

International Journal of Wavelets, Multiresolution and Information Processing
© World Scientific Publishing Company

Automatically Detecting “Significant Events” on SenseCam

NA LI* and MARTIN CRANE and HEATHER J. RUSKIN

*School of Computing, Dublin City University
Gasnevin 9, Dublin, Ireland
na.li@dcu.ie
mcrane@computing.dcu.ie
hruskin@computing.dcu.ie*

Received (Day Month Year)
Revised (Day Month Year)
Communicated by (xxxxxxxxxx)

SenseCam is an effective memory-aid device that can automatically record images and other data from the wearer’s whole day. The main issue is that, while SenseCam produces a sizeable collection of images over the time period, the vast quantity of captured data contains a large percentage of routine events, which are of little interest to review. In this article, the aim is to detect “Significant Events” for the wearers. We use several time series analysis methods such as Detrended Fluctuation Analysis (DFA), Eigenvalue dynamics and Wavelet Correlations to analyse the multiple time series generated by the SenseCam. We show that Detrended Fluctuation Analysis exposes a strong long-range correlation relationship in SenseCam collections. Maximum Overlap Discrete Wavelet Transform (MODWT) was used to calculate equal-time Correlation Matrices over different time scales and then explore the granularity of the largest eigenvalue and changes of the ratio of the sub-dominant eigenvalue spectrum dynamics over sliding time windows. By examination of the eigenspectrum, we show that these approaches enable detection of major events in the time SenseCam recording, with MODWT also providing useful insight on details of major events. We suggest that some wavelet scales (e.g. 8 minutes -16 minutes) have the potential to identify distinct events or activities.

Keywords: Lifelogging; SenseCam; equal-time Correlation Matrices; Maximum Overlap Discrete Wavelet Transform

1. Introduction

A SenseCam¹ can be thought of as a “black box flight recorder” for human beings. Developed by Microsoft Research in Cambridge, UK, SenseCam is a small, wearable camera worn around the neck to capture, automatically, images and other sensor readings, in order to document the events of a wearer’s day. Images can be periodically reviewed to refresh and strengthen the wearer’s memory of an event.

Besides a camera, the SenseCam also contains several electronic sensors, including those which record light-intensity and light-colour, a passive infrared (body

*Corresponding author.

heat) detector, a temperature sensor, and a multiple-axis accelerometer for monitoring changes in movement in the x , y , z directions of the wearer's environment.

The device takes pictures at VGA resolution, (480x640 pixels), and stores these as compressed JPEG files on internal flash memory. SenseCam can collect a large amount of data, even over a short period of time, with a picture typically taken every 30 seconds. Hence there are about 4,000 images captured in any one day, or of the order of 1 million images captured per year.

Although research shows that the SenseCam can be an effective memory-aid device^{2,3}, as it helps to improve retention of an experience, wearers seldom wish to review life events by browsing large collections of images manually^{4,5,6,7}. The challenge then is to manage, organise and analyse these large image collections in order to automatically highlight *key episodes* and, ideally, classify them in order of importance to the wearer. Previously, the lifelog of SenseCam images was segmented into approximately 20 distinct events in a wearer's day (about 7,000 events per year)⁸. Nevertheless, this large collection of personal information still contains a significant percentage of routine events. The objective is to determine which events are the most important or unusual to the wearers.

In this paper, several methods are used to analyse the multiple time series recorded by the Microsoft SenseCam, in order to highlight "significant events" in the data. Such important or unusual events should exhibit *atypical* or *non-stationary*^a characteristics.

This paper is organized as follows: in Section 2 we review the methods, in Section 3 we describe the data used, while Section 4 details the results obtained. Conclusions are given in Section 5.

2. Methods

In this section, we first considered the Detrended Fluctuation Analysis (DFA) method to try to detect a *long-range correlation* relationship in SenseCam image time series. Equal-time Cross-Correlation Matrices have also been analysed to characterise dynamical changes in non-stationary multivariate SenseCam time-series. Finally, the Maximum Overlap Discrete Wavelet Transform (MODWT) has been used to calculate equal-time Correlation Matrices over different time scales and to examine the details of the eigenvalue spectrum.

2.1. Time and Scale Hurst exponent

Long-range data correlation is found in a variety of systems, with examples including physiology reading⁹, financial time series for stock market data¹⁰ and others as diverse as DNA sequences and ozone data readings in meteorology^{11,12}. Detrended

^aA non-stationary process is a stochastic process for which the mean, variance and covariance change over time. Non-stationary behaviour can include trends, cycles, random walks or combinations of these.

Fluctuation Analysis (DFA) is suitable for studying long-range exponential correlations in non-stationary time series. The advantage of DFA is that it can remove the different trends from external factors in the data and reduce noise level measurement. The images captured from the SenseCam are determined by multiple sensor data: an accelerometer to detect motion, sensors to detect changes in light levels and so on, thus involving many parameters. The generated image time series are complex and composed of many interacting units. The DFA method is adopted initially in order to analyse these and remove stationary trends, thus helping to highlight non-stationary events, which could be of importance.

2.1.1. Detrended Fluctuation Analysis

To illustrate the DFA algorithm⁹, we choose a unit pixel (1×1) from each image and with numbers of images integrated to give the images pixel time series $X(t)$. Next the time series $X(t)$ is divided into boxes of equal length, n . In each box of length n , a least-squares line $z(t)=at+b$ is fitted to the data representing the trend in that box. The root-mean-square fluctuation of this integrated and detrended time series is calculated, as in⁹:

$$F(n) = \sqrt{\frac{1}{N} \sum_{t=1}^N [X(t) - z(t)]^2}, \quad (2.1)$$

If the sequence $X(t)$ are random, uncorrelated variables or short-range, correlated variables, then, in general, $F(n)$ will obey a power-law scaling behaviour

$$\langle F(n) \rangle \sim n^H \quad (2.2)$$

This gives a relationship between $F(n)$ and the box size n . Typically, $F(n)$ will increase with box size⁹. Under such conditions, the fluctuations can be characterized by a scaling exponent H , the slope of the line relating $\log F(n)$ with $\log n$. The exponent H is called the Hurst exponent. If $H < 0.5$, this indicates *anti-persistent* time series, that is, deviations of one sign are generally followed by deviations with opposite sign. The limiting case $H=0$, corresponds to white noise, where fluctuations at all frequencies are equally present. If $H > 0.5$, this indicates that a *persistent* time series is found, i.e. deviations tend to keep the same sign. The limiting case $H=1$, reflects $X(t) \propto t$, a smooth signal.

2.2. Correlation Dynamics

The DFA method provides the initial background analysis for SenseCam time series data. The equal-time Cross-Correlation Matrices can be formed, and used, to characterise dynamical changes in non-stationary multivariate time-series.

The behaviour of the largest eigenvalue of a cross-correlation matrix over small windows of time, has been studied for financial series^{13,14,15,16,17,18,19,20,21},

electroencephalographic (EEG) recordings²², magnetoencephalographic (MEG) recordings²³ and a variety of other multivariate data. Similar techniques are used here to investigate the dynamics of SenseCam images.

To reduce the size of the calculation further, and thus the amount of memory used, we first adopt an averaging method to decrease image size from 480x640 pixels to 6x8 pixels. Hence the correlation matrix is made up of 48 time series for over 10,260 images. The equal-time cross-correlation matrix, between time series of images, is calculated using a sliding window, where the number of pixels in one image, N , is smaller than the window size T . Given pixels $G_i(t)$, $i=1,\dots,N$, of a collection of images, we normalise G_i within each window in order to standardise the different pixels for the images as follows:

$$g_i(t) = \frac{G_i(t) - \overline{G_i(t)}}{\sigma_{(i)}} \quad (2.3)$$

where $\sigma_{(i)}$ is the standard deviation of G_i for image numbers $i=1,\dots,N$, and $\overline{G_i}$ is the time average of G_i over a time window of size T . Then the equal-time cross-correlation matrix may be expressed in terms of $g_i(t)$

$$C_{ij} \equiv \langle g_i(t)g_j(t) \rangle \quad (2.4)$$

The elements of C_{ij} are limited to the domain $-1 \leq C_{ij} \leq 1$, where $C_{ij} = \pm 1$ defines perfect positive/negative correlation and $C_{ij} = 0$ corresponds to no correlation. In matrix notation, the correlation matrix can be expressed as $C = \frac{1}{T}GG^t$ where t is the transpose of a matrix and G is an $N \times T$ matrix with elements g_{it} .

The eigenvalues λ_i and eigenvectors \bar{v}_i of the correlation matrix C are found from the eigenvalue equation $C\bar{v}_i = \lambda_i\bar{v}_i$.

The eigenvalues are then ordered by size, such that $\lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_N$. Given that the sum of the diagonal elements of a matrix (the Trace) remains constant under linear transformation²², $\sum_i \lambda_i$ must always equal the Trace of the original correlation matrix. Hence, if some eigenvalues increase then others must decrease, to compensate, and vice versa, (a feature known as *Eigenvalue Repulsion*).

There are two limiting cases for the distribution of the eigenvalues: (i) with perfect correlation, $C_i \approx 1$, when the largest is maximised with value N , (all others taking value zero). (ii) when each time series consists of random numbers with average correlation $C_i \approx 0$ and the corresponding eigenvalues are distributed around 1, (where any deviation is due to spurious random correlations). Between these two extremes, the eigenvalues at the lower end of the spectrum can be much smaller than λ_{max} . To study the dynamics of each of the eigenvalues using a sliding window, we normalise each eigenvalue in time using

$$\tilde{\lambda}_i(t) = \frac{(\lambda_i - \bar{\lambda})}{\sigma^\lambda} \quad (2.5)$$

where $\bar{\lambda}$ and σ^λ are the mean and standard deviation of the eigenvalues over a particular reference period. This normalisation allows us to visually compare

eigenvalues at both ends of the spectrum, even if their magnitudes are significantly different. The reference period used to calculate the mean and standard deviation of the eigenvalue spectrum can be chosen to be a low volatility sub-period, (which helps to enhance the visibility of high volatility periods), or the full time-period studied.

2.3. Wavelet Multiscale Analysis

The wavelet transform (WT) is a mathematical tool that can be applied to many applications such as image analysis, meteorology, signal processing and financial time series²⁴ and is used to decompose a signal into different time horizons. For example, wavelets allow us to decompose a signal on a Scale-by-Scale basis, e.g. in measuring the correlation between Equities over different time scales (values at hourly intervals, two hourly intervals etc.). This allows characterization of the impact of the different trading strategies or horizons on correlations between equities. In particular, the discrete wavelet transform (DWT)²⁵ is useful in dividing the data series into components of different frequency, so that each component can be studied separately in order to investigate the data series in depth. In our case, where we wish to compare different pixel time series values, we may do so over a variety of time scales.

2.3.1. MODWT

The Maximum Overlap Discrete Wavelet Transform, (MODWT)²⁵, is a linear filter that transforms a series into coefficients related to variations over a set of scales. Like the DWT it produces a set of time-dependent wavelet and scaling coefficients with basis vectors associated with a location t and a unitless scale $\tau_j=2^{j-1}$ for each decomposition level $j=1,\dots,J_0$. Unlike the DWT, the MODWT, has a high level of redundancy. However, it is *non-orthogonal* and can handle any sample size N , whereas the DWT restricts the sample size to a multiple of 2^j . MODWT retains downsampled^b values at each level of the decomposition that would be discarded by the DWT. This reduces the tendency for larger errors at lower frequencies when calculating frequency dependent variance and correlations, as more data are available.

Decomposing a signal, using the MODWT to J levels, theoretically involves the application of J pairs of filters. The filtering operation at the j^{th} level consists of applying a rescaled father wavelet to yield a set of detail coefficients

$$\tilde{D}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{\varphi}_{j,l} f_{t-l} \quad (2.6)$$

^bDownsampling or decimation of the wavelet coefficients retains half of the number of coefficients that were retained at the previous scale. Downsampling is applied in the Discrete Wavelet Transform

6 *N. Li, M. Crane and H. J. Ruskin*

and a rescaled mother wavelet to yield a set of scaling coefficients

$$\tilde{S}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{\phi}_{j,l} f_{t-l} \quad (2.7)$$

for all times $t = \dots, -1, 0, 1, \dots$, where f is the function to be decomposed²⁶. The rescaled mother, $\tilde{\varphi}_{j,t} = \frac{\varphi_{j,t}}{2^j}$, and father, $\tilde{\phi}_{j,t} = \frac{\phi_{j,t}}{2^j}$, wavelets for the j^{th} level are a set of scale-dependent localised differencing and averaging operators and can be regarded as rescaled versions of the originals. The j^{th} level equivalent filter coefficients have a width $L_j = (2^j - 1)(L - 1) + 1$, where L is the width of the $j = 1$ base filter. In practice the filters for $j > 1$ are not explicitly constructed because the detail and scaling coefficients can be calculated, using an algorithm that involves the $j = 1$ filters operating recurrently on the j^{th} level scaling coefficients, to generate the $j + 1$ level scaling and detail coefficients²⁶. Each of the sets of coefficients in wavelet is called a ‘crystal’.

2.3.2. Wavelet Variance

The wavelet variance $\nu_f^2(\tau_j)$ is defined as the expected value of $\tilde{D}_{j,t}^2$ if we consider only the non-boundary coefficients^c. An unbiased estimator of the wavelet variance is formed by removing all coefficients that are affected by boundary conditions and is given by

$$\nu_f^2(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \tilde{D}_{j,t}^2 \quad (2.8)$$

where $M_j = N - L_j + 1$ is the number of non-boundary coefficients at the j^{th} level²⁶. The wavelet variance decomposes the variance of a process on a scale-by-scale basis (at increasingly higher resolutions of the signal) and allows us to explore how a signal behaves over different time horizons.

2.3.3. Wavelet Covariance and correlation

The wavelet covariance between functions $f(t)$ and $g(t)$ is similarly defined to be the covariance of the wavelet coefficients at a given scale. The unbiased estimator of the wavelet covariance at the j^{th} scale is given by

$$\nu_{fg}(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \tilde{D}_{j,t}^{f(t)} \tilde{D}_{j,t}^{g(t)} \quad (2.9)$$

^cThe MODWT treats the time-series as if it were periodic using “circular boundary conditions”. There are L_j wavelet and scaling coefficients that are influenced by the extension, and which are referred to as the boundary coefficients.

where all the wavelet coefficients affected by the boundary are removed²⁶, and $M_j = N - L_j + 1$.

The MODWT estimate of the wavelet cross-correlation between functions $f(t)$ and $g(t)$ may be calculated using the wavelet covariance and the square root of the wavelet variance of the functions at each scale j . The MODWT estimator, of the wavelet correlation is given by

$$\rho_{fg}(\tau_j) = \frac{\nu_{fg}(\tau_j)}{\nu_f(\tau_j)\nu_g(\tau_j)} \quad (2.10)$$

where, at scale j , $\nu_{fg}(\tau_j)$ is the covariance between $f(t)$ and $g(t)$, $\nu_f(\tau_j)$ is the variance of $f(t)$ and $\nu_g(\tau_j)$ is the variance of $g(t)$.

3. Data

In this study, the data were generated from one person wearing the SenseCam over a six day period, from a Saturday to a Thursday. These particular days were chosen in order to include a weekend, where normal home activity varied in comparison to events on weekdays or a working week. Data statistics are reported in Table 1. To create a *ground truth*^d, the user reviewed her collection and manually marked the boundary image between all events.

3.1. Weekend

Saturday, a typical example to illustrate the difference mentioned above, involved the subject walking to the nearest bus stop from home, a bus journey to the city centre, walking through local streets as well as a visit to a shopping centre. This day also involved dinner with a friend and a bus journey back to the original bus stop.

On the next day, the subject only wore the SenseCam during the afternoon to the office: thus images described the journey of the subject from the home to office, a period spent working in front of the laptop and the return journey back home.

3.2. Weekdays

Over the next four days, these images described a typical day for the subject: sitting in the office, talking with a colleague and sharing lunch in the cafeteria and so on. On some days, the subject wore the SenseCam home, so that it recorded the wearer’s journey from the office to home, and the next morning from home to the office. Overall, the record formed a total lifelong of 10,260 images over the six day period, with average wearing time varying from about 11 hours on Saturday to about 6 hours on Tuesday.

^dIn machine learning, the term “ground truth” refers to the accuracy of the training set’s classification for supervised learning techniques. This is used in statistical models to support or reject research hypotheses.

Table 1. Data Statistics

User	Events Catalogue	Groundtruthed Events	Images
1	Working	15	6146
1	Walking Outside	32	1494
1	Shopping	12	826
1	Eating	3	658
1	Taking Bus	2	297
1	Others	5	839
		Total: 69	Total: 10,260

4. Results

4.1. Detrended Fluctuation Analysis

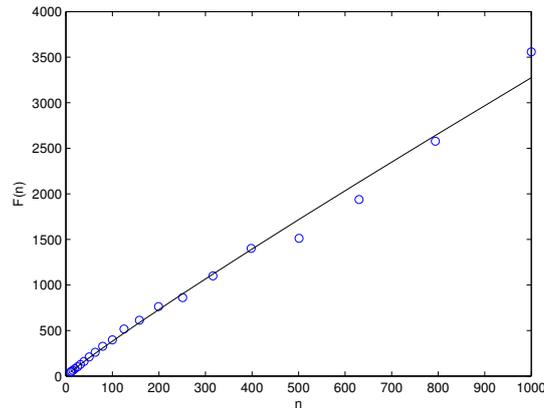


Fig. 1. Plot of $\log F(n)$ vs box size $\log n$ from 10 to 1000 for total 10260 images.

In the plot of $\log F(n)$ vs $\log n$ for different box sizes (Figure 1), the exponent $H=0.93203$ is clearly greater than 0.5, and reflects strong long-range correlation on images from the SenseCam, i.e. indicates that the time series is not a random walk^e, but is cyclical, implying that continuous low levels of background information are picked up constantly by the device. Consequently, the DFA provides a measure of many similar ‘typical’ backgrounds or environments.

^eA random walk is a mathematical formalisation of a path that consists of a succession of random steps.

4.2. Wavelet Analysis

For the present study, we selected the least asymmetric (LA) wavelet, (known as the Symmlet²⁵), which exhibits near symmetry about the filter midpoint. LA filters are defined in even widths and the optimal filter width is dependent on the characteristics of the signal and the length of the data series. The filter width chosen for this study was the LA8, (where 8 refers to the width of the scaling function), since this enables accurate calculation of wavelet correlations to the 10th scale, which is appropriate given the length of data series available. Although the MODWT can accommodate any level, the highest level, J_0 , is chosen in practice so as to prevent decomposition at scales longer than the total length of the data series. Hence the choice of the 10th scale here, which still contains enough detail to capture subtle changes in the signal²⁶.

Before examining the image time series in detail, it is important to introduce the *gray scale pixel values* concept. In a gray scale image, a pixel with a value of 0 is completely black and a pixel with a value of 255 is completely white. The images captured from SenseCam are coloured, but are converted to gray-scale images in order to simplify the calculation.

First, the MODWT of each image’s pixels was calculated within each window and the correlation matrix between pixels at each scale found. The eigenvalues of the correlation matrix in each window were determined, and the eigenvalue time series were normalised over time. Then the Largest Eigenvalue for different window sizes was analysed. With increased window size comes increased smoothing-as expected. This removes some of the high frequency small-scale changes, typically associated with noise. As the window size is increased, the peaks in the series become more pronounced. These peaks reflect large changes in grayscale of the images. Typically, SenseCam captures two images every minute, so we can measure wavelet eigenvalue dynamics in time (minutes). These results as shown in a heat map in Figure 2 and discussed below.

4.2.1. Dynamics of the Largest Eigenvalue at various wavelet scales

In financial data, it has been known for some time that the largest eigenvalue (λ_1) contains information on the risk associated with the particular assets of which the covariance matrix is composed, (i.e. the ‘market’ factor)²⁴. Similarly, in the case of SenseCam images, we would expect the largest eigenvalue to contain information from the image that reflects the largest change in the SenseCam recording.

Figure 2 show the time series of the Largest Eigenvalue dynamics across different wavelet scales. From these, we note the following features:

- At increased scale, (i.e for longer time periods), of the wavelet crystal components, increased smoothing was observed- as expected. Increasing the scale has the effect of removing some of the high frequency small-scale changes, typically associated with noise.

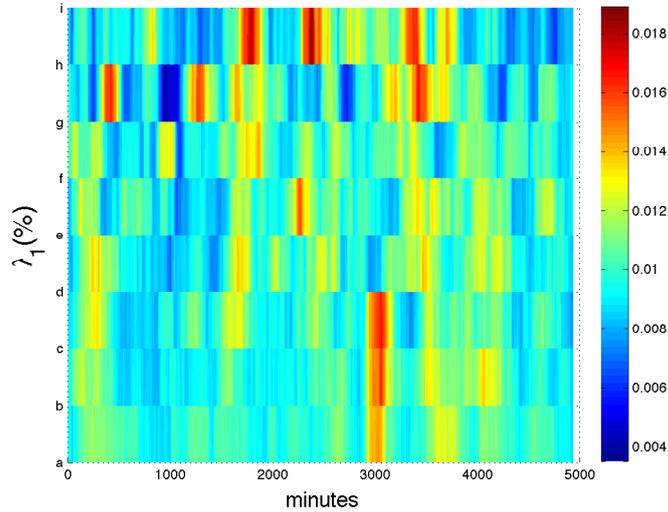


Fig. 2. Heatmap diagram showing the dynamics of the largest Eigenvalue λ_1 across 9 wavelet scales. Scales 1 (*a*) to 9 (*i*) correspond to a 1-2 minute period, a 2-4 minute period, a 4-8 minute period, a 8-16 minute period, a 16-32 minute period, a 32-64 minute period, a 64-128 minute period, a 128-256 minute period, a 256-512 minute period, respectively.

- The different features, found at various scales, suggest that the correlation matrix captured different major events with different time horizons. This will be examined in more detail in the next subsection.

4.2.2. *The Largest Eigenvalue λ_1 compared with the ratio of λ_1/λ_2 dynamics*

We also wished to determine whether the sub-dominant eigenvalues λ_2 hold further information on the key sources of major events and what information these contribute additionally to the images. In Figure 3, the dynamics of the series for the largest eigenvalue and changes of the eigenvalue ratio λ_1/λ_2 were examined from the MODWT analysis. Here, we detail several scenarios for the peaks in the largest eigenvalue and the ratio of the largest to the next largest eigenvalue for a window size of 400 images. We have tried to identify the position and nature of peaks of major events from the real images generated from SenseCam collections.

Analysis of Scenarios: Peaks in the largest eigenvalue correspond to key sources of major events in the SenseCam recording. We have studied the largest eigenvalue λ_1 and the ratio of λ_1/λ_2 time series for a window size of 400 images to try to identify the position and nature of such events from the real images generated from SenseCam. The different features, found at various scales, suggest that the correlation matrix captured different major events with some features consistent

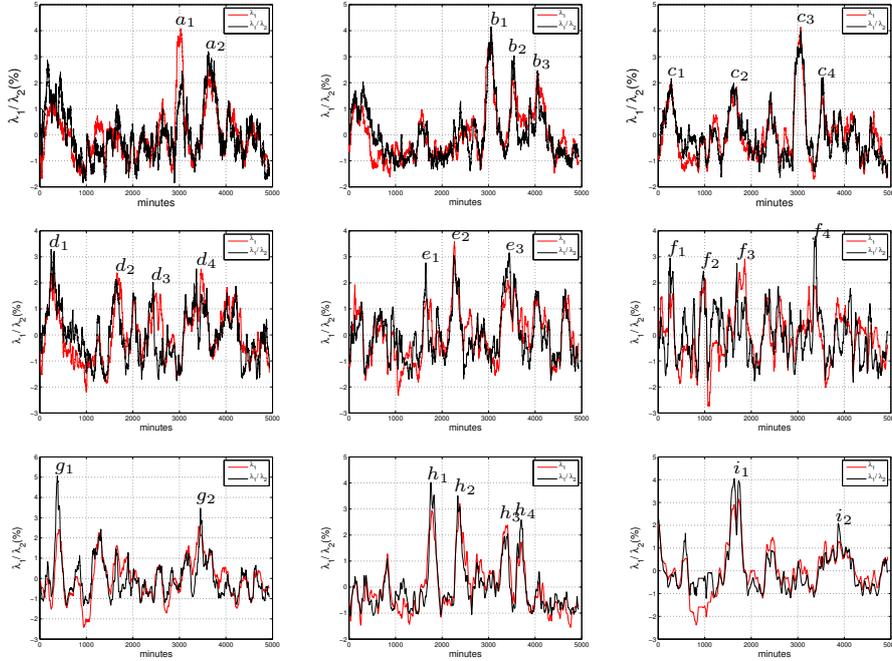


Fig. 3. The Largest Eigenvalue λ_1 (red) and the ratio of λ_1/λ_2 (black) dynamics across 9 wavelet scales. The Top 3 figures (a-c) are for scales 1 to 3 corresponding to a 1-2 minute period, a 2-4 minute period, a 4-8 minute period, respectively, Middle 3 figures (d-f) are for scales 4 to 6 corresponding to a 8-16 minute period, a 16-32 minute period, a 32-64 minute period, respectively and Bottom 3 figures (g-i) are for scales 7 to 9 corresponding to a 64-128 minute period, a 128-256 minute period, a 256-512 minute period, respectively.

and others specific to certain scales. We group similar peaks in reporting, where scenarios are very similar, with more details given below.

1. Peaks a_1, b_1, c_3

This group of fluctuations in the signal relate to the subject arriving at the office, and switching on the lights and the laptop. The laptop colour is white, with the screen the largest object in the field of view of the SenseCam. Thus, the lights and the laptop mainly contribute to these peaks.

Peak a_1 was less noticeable for λ_1/λ_2 due to an increase in λ_2 . The subject was moving during this event. Peak a_1 corresponds to the wavelet scale of a 1-2 minute period. Over this short time period, λ_2 carries other information on the environment, such as the white wall, desk and chairs in the office (peak a_1). As the time period increases, the subject stopped moving and sat down so that the λ_1 and λ_1/λ_2 peaks coincide at peak c_3 .

2. Peaks a_2, b_2, b_3, c_4

These peaks occurred when the subject was working in front of the laptop, with slight movements, partially obscuring the camera, (e.g. hair or hand interruption).

Here, peaks are due to the differences in light level from the laptop, ceiling lights and another desktop. Thus, the laptop screen and ceiling light register higher pixel values than other objects, for which changes are picked up by the SenseCam.

3. Peaks c_1, d_1, f_1, g_1

These peaks refers to the period while the subject was walking in the city centre during the day time. During this period, the subject visited several shops, with lights in the shops reflecting higher pixel values than other objects that contribute to the peaks. A point to note was that a small dip in peaks d_1 and f_1 can be observed, corresponding to the subject picking up clothes, i.e. introducing a new major object into the image field at that time.

In peaks d_1, f_1, g_1 , the ratio λ_1/λ_2 implies that λ_1 increased with λ_2 . The scenario involved the subject moving along a street, visiting shops, with the strong light in the shops dominating during less movement and less change in other objects.

4. Peaks $c_2, d_2, e_1, f_3, h_1, i_1$

These peaks involved the subject walking from home to office in the morning, working in front of the laptop and talking with her colleagues. Note that the office was dark until the subject switched on the lights and laptop, consequently the lights and laptop introduction are highlighted by the peaks. While the subject was standing and talking with colleagues, the camera captured office lights which contributed to the peaks.

At peak f_3 in wavelet scale 6, corresponding to a 32-64 minute period, the highest value of λ_1/λ_2 ratio occurs earlier in time than that for the largest eigenvalue λ_1 , which may imply capture of other effects compared to that of the largest eigenvalue λ_1 ; This shift in the peak of λ_1 relative to that in λ_1/λ_2 occurred when the subject was talking with colleagues (so during localised position change relative to lighting). These movements may contribute to difference in peak position.

For peaks h_1, i_1 , the ratio λ_1/λ_2 was less affected, implying that λ_1 increased with λ_2 . These features involved the subject sitting in front her desk and standing in front of a colleague's desk, talking to colleagues. During the standing period, more objects were captured by the SenseCam especially the lights, with light changes also featuring during localised movement, which may have caused increases in λ_1 and λ_2 .

5. Peak d_3

The d_3 peak ratio of λ_1/λ_2 occurs slightly earlier in time than that for the largest eigenvalue λ_1 , corresponding to a period when the subject was sitting in front of her laptop, with the exception of a short period when the subject moved to another PC on the desk. This produced a small dip in the largest eigenvalue λ_1 .

6. Peaks d_4, e_3

These involved the subject sitting in front of her laptop, leaving to have lunch in the cafeteria with colleagues and then returning to the office. The static period in front of the laptop is reflected in a higher pixel value for this peak. When the subject was moving the images were less clear, hence a small dip around this peak was observed.

For peak d_4 , the peak value for λ_1/λ_2 ratio occurs slightly earlier in time than that for the largest eigenvalue λ_1 , which may be caused by the subject changing position from sitting to moving.

7. Peaks e_2, h_2

These describe the movement of the subject from outside coming into the office and sitting in front of the laptop. The white laptop and screen are the major objects in these images.

8. Peak f_2

The peak of the ratio of the λ_1/λ_2 series is shifted slightly in time in comparison to that for the largest eigenvalue λ_1 . The peak involved the subject walking down an urban street during the evening. The camera captured the lights on the road and in the shops.

9. Peaks f_4, g_2, h_3

These peaks occurred while the subject was sitting in front of the laptop; it is important to note that the camera was inadvertently blocked by the subject on numerous occasions.

10. Peak h_4

The subject was sitting in front of the laptop and then visited the cafeteria with colleagues. While there, she ordered lunch, after which she returned to the office and sat in front of the laptop. Some partial or total blocking of the laptop occurs during the seated periods.

11. Peak i_2

This peak was a typical case for the subject's activities. She was working in front of laptop, with the camera sometimes obscured and sometimes fully capturing the ceiling light.

MODWT gives a clear picture of the movements in the image time series by reconstructing them using each wavelet component. MODWT captured the features markedly apparent at specified scales. A number of features from the image are reproduced and can be examined by studying these eigenvalue series.

For lower scales (1-8 minutes), most of the peaks highlight dramatic light level changes, such as the subject switching on the lights in the dark office or the camera being totally blocked, etc. The subject activity changes, where more people were involved in an event, such as (i) visiting the cafeteria and subsequently returning to the office (about 1 hour) or (ii) travelling from home to the office (about 15 minutes) were highlighted by the middle wavelet scales (8 minutes -1 hour). We found that most distinct events or activities were highlighted by wavelet scale 4 (8 minutes -16 minutes), e.g. in peak d_2 , which describes the journey for the subject from the home to the office, (about 15 minutes), and is sufficiently fine-grained to pick up other activities. At higher scales (larger periods 1- 8.5 hours), the SenseCam mostly captured the subject maintaining a single activity, which may generally be of less interest to wearers or analysts.

For the majority of images, the subject was sitting in front of her laptop, with

laptop, lights and seating position unchanged over an extended period, contributing high pixel values in a sequence of images. This typical case was always marked by a peak in the SenseCam signal. The signal fluctuation is caused by light level changes, such as the subject moving from indoors to outdoors, the subject changing position from sitting to moving, movement increase and more people joining in the scene. Note that a movement or multiple person interactions can be captured by specific scales, using the MODWT method. The ratio analysis strongly reinforces observations on the largest eigenvalue over time. The ratio of λ_1/λ_2 has smaller variation compared to that for the largest eigenvalue λ_1 . This implies that the second largest eigenvalue (λ_2) carries additional information to describe events, but does not decide the occurrence of major events for SenseCam. It appears, however, that it does carry information for events surrounding the major ones, e.g. possible lead-in, lead-out.

4.3. Evaluation

We evaluate the different wavelet scales performance using the precision (P) and recall (R) metrics, as defined below. Moreover, we compute the F_1 score as a measure of the method accuracy²⁷.

$$Precision = \frac{|\text{determined boundaries}| - |\text{wrong boundaries}|}{|\text{determined boundaries}|} \quad (4.1)$$

$$Recall = \frac{|\text{detected reference boundaries}|}{|\text{determined boundaries}|} \quad (4.2)$$

$$F_1 = 2 * \frac{P * R}{P + R} \quad (4.3)$$

Table 1 shows more than 60 ground truth events manually segmented by a user. In order to determine accurate boundaries, each peak point boundary is calculated, (for the difference between neighbouring left and right most trough values)⁸. This is obviously a crude boundary designation; All values within a peak area are combined so that a signal value is less informative. Significant peaks are determined (distinct events or activities) by λ_1/λ_2 percentage pixel values that are larger than zero. Table 2 shows the precision, recall and F_1 measure for λ_1/λ_2 at different wavelet scales. As we can see, most scales appear with high precision and all with very low recall. The main weakness as well as strength for wavelet scales is that different scales highlight different distinct events dependent on the time horizons. Some events at certain scales will be missed, so that the overall recall values are low for this approach. In addition, some activities, such as working in front of the laptop, last for several hours. In manually segmenting 69 events of 10,260 images only, the detection probability for a given event is quite low. In consequence this approach is quite crude and we would suggest that further modifications are needed, such as incorporating other than peak distance and weighting scale combinations.

Table 2. Precision, Recall and F_1 measures for MODWT method

Wavelet Scales	λ_1/λ_2		
	<i>Precision</i>	<i>Recall</i>	F_1
Scale1 (1-2 minute period)	0.3929	0.4058	0.3992
Scale2 (2-4 minute period)	0.7857	0.2029	0.3225
Scale3 (4-8 minute period)	0.5000	0.3188	0.3894
Scale4 (8-16 minute period)	0.4783	0.3333	0.3929
Scale5 (16-32 minute period)	0.5238	0.3043	0.3850
Scale6 (32-64 minute period)	0.5789	0.2754	0.3732
Scale7 (64-128 minute period)	0.7333	0.2174	0.3354
Scale8 (128-256 minute period)	0.9167	0.1739	0.2924
Scale9 (256-512 minute period)	1	0.1594	0.2750

5. Conclusions

We have applied Detrended Fluctuation Analysis (DFA) to image time series, recorded by the SenseCam and have calculated the values of the DFA exponent for these series for a total of 10260 images. The results show evidence of strong long-range correlations in the time series, which means that some information is always picked up by the device, even during relatively static periods. Consequently, DFA provides a useful background summary.

The Maximum Overlap Discrete Wavelet Transform (MODWT) method, calculating equal-time Correlation Matrices over different time scales, was used to investigate the largest eigenvalue and the changes in the dominant/sub-dominant eigenvalue ratio spectrums. As shown in Figure 2, the different features, found at various scales, suggest that the correlation matrix captured different major events with different time horizons. We note that these “jitters” may contain additional information surrounding the major events. This suggests that the correlation matrix for different information captured from SenseCam can be filtered by different time horizons, while consistently occurring peaks should help us to identify major events captured by the SenseCam. By examining the behaviour of the largest eigenvalue and the change in eigenvalue ratios over time, the eigenvalue ratio analysis confirmed that the largest eigenvalue carries most of the major event information, whereas sub-dominant eigenvalues carry information on supporting or lead in/ lead out events. On analysing events, (with a sliding window set to 400 images), we identified the light level as a major event delineator during static periods of image sequence. We have shown that the wavelet scale from 8 minutes to 16 minutes could be said to be the most important time horizon for identification of the distinct events or activities of the user, e.g. the subject changing position from sitting indoors to walking outdoors, as graining is sufficiently fine to enable recording of short term activities, (over a window of say 10 minutes, such as seeking and taking medication for example). This time scale, on the other hand, is not so fine grained as to be

swamped by noise. The MODWT method provides a powerful tool for examination of the nature and quality of the captured SenseCam data for categories of users, such as the case described.

Work is ongoing for larger datasets which should help to confirm initial findings, but may also offer further insight on classification of event type in SenseCam data.

6. Acknowledgements

NL would like to acknowledge generous support from the Sci-Sym Centre Small Scale Research Fund, as well as additional support from the School of Computing and CLARITY Centre. The authors are also indebted to Dr.Cathal Gurrin of CLARITY, for access to the initial work on this topic, for facilitating data recording by providing SenseCam equipment and for a number of useful discussions.

References

1. S. Hodges, L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Butler, G. Smyth, N. Kapur and K. Wood, SenseCam: A Retrospective Memory Aid. In: Dourish, P., Friday, A.(eds.) *UbiComp 2006. Lecture Notes in Computer Science*. **4206** (Springer- Heidelberg, 2006) pp.177-193.
2. R. Harper, D. Randall, N. Smyth, C. Evans, L. Heledd and R. Moore, Thanks for the Memory, *In HCI 2007-Proceedings of the 21st BCS HCI Group Conference, Lancaster, U.K.* (2007) 39-42.
3. R. Harper, D. Randall, N. Smyth, C. Evans, L. Heledd and R. Moore, The Past is a Different Place: They Do Things Differently There. *In Designing Interactive Systems, Cape Town, South Africa.* (2008) 271-280.
4. D. Ashbrook, K. Lyons and J. Clawson, Capturing Experiences Anytime, Anywhere? *IEEE Pervasive Computing.* (2006) 8-11.
5. G. Bell and J. Gemmell, A Digital Life? *Scientific American.* (2007) 58-65.
6. M.L. Lee and A.K. Dey, Providing good memory cues for people with episodic memory impairment, *In Assets'07: Proceedings of the 9th international ACM SIGACCESS conference on Computers and accessibility, Tempe, Arizona, USA.* (2007) 131-138.
7. W.H. Lin and A. Hauptmann, Structuring Continuous Video Recordings of Everyday Life Using Time-Constrained Clustering, *In Multimedia Content Analysis, Management and Retrieval SPIE-IST Electronic Imaging, San Jose, California, USA.* **6073** (2006) 111-119.
8. A.R. Doherty and A. F. Smeaton, Automatically Segmenting Lifelog Data into Events, *Proc. WIAMIS' 08.* (2008) 20-23.
9. C. K. Peng, S. V. Buldyrev, S. Havlin, M. Simons, H. E. Stanley and A. L. Goldberger, Mosaic Organization of DNA Nucleotides, *Physical Review E.* **49** (1994) 1685-1689.
10. J.A.O. Matos, S.M.A. Gama, H.J. Ruskin, A.A. Sharkasi and M. Crane, An econo-physics approach to the Portuguese Stock Index-PSI-20, *Physica A.* **342** (2004) 665-676.
11. S.V. Buldyrev, A.L. Goldberger, S. Havlin, R.N. Mantegna, M.E. Malsa, C.-K. Peng, M. Simons and H.E. Stanley, Long-range correlation properties of coding and noncoding DNA sequences: GenBank analysis *Physical Review E.* **51** (1995) 5084-5091.
12. C. Heneghan and G. McDarby, Establishing the relation between detrended fluctuation analysis and power spectral density analysis for stochastic processes, *Physical Review E.* **62** (2000) 6103-6111.

13. L. Laloux, P. Cizeau, J.-P Bouchaud and M. Potters, Noise dressing of financial correlation matrices, *Physical Review Letters*. **83(7)** (1999) 1467-1470.
14. V. Plerou, P. Gopikrishnan, B. Rosenow, L.A. Nunes Amaral and H.E. Stanley, Universal and non-universal properties of cross-correlations in financial time series, *Physical Review Letters*. **83(7)** (1999) 1471-1474.
15. P. Gopikrishnan, B. Rosenow, V. Plerou and H. E. Stanley, Identifying business sectors from stock price fluctuations, *Physical Review E*. **64** (2001) 035106R/1-035106R/4.
16. J.P. Bouchaud and M. Oshikaw, Theory of Financial Risk and Derivative Pricing, *Cambridge University Press*, (2003).
17. D. Wilcox and T. Gebbie, On the analysis of cross-correlations in South African market data, *Physica A*. **344 (1-2)** (2004) 294-298.
18. S. Sharifi, M. Crane, A. Shamie and H.J. Ruskin, Random matrix theory for portfolio optimization: A stability approach, *Physica A*. **335 (3-4)** (2004) 629-643.
19. T. Conlon, H.J. Ruskin and M. Crane, Random matrix theory and fund of funds portfolio optimisation, *Physica A*. **382 (2)** (2007) 565-576.
20. T. Conlon, H.J. Ruskin and M. Crane, Wavelet multiscale analysis for Hedge Funds: Scaling and Strategies, *Physica A*. **387 (2)** (2008) 5197-5204.
21. B. Podobnik and H.E. Stanley, Detrended cross-correlation analysis: A new method for analysing two non-stationary time series, *Physical Review Letters*. **100 (8)** (2008) 084102/1-084102/11.
22. K. Schindler, H. Leung, C. E. Elger and K. Lehnertz, Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial EEG, *Brain*, **130** (2007) 65-77.
23. J. Kwapien, S. Drozda and A.A. Ionannides, Temporal correlations versus noise in the correlation matrix formalism: An example of the brain auditory response, *Physical Review E*. **62** (2000) 5557-5564.
24. A. Sharkasi, M. Crane, H.J. Ruskin and J.A. Matos, The reaction of stock markets to crashes and events: A comparison study between emerging and mature markets using wavelet transforms, *Physica A*. **368 (2)** (2006) 511-521.
25. C.S. Burrus, R.A. Gopinath and H. Gao, Introduction to Wavelets and Wavelet Transforms, *Prentice Hall*. (1997).
26. D.B. Percival and A.T. Walden, Wavelet methods for time series analysis, *Cambridge University press*. (2000).
27. H. Misra, F. Hopfgartner, A. Goyal, P. Punitha and J. M. Jose, TV News Story Segmentation based on Semantic Coherence and Content Similarity, In: *Boll, S., Tian, Q., Zhang, L., Zhang, Z., Chen, Y.-P.P(eds.) MMM 2010. Lecture Notes in Computer Science*. **5916** (Springer- Heidelberg 2010) pp.347-357.