

NTCIR-12 Lifelog Data Analytics

Na Li, Cathal Gurrin, Martin Crane and Heather J. Ruskin
School of Computing, Dublin City University, Ireland
na.li3@mail.dcu.ie, {cathal.gurrin,martin.crane,heather.ruskin}@dcu.ie

ABSTRACT

Lifelogging is the process of automatically, ambiently and digitally recording episodes of one's life experiences. NTCIR-12 Lifelog¹ test collection was initially created, as support for the Information Retrieval (IR) community, to develop new and novel lifelogging retrieval and visualisation systems. In this paper, our goal is organising and analysing the NTCIR-12 Lifelog dataset by using a time series approach to facilitate automatic discovery of repeat events.

Keywords

Lifelogging; Time Series Analysis; Motifs; Minimum Description Length; equal-time Correlation Matrices; Maximum Overlap Discrete Wavelet Transform

1. INTRODUCTION

Lifelogging is the act of recording aspects of life in digital format. Lifelogs can consist of heterogeneous sensor data such as images, GPS coordinates, WiFi streams, accelerometer and light level measurements amongst others. Lifelogging can take many forms, such as capturing everything seen through wearable cameras, detection of people met (through bluetooth devices), identifying places visited (through GPS), calculating distances or speed (by accelerometer readings), etc. The lifelogging vision is that the user never forgets anything since everything is being tracked, photographed and recorded. The added value is comprehensive memory reinforcement [16].

The Lifelogging concept has existed for more than 7 decades. The concept of logging and storing all of an individual's accumulated digital profile was first proposed in 1945, by Bush [7], using the Memex concept. More recently, Bell's "MyLifeBits" project [11], captured all of a subject's personal data in digital form and created software that allowed access to these data. The goal technologically is to create a personal archive, or a "portable, infallible, artificial memory" that can be exploited to increase job productivity, serve as a basis for medical treatment or to improve performance in

¹<http://ntcir-lifelog.computing.dcu.ie>

school and in many other scenarios.

Visual lifelogging is the process of automatically capturing *images* and storing these in a personal repository. Although technologies for visual lifelogging have existed for some time, the development of many generations of wearable camera technologies has been pioneered by Mann since the 1990s and has addressed, from an early stage, many of the fundamental challenges found in wearable lifelogging (personal imaging [33]). Visual lifelogging has gained in popularity due to projects such as the Microsoft Sensecam [17], which has shown how wearable camera data can aid human memory. The use of wearable cameras has been explored so far mainly in this role, but also for health monitoring and well-being, in social iterations and for leisure activity records amongst others. However, a recent study [4] suggests that lifelogging has even wider application and that an increasing number of devices will be available in the near future. A prominent example is Google Glass², which has received considerable media attention since its first announcement to the public.

To date, much of the research on lifelogging has focused on developing sensors [1], capturing and storage of data [1], processing data into annotated events [9], search and retrieval of information [9], assessing user experience and designing user interfaces for applications of the memory aids [17, 16], as well as diet monitoring [39], analysis of activities of daily living (ADL) [34] and similar. However, these features are now comparatively well understood [17], resulting in improved wearability of devices and inexpensive storage [11]. The challenge has now shifted to that of retrieving the most *relevant* information from the vast quantities of captured data [3, 31], where relevance is determined by the specific need.

The NTCIR-12 Lifelog dataset was initially created for supporting the Information Retrieval (IR) community to develop new and novel lifelogging retrieval and visualisation systems, i.e. as a test baseline. In this paper, we provide an in-depth analysis of the NTCIR-Lifelog dataset. We focus in particular on identifying *repeat* events (motifs) that lifeloggers have experienced, since these can help establish typical user behaviour patterns. The paper is organised as follows: In Section 2, we provide an overview of Lifelogging data collections and review the background to the time series approach. In Section 3 we discuss our methodology in detail, while in Sections 4 and 5, we describe, respectively, the data sets used in our experiment and the results obtained. Conclusions and future direction are discussed in Section 6.

²<https://www.google.com/glass/start/>

2. BACKGROUND

2.1 Lifelogging Data Collections

The vast majority of published Information Retrieval (IR) research assesses effectiveness using resources known as *test collections*, in conjunction with evaluation measures. Test collections have a history dating from the 1990s: TREC in US, CLEF in Europe and NTCIR in Asia run evaluation campaigns aimed at supporting the development, testing and evaluation of IR systems. In 2016, for the first time, NTCIR has included The NTCIR12-Lifelog as a pilot task [20]. This represents the first dataset test collection in Lifelogging for the Information Retrieval community. The purpose of the NTCIR-12 Lifelog pilot task is to explore methods of searching through large lifelog archives. Two subtasks are, 1) the Lifelog Semantic Access Task (LSAT); to examine search and retrieval from lifelogs and 2) the Lifelog Insight Task (LIT); to investigate knowledge mining and visualisation of lifelogs [13].

NTCIR-12 Lifelog data were generated by three individuals, wearing the Autographer camera (as shown in Fig.1) for periods of about one month, capturing on average 1000-15000 images per day, where a sample image is also shown (Fig.1). This camera uses 5 built-in sensors which include an accelerometer, magnetometer, temperature, color, PIR (infrared motion detector), and GPS. It has a 5 megapixel low light image sensor and offers a 136 degree wide-angle lens. It captures images and other sensor readings automatically, recording the wearer's every moment. In the NTCIR-12 Lifelog, every image was resized down to 1024 x 768 resolution and all faces were blurred manually. Data statistics are reported in Table 1.

Although the concept of lifelogging is not novel, the quality and number of quantifying activities is currently burgeoning. This is principally a result of a rise in the profile of lifelogging and its being seen as a key aspect of self-awareness and personal archiving. Given the nature of the data i.e., personal information content, privacy and security issues have been carefully considered by all lifeloggers [28]. To the best of our knowledge, there have been only two public lifelogging datasets which focus on *Image* collections, "All I have Seen" (AIHS) [18] and "The Egocentric Dataset of the University of Barcelona" (EDUB) [5] before publication of the NTCIR12-Lifelog data set. Table 2 provides a short summary of comparison statistics for the three visual lifelogging datasets.

From the table, it is clear that the NTCIR-12 Lifelog collection is much larger compared to the previous two visual Lifelogging datasets. We have examined AIHS datasets in our previous research [22] and have noted there that these are far from perfect, due to incompatible Lifelogging devices and errors occurring during the lifelogging process. The AIHS dataset reportedly includes 19 days with a total of 45612 images, but if each image is examined, actual coverage is 29 days with a total of 43,399 images. Similarly, for the NTCIR-12 Lifelog dataset; e.g. for Lifelogger 2 on the tenth day of collection, some images are not in the correct order while the ninth day collection for Lifelogger 3 shows most images to be the same. Such anomalies can obscure formal data analysis and must be detected at the exploratory stage.

2.2 Time Series Approach

Authors define IR in different ways: Some state that as a field it is concerned with organising information [40]; others emphasise the range of different materials that need to be searched [46]. The pilot

Table 1: Summary of NTCIR-12 Lifelog dataset

Lifeloggers	Number of Collection	Total of Collection
	Days	Images
1	26	37,140
2	25	24,401
3	28	26,583
	Total: 79	Total: 88,124



Figure 1: Autographer Wearable Camera and Sample an image from the NTCIR12-Lifelog Test Collection

Lifelog task aims at prototyping comparative evaluation of information access and retrieval systems operating over personal lifelog data. In this paper, we are concerned with both organisation and analysis of the information.

Exploring a range of techniques and using a variety of lifelogging devices has been well explored by the community. However, lifelogging is concerned with more than just the technology used to capture the record; it needs sound organisation and analysis [14]. Recently, following the development of computer networks, large volume databases, machine learning technologies and the wide choice of computing devices, it has become realistically practical for researchers to investigate the underlying patterns of our daily lives

Current applications address this e.g. by employing automatic classifiers for segmenting a whole day's recording into events and then searching the historical record [9]; or by building ontology-based multi-concept classifiers and searching for specific events [44]. Such concept recognition relies on the objects already being identified in the system and on training classifiers from a set of defined objects. While supervised methods can yield better (more accurate) results in the form of patterns known a priori, they need an expert to input their prior knowledge into the system. Also the results for the classifier depends heavily on there being sufficient good quality training data (i.e. is biased to detect activities which the expert fees in beforehand). Given that visual lifelogs usually consist of large and often unstructured collections of multimedia information, such a 'concept-based' and 'rule-based method' for analysing lifelogging data is not suitable for all use-cases. Ideally, an algorithm should be able to detect unknown phenomena occurring at different frequencies in such data. In our previous work, we introduced and evaluated the use of sophisticated time series analysis methods for organisation and analysis of large lifelog data sets [24, 25, 26, 27].

Over the last several years there has been considerable research into the role of the maximal eigenvalue of the correlation matrix over small time windows. This research has been in the areas of Financial time series [38], electroencephalographic (EEG) data [41], magnetoencephalographic (MEG) data [21] and various other forms of multivariate time series. We apply a similar approach

Table 2: Summary of Three Lifelog dataset

Name	Publish Year	Camera	Description	Annotations	Remarks
AIHS	2010	SenseCam	1 Lifelogger; 19 Days Collections; Total 45,612 Images;	labels 4500 images	We examined the dataset found this dataset includes 29 days with a total of 43,399 Images;
EDUB	2015	Narrative	4 Lifeloggers; 2 days each; Total 8 Days Collections; Total 4912 Images;	labels objects	n/a
NTCIR-12 Lifelog	2016	Autographer	3 Lifeloggers; about one month time period collections each; Total 79 Days Collections; Total 88, 124 Images;	Using the CAFFE visual concept detector to identify 1,000 visual concepts;	n/a

Table 3: The NTCIR12-Lifelog Data Set

Lifelogger	Day Number	Date	Images	Lifelogger	Day Number	Date	Images	Lifelogger	Day Number	Date	Images
1	1	23-02-2015	1,392	2	1	20-04-2015	1,402	3	1	15-06-2015	1,130
1	2	24-02-2015	1,506	2	2	21-04-2015	1,038	3	2	16-06-2015	761
1	3	25-02-2015	1,570	2	3	22-04-2015	1,211	3	3	18-06-2015	985
1	4	26-02-2015	1,454	2	4	23-04-2015	1,100	3	4	19-06-2015	192
1	5	27-02-2015	1,280	2	5	24-04-2015	948	3	5	20-06-2015	460
1	6	28-02-2015	1,375	2	6	25-04-2015	1,004	3	6	21-06-2015	1,094
1	7	01-03-2015	1,527	2	7	26-04-2015	1,034	3	7	24-06-2015	1,299
1	8	02-03-2015	1,423	2	8	27-04-2015	1,389	3	8	28-06-2015	1,066
1	9	03-03-2015	1,129	2	9	28-04-2015	864	3	9	01-07-2015	956
1	10	04-03-2015	1,450	2	10	30-04-2015	1,089	3	10	11-07-2015	1,690
			Total: 14,106				Total: 11,079				Total: 9,573

Table 4: Summary of Ground Truth

Lifelogger	Events Catalogue	No. of Events	Lifelogger	Events Catalogue	No. of Events	Lifelogger	Events Catalogue	No. of Events
1	Attending Lecture	2	2	Attending Lecture	1	3	Cooking	12
1	Cooking	35	2	Cooking	17	3	Drinking	33
1	Doing Laundry	3	2	Cycling	7	3	Driving	13
1	Drinking	43	2	Doing exercise	3	3	Eating	23
1	Driving	37	2	Doing Laundry	3	3	Ordering Food	10
1	Eating	34	2	Drinking	52	3	Playing Lotto	2
1	Fixing Car	3	2	Eating	27	3	Playing with Phone	99
1	Giving Lecture	5	2	Ordering Food	15	3	Reading paper	20
1	Ordering food	20	2	Playing Guitar	6	3	Shopping	5
1	Playing with Phone	158	2	Playing with Phone	86	3	Sitting	29
1	Reading Paper	37	2	Reading Paper	29	3	Standing	32
1	Shopping	3	2	Shopping	4	3	Talking with People	89
1	Sitting	228	2	Sitting	135	3	Using Coffee Machine	2
1	Standing	107	2	Standing	68	3	Walking	130
1	Talking with People	162	2	Talking with People	99	3	Watching TV	47
1	Using Coffee Machine	5	2	Using Coffee Machine	1	3	Working in front of Computer	55
1	Walking	236	2	Walking	170			
1	Watching TV	80	2	Watching TV	6			
1	Working in front of Computer	163	2	Working in front of Computer	114			

therefore, to initial analysis of lifelog image data streams. The aim, underlying application of this multiscaled cross-correlation matrix technique, is to extract dynamic features from time series, for which atypical or non-stationary characteristics are exhibited, symptomatic of "Distinct Significant Events" in the data.

The **wavelet transform** (WT) is a mathematical tool applicable to several areas including image analysis [47], meteorology [8], signal processing [32] and financial time series[6] and can be used to decompose a signal into different time horizons. It is useful in dividing the data series into components on a scale-by-scale basis, to enable to study each component individually to examine the time series in details. As we want to compare values for different image pixel values, we may do so over a range of time scales.

Motifs or recognisable sub sequences of time series, occur frequently and can be linked to specific, previously unknown, cases [29]. Motif analysis can be formed in many medical applications, e.g., measuring data from body sensors [35] and selecting maximally informative genes [2], protein sequence identification [37] and others. Time series motifs are used also for finding patterns in sports in motion capture data [43] as well as in video surveillance applications [15]. Many researchers have studied the extraction of characteristic features from multi-dimensional time series data. In Tanaka et al. [43], Principle Component Analysis (PCA) is used to transform multi-dimensional time series into a one dimensional realisation to detect motifs that are common to all. More recently, Minnen et al. [36] extended a motif discovery method for single time series to detection of motifs occurring across several dimensions of a multi dimensional signal. Visual lifelogs contain records of a wearer's activities and events that occur over different time periods. Consequently, we argue that motifs can represent activities of different length and timing, representing different attributes of big data. We explore this idea by analysing high frequency patterns in multi-dimensional visual lifelogging data.

3. DATA SET

The NTCIR12-Lifelog data was generated using wearable cameras for images and the Moves smartphone application for location & activities. Descriptive statistics of this data set are reported in Table 3. As mentioned previously, the organisers of the NTCIR12-Lifelog also identified about 1,000 concepts by using the CAFFE visual concept detector. The output of this concept detector, however, can not be used to evaluate our method since it includes only simple visual objects such as desktop computer, beer glass, banana, car wheel and so on. Moreover, the annotations provided are not very accurate. In order to assess our methodology, we therefore first had to create a gold standard by identifying key lifelogger activities. We manually annotated activities that the three lifeloggers performed over ten days, resulting in annotations for a total of 34,758 images³. Descriptive statistics of these extracted data are reported in Table 4.

4. METHODS

Our previous results[23, 24] indicate that different distinct events or activities can be detected at different scales through wavelet analysis. Building on this observation, we aim in this paper to extract the motifs of different wavelet scales using the Minimum Description Length (MDL) principle. In this section, therefore, we first give a

review of the Cross-correlation matrix structure and the Maximum Overlap Discrete Wavelet Transform (MODWT). Then, we introduce the Symbolic Aggregate appRoXimation (SAX) algorithm for discretization of time series data into symbolic strings. Finally, we detail our motif extraction algorithm, based on the MDL principle.

4.1 Cross-Correlation Matrix and Eigenvalues Dynamics

The equal-time cross-correlation matrix, capturing between a time series, is calculated using a sliding window approach. The equal-time cross-correlation matrix may be expressed by:

$$C \equiv C_{ij} \equiv \frac{\langle G_i(t)G_j(t) \rangle - \langle G_i(t) \rangle \langle G_j(t) \rangle}{\sigma_{(i)}\sigma_{(j)}} \quad (1)$$

Here, time series $G_i(t)$, $i=1,\dots,N$ of a collection of images, where N is the number of pixels in one image and is smaller than the window size T , and $\sigma_{(i)}$ is the standard deviation of G_i .

The cross-correlation matrix C is limited to the domain $-1 \leq C \leq 1$, where $C = \pm 1$ defines perfect positive/negative correlation and $C = 0$ corresponds to no correlation.

The diagonalization of C , gives its eigenvalues λ_i , with $i=1,\dots,N$ with rank order from the largest to the smallest. Given that the elements on the main diagonals of a matrix must sum to a constant value under linear transformation [12], this value (the Trace) must equal $\sum_i \lambda_i$. So, if some values of λ increase, others must decrease (a feature known as *Eigenvalue Repulsion*[10]).

Two limiting cases for eigenvalue distribution: (i) perfect correlation, $C_i \approx 1$, when the maximal eigenvalue has the value N , (all others taking value zero). (ii) where the time series comprise random values of average correlation $C_i \approx 0$ with their respective eigenvalues distributed around unity (any deviation being down to random correlations). Between these extremes, those eigenvalues on the lower end of the spectrum are often much less than the maximal eigenvalue. In order to examine the dynamics of the each value of λ with a sliding window, each can be normalised in time to:

$$\tilde{\lambda}_i(t) = \frac{(\lambda_i - \bar{\lambda})}{\sigma^\lambda} \quad (2)$$

where $\bar{\lambda}$ and σ^λ denote the mean and standard deviation of the eigenvalues over a particular period of interest. Such normalisation allows for visual comparison of eigenvalues at both extremes of the spectrum, even in cases where there is substantial difference numerically. The reference period for evaluation of $\bar{\lambda}$ and σ^λ of the eigenvalue spectrum can be taken to be a sub-period of low volatility (so aiding visibility of high volatility periods), or the entire time period studied.

4.2 Wavelet Multiscale Analysis

The wavelet transform (WT) is used to decompose a signal into different time horizons. In particular, the discrete wavelet transform (DWT)[45] is useful in dividing the data series into components of different frequencies, so that each component can be studied separately in order to investigate the data series in depth.

The **Maximum Overlap Discrete Wavelet Transform**, (MODWT)[45], 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

³<https://enaliblog.wordpress.com/downloads/>

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 downsampled⁴ 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 to J levels, using the MODWT, theoretically involves the application of J pairs of filters. The filtering operation at the j^{th} level consists of applying a rescaled *father wavelet*⁵ to yield a set of detail coefficients

$$\tilde{D}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{\varphi}_{j,l} f_{t-l} \quad (3)$$

and a rescaled *mother wavelet*⁶ to yield a set of scaling coefficients

$$\tilde{S}_{j,t} = \sum_{l=0}^{L_j-1} \tilde{\phi}_{j,l} f_{t-l} \quad (4)$$

for all times $t = \dots, -1, 0, 1, \dots$, where f is the function to be decomposed [42]. The rescaled mother, $\tilde{\phi}_{j,t} = \frac{\phi_{j,t}}{2^j}$, and father, $\tilde{\varphi}_{j,t} = \frac{\varphi_{j,t}}{2^j}$, 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 $(j+1)$ th level scaling and detail coefficients[42]. Each of the sets of coefficients in a wavelet is called a ‘crystal’.

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 coefficients⁷. An unbiased estimator of the wavelet variance is formed by removing all coefficients that are affected by boundary conditions and is given by

$$\nu_f^2(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \tilde{D}_{j,t}^2 \quad (5)$$

where $M_j = N - L_j + 1$ is the number of non-boundary coefficients at the j^{th} level[42]. The wavelet variance decomposes the variance of a process on a scale-by-scale basis (at increasingly

⁴Downsampling 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

⁵The Father wavelet is also known as the scaling function; this function filters the lowest level of the transform and ensures all the spectrum is covered.

⁶The Mother wavelet, also known as the wavelet function; this wavelet function is, in effect, a band-pass filter and scaling it for each level halves its bandwidth.

⁷The 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.

higher resolutions of the signal) and allows us to explore how a signal behaves over different time horizons.

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

$$\nu_{fg}(\tau_j) = \frac{1}{M_j} \sum_{t=L_j-1}^{N-1} \tilde{D}_{j,t}^{f(t)} \tilde{D}_{j,t}^{g(t)} \quad (6)$$

where all the wavelet coefficients affected by the boundary are removed[42], 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

$$\rho_{fg}(\tau_j) = \frac{\nu_{fg}(\tau_j)}{\nu_f(\tau_j)\nu_g(\tau_j)} \quad (7)$$

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 $g(t)$.

4.3 Dimensionality Reduction and Discretization

We use a dimensionality reduction algorithm based on Piecewise Aggregate Approximation (PAA) [29, 19] called Symbolic Aggregate appRoXimation (SAX) [30]. We apply this technique to transform the Largest eigenvalue time series into a sequence of symbols. For the largest eigenvalue time series λ_1 with number of images n , this time series can be reduced to a string of arbitrary length w , (where $w < n$) and the alphabet size of arbitrary length a , (where $a > 2$). The Largest eigenvalue time series $\lambda_1(t) = \{x_1, \dots, x_n\}$ of length n can be represented as a w -dimensional space by a vector $\bar{\lambda} = \{\bar{x}_1, \dots, \bar{x}_w\}$:

$$\bar{x}_i = \frac{w}{n} \sum_{j=\frac{wn}{n}(i-1)+1}^{\frac{wn}{n}i} x_j \quad (8)$$

SAX allows us to reduce the time series from n dimensions to w dimensions, where the data are divided into w equal sized ‘frames’. The mean value of the data falling within a frame is calculated and a vector of these values becomes the data-reduced representation. Finally, a ‘Behaviour Symbol’(BS) is assigned for every subsequence of PAA symbols. Table 5 summarizes the major notation used in this and the next subsections.

4.4 Estimating Extracted Motif Candidate Based on MDL Principle

Several theoretical information theory principles from literature are relevant to the current analysis, including AIC (Akaike’s Information Criterion), BIC (Bayesian Information Criterion) and MDL (Minimum Description Length).

The AIC estimates the best model based on ‘prediction capability’, while BIC estimates the best model based on Bayesian principles, i.e. can be refined by additional data. Our approach is focused on finding frequent patterns, however, rather than prediction for the time series. The MDL principle seeks the ‘best’ model to describe a set of data defined as that which minimises the description length

Table 5: Summary of the notation

λ_1	A time series $\lambda_1(t) = \{x_1, \dots, x_n\}$
λ	A Piecewise Aggregate Approximation of a time series $\lambda = \{\hat{x}_1, \dots, \hat{x}_w\}$
$\hat{\lambda}$	A symbol representation of a time series $\hat{\lambda} = \{\hat{x}_1, \dots, \hat{x}_w\}$
w	The number of PAA segments representing time series C
a	Alphabet size(e.g., for the alphabet= a,b,c , $a=3$)
T_{min}	Analysis window (e.g., for the alphabet $A=bacc$, $T_{min}=4$)
BS	Behaviour Symbol (e.g., for the alphabet $A=bacc$)
BSS	Behaviour Symbol Sequences (e.g., ABCBBBCBDEBCBDE....in Fig. 2)
DL	Description Length

of the entire data set. The underlying concept is selection of the best model to compress the data in a meaningful way.

The ‘data encoding cost’ is the lower bound of description length that is required to encode each segment. The ‘parameter encoding cost’ is the description length that is required to describe the order of BS in each segment. Finally, the ‘segmentation cost’ is required to describe the location of all segments. The work-flow of the MDL pattern algorithm can be visualized in Fig. 2. For example, in this figure, the length of the first segment is $s_1=7$, the length of the second segment is $s_2=3$ and so on. In addition, we assume that the j_{th} BS has a length l_{ij} . A data encoding cost for the j_{th} BS in the i_{th} segment is calculated then as:

$$-l_{ij} \log_2 \frac{l_{ij}}{t_i} \quad (9)$$

By calculating this cost for all unique Behaviour Symbol Sequences (BSS) in the i -th segment, we obtain the data encoding cost of the whole segment as:

$$\sum_j -l_{ij} \log_2 \frac{l_{ij}}{t_i} \quad (10)$$

Using the following equation, we then calculate the data encoding cost $DL1(\tilde{C}|SC)$ of \tilde{C} that is segmented by the pattern SC :

$$DL1(\tilde{C}|SC) = \sum_i^m \sum_j -l_{ij} \log_2 \frac{l_{ij}}{t_i} \quad (11)$$

We calculate the complementary parameter encoding cost of each segment as $\log_2 t_i$.

Then, the second segment cost $DL2(\tilde{C}|SC)$ of \tilde{C} is calculated as:

$$DL2(\tilde{C}|SC) = \sum_i^m \log_2 t_i \quad (12)$$

and the segmentation cost $DL3(\tilde{C}|SC)$ as:

$$DL3(\tilde{C}|SC) = m \log_2 \left(\sum_i^m t_i \right) \quad (13)$$

Finally, based on this table, we obtain the description length of \tilde{C}

that is segmented by the pattern SC as follows:

$$MDL(\tilde{C}|SC) = DL1(\tilde{C}|SC) + DL2(\tilde{C}|SC) + DL3(\tilde{C}|SC) \quad (14)$$

We use Eq. (15) as the MDL estimation function for the MDL pattern detection algorithm.

(a) a BS sequence:

A C **D F H** C B **D F H** D A **D F H** B E D
 (BS length) 2 5 **1 1 1** 5 3 **1 1 1** 1 2 **1 1 2** 2 3 1

(b) A C | D F H | C B | D F H | D A | D F H | B E D
 2 5 | 1 1 1 | 5 3 | 1 1 1 | 1 2 | 1 1 2 | 2 3 1
 Segment s1 s2 s3 s4 s5 s6 s7

(c)

Segment	Length	Data encoding cost	Parameter encoding cost
s1	7	A x 2 : $-2 \log_2 \frac{2}{7}$ C x 5 : $-2 \log_2 \frac{5}{7}$	$\log_2 7$
s2	3	D x 1 : $-1 \log_2 \frac{1}{3}$ F x 1 : $-1 \log_2 \frac{1}{3}$ H x 1 : $-1 \log_2 \frac{1}{3}$	$\log_2 3$
s3	8	C x 5 : $-5 \log_2 \frac{5}{8}$ B x 3 : $-3 \log_2 \frac{3}{8}$	$\log_2 8$
		.	
		.	
		.	
s7	6	B x 2 : $-2 \log_2 \frac{2}{6}$ E x 3 : $-3 \log_2 \frac{3}{6}$ D x 1 : $-1 \log_2 \frac{1}{6}$	$\log_2 6$
Sum	34	DL1	DL2

And with Segmentation cost: **DL3 = 7 log 34**

(d) **MDL = DL1 + DL2 + DL3**

Figure 2: Calculation of the MDL pattern algorithm

5. EXPERIMENTAL EVALUATION

First, the MODWT of the pixels for each image was calculated within each window of size 400 images 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 in time. Then, the largest Eigenvalue for different window sizes was calculated. Finally, the SAX algorithm was applied to transform the time series to PAA symbols.

Fig. 3 shows the Lifelogger 1 time series of the Largest Eigenvalue for different wavelet scales. The wavelet scales 1-9 correspond respectively to periods of 1-2 minutes, 2-4 minutes, 4-8 minutes, 8-16 minutes, 16-32 minutes, 32-64 minutes, 64-128 minutes, 128-256 minutes and 256-512 minutes. From this figure, we note that the

different features, found at various scales, suggest that the correlation matrix captures different major events with different time horizons. From high frequency to low frequency (top to bottom), the time series becomes smoother. Low frequency analysis removes small-scale changes, typical at high frequency and associated with noise. The peaks in the series become more pronounced of low frequency (large-scale). These peaks reflect larger changes in the images.

Fig. 4 shows the time series of the largest Eigenvalue dynamics across different wavelet scales for the three Lifeloggers. The Heatmaps show that different lifeloggers have different life patterns. For example, for Lifelogger 1, some areas are consistently captured by the camera at certain scales, such as the section of map around Day 8, (captured by wavelet scales 4, 5, 6 and corresponding to periods of 8-16, 16-32 and a 32-64 minutes). These peaks refer to periods when the Lifelogger was sitting in the living room; the camera consequently captures the ceiling lights. For Lifelogger 3, the section of the map around Day 1 at high frequency scales and section of the map around Day 7 at middle frequency scales are consistently captured by the camera. These peaks involve the Lifelogger sitting in the living room and the camera capturing lights of the ceiling and TV. By examining the data sets, we note that Lifelogger 3's lifestyle is more active compared to the other two Lifeloggers. Activities include passing through airport and train stations, visiting the pub, attending a party in a friend's home and so on. The Fig.4 Heatmap reflects the increased number of peaks in Lifelogger 3's time series compared to the others two. The Heatmap shows that some features are consistently captured at certain scales and others across scales, suggesting that the correlation matrix successfully highlight different major events with different time horizons.

As shown in Table 6, our approach extracted key "motifs" such as Working in front of a computer. We defined a 'match' pair as motif similarity identified for this event, while a 'mismatch' was recorded if the event occurred but was not identified by the time series motif technique. The results suggest that the high frequency wavelet scales perform better for Lifelogger 1 &3 while the middle frequency wavelet scales are more accurate for Lifelogger 2. Unfortunately, the number of 'match' pairs is less than that of 'mismatch' pairs for each wavelet scale. The main weakness, as well as strength, for wavelet scales is that different scales highlight different distinct events dependent on the time horizons. In our example, working in front of the computer can last for several hours or a few seconds, so that some additional measure indicating event duration is required. In the wavelet approach also, some events are missing from some scales but prominent in others. We found that a few extracted "motifs" among in the consecutive wavelet scales do overlap. We can roughly calculate our approach accuracy by adding the 'match' pairs of wavelet scales 1, 3, 5, 7 &9 and dividing by the total number of events of identified type. Identification accuracy for Lifeloggers 1, 2 &3 is $\sim 40\%$, 76% and 65% of total events, respectively. We note that this accuracy is quite crude and suggest that further modification is still needed to allow for motifs consecutively repeated or persistent over time.

6. CONCLUSIONS

In this paper, we propose a novel approach based on time series methods for managing large lifelogging data sets. The major contributions of this paper include exploration of a time series approach for investigation of three lifelogger data sets. This approach demonstrates strengths in the organisation, structuring and interpretation of vast amount of heterogeneous streams of visual data.

In particular, through application of the Maximum Overlap Discrete Wavelet Transform (MODWT) on equal-time Cross Correlation Matrices, we find that different features occur at different wavelet time-scales. This suggests that the correlation matrix captures different major events corresponding to different time horizons. Further, the discovery of distinct behavioural motifs provides a basis for prototype templates for identification of similar scenarios at specific time scales e.g., 'typical' lifestyle patterns of lifeloggers. Nevertheless, the issue of 'sustained activity' or motif persistence rather than short-term events needs to be addressed by future of refinement of the method.

7. REFERENCES

- [1] R. Albatal, C. Gurrin, J. Zhou, Y. Yang, D. Carthy, and N. Li. Senseseer mobile-cloud-based lifelogging framework. In *Technology and Society (ISTAS), 2013 IEEE International Symposium on*, pages 144–146. IEEE, 2013.
- [2] I. P. Androulakis. Selecting maximally informative genes. *Computers & chemical engineering*, 29(3):535–546, 2005.
- [3] D. Ashbrook, K. Lyons, and J. Clawson. Capturing experiences anytime, anywhere. *IEEE Pervasive Computing*, 5(2):8–9, 2006.
- [4] I. Askoxylakis, I. Brown, P. Dickman, M. Friedewald, K. Irion, E. Kosta, M. Langheinrich, P. McCarthy, D. Osimo, S. Papiotis, A. Pasic, M. Petkovic, S. Spiekermann, and D. Wright. To log or not to log? - Risks and benefits of emerging life-logging applications. Technical report, European Network and Information Security Agency (ENISA), Nov. 2011.
- [5] M. Bolaños and P. Radeva. Ego-object discovery. *arXiv preprint arXiv:1504.01639*, 2015.
- [6] J.-P. Bouchaud and M. Potters. *Theory of financial risk and derivative pricing: from statistical physics to risk management*. Cambridge university press, 2003.
- [7] V. Bush and A. W. M. Think. The atlantic monthly. *As we may think*, 176(1):101–108, 1945.
- [8] Z. Can, Z. Aslan, O. Oguz, and A. Siddiqi. Wavelet transforms of meteorological parameters and gravity waves. In *Annales Geophysicae*, volume 23, pages 659–663, 2005.
- [9] A. R. Doherty, C. J. Moulin, and A. F. Smeaton. Automatically assisting human memory: A sensecam browser. *Memory*, 19(7):785–795, 2011.
- [10] A. Dumitru and D. Smith. Eigenvalue repulsion in an effective theory of su (2) wilson lines in three dimensions. *Physical Review D*, 77(9):094022, 2008.
- [11] J. Gemmell, G. Bell, and R. Lueder. Mylifebits: a personal database for everything. *Communications of the ACM*, 49(1):88–95, 2006.
- [12] P. Gopikrishnan, B. Rosenow, V. Plerou, and H. E. Stanley. Identifying business sectors from stock price fluctuations. *arXiv preprint cond-mat/0011145*, 2000.
- [13] C. Gurrin, H. Joho, F. Hopfgartner, L. Zhou, and R. Albatal. Overview of ntcir-12 lifelog task. In *Proceedings of the NTCIR Conference on Evaluation of Information Access Technologies*, pages 354–360, 2016.
- [14] C. Gurrin, A. F. Smeaton, and A. R. Doherty. Lifelogging: Personal big data. *Foundations and trends in information retrieval*, 8(1):1–125, 2014.
- [15] R. Hammid, S. Maddi, A. Johnson, A. Bobick, I. Essa, and C. L. Isbell. Unsupervised activity discovery and characterization from event-streams. *arXiv preprint*

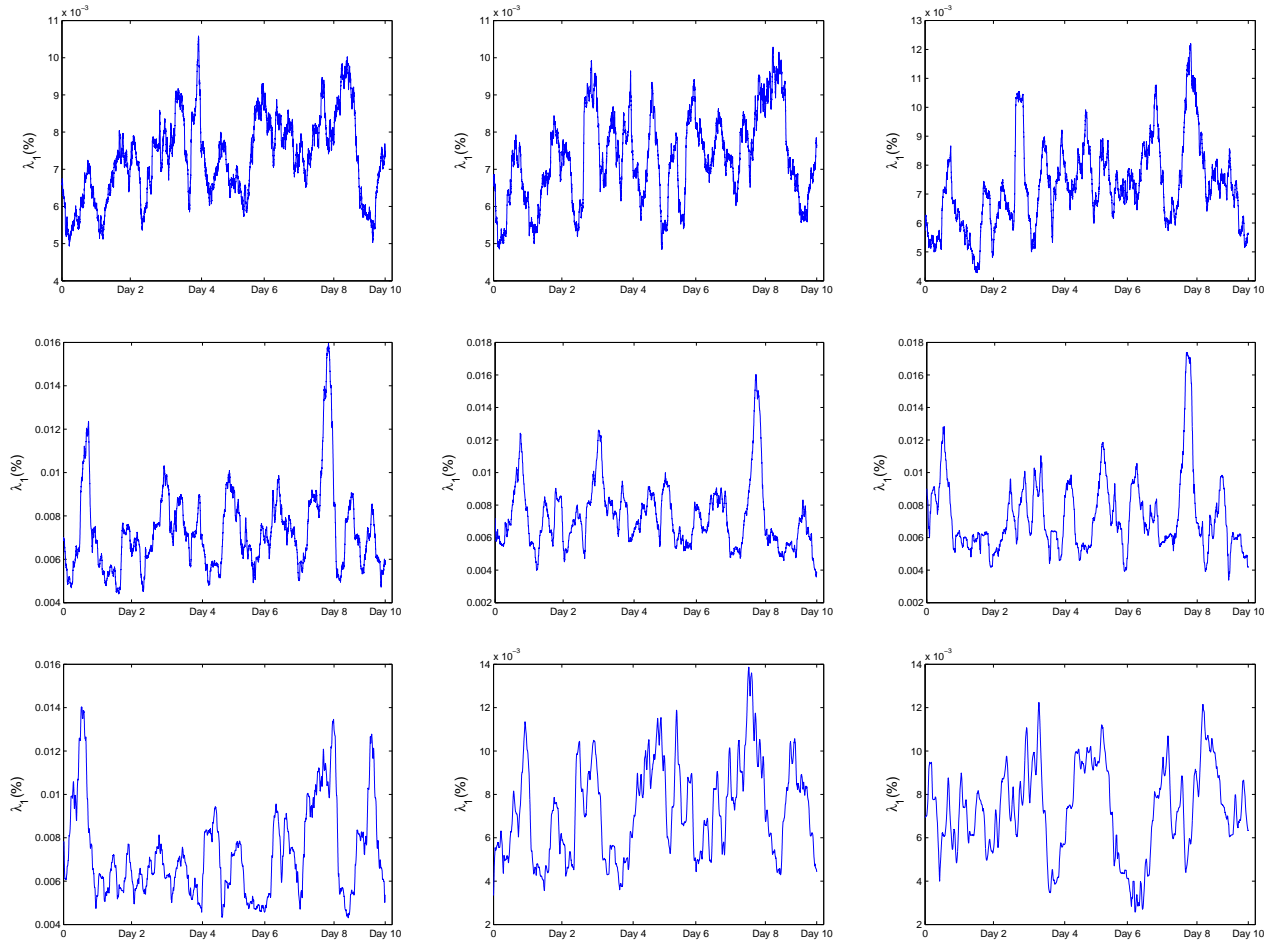


Figure 3: The Largest Eigenvalue λ_1 across 9 wavelet scales. Scales 1 to 9 correspond respectively to periods of 1-2 , 2-4, 4-8, 8-16, 16-32, 32-64, 64-128, 128-256 and 256-512 minutes, for Lifelogger 1

Table 6: Motif: ‘Working in front of Computer’ Event

Wavelet Scales	Lifelogger 1		Lifelogger 2		Lifelogger 3	
	Match	MisMatch	Match	MisMatch	Match	MisMatch
W1	9	154	7	107	5	50
W2	13	150	8	106	11	44
W3	17	146	11	103	7	48
W4	24	139	12	102	15	40
W5	27	136	22	92	19	36
W6	13	150	37	77	11	44
W7	9	154	43	69	5	50
W8	3	160	20	94	0	55
W9	3	160	4	110	0	55

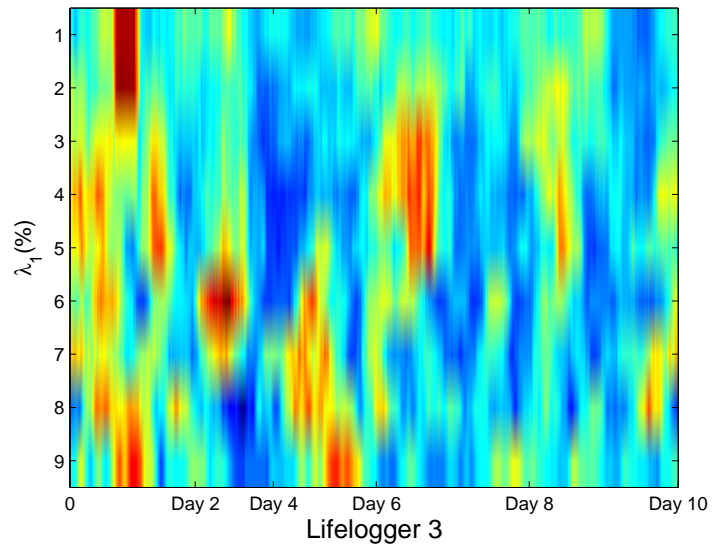
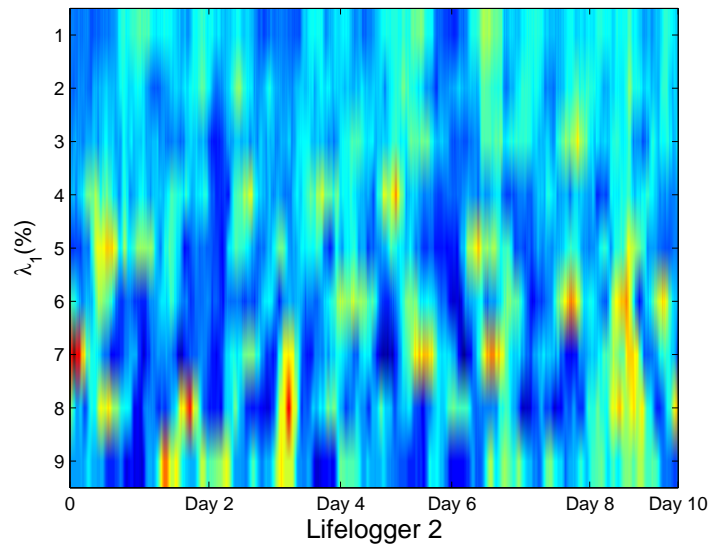
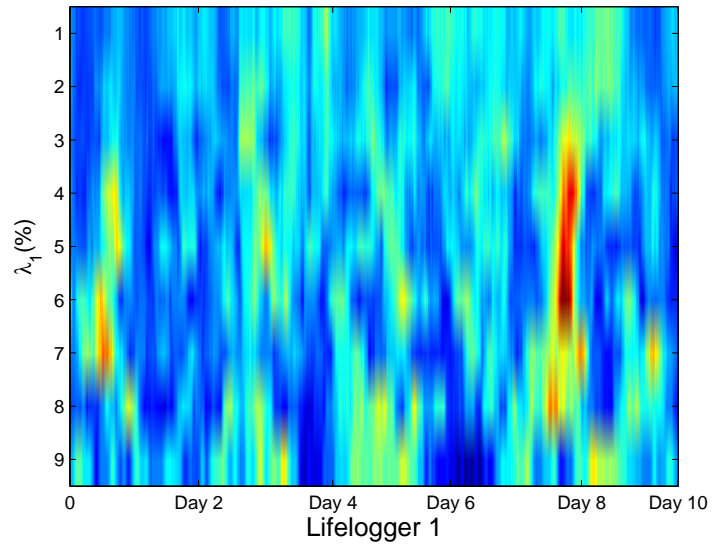


Figure 4: Heatmap diagram showing the dynamics of the largest Eigenvalue λ_1 across 9 wavelet scales. Scales 1 to 9 correspond respectively to periods of 1-2 , 2-4, 4-8, 8-16, 16-32, 32-64, 64-128, 128-256 and 256-512 minutes.

arXiv:1207.1381, 2012.

- [16] M. Harvey, M. Langheinrich, and G. Ward. Remembering through lifelogging: A survey of human memory augmentation. *Pervasive and Mobile Computing*, 27:14–26, 2016.
- [17] 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 *International Conference on Ubiquitous Computing*, pages 177–193. Springer, 2006.
- [18] N. Jovic, A. Perina, and V. Murino. Structural epitome: a way to summarize ones visual experience. In *Advances in neural information processing systems*, pages 1027–1035, 2010.
- [19] E. Keogh, K. Chakrabarti, M. Pazzani, and S. Mehrotra. Dimensionality reduction for fast similarity search in large time series databases. *Knowledge and information Systems*, 3(3):263–286, 2001.
- [20] K. Kishida and M. Kato. Overview of ntcir-12. In *Proceedings of the NTCIR Conference on Evaluation of Information Access Technologies*, pages 1–7, 2016.
- [21] J. Kwapien, S. Drozd, and A. Ioannides. Temporal correlations versus noise in the correlation matrix formalism: an example of the brain auditory response. *Physical Review E*, 62(4):5557, 2000.
- [22] N. Li, M. Crane, C. Gurrin, and H. J. Ruskin. Finding motifs in large personal lifelogs. In *Proceedings of the 7th Augmented Human International Conference 2016*, page 9. ACM, 2016.
- [23] N. Li, M. Crane, and H. J. Ruskin. Visual experience for recognising human activities. In *International Competition on Evaluating AAL Systems through Competitive Benchmarking*, pages 173–185. Springer, 2012.
- [24] N. Li, M. Crane, and H. J. Ruskin. Automatically detecting "significant events" on sensecam. *International Journal of Wavelets, Multiresolution and Information Processing*, 11(06):1350050, 2013.
- [25] N. Li, M. Crane, H. J. Ruskin, and C. Gurrin. Application of statistical physics for the identification of important events in visual lifelogs. In *Bioinformatics and Biomedicine (BIBM), 2013 IEEE International Conference on*, pages 589–592. IEEE, 2013.
- [26] N. Li, M. Crane, H. J. Ruskin, and C. Gurrin. Multiscaled cross-correlation dynamics on sensecam lifelogged images. In *International Conference on Multimedia Modeling*, pages 490–501. Springer, 2013.
- [27] N. Li, M. Crane, H. J. Ruskin, and C. Gurrin. Random matrix ensembles of time correlation matrices to analyze visual lifelogs. In *International Conference on Multimedia Modeling*, pages 400–411. Springer, 2014.
- [28] N. Li and F. Hopfgartner. *To Log or Not to Log? SWOT Analysis of Self-Tracking*, pages 305–325. Springer Fachmedien Wiesbaden, Wiesbaden, 2016.
- [29] J. Lin, E. Keogh, S. Lonardi, and P. Patel. Finding motifs in time series. In *Proceedings of the Second Workshop on Temporal Data Mining*, pages 53–68, 2002.
- [30] J. Lin, E. Keogh, L. Wei, and S. Lonardi. Experiencing sax: a novel symbolic representation of time series. *Data Mining and knowledge discovery*, 15(2):107–144, 2007.
- [31] W.-H. Lin and A. Hauptmann. Structuring continuous video recordings of everyday life using time-constrained clustering. In *Electronic Imaging 2006*, pages 60730D–60730D. International Society for Optics and Photonics, 2006.
- [32] S. Mallat. *A wavelet tour of signal processing: the sparse way*. Academic press, 2008.
- [33] S. Mann. Continuous lifelong capture of personal experience with eyetap. In *Proceedings of the 1st ACM workshop on Continuous archival and retrieval of personal experiences*, pages 1–21. ACM, 2004.
- [34] R. Megret, V. Dovgalecs, H. Wannous, S. Karaman, J. Benois-Pineau, E. El Khoury, J. Piquier, P. Joly, R. Andre-Obrecht, Y. Gaestel, et al. The immed project: wearable video monitoring of people with age dementia. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 1299–1302. ACM, 2010.
- [35] D. Minnen, T. Starner, I. Essa, and C. Isbell. Discovering characteristic actions from on-body sensor data. In *2006 10th IEEE international symposium on wearable computers*, pages 11–18. IEEE, 2006.
- [36] D. Minnen, T. Starner, I. A. Essa, and C. L. Isbell Jr. Improving activity discovery with automatic neighborhood estimation. In *IJCAI*, volume 7, pages 2814–2819, 2007.
- [37] C. G. Nevill-Manning, T. D. Wu, and D. L. Brutlag. Highly specific protein sequence motifs for genome analysis. *Proceedings of the National Academy of Sciences*, 95(11):5865–5871, 1998.
- [38] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, and H. E. Stanley. Universal and nonuniversal properties of cross correlations in financial time series. *Physical Review Letters*, 83(7):1471, 1999.
- [39] S. Reddy, A. Parker, J. Hyman, J. Burke, D. Estrin, and M. Hansen. Image browsing, processing, and clustering for participatory sensing: lessons from a dietsense prototype. In *Proceedings of the 4th workshop on Embedded networked sensors*, pages 13–17. ACM, 2007.
- [40] G. Salton. Automatic information organization and retrieval. 1968.
- [41] K. Schindler, H. Leung, C. E. Elger, and K. Lehnertz. Assessing seizure dynamics by analysing the correlation structure of multichannel intracranial eeg. *Brain*, 130(1):65–77, 2007.
- [42] S. Sharifi, M. Crane, A. Shamaie, and H. Ruskin. Random matrix theory for portfolio optimization: a stability approach. *Physica A: Statistical Mechanics and its Applications*, 335(3):629–643, 2004.
- [43] Y. Tanaka, K. Iwamoto, and K. Uehara. Discovery of time-series motif from multi-dimensional data based on mdl principle. *Machine Learning*, 58(2-3):269–300, 2005.
- [44] P. Wang and A. F. Smeaton. Using visual lifelogs to automatically characterize everyday activities. *Information Sciences*, 230:147–161, 2013.
- [45] D. Wilcox and T. Gebbie. On the analysis of cross-correlations in south african market data. *Physica A: Statistical Mechanics and its Applications*, 344(1):294–298, 2004.
- [46] I. H. Witten, A. Moffat, T. Bell, and M. Gigabyte. Compressing and indexing documents and images. *San Diego, CA*, 1999.
- [47] Z. Xizhi. The application of wavelet transform in digital image processing. In *MultiMedia and Information Technology, 2008. MMIT'08. International Conference on*, pages 326–329. IEEE, 2008.