A Wavelet-based Approach to Image Feature Stability Assessment

Antonio Robles-Kelly[1,2] and Roland Goecke[1,2]
[1]National ICT Australia, Canberra, Australia
[2]Dept. of Information Engineering, RSISE, Australian National University, Canberra, Australia
Email: {antonio.robles-kelly, roland.goecke}@nicta.com.au

Abstract

In this paper, we present a novel method for assessing image-feature stability. The method hinges on applying the discrete wavelet transform to the image features under study throughout a number of video frames in an image sequence. For purposes of stability assessment, we recover the image-feature vectors for each video frame and then track them trough a series of consecutive frames in the image sequence. We apply the discrete wavelet transform to the time series constructed from the pairwise Euclidean distances for each of the image features under study and use the wavelet transform coefficients to assess their stability. We then recover the stable features by clustering together those time series which exhibit largely constant low-pass wavelet coefficients. We present results of the stability analysis for Harris corners, Maximally Stable Extremal Regions, and Scale Invariant Feature Transform regions extracted from two real-world video sequences. We also elaborate on the applications of our method to indexing, retrieval, and compression of stable image feature vectors.

1. Introduction

One of the central issues in many computer vision problems is the extraction of descriptive local image features (or interest points) which can then be used to represent the information in the image in a compact way in applications such as object recognition or image retrieval. For the image features (or local descriptors) to be useful in such applications, they must be invariant to affine transformations and changes in the viewing conditions. Furthermore, they need to be distinctive for different objects. Many algorithms have been suggested in recent years, for example, Harris corners [9], scale invariant feature transform (SIFT) regions [14], maximally stable extremal regions (MSER) [16], affine invariant interest points [17], visual salency [12], or local grayvalue invariants [22]. Such local image features have been applied to object tracking, object recognition, and image retrieval, *e.g.* [12, 21]. Several authors have evaluated the performance of such features, *e.g.* [8, 18, 23]. Common to all these features is that they work in the spatial domain of single images, i.e. they do not consdier spatio-temporal relationships which occur in video sequences. To include such relationships, [13] has proposed the space-time interest points which extend Harris corners to 3D using a scale-space representation. Another extension to spatio-temporal features has been provided by [7] who build cuboids from spatio-temporally windowed pixel values, which are then clustered into prototypes that are used during recognition.

In this paper, we present a novel method for the assessment of image feature stability which takes into account spatio-temporal relationships. Further, the image features assessed as stable by the algorithm are used to represent the same part of an object over a number of consecutive frames. The method is based on a wavelet analysis of the pairwise Euclidean distance between image feature vectors. It turns out that the wavelet transform (WT) coefficients offer a way to cluster image feature vectors with an akin behaviour over time. Maximising WT coefficient similarity, a normalised graph cut is used to separate stable from unstable image features. If the WT coefficients in a cluster are largely low-frequency components, the image feature vectors exhibit only small variation over time, which means that the image features are stable. If the coefficients show largely high-frequency components, the image features are unstable. We perform stability analyses for Harris corners, MSERs, and SIFT regions extracted from two real-world video sequences. We also outline applications to indexing, retrieval and compression of stable image feature vectors.

As mentioned earlier, throughout the paper, we use the WT. The WT has been shown to be a very useful tool in a number of signal processing application areas, such as multiresolution analysis and denoising in image processing, image and video compression, and speech analysis, *e.g.* [11, 24]. In the computer vision community, image processing and filtering techniques akin to wavelet analysis can be found as early as the early 1980s [4, 5]. The WT bears similarities with the Fourier transform (FT) in the sense that both analyze the frequency spectrum of a signal. The main

Proceedings of the 2006 Conference on Computer Vision and Pattern Recognition Workshop (CVPRW'06)
0-7695-2646-2/06 $20.00 © 2006 IEEE

difference between them resides in the fact that the FT assumes stationarity of the time series. The WT, in contrast, can handle non-stationary time series and capture both, frequency and time information [6, 20]. The WT exists in both a continuous as well as a discrete form. For practical applications with sample values at discrete time points, generally the Discrete Wavelet Transform (DWT) is used [15].

The remainder of this paper is structured as follows. Section 2 provides an overview of wavelet analysis in general and the DWT in particular, as the latter is used in this work. In Section 3, we present the application of wavelet analysis to assessing image feature stability in video sequences. The experimental results are presented and discussed in Section 4, where we also discuss applications for the feature stability analysis. Finally, conclusions are drawn in Section 5.

2. Wavelet Analysis

In this section, before explaining in detail how we can apply wavelet analysis to image feature stability in video sequences, we introduce the necessary signal processing background on wavelets. We commence, in Section 2.1, by motivating the use of the wavelet analysis presented here and setting up the background for the discrete domain treatment pursued thereafter. We then introduce the discrete wavelet transform in Section 2.2.

2.1. A General Overview

Various techniques for the analysis of time series exist, the FT being one frequently used. In many application areas it is common for the signal to represent a (continuous or discrete) function over time, which intuitively leads to a time-amplitude representation. However, such a representation is not always the best representation for the analysis because it is often the case that the most distinguished information is contained in the frequency domain. The FT transforms a time-amplitude representation into a frequency representation [19]. In this frequency representation, only the frequency information is preserved, while all time information is lost. If one is only interested in the frequencies which occur in the signal, then the FT is sufficient. However, if one also wants to know when a frequency occurred in a signal, an alternative to the FT must be used.

One way of overcoming this shortfall is to use a short time window [19], treating the time series as stationary for the duration of the time window, and to assign the frequency occurrence during the time window to some point in the time window, *e.g.* the first time point. However, for many practical applications, the assumption of the time series being stationary throughout the time window does not hold.

The choice of an appropriate time window is therefore very important. From Heisenberg's uncertainty principle – applied to signal processing – it can be deducted that it is impossible to know the exact frequency and the exact time of occurrence of this frequency in the time series [10]. The WT offers an approximate solution of arbitrary accuracy by using a fully scalable window, which is shifted along the time axis and the frequency spectrum is calculated for every position. This process is then repeated multiple times with a changing window size, giving rise to a multiresolution analysis. The result of the WT is a time-scale representation.

For completeness, let us first briefly look at the case of the continuous wavelet transform (CWT)

$$CWT_f^{\psi}(\tau, s) = \frac{1}{\sqrt{|s|}} \int f(t)\ \psi^*\left(\frac{t-\tau}{s}\right)\ dt \quad (1)$$

where $f(t)$ denotes the time series function, ψ^* the wavelet function, t the time, s the scale, and τ the translation of the wavelet. It is important to understand that ψ^* describes a family of functions derived from one basic wavelet function $\psi(t)$ – the mother wavelet – by scaling and translation

$$\psi_{\tau,s}(t) = \frac{1}{\sqrt{|s|}}\ \psi\left(\frac{t-\tau}{s}\right) \quad . \quad (2)$$

Hence, the wavelet analysis is self-similar at all scales and thus does not privilege any particular scale. It can be shown that $\psi(t)$ can be any band-pass function of finite energy and the scheme holds [25]. As a result, the CWT acts like a band-pass filter.

The basic wavelet function $\psi(t)$ and discrete time-scale parameters s, τ can be chosen such that the wavelets form an orthonormal basis [6, 20]. This leads us to the discrete case, which shall be looked at in more detail in the next section, as the remainder of the paper assumes a DWT.

2.2. The Discrete Wavelet Transform

The CWT is not practical to use because (a) continuously shifting a continuously scalable function over a signal leads to a large amount of redundancy in the wavelet coefficients, (b) the number of wavelets is infinite, and (c) for most functions the CWT has no analytical solution and can only be computed numerically. The DWT overcomes these problems by firstly choosing discrete wavelets, which only allow discrete translation and scale steps, and secondly by the multiresolution formulation

$$f(2^s t) = \sum_{\tau} \psi_{s+1}(\tau)\ f(2^{s+1}t - \tau) \quad . \quad (3)$$

The new time series $f(2^s t)$ has half the bandwidth of the previous one $f(2^{s+1}t - \tau)$ but twice the frequency.

To avoid having to use an infinite number of wavelets to cover the entire spectrum, a scaling function $\phi(t)$ is introduced, which covers the frequency spectrum otherwise taken care of by wavelets up to scale s

$$\phi(t) = \sum_{s,\tau} \gamma(s,\tau)\ \psi_{s,\tau}(t) \quad . \quad (4)$$

The scaling function $\phi(t)$ is also known as an averaging filter, as it is essentially a low-pass filter.

A time series is then analysed by using a combination of the scaling function, acting as a low-pass filter, and the wavelets, which act as a high-pass filter [15]. An iterative filter bank can be built by using a series of wavelets at different scales together with a scaling function, which provides a simple way of computing the DWT. The outputs of the different filter bank stages are the wavelet and scaling function coefficients. The filter bank iteratively splits the signal into a high-pass and a low-pass part. The former contains the high-frequency part of the time series, or noise, and is commonly labelled as *detail*. The low-pass part, containing the low-frequency part of the time series at the current scale, is the approximation of the time series once the high-frequency information has been removed. Generally, and also in the image feature stability analysis in this paper, it is this noise-removed approximation that is of most interest.

The multiresolution formulation for the scaling function can now be written as

$$\psi(2^s t) = \sum_{\tau} \lambda_{s+1}(\tau)\ \phi(2^{s+1}t - \tau) \tag{5}$$

where $\psi(2^s t)$ is the wavelet at scale s and $\phi(2^{s+1}t - \tau)$ is the scaling function at scale $s+1$. The time series $f(t)$ can then be expressed as a sum of the scaling function and the wavelets

$$f(t) = \sum_{\tau} \lambda_s(\tau)\ \phi(2^s t - \tau)\ +\ \sum_{\tau} \gamma_s(\tau)\ \psi(2^s t - \tau) \tag{6}$$

with $\lambda_s(\tau)$ and $\gamma_s(\tau)$ being the coefficients of the scaling function $\phi(2^s t - \tau)$ and wavelets $\psi(2^s t - \tau)$, respectively, at scale s. Since the scaling function and the wavelets form an orthonormal basis [15], the coefficients are found by taking the inner products $\langle \cdot \rangle$ of the time series and the scaling function and wavelets, respectively,

$$\lambda_s(\tau) = \langle f(t),\ \phi_{s+1,\tau}(t) \rangle \tag{7}$$

$$\gamma_s(\tau) = \langle f(t),\ \psi_{s+1,\tau}(t) \rangle \quad . \tag{8}$$

Without going into further detail (see, for example, [15] for further details), these equations can be rewritten using suitably scaled versions of equations 3 and 5 as

$$\lambda_s(\tau) = \sum_{t} \phi(t - 2\tau)\ \lambda_{s+1}(t) \tag{9}$$

$$\gamma_s(\tau) = \sum_{t} \psi(t - 2\tau)\ \gamma_{s+1}(t) \tag{10}$$

where $\phi(t - 2\tau)$ is the (low-pass) scaling or approximation function, $\psi(t-2\tau)$ is the (high-pass) wavelet or detail function. These two equations state that the coefficients $\lambda(\tau)$ and $\gamma(\tau)$ at a certain scale can be determined by calculating a weighted sum of the coefficients at the previous scale. It is these coefficients that we use in the image feature stability analysis. Details are given in the next Section.

3. Feature Stability in the Wavelet Domain

As mentioned earlier, we aim to assess the stability over time of a set of feature vectors making use of wavelet analysis. The rationale here is that stable features will have a low-pass wavelet coefficient set which is largely constant. So far, we have taken this property for granted and presented an overview of the machinery used to model the behaviour of the time series corresponding to the image feature under study. Now, we turn our attention to the formal relationship between the image feature vectors and the wavelet coefficients. We also exploit this relationship for purposes of stable feature recovery.

Thus, in this section, our aims are twofold. Firstly, we make use of the equations in the previous section to provide a direct link between the image feature vector under study, its time series and the wavelet coefficients. Secondly, we introduce the use of graph cuts for the purpose of separating the stable features from the unstable ones.

3.1. Feature Stability Analysis

To provide a link between the behaviour of the image features over a set of consecutive video frames, we employ the DWT introduced in Section 2.2. In particular, Equations 7 to 10 are used in the stability analysis.

Suppose we have a sequence of N consecutive images (frames) $S = \{I_1, I_2, \ldots, I_i, \ldots, I_N\}$ from a video sequence. Each image is described by its feature vectors. For r consecutive images from S, a 1D DWT is performed on the pairwise Euclidean distances between a pair of image feature vectors $\alpha(t)$ and $\alpha(t+1)$ corresponding to the feature vector α at two consecutive frames in S indexed to the time t. Hence, we have

$$f(t) = \langle \alpha(t) - \alpha(t+1),\ \alpha(t) - \alpha(t+1) \rangle \quad . \tag{11}$$

As a result of the DWT, we get the wavelet and scaling function coefficients $\gamma_s(\tau)$ and $\lambda_s(\tau)$, respectively. We use the $\lambda_s(\tau)$, as they describe the noise-removed time series. Stable features show a smaller amount of noise than unstable features. We can write

$$\begin{aligned} \lambda_s(\tau) &= \langle f(t),\ \phi_{s+1,\tau}(t) \rangle \\ &= \langle \langle \alpha(t) - \alpha(t+1),\ \alpha(t) - \alpha(t+1) \rangle,\ \phi_{s+1,\tau}(t) \rangle \ . \end{aligned} \tag{12}$$

From this equation, we can derive

$$f(t) = \frac{\langle \langle \alpha(t) - \alpha(t+1),\ \alpha(t) - \alpha(t+1) \rangle,\ \lambda_s(\tau) \rangle}{\phi_{s+1,\tau}(t)} \tag{13}$$

which gives us a relationship between the image features and the wavelet analysis. We use this relationship in the following section for purposes of separating the stable features from the unstable ones.

3.2. Stability-based Feature Separation via Graph Cuts

As mentioned earlier, our aim of computation is to assess the stability of the time series $f_i(t)$ for the i^{th} image feature vector throughout the consecutive frames S under study. This, ultimately, implies classifying image features into stable and unstable ones. Due to the lack of training data or ground truth to start with, we cast this problem as an unsupervised learning one which we aim to solve making use of the normalised cut [26].

Recall that Shi and Malik [26] have posed the problem of pairwise clustering as that of finding the optimal partitioning of a weighted graph $G = (E, V)$ by recovering the graph cut whose cost is normalised by the sum of total edge connections of all nodes in the graph. The clustering problem, hence, becomes that of finding the partition that minimises the 'normalised' cost given by

$$\text{NCut}(B_1, B_2) = \frac{\text{Cut}(B_1, B_2)}{\text{Assoc}(B_1, V)} + \frac{\text{Cut}(B_1, B_2)}{\text{Assoc}(B_2, V)} \quad (14)$$

where B_1 and B_2 are two disjoint, connected clusters such that $V = B_1 \cup B_2$, $\text{Cut}(B_1, B_2)$ is the cost of the cut and $\text{Assoc}(B_h, V)$ is the association for the cluster indexed h.

In their mathematical analysis, Shi and Malik pose the problem as a generalised eigensystem and arrive at an expression that is reminiscent of the Rayleigh quotient. They show that the optimal solution to the normalised cut problem is given by the second smallest eigenvector (i.e. the eigenvector corresponding to the second smallest eigenvalue) of the matrix $C = D^{-\frac{1}{2}}(D - A)D^{-\frac{1}{2}}$, where $D = \text{diag}(\deg(1), \deg(2), \ldots, \deg(\mid V \mid))$ is a diagonal matrix and $\deg(i) = \sum_{j=1}^{|V|} A(i, j)$ is the i^{th} row-degree of the adjacency matrix A.

Thus, to make use of the normalised cut, we first require a graph-based representation of the similarity between time series so as to capture their stability, as reflected by the DWT coefficients. As mentioned earlier, we consider the feature i to be stable if its time series $f_i(t)$ exhibits small variation over time, i.e. its DWT low-pass coefficients $\lambda_s(\tau)$ are largely constant. Based on this characterisation of the stability of $f_i(t)$, we abstract the clustering task to an all-connected graph G whose adjacency matrix entries $A(i, j)$ are given by

$$A(i, j) = \begin{cases} \exp(-\kappa \parallel \Lambda_i - \Lambda_j \parallel) & \text{if } i \neq j \\ 0 & \text{otherwise} \end{cases} \quad (15)$$

where κ is a constant and Λ is the matrix of DWT coefficients whose entry indexed i and j is $\Lambda(i, j) = \lambda_j(i)$.

Defined in this manner, the adjacency matrix entries $A(i, j)$ can be viewed as a similarity measure between DWT coefficients for every pair of image feature vectors indexed i and j. As a result, if the DWT coefficients for two image feature vectors are close to one another, the corresponding entry of the adjacency matrix is close to one. If they are far apart from one another, their adjacency matrix entry is close to zero. Furthermore, we can write the cost of the cut for two clusters B_1 and B_2, $V = B_1 \cup B_2$, making use of the adjacency matrix entries and Equation 13 as follows

$$\begin{aligned} \text{Cut}(B_1, B_2) &= \sum_{i \in B_1; j \in B_2} A(i, j) \\ &= \sum_{i \in B_1; j \in B_2} \exp\Bigg(- \kappa \sum_{s+1, \tau} \big(\langle f_i(t), \phi_{s+1,\tau}(t) \rangle \\ &\qquad - \langle f_j(t), \phi_{s+1,\tau}(t) \rangle \big)^2 \Bigg) \quad . \end{aligned} \quad (16)$$

Similarly, we can express the association as

$$\begin{aligned} \text{Assoc}(B_h, V) &= \sum_{i \in B_h; j \in V} A(i, j) \\ &= \sum_{i \in B_h; j \in V} \exp\Bigg(- \kappa \sum_{s+1, \tau} \big(\langle f_i(t), \phi_{s+1,\tau}(t) \rangle \\ &\qquad - \langle f_j(t), \phi_{s+1,\tau}(t) \rangle \big)^2 \Bigg) \quad . \end{aligned} \quad (17)$$

From the equations above, we can conclude that, in maximising the DWT coefficient similarity between groups, the normalised cut is encouraging the formation of clusters of feature vectors with an akin behaviour over time. At the same time, the normalised cut is minimising inter-cluster proximity. This bipartition of the image feature vectors can be viewed as a separation between those features whose low-pass wavelet coefficients are large and those that exhibit a dominant high-frequency wavelet component.

This observation is important since it allows us to use the average norm for the matrices of DWT coefficients for the image features in each cluster to determine whether B_1 or B_2 corresponds to those feature vectors which are stable. The average norm for the matrices of wavelet coefficients corresponding to the image features in the cluster indexed h is given by

$$\beta(B_h) = \frac{1}{\mid B_h \mid} \sum_{i \in B_h} \parallel \Lambda_i \parallel \quad . \quad (18)$$

If the average norm is large, then the coefficients for the image features in the cluster are largely low-pass and, hence, the feature vectors have little variation over time and are stable. If the quantity above is small, then the variation is large and the image feature vectors are unstable. As a result, we can separate the stable features from the unstable ones by performing a normalised cut over the set of image feature vectors and, once the two clusters are at hand, we select those features in the cluster B^* such that

$$B^* = \left\{ B_i \mid \beta(B_i) = \max_{h=\{1,2\}} \big(\beta(B_h) \big) \right\} \quad . \quad (19)$$

Figure 1. Stable features on sample frames of the PETS 2000 (left) and PETS 2001 (right) sequences. From top-to-bottom: Harris corners, MSERs and SIFT regions.

4. Experiments and Applications

In this section, we present results on two real-world video sequences and elaborate on the applications of our stability assessment method. To do this, we commence by providing results on the separation between stable and unstable image feature vectors. We then explore the applications of the method to image feature compression and video indexing and retrieval.

As experimental vehicles, we use two 100-frame fragments of video sequences from the PETS data sets [1]. The first of these is a fragment of the outdoor and vehicle tracking sequence from PETS 2000. The second of these is a section of the first data set of PETS 2001. For

all our experiments, we have recovered the image features and, once these are at hand, matched them over 15 consecutive frames using the KD-Tree relational matching algorithm [3]. Throughout the section, we make use of three alternatives for the image feature vectors whose stability is being assessed. These are the Harris corner detector, the MSERs, and the SIFT regions. Also, in all our experiments, we employ Daubechies wavelets.

4.1. Recovery of Stable Image Feature Vectors

We first present results on stable image feature vector separation. As mentioned earlier, we use 100 frames of the video sequences and apply our wavelet-based approach to the image features in every frame along a 15-frame window (≈ 0.62 seconds). To compute the DWT coefficients corresponding to the features in the frame i, we use all frames between the frames $i-7$ and $i+7$. After recovering the time series for the features, we compute four DWT coefficients using a single scale. Thus, in our experiments, the DWT coefficients reflect the behaviour of the image features at frame i throughout a time window of 15 frames.

In Figure 1, we show, from top-to-bottom, the stable features for a sample frame of the PETS 2000 (left panels) and the PETS 2001 (right panels) sequences, respectively, when Harris corners, MSERs and SIFT regions are used as feature vectors. In the figure, we use a red 'dot' to denote the point at which the image feature is in the sample frame. The arrow shows the trajectory on the image plane of the feature over time. The tail of the arrow is at the point at which the image feature was at frame $i-7$. The head denotes the point at which the feature will be at frame $i+7$. For the sake of clarity, we only plot the dots and the arrows for the Harris corners and the SIFT regions, whereas for the MSERs, we also fit ellipses to the maximally stable regions.

From Figure 1, we can conclude that the features classified as stable by the algorithm are in good accordance with the overall structure of the motion in the scene. For instance, in the PETS 2000 sequence, the arrows on the red car capture well the nature of its forward motion, towards the parking at the cul-de-sac. In the PETS 2001 sequence, the MSERs also capture both, the motion of the green car and the stationary background regions.

Next, we perform a more quantitative analysis on the quality of the separation between stable and unstable features. Here, we plot the average norm for the matrices of wavelet coefficients for both clusters, i.e. $\beta(B_h)$, $h = \{1, 2\}$, as a function of frame index. In Figure 2, we show, from left-to-right, the plots for the PETS 2000 (top row) and PETS 2001 (bottom row) sequences when Harris corners, MSERs and SIFT regions are used as feature vectors.

From the plots, we note that, in general, the traces for the mean wavelet coefficient norms corresponding to the two clusters B_1 and B_2, i.e. the cluster of stable and unstable features, are well separated. In all the panels, the upper trace corresponds to the mean norm of the stable features.

4.2. Feature Vector Compression

As mentioned earlier, there is a wide variety of applications for our algorithm. One of the main consequences of the treatment of the image feature vectors as a time series is that, modelled in this way, the feature vectors in the image frame can be compressed making use of the wavelet coefficients. This is due to the fact that, given the feature vector at the frame indexed i and its wavelet coefficients, we can recover the time series $f_i(t)$ making use of Equation 13. With the time series at hand, we can recover the feature vectors solving a system of linear equations governed by the time series $f_i(t)$ and the image feature vector at the frame i.

As a result of this treatment, we only need to store a single feature vector and the wavelet coefficients for every stable features in the scene. Further, in the case of the SIFT regions, we view the 128-order vector of the image feature at the frame of reference as a one-dimensional function that can be decomposed using a DWT and store it in its wavelet coefficient form. In our experiments, this yields a compression rate that is on average 5 for the Harris corners, 8.33 for the MSERs, and 32 for the SIFT regions.

4.3. Video Indexing and Retrieval

Another application is video indexing and retrieval. The fact that our method captures the behaviour of the features in the scene over a given amount of time makes it ideal for tasks aiming at indexing or retrieving frames or regions of interest. Furthermore, its utility for compression allows to recover 'compressed' descriptors that can then be stored and indexed to time along the video sequence.

Once the feature vectors for the regions of interest (ROI) have been recovered, we separate the stable from the unstable ones and compress them. These compressed feature vectors are then descriptors which are indexed to the ROI to whom they belong. Since the stability assessment captures the behaviour of the feature vectors over time, we only store one descriptor every seven frames. With the wavelet representation of the time-subsampled image descriptors at hand, we follow Ancona *et al*. [2] and perform indexing and retrieval based on a tree search over the cross-correlation between the coefficients of the wavelet decomposition of the descriptors corresponding to the ROIs in the database.

Thus, the indexing is not only efficient in terms of description length, but also the retrieval is faster. This is a result of not having to search through the whole sequence, but only on the compressed, time-subsampled descriptors. In Figure 3, we show a sample frame for the PETS 2000 sequence in the top row and for the PETS 2001 sequence in the bottom row. The background has been subtracted in both sequences. In the top, left-hand panel, Harris corners

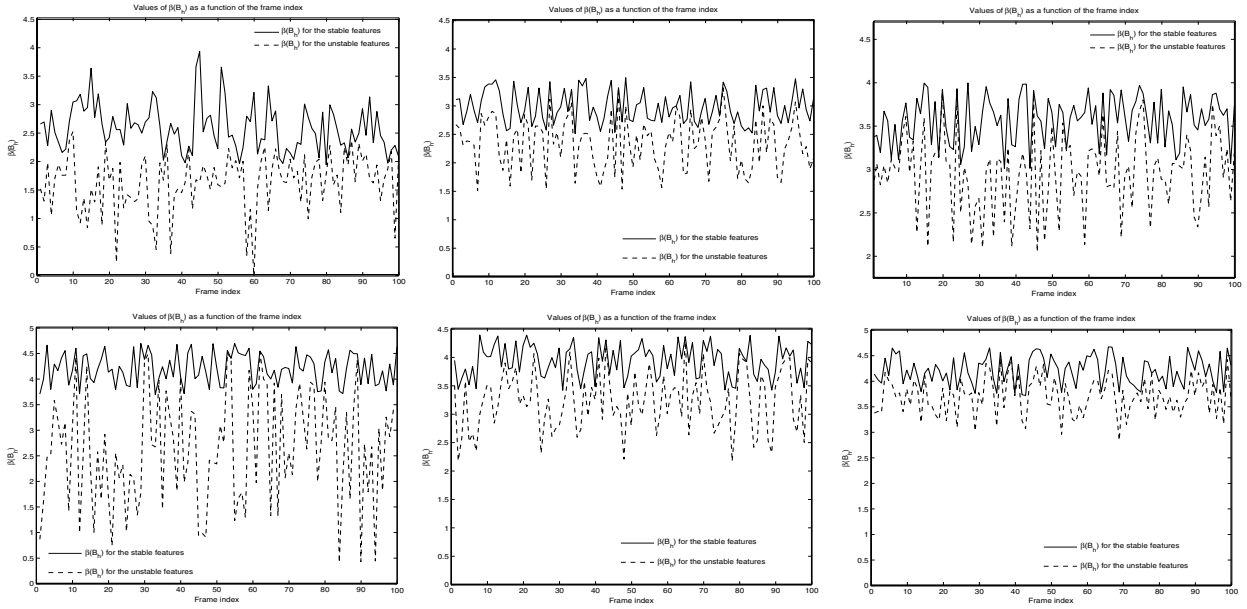

Figure 2. Mean wavelet coefficient norms $\beta(B_h)$, $h = \{1, 2\}$, for 100 frames of the PETS 2000 (top) and PETS 2001 (bottom) sequences. From left-to-right: Plots for the Harris corners, MSERs and SIFT regions.

are used as image features. In the top, right-hand panel, we show the results yield by the algorithm when MSERs are used. In the bottom, left-hand panel, Harris corners are used again and in the bottom, right-hand panel, SIFT regions are used. It is worth noting that none of the stable features are located on the spurious regions. This is in good accordance with the notion that, for the regions of interest, the features should remain stable over time.

5. Conclusions

In this paper, we have cast the problem of assessing image feature stability as that of characterising the behaviour over time of the feature vectors via a discrete wavelet transform. We have also shown how the normalised cut can be used for separating stable features from unstable ones in video sequences. The formulation of feature stability presented here has a number of advantages and allows the use of wavelet coefficients to tackle a variety of problems. In this paper, we have presented two applications. The first of these concerns the utility of the wavelet analysis for compressing stable features. The second one, which is related to the first one, employs the compressed features for indexing and retrieving regions of interest in video sequences.

6. Acknowledgement

National ICT Australia is funded by the Australian Government's *Backing Australia's Ability* initiative, in part through the Australian Research Council.

References

[1] PETS Datasets. http://ftp.pets.rdg.ac.uk/.

[2] M. Ancona, W. Cazzola, P. Raffo, and M. Corvi. Image database retrieval using wavelet packets compressed data. In *Proc. Sixty SIMAI National Conference*, May 2002.

[3] J. Bentley. Multidimensional Binary Search Trees Used for Associative Searching. *Comm. of the ACM*, 8(9), 1975.

[4] P. Burt and E. Adelson. A Multiresolution Spline with Application to Image Mosaics. *ACM Transactions on Graphics*, 2(4):217–236, 1983.

[5] J. Crowley and A. Parker. A Representation for Shape Based on Peaks and Ridges in the Difference of Low-pass Transform. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 6(2):156–170, 1984.

[6] I. Daubechies. The Wavelet Transform, Time-Frequency Localization and Signal Analysis. *IEEE Transactions on Information Theory*, 36(5):961–1005, Sept. 1990.

[7] P. Dollár, V. Rabaud, G. Cottrell, and S. Belongie. Behaviour Recognition via Sparse Spatio-Temporal Features. In *Proc. IEEE Int. Workshop Vis. Surveill. and Perf. Eval. of Tracking and Surveill. VS-PETS 2005*, Beijing, China, Oct. 2005.

[8] F. Furesjö and H. Christensen. Evaluation of the influence of feature detectors and photometric desriptors in object recognition. Technical Report TRITA-NA-P0406, KTH Dept. of Numerical Analysis and Computer Science, 2004.

[9] C. Harris and M. Stephens. A combined corner and edge detector. In *Proceedings of the 4th Alvey Vision Conference*, pages 147–151, Manchester, UK, 1988.

[10] W. Heisenberg. Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik. *Zeitschrift für Physik*, 43:172–198, 1927.

Figure 3. Top: Stable Harris corners (left) and MSERs (right) on the regions of interest in a sample frame of the PETS 2000 sequence. Bottom: Stable Harris corners (left) and SIFT regions (right) on the regions of interest in a sample frame of the PETS 2001 sequence.

[11] T.-C. Hsung, D. P.-K. Lun, and W.-C. Siu. Denoising by singularity detection. *IEEE Transactions on Signal Processing*, 47(11):3139–3144, Nov. 1999.

[12] T. Kadir and M. Brady. Saliency, Scale and Image Description. *Int. J. Comp. Vis.*, 45(2):83–105, Nov. 2001.

[13] I. Laptev. On Space-Time Interest Points. *International Journal of Computer Vision*, 64(2/3):107–123, Sept. 2005.

[14] D. Lowe. Distinctive Image Features from Scale-Invariant Keypoints. *Int. J. Comp. Vis.*, 60(2):91–110, Nov. 2004.

[15] S. Mallat. A Theory for Multiresolution Signal Decomposition: The Wavelet Representation. *IEEE Trans. PAMI*, 11(7):674–693, 1989.

[16] J. Matas, O. Chum, U. Martin, and T. Pajdla. Robust Wide Baseline Stereo from Maximally Stable Extremal Regions. In D. Marshall and P. Rosin, editors, *Proceedings of the 13th British Machine Vision Conference BMVC2002*, volume 1, pages 384–393, Cardiff, UK, Sept. 2002.

[17] K. Mikolajczyk and C. Schmid. An affine invariant interest point detector. In *Proc. ECCV2002*, volume I, pages 128–142, Copenhagen, Denmark, May 2002. Springer.

[18] K. Mikolajczyk and C. Schmid. A performance evaluation of local descriptors. *IEEE Transactions on Pattern Analysis & Machine Intelligence*, 27(10):1615–1630, Dec. 2005.

[19] A. Oppenheim and R. Schafer. *Discrete-Time Signal Processing*. Prentice Hall Signal Processing Series. Prentice Hall, Upper Saddle (NJ), USA, 2nd edition, 1999.

[20] O. Rioul and M. Vetterli. Wavelets and Signal Processing. *IEEE Signal Proc. Mag.*, 8(4):14–38, Oct. 1991.

[21] F. Schaffalitzky and A. Zisserman. Automated Scene Matching in Movies. In *Proc. CIVR 2002*, volume 2383 of *LNCS*, pages 186–197, London, UK, July 2002. Springer.

[22] C. Schmid and R. Mohr. Local grayvalue invariants for image retrieval. *IEEE Trans. PAMI*, 19(5):530–535, May 1997.

[23] C. Schmid, R. Mohr, and C. Bauckhage. Evaluation of Interest Point Detectors. *International Journal of Computer Vision*, 37(2):151–172, June 2000.

[24] J. Shapiro. Embedded image coding using zerotrees of wavelet coefficients. *IEEE Transactions on Signal Processing*, 41(12):3445–3462, Dec. 1993.

[25] Y. Sheng. Wavelet Transform. In A. Poularikas, editor, *The Transforms and Applications Handbook*, The Electrical Engineering Handbook Series, pages 747–827. CRC Press, Boca Raton (FL), USA, 1996.

[26] J. Shi and J. Malik. Normalized Cuts and Image Segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 22(8):888–905, Aug. 2000.