SSD Matching Using Shift-Invariant Wavelet Transform

Fangmin Shi, Neil Rothwell Hughes and Geoff Roberts
Mechatronics Research Centre
University of Wales College, Newport
Allt-Yr-Yn Campus
PO Box 180
Newport NP20 5XR, UK
fangmin.shi@newport.ac.uk
neil.rothwell-hughes@newport.ac.uk
geoff.roberts@newport.ac.uk

Abstract

The conventional area-based stereo matching algorithm suffers from two problems, the windowing problem and computational cost. Multiple scale analysis has long been adopted in vision research. Investigation of the wavelet transform suggests that -- dilated wavelet basis functions provide changeable window areas associated with the signal frequency components and hierarchically represent signals with multiresolution structure. This paper discusses the advantages of applying wavelet transforms to stereo matching and the weakness of Mallat's multiresolution analysis. The shift-invariant dyadic wavelet transform is exploited to compute an image disparity map. Experimental results with synthesised and real images are presented.

1 Introduction

Finding correspondence is an ill-posed problem in stereo vision. Area-based stereo matching is one of the conventional solutions. It compares the intensity similarity between windowed areas of two stereo images. The sum of squared difference (SSD) [1] is commonly used as the similarity measure:

$$ssd(x) = \sum_{t=x-s}^{x+s} \left| L(t) - R(t) \right|^2 \qquad (1)$$

where x is the pixel index over the stereo images L and R, t indexes over the local area around x within $\pm\sigma$. It is well known that this method suffers from the windowing problem and computational cost [1].

In order to alleviate these problems, multistage strategies were developed by vision

BMVC 2001 doi:10.5244/C.15.13

researchers such as multistage matching by dividing images into small blocks of equal size [2], multiscale matching based on Gaussian filtered images [3], [4], [5], and the pyramid structure that generates sets of low-pass and band-pass filtered images [6]. A more general hierarchical architecture and fast implementation was created by Mallat in 1989 following a study of wavelet concepts. This is known as wavelet multiresolution analysis (MRA) [7].

The advantage of the wavelet transform is that it uses wide windows for low-frequency components and narrow windows for high-frequency components [8]. These windows are formed by dilations and translations of a prototype (or mother) wavelet. Thus, if SSD matching is performed on the wavelet transforms of signals, the windowing problem with the conventional SSD approach is naturally solved. However, Mallat's multiresolution analysis lacks shift-invariance, which will be discussed in the following section. Stereo matching requires a shift-invariant transform because stereo image pairs can be considered as the shifted versions of each other (with distortions). This makes MRA unsuitable for matching.

This paper discusses the strengths of wavelet transforms when applied to stereo matching and some alternative wavelet methods to Mallat's MRA. The Dyadic wavelet transform is exploited to compute a dense disparity map. Experimental results using synthesised and real images are presented.

2 Properties of the Wavelet Transform for Stereo Matching

Modern wavelet theory was motivated initially for the sake of a better time-frequency signal representation than the short time Fourier transform (STFT) and to overcome its drawbacks. In contrast with the STFT that uses a constant window for the whole signal, the wavelet transform uses wide windows for low-frequency components and narrow windows for high-frequency components [8]. It achieves this by decomposing a signal into the dilations and translations of a mother wavelet.

Let $j(t)$ denote a mother wavelet, which is a small oscillatory function with finite support. A family of wavelets $j_{a,b}(t)$ is then represented by

$$j_{a,b}(t) = a^{-1/2} j((t-b)/a) \tag{2}$$

where a is the scale parameter, b is the translation parameter, and $a, b \hat{I} R$. The wavelet transform represents a signal $x(t)$ by an infinite set of such basis functions:

$$WT(b,a) = \int x(t) a^{-1/2} j((t-b)/a) dt \tag{3}$$

2.1 Automatic Windowing Analysis

In order to illustrate the time-frequency resolution of a wavelet transform, Figure 1 shows the coverage of a wavelet in the time-frequency plane. It is evident that when the frequency interval goes up by a scale factor, the time interval goes down by the same factor. Let Dt and Df denote the window width of the mother wavelet in time and in the spectral domain, and Dt_{ab} and Df_{ab} are the corresponding denotations of the scaled and shifted wavelet. That is:

$$Dt \times Df = Dt_s \times Df_s \quad (4)$$

This reveals that the product of the window width of time and frequency is constant at all scales [8]. This property is one of the most important advantages that wavelet transforms provide. In contrast with the STFT applying either narrow or wide window (but not both) to the whole signal, wavelet transforms are able to analyse high-frequency components using small windows and low-frequency components using big windows. This property is ideal when dealing with non-stationary signals that contain both short high-frequency components and long low-frequency components.

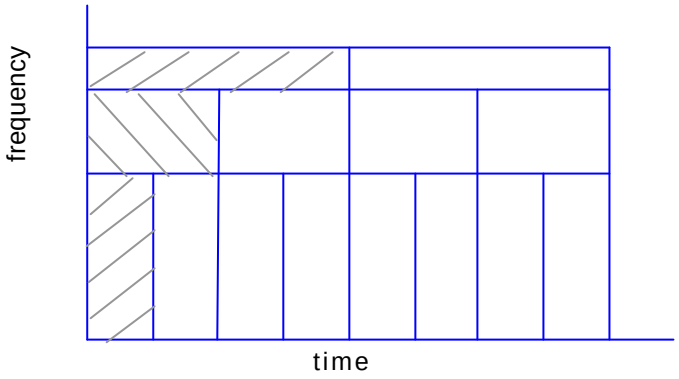

Figure 1 Time-frequency plane of wavelet transform: window area is constant at all scales

An image is a typical non-stationary signal, which consists of a slowly changing background and rapidly changing details. After decomposing an image, a set of images at different resolutions is obtained. At coarser resolutions, the matching is performed by comparing wider areas leading to larger uncertainty in disparity localisation. At finer scales, the compared areas tend to be more localised and smaller uncertainty in disparity localisation is expected. The windowing problem that occurs with SSD matching is then naturally solved by choosing the right wavelet.

2.2 Why MRA Is not Suitable for Matching

In equation (3), if parameter a and b are continuous real values, then the transform is called a continuous wavelet transform. The wavelet basis functions constitute an overcomplete representation in which information is highly redundant. This redundancy can be reduced by discretising a and b. Daubechies [9] found that when $a=2^n$, $b=k2^n$, n, $k \in \mathbf{Z}$, the basis functions $\{\psi_{kn}(x)=2^{-n}\psi(2^{-n}x-k)\}$ are orthogonal for certain choices of wavelet. Stimulated by the pyramidal approach in vision, Mallat, as a former vision researcher, proposed a fast implementation for the wavelet orthogonal decomposition [7]. This is the well known multiresolution analysis (MRA).

MRA decomposes a signal into the same size subimages at dyadic scales. At each scale an approximation part and a detail part are formed by passing the signal through a half-band low-pass filter and a half-band high-pass filter, and subsequently downsampling them by two. The approximation part is then hierarchically decomposed. Downsampling a signal simply discards every other sampling point. This

operation reduces the number of signal samples by a half when the scale is doubled.

Shift-invariance (or *time-invariance*) means that if a signal is delayed in time, its transform result is delayed as well. Downsampling is not shift-invariant. Neither, therefore, is MRA. This issue was discussed by Strang [10] and the shift-invariance problem was considered to be the main drawback of MRA.

For stereo matching, intuitively, one image can be assumed to be the shifted version of the other. The shifted value with respect to each pixel is the disparity, which is dependent on the pixel position. Only shift-invariant wavelet transforms can be used for matching.

2.3 Shift-Invariant Wavelet Transforms

The continuous wavelet transform possesses the property of shift-invariance. However, its high redundancy gives rise to high computational cost. Approximate shift-invariance with effective computation needs to be achieved. Mallat used the dyadic wavelet transform and the zero-crossings of the dyadic wavelet transform to reduce the representation size [11]. Simoncell [12] built steerable filters to achieve a shiftable transform that is jointly invariant in position, scale and orientation. More recently, a shift-invariant complex wavelet transform [13] with perfect reconstruction has been constructed and applied to image processing and computer vision.

The wavelets used in these papers could be applied to the matching problem. This paper will discuss the application of the dyadic wavelet transform to disparity computation. The motivation for this is discussed in the next section.

3 Correspondence Matching Using the Dyadic Wavelet Transform

To simplify the numerical computations and maintain shift-invariance, the scale parameter a of equation (3) is discretised along a dyadic sequence $\{2^n, n \hat{I} \boldsymbol{Z}\}$ while leaving the shift parameter b continuous. The dyadic wavelet transform ($DWT(b, j)$)has the following form:

$$DWT(b,n) = \int x(t) 2^{-n/2} j((t-b)/2^{-n}) dt \qquad (5)$$

Mallat [14] proved that under certain condition dyadic wavelet transform defines a complete and stable representation. The algorithmic efficiency if gretly improved compared with the continuous wavelet transform. The information is still redundant due to the continuous translation parameter. However, it is good for matching task aiming at dense disparity map output.

Figure 2 gives two signals (epipolar lines from two synthesised stereo images). The decomposition results at three scales, 2^1, 2^2 and 2^3, are shown in Figure 3.

The sum of squared difference (SSD) [15] is applied to the transformed signals to measure the similarity of the corresponding points. The smaller the SSD value, the more likely it is that the points correspond to each other. Unlike the conventional SSD, directly applied to the image intensity value, the SSD here is applied to the wavelet coefficients.

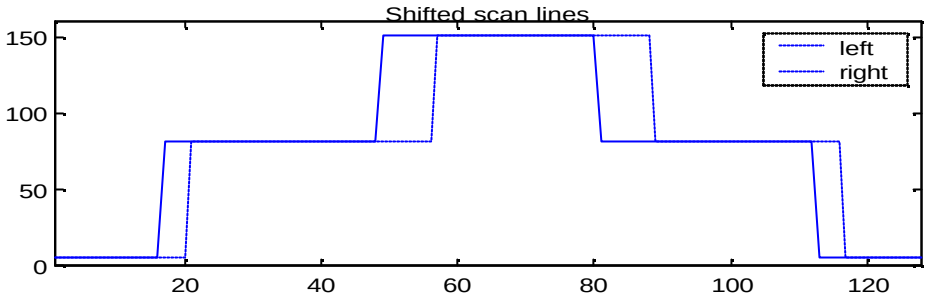

Figure 2 Scan lines from stereo images

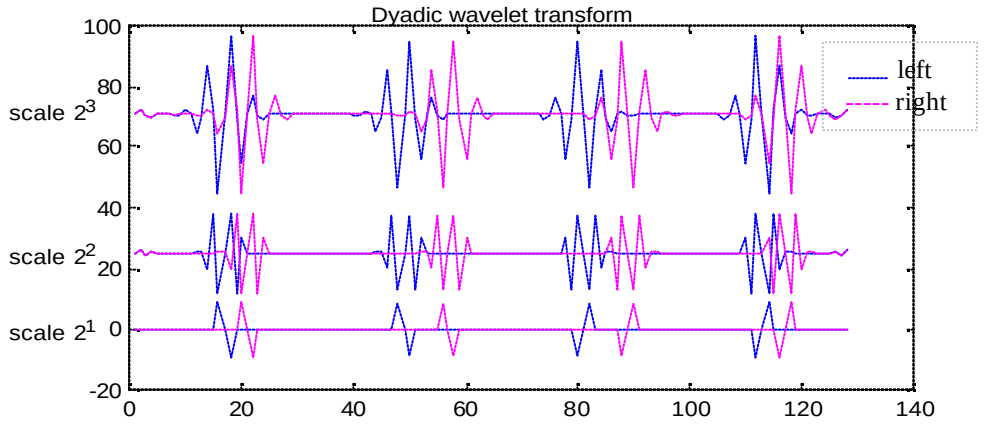

Figure 3 1D dyadic wavelet transform at three scales

Let $x_1(t)$ and $x_2(t)$ be two scan lines of stereo images, their dyadic wavelet transforms are denoted by $DWT1(2^n, t)$ and $DWT2(2^n, t)$, respectively, where $n \hat{I} N$. The SSD measure ($ssd(n,x)$) is defined as:

$$ssd(n,x) = \sum_{t=x-2^n s}^{x+2^n s} \left| DWT1(2^n, t) - DWT2(2^n, t) \right|^2 \qquad (6)$$

where s is the size of an interval where the energy of the mother wavelet is mostly concentrated [11].

From equation (6), it can be seen that at each scale the searching area is ($-2^n s$, $2^n s$). Matching should be taken at as much scales as possible. The maximum scale ($nmax$) should be determined by: $2^{n\max} s \le signal\ \ length$.

Besides the *epipolar constraint* [16], other constraints e.g. s*imilarity*, *uniqueness*, *ordering* and c*ontinuity* [17] are also applied along with equation (6). Figure 4 gives the computed disparity result at three scales. The corresponding SSD values are also recorded as a measure of the matching confidence. The smaller the SSD value is, the higher is the confidence of the matching. For comparison, the parameter at three scales is ploted in one figure, see Figure 5.

After the computation from the above steps, each pixel corresponds to two parameters, its disparity value, *d(n,x)*, and its SSD value, SSD *(n,x)*. The most intuitive way is to choose the right scale *N* for each pixel disparity so that at that scale its SSD value is a minimum of all the scales. That is, if $N = \min_{n} \{ssd(n, x)\}$, then *d(x)=d(N,x)*.

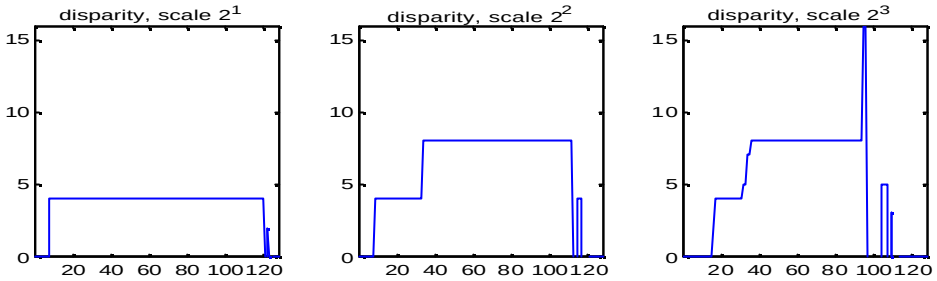

Figure 4 Disparity at three scales

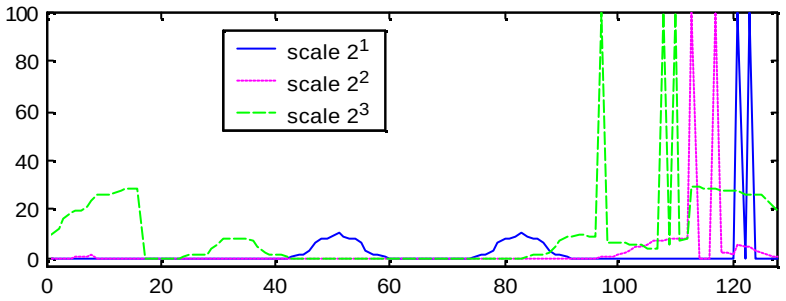

Figure 5 SSD at three scales

One of the big problems of SSD matching is its noise characteristics. Figure 6(a) shows the computational result using above method. It can be seen that the disparity value at the right end, i.e. around pixel 120 is not very good. This is the general case when the same program is tested with some other more complicated image pairs, which show worse results at some points. In some cases, for example, the points at the border of the image may not have matches, or images corrupted with noise give rise to unstable fluctuating results. Thus noise reduction is needed.

The noise problem can be dealt with using one of two possible methods, both of which use an additional matching constraint to remove the points with higher SSD value than a threshold. Hard thresholding and soft thresholding [18] are employed in the two methods, respectively. The standard deviation is adopted as a soft threshold in this paper.

Figure 6(b) gives the computational results using soft thresholding, which shows sharp edges and stable disparity values.

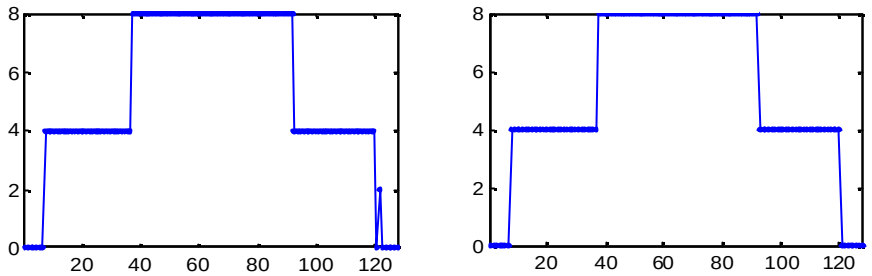

Figure 6 Computed disparity, (a) without threshold, *(b)* soft threshold

4 Experimental Results with Images

For the initial test, the commonly used random dot stereograms [19] are constructed. Figure 7 shows the synthesised *'Squares'* images of size 128*128. The central two squares are right shifted by 4 and 8 pixels, respectively, between the two images.

The image matching is performed along the theoretical epipolar lines of the images. The disparity map and the depth map for *Squares* are given in Figure 8.

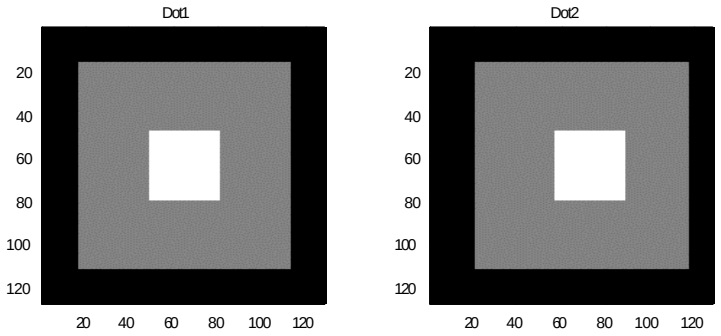

Figure 7 Stereo pair: *Squares*

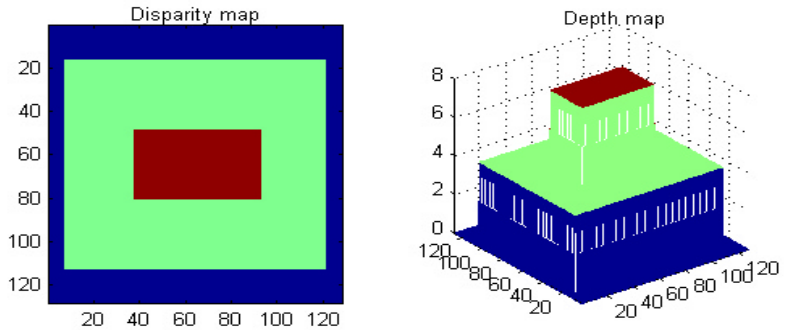

Figure 8 Disparity map and depth map: *Squares*

To increase the complexity of the image features, another pair of random dot stereograms, *Random Square* showed in Figure 9 , is tested. In contrast with the *Squares* images, the pixels in *Random Squares* are random values between *0* and *1*. In Figure 10, the left figure gives the ground truth disparity map and right shows the computational results using the above method.

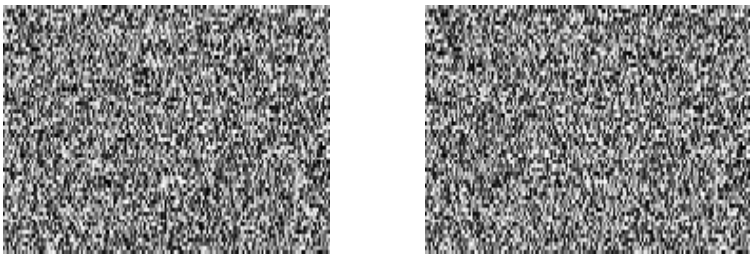

Figure 9 Stereo pair 2: *Random Squares*

Figure 10 A comparison of the ground truth data and computed data (I)
Left: ground truth disparity map, Right: disparity map using wavelets

Computation with real image pairs is also tested. One imagery popularly used is shown in Figure 11, which can be downloaded from the web site, http://www.research.microsoft.com/~szeleski/stereo. The ground truth and estimated disparity maps using dyadic wavelet transform are displayed in Figure 12.

Figure 11 Stereo pair 3: Tsukuba images

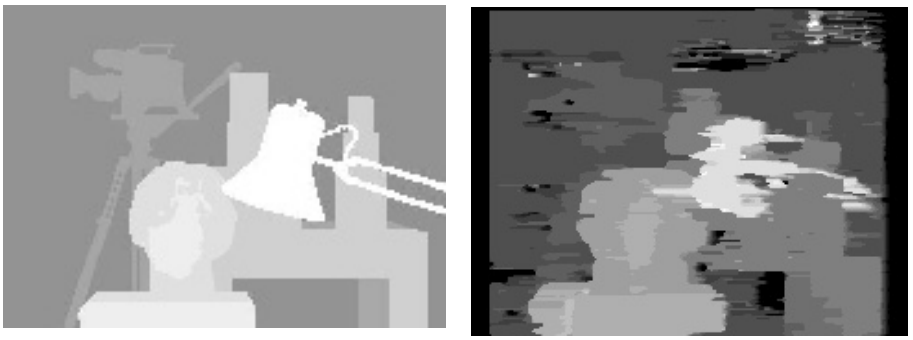

Figure 12 A comparison of the ground truth data and computed data (II)
Left: ground truth disparity, right: estimated disparity result using wavelets

Another pair of real images was taken in the authors' laboratory and used in [20] is shown in Figure 13, the right figure of which gives the computed disparity map.

Figure 13 Real stereo pairs 4 and the disparity map

5 Conclusion and Future Work

This paper has presented a wavelet approach to computing disparity maps. It is of vital importance to apply shift-invariant wavelet transforms when using wavelet techniques for stereo matching. As an initial approach to the application of wavelet transforms to stereo matching, the dyadic wavelet transform was used to develop a matching algorithm. The sum of squared difference is defined based on the values of dyadic wavelet transform coefficients. The experimental results give rise to promising disparity maps. This demonstrates the viability of applying the wavelet transform approach to stereo matching.

However, a better compromise between the algorithmic efficiency and information redundancy could be made because the translation parameter of the dyadic wavelet transform remains continuous. Further investigation of other wavelet transform techniques such as wavelet zero-crossings and dual-tree complex wavelet transforms applied to the matching problem is therefore being carried out. And the comparison of wavelet-based algorithms with standard matching algorithms will be made in the future work.

6 References

[1] Trucco, E. and Verri, A., *Introductory Techniques for 3-D Computer Vision*, Prentice Hall, 1998.

[2] Rosenfeld, A. and Thurston, M., Coarse-fine Template Matching, *IEEE Trans. System, Man, and Cybernetics*, vol. 7, pp. 104-107, 1977.

[3] Marr, D. and Poggio, T., A Computational Theory of Human Stereo Vision, *Proc. R. Soc. Lond.*, vol. 204, pp. 301-328, 1979.

[4] Grimson, W., A Computer Implementation of a Theory of Human Stereo Vision, *Phil. Trans. Royal Soc. London*, vol. V292, pp. 217-253, 1981.

[5] Grimson, W. E. L., Computational Experiments with a Feature-Based Stereo Algorithm, *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 7, pp. 17-34, 1985.

[6] Burt, P. and Adelson, E. H., The Laplacian Pyramid as a Compact Image Code, *IEEE Trans. Communications*, vol. 31 , pp. 532-540, 1983.

[7] Mallat, S., A Theory for Multiresolution Signal Decomposition: the Wavelet Representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 11, pp. 674-693, 1989.

[8] Chui, C. K., *Wavelets: A Tutorial in Theory and Applications*, Academic Press, 1992.

[9] Daubechies, I., Orthonormal Bases of Compactly Supported Wavelets, *Communications on Pure Applied Mathematics*, vol. 41, pp. 906-996, 1988.

[10] Strang, G. and Nguyen, T., *Wavelets and Filter Banks*, Wellesley-Cambridge Press, Second Ed., 1997.

[11] Mallat, S., Zero-crossings of a Wavelet Transform, *IEEE Transactions on Information Theory*, vol. 37, pp. 1019-1033, 1991.

[12] Simoncell, E. P., Freeman, W. T., Adelson, E. H., and Heeger, D. J., Shiftable Multiscale Transforms, *IEEE Trans. Information Theory*, vol. 38, pp. P587-607, 1992.

[13] Kingsbury, N. G., The Dual-tree Complex Wavelet Transform: A New Technique for Shift Invariance and Directional Filters, *IEEE Digital Signal Processing Workshop, DSP 98*, Bryce Canyon, pp. Paper no 86, 1998.

[14] Mallat, S., *A Wavelet Tour of Signal Processing*, Academic Press, 1998

[15] Anandan, P., Computing Dense Displacement Fields with Confidence Measures in Scenes Containing Occlusion, *Proceedings DARPA Image Understanding Workshop*, pp. 236-246, 1984.

[16] Faugeras, O., *Three Dimensional Computer Vision: a Geometric Viewpoint*, The MIT Press, 1993.

[17] Marr, D., *Vision*, W. H. Freeman and Company, 1982.

[18] Chambolle, A., Devore, R. A., Lee, N. Y., and Lucier, B. J., Nonlinear Wavelet Image Processing: Variational Problems, Compression, and Noise Removal Through Wavelet Shrinkage, *IEEE Transactions on Image Procesing*, vol. 7, pp. 319-334, 1998.

[19] Julesz, B., *Foundations of Cyclopean Perception*, University of Chicago Press, 1971.

[20] Rothwell Hughes, N., *Fuzzy Filters for Depth Map Smoothing*, PhD Thesis, University of Wales, 1999.