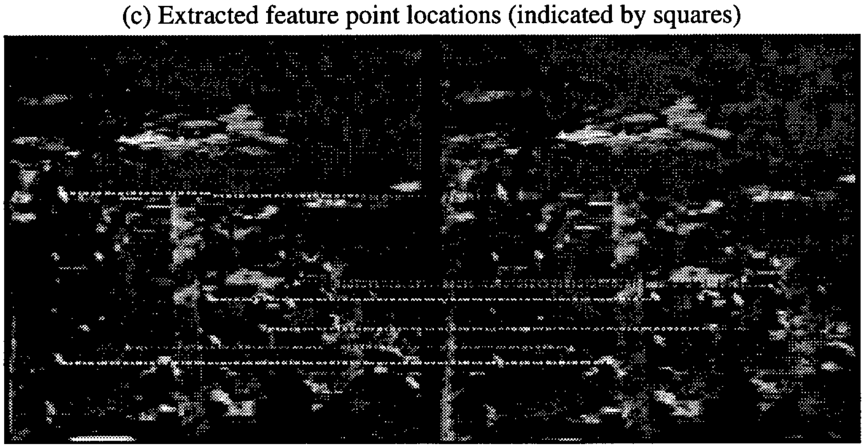
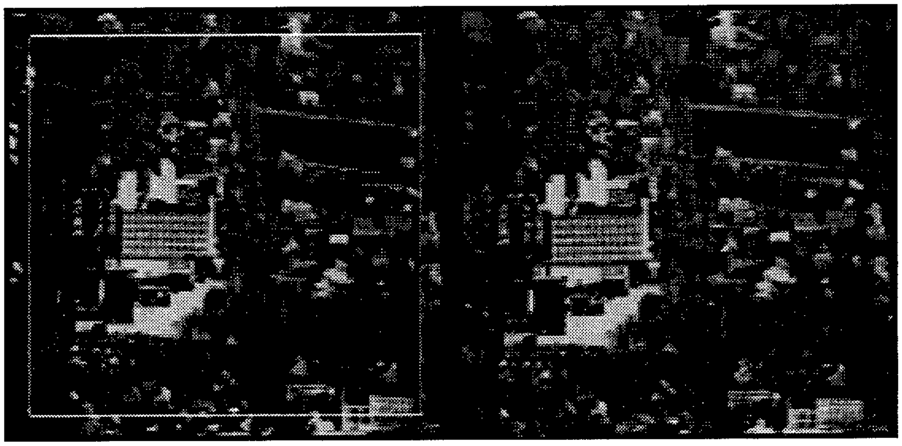
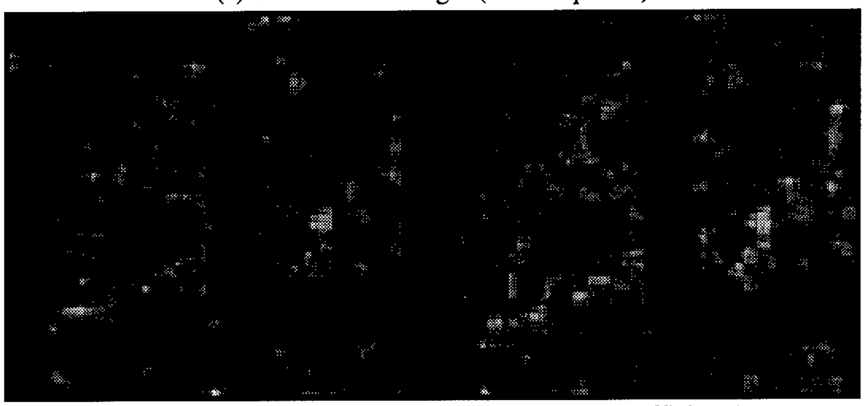

(d) Matched features (indicated by dotted lines)

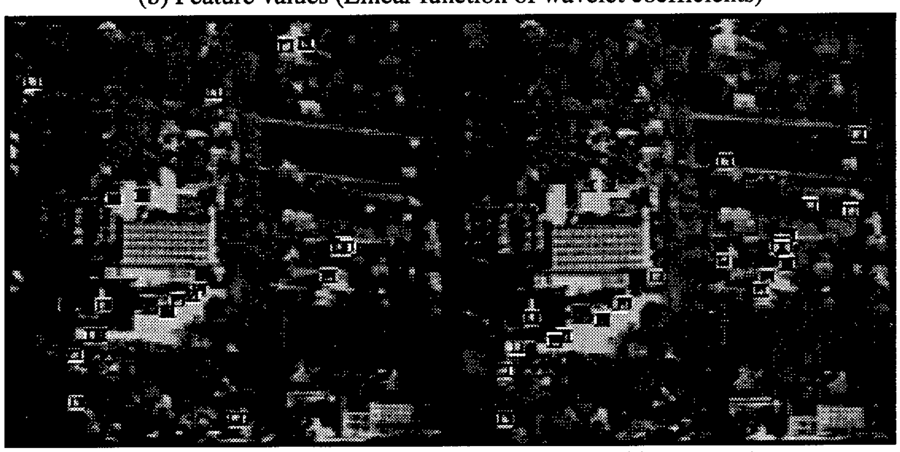
Figure 4. Wavelet Feature Matching in Urban Clutter



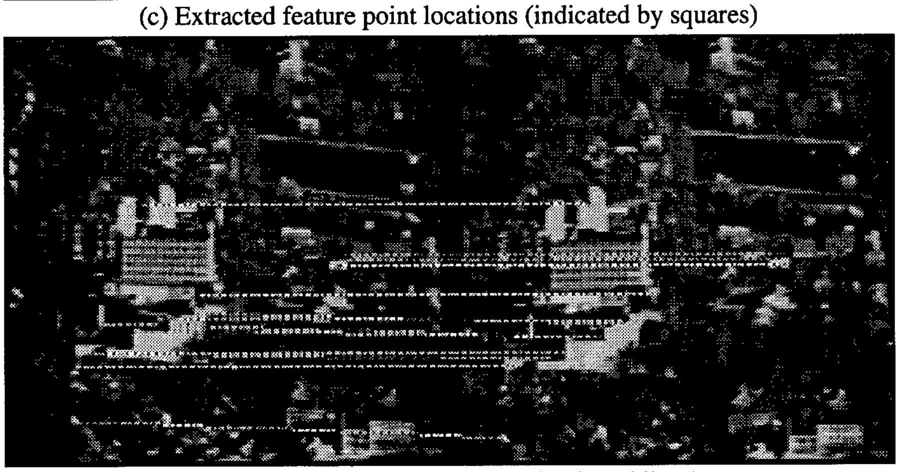
(a) Sensed FLIR images (from sequence)



(b) Feature values (Linear function of wavelet coefficients)



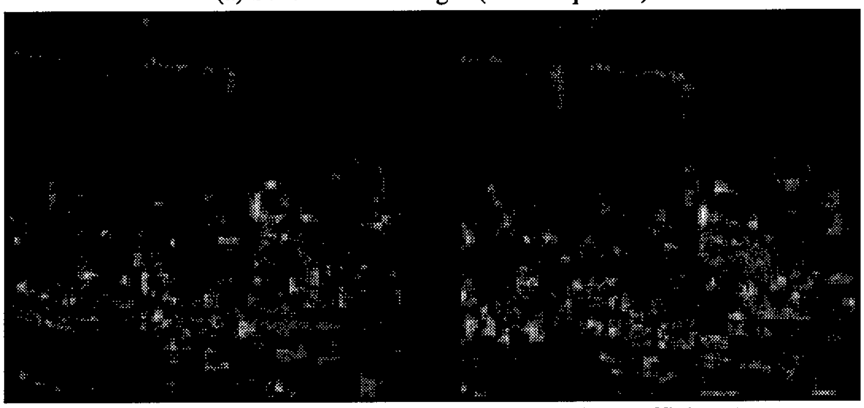
(c) Extracted feature point locations (indicated by squares)



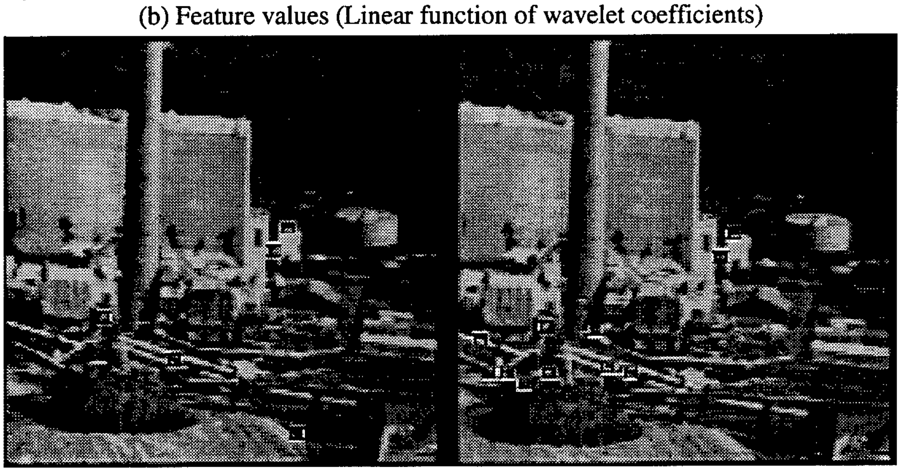
(d) Matched features (indicated by dotted lines)

Figure 5. Wavelet Feature Matching on Hospital Complex

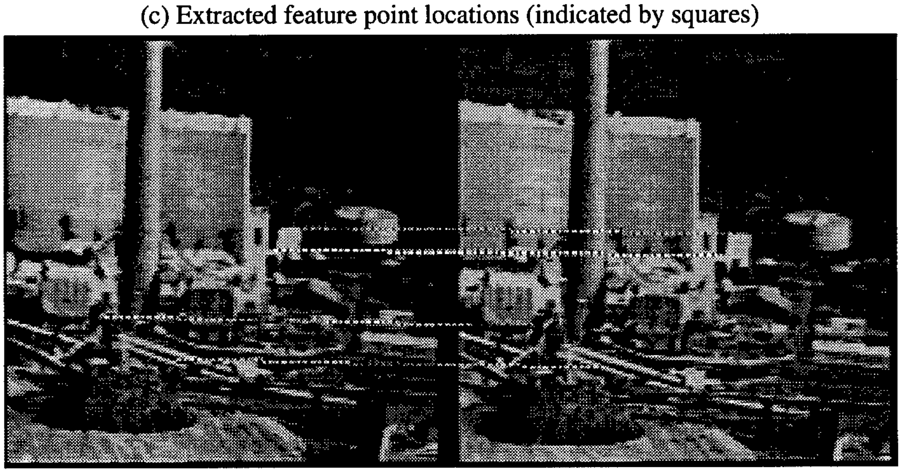
(a) Sensed FLIR images (from sequence)



(b) Feature values (Linear function of wavelet coefficients)



(c) Extracted feature point locations (indicated by squares)



(d) Matched features (indicated by dotted lines)

Figure 6. Wavelet Feature Matching on Power Plant