

Khalid El Asnaoui* and Petia Radeva

Automatically Assess Day Similarity Using Visual Lifelogs

<https://doi.org/10.1515/jisys-2017-0364>

Received July 24, 2017; previously published online February 8, 2018.

Abstract

Today, we witness the appearance of many lifelogging cameras that are able to capture the life of a person wearing the camera and which produce a large number of images everyday. Automatically characterizing the experience and extracting patterns of behavior of individuals from this huge collection of unlabeled and unstructured egocentric data present major challenges and require novel and efficient algorithmic solutions. The main goal of this work is to propose a new method to automatically assess day similarity from the lifelogging images of a person. We propose a technique to measure the similarity between images based on the Swain's distance and generalize it to detect the similarity between daily visual data. To this purpose, we apply the dynamic time warping (DTW) combined with the Swain's distance for final day similarity estimation. For validation, we apply our technique on the Egocentric Dataset of University of Barcelona (EDUB) of 4912 daily images acquired by four persons with preliminary encouraging results.

Methods: The search strategy was designed for high sensitivity over precision, to ensure that no relevant studies were lost. We performed a systematic review of the literature using academic databases (ACM, Scopus, etc.) focusing on themes of day similarity, automatically assess day similarity, assess day similarity on EDUB, and assess day similarity using visual lifelogs. The study included randomized controlled trials, cohort studies, and case-control studies published between 2006 and 2017.

Keywords: Lifelogging, day similarity, similarity measure, EDUB, DTW.

PACS: 86U10.

1 Introduction

Lifelogging consists of taking over a long period of time images that capture the daily experiences and activities of the user wearing a camera. Lifelogging is an active research field for which several devices are spreading faster every day. This growth can represent great advantages to develop new methods for extraction of meaningful information about the user wearing the device and his/her environment. The pictures taken provide considerable potential for knowledge of how people live their lives; hence, they provide new opportunities for many applications in several fields including security, leisure, healthcare, and the quantified self. The data can be images, texts, sounds, numerical measures such as cardiac rhythm, sleep time, or calories ingested and all biological data from sensors on the body and all the moments that matter to you, your workouts to keep fit at your hobbies. So running, walking in the park, browsing the web, or watching a movie – you keep track of everything you do.

The collection of these data tools goes from smartphone to automated digital camera via electronic bracelets and, more generally, the wearable instruments that are carried themselves. The lifelogging is closely

***Corresponding author: Khalid El Asnaoui**, Moulay Ismail University, Faculty of Sciences and Techniques, Department of Computer Sciences, BP 509 Boutalamine, 52000 Errachidia, Morocco, e-mail: khalid.elasnaoui@gmail.com.

<https://orcid.org/0000-0003-4260-6898>

Petia Radeva: Universitat de Barcelona, Barcelona Perceptual Computing Laboratory (BCNPCL), Department of MAIA, Barcelona, Spain

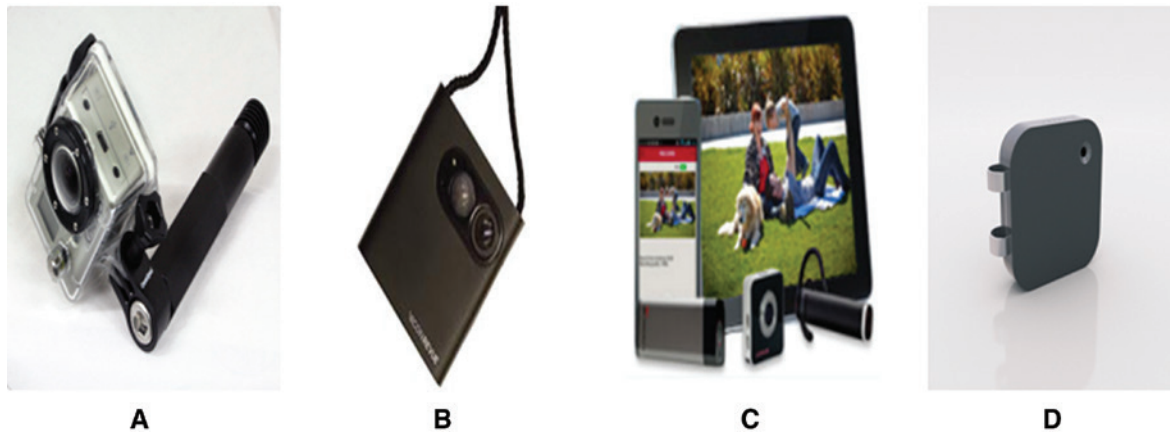


Figure 1: A variety of life-logging wearable devices: (A) GoPro (2002). (B) SenseCam (2005). (C) Looxcie (2011). (D) Narrative (2013).

related to the movement of the “quantified self”. The information used to be archived for the benefit of the lifelogger and shared with others in various degrees.

The arrival of the smartphone and, more recently, technologies of wearable technologies have really created the opportunity for mass participation in lifelogging, as prior to that, all of the hardware required was very specific or proprietary. Nowadays, wearable cameras are very small, discrete cameras housed as watches, glasses, and other subtle wearable devices that can be worn all the day and automatically record a person’s everyday activities in a passive fashion. Most wearable cameras in the market such as MeCam, GoPro, Google Glass, or Looxcie (see Figure 1A and C) are called wearable video cameras, and have a relatively high temporal resolution (HTR). They capture around 35 frames per second and are mostly used for recording the user experience during a few hours in sports and entertainment over the last years.

Instead, lifelogging photo cameras, such as the Narrative Clip (see Figure 1D), are prepared to take pictures in a time-lapse mode that can cover all the day. In this work, we use a Narrative camera that has gone on sale in the late 2012 and incorporates a small wearable camera that clips onto the clothes of the wearer to capture over a thousand images per day using the in-built optical sensor. Usually, these cameras upload their images to their corresponding cloud-based server for online display, event segmentation, and analysis. SenseCam (see Figure 1B), that was initially created by Microsoft, is a wearable camera, worn about the neck that can capture thousands of photos daily, have low temporal resolution (LTR), and capture only two to three frames per minute, being suitable for image acquisition during long periods of time.

Figure 2 presents an example of a walk of a person wearing a wearable camera on a street [5].

The lifelogs formed by the data collected, over long periods of time, by continuously recording the user life, provide a large potential of mining or inferring knowledge about how people live [5], hence, enabling a large amount of applications. Indeed, a collection of studies published in a special issue of the *American Journal of Preventive Medicine* [11] has proven the potential of visual lifelogs captured through a SenseCam from several viewpoints. In particular, it has been demonstrated that, used as a tool to understand and track lifestyle behavior, visual lifelogs would enable the prevention of non-communicable diseases associated to unhealthy trends and risky profiles (such as obesity or depression, among others). In addition, the lifelogs can be used as a tool for re-memory cognitive training; visual lifelogs would enable the prevention of cognitive and functional decline in elderly people [10, 22, 29].

When analyzing several days of a person (lifelogger) trying to characterize his behavior, habits, or lifestyle, a natural question arises: how to automatically measure the day’s similarity in order to assess the diary routines. Such information can be of high interest for different health applications: for example, the days when the wearer of the camera is less active could predict the beginning of depression or physical pain, while the days that are too busy could lead to stress and fatigue.



Figure 2: Example of sequence when walking in the street acquired by the Narrative Clip wearable camera.

Our main goal in this work is to find an algorithm that can assess the similarity of days of a person for a consequent extraction of daily patterns. Toward this end, we developed and tested our algorithm for assessing the similarity of a person's days based on 4912 daily images acquired by four persons using a wearable camera Narrative (see Figure 1D). This set is divided into eight different days. Hence, we successfully applied our proposed method to assess day similarity using visual lifeloggging image content.

The rest of the paper is organized as follows: we analyze the privacy in visual lifeloggging in Section 2. Section 3 deals with some related work. In Section 4, we will develop our proposed contribution. Section 5 introduces the tool of histogram and histobin used in this work. We present in Section 6 the dynamic time warping (DTW) that we have used. Section 7 presents some results obtained using our proposed method and interpreting the results. The conclusion is given in the last section.

2 Analysing Privacy in Visual Lifeloggging

When we talk about privacy in visual lifeloggging, it is necessary to say that the right to privacy is one of the fundamental human rights in any modern society. It advocates and facilitates mechanisms to uphold the privacy of all individuals within the society. However, what is private is highly debated. This is because privacy has social, legal, psychological, political, and technical connotations. Even more, privacy is of dynamic nature. What is considered private in a society can change considerably with time. Many of these changes are driven by technological advancements. This work is extended in Ref. [20].

3 Related Work

In recent years, many interesting applications for lifeloggging and human behavior have appeared and are being actively researched. Visual lifeloggging data have been used to address different computer vision problems: informative image detection [30, 46], egocentric summarization [23, 40], content-based search and retrieval [7, 45], interaction analysis [1, 9], scene understanding [26], concept recognition [6], body move-

ments [27], object–hand recognition [18, 41], just to mention a few. Especially interesting is the work on behavior analysis from egocentric data. Fathi et al. [18] presented a new model for human activity recognition in short egocentric videos. This work is extended in Ref. [19] by a generative model incorporating the gaze features. Pirsiavash et al. [37] presented a temporal pyramid to encode spatio-temporal features along with detected active object knowledge. Ma et al. [31] proposed a twin stream convolutional neural network (CNN) architecture for activity recognition from videos.

When analyzing the daily activities and characterizing the day of a person, a natural question arises: can we extract patterns of days: weekend vs. week day, busy vs. relaxed day, active vs. passive day, social vs. lonely day, etc.? Following this context, one can ask himself how many patterns of days we can detect related to a person. To this purpose, we need to define a similarity measure for the day data and later apply it, for example, a clustering approach. This paper addresses the first part of the problem: how to define a measure to estimate the similarity of days. Biagioni and Krumm [4] has developed a method for assessing the day similarity from location traces recorded from GPS data. An accurate similarity measure is used to find anomalous behavior, to cluster similar days, and to predict future travel. In the work, the authors collected an average of 46 days of GPS traces from 30 volunteer subjects. Each subject is shown random pairs of days and asked to assess their similarity. They tested eight different similarity algorithms in an effort to reproduce human subjects' assessments. Finally, they applied the best similarity algorithm to cluster days using location traces.

Following the same objective of the day similarity characterization, the question is how to use visual lifelogs for it as lifelogging images give richer information about human behavior compared to GPS: visual lifelogging images contain information about the environment of the person, the events he/she is involved, interactions and daily activities, etc. To the best of our knowledge, this is the first work suggesting an automatic assessment method for day similarity using lifelogging images.

4 Proposed Contribution

The contribution of this paper is the introduction of the concept of visual lifelog day similarity and its automatic assessment using visual image content. In this section, we propose our contribution in which we integrate the Swain's distance into the DTW algorithm in order to compute the similarity between pairs of days from egocentric visual lifelogs. Figure 3 shows the main pipeline of the proposed method.

To discover the day similarity of egocentric visual lifelogs, we introduce the following annotations: let us consider 2 days S_a and S_b , respectively. We convert each image from the sequences S_a and S_b to the HSV

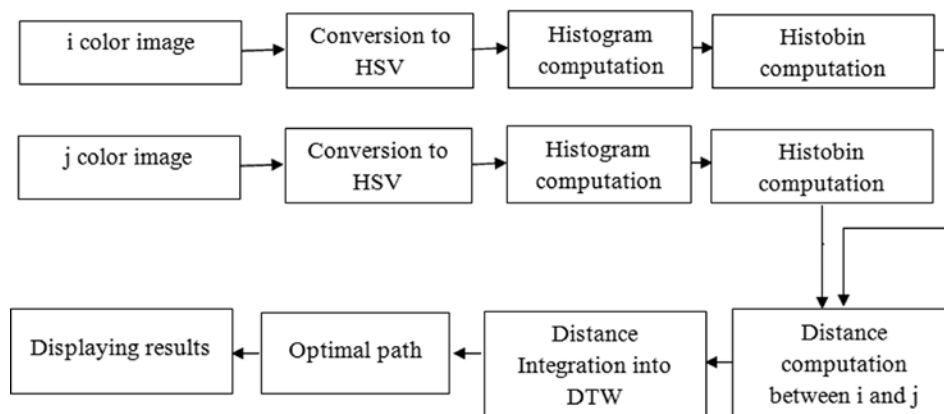


Figure 3: The schema of the proposed method.

color space, and then, we compute its histogram and histobins [13–16]. After that, we compute the distance between each pair of images from both sequences using the intersection of histograms in the HSV color space. Finally, we apply the DTW with what integrates the Swain's distance between the images. Once the DTW matrix of all day images of S_a and S_b is computed, we calculate the optimal path using the algorithm of "Backtracking".

5 Histogram and Histobin

In relation to our goal, first, we need to determine the visual features to compare the images. A technique widely used for color description is the histogram intersection of all image channels. Within this method, we first should compute the image histograms. Then, a histobin will be created from the histogram as follows: each bin (hole) in the histobin is the sum of a few neighboring elements of the histogram. The number of neighbors is determined by the number of the bin histobin. Let N be the number of bins for each component, the histobin has $3 \times N$ bins in total. The number of neighbors is $256/N$. One can see that the histobin is more compact than the histogram. In our system, we calculate the histobin for each color component and concatenate them. When we have the histobin image, the distance between the two images I and J is the distance between their corresponding histobins. This distance, called Swain's distance, can be obtained by the following formula [42]:

$$d(I, J) = \frac{\sum_k |h_k(I) - h_k(J)|}{\sum_k h_k(I)} \quad (1)$$

where I and J are two images, $h(I)$ is the histobin of I , $h(J)$ is the histobin of J , $h_k(I)$ is the bin k of the histobin $h(I)$, $h_k(J)$ is the bin k of the histobin $h(J)$, and $d(I, J)$ is the distance between both images in terms of the intersection of histograms.

We experimented using $k = 8, 16$, and 32 bins to create our histobins and different color spaces. We found out that 24 bins and the HSV color space provide the best results.

6 Lifelog Matching

In order to make lifelog matching, we used the DTW that allows to measure the similarity between two sequences, which can have different lengths. DTW is an algorithm for finding the best match between a stored reference and a signal to be recognized by calculating the difference between the respective feature vectors.

The DTW algorithm is a well-known algorithm in many areas. While first introduced in the 1960s [3] and extensively explored in the 1970s by application to speech recognition [35, 39], it is currently used in many fields: handwriting and online signature matching [12, 43], sign language recognition [28] and gesture recognition [8, 36], data mining and time series clustering (time series databases search) [2, 17, 21, 24, 25, 33], computer vision and computer animation [47], surveillance [44], protein sequence alignment and chemical engineering [34], music and signal processing [32], and face detection [38].

Generally, DTW is a method that looks for an optimal matching between two time series, under certain restrictions. Time series can be distorted by non-linear transformation of the time variable. This alignment method of time series is often used in the context of hidden Markov models. The DTW algorithm has gained significant attention and popularity in the last decade, by being extremely effective as a measure of similarity technique, for different applications in different domains. Given two time series, it minimizes the effects

of shift and distortion over time allowing the transformation of the time series data in order to detect similar forms with different phases.

Let us consider the two sequences of images to compare $S_a = I_1, I_2, \dots, I_N$ of length N and $S_b = J_1, J_2, \dots, J_M$ of length M . In order to compare these sequences given the feature space denoted by $h(\cdot)$, we need to use the local distance measure between images I and J that is defined as a function:

$$d: h(I) \times h(J) \rightarrow \mathbb{R}^+ \quad (2)$$

Typically, the distance d has a small value if the sequences are similar and otherwise of large values.

The algorithm starts by building a distance matrix $C \in (\mathbb{R}^{N \times M})$ representing all pairwise distances between S_a and S_b . This distance matrix called local cost matrix for the alignment of the two sequences S_a and S_b is defined in our algorithm as follows:

$$C \in \mathbb{R}^{N \times M} : c_{kl} = d(I_k, J_l), k \in [1:N], l \in [1:M] \quad (3)$$

The DTW algorithm is defined in the framework of dynamic programming to align the time series so that a total distance measure is minimized. It is defined as the DTW distance function:

$$dtw(S_a, S_b) = c_{p^*}(S_a, S_b) = \min\{c_p(S_a, S_b), p \in P^{N \times M}\} \quad (4)$$

where $P^{N \times M}$ is the set of all possible warping paths. The algorithm builds the accumulated cost matrix or global cost matrix D , which is defined by the following algorithm 1:

Algorithm 1: AccumulatedCostMatrix (S_a, S_b, c).

Input: S_a, S_b, c
 $dtw[] \rightarrow new[N \times M]$
 $dtw(1, 1) \leftarrow 0$
For $i = 2; i < N; i++$ do
 $dtw(i, 1) \leftarrow dtw(i-1, 1) + c(i, 1)$
EndFor
For $j = 2; j < M; j++$ do
 $dtw(1, j) \leftarrow dtw(1, j-1) + c(1, j)$
EndFor
For $i = 2; i < N; i++$ do
 For $j = 2; j < M; j++$ do
 $dtw(i, j) \leftarrow c(i, j) + \min\{dtw(i-1, j), dtw(i-1, j-1), dtw(i, j-1)\}$
 EndFor
EndFor
Output: dtw

The time cost of building this matrix is $O(NM)$, which equals the cost of the following algorithm, where S_a and S_b are the input time series composed of the image sequences to be compared, and c is the local cost matrix representing all the pairwise distances between the images from S_a and S_b . After computing and building, the local cost matrix that represents all the pairwise distances between S_a and S_b , the alignment path (or warping path or warping function) that defines the correspondence between each image $I_k \in S_a$ and the corresponding image $I_l \in S_b$ is constructed using backtracking. This warping path starts at the end at the right bottom of the DTW matrix and goes to the beginning of the matrix following the strategy shown by algorithm 2.

The principal objective is to search for an optimal path for which a least cost is associated. This path offers the low-cost areas between the two sequences S_a and S_b .

Algorithm 2: OptimalWarpingPath (*dtw*).

Input: Matrix *dtw*
 Path[] $\leftarrow$ new array
 $i \leftarrow \text{RowCount}(dtw)$
 $j \leftarrow \text{ColumnCount}(dtw)$
 While ($i > 1$) and ($j > 1$) do
 if $i = 1$ Then $j \leftarrow j - 1$
 else if $j = 1$ Then $i \leftarrow i - 1$
 else
 if $dtw(i - 1, j) = \min\{dtw(i - 1, j), dtw(i, j - 1), dtw(i - 1, j - 1)\}$
 Then $i \leftarrow i - 1$
 else if $dtw(i, j - 1) = \min\{dtw(i - 1, j), dtw(i, j - 1), dtw(i - 1, j - 1)\}$
 Then $j \leftarrow j - 1$
 else
 $i \leftarrow i - 1$
 $j \leftarrow j - 1$
 EndIf
 Path.Add(i, j)
 End If
 End While
Output: Path

7 Experimental Results

Our work is aimed at measuring the similarity between days, usually captured with a lifelogging wearable camera.

7.1 Data

For this purpose, our experiments were performed on the EDUB public dataset of images acquired with a Narrative wearable camera (see Figure 1D). This device is typically clipped around the chest area or on the users' clothes under the neck. For this purpose, we used the public image dataset EDUB <http://www.ub.edu/cvub/dataset/> on which we validate our results (Figure 4). This dataset is a set composed of 4912 images, their sizes are 38,512, and is acquired by four persons. This set is acquired during eight different days, 2 days per person. Figure 4 shows an example of the images in the EDUB dataset [5].

The Table 1 below shows how EDUB 2015 dataset [5] is structured. For example subject1_1 refers to day 1.

7.2 Assessing Day Distance

The results of the day similarity depend strongly on the goodness of the algorithm to compare 2 days. To illustrate it, first, we checked if the chosen Swain's distance detects correctly similar images. To this purpose, we did several tests where we chose a query image and retrieved the most similar images of it. Figure 5 shows an example of the retrieval process. Images correspond to the 14 most similar images to the query image at the top-left corner sorted and displayed according to the score given by the histobin intersection [see equation (1)] in descending order, from left to right and top to bottom. We note that the results are very similar to what a user visually could find; in particular, the algorithm recovered images from the same scene than the query image. Several experimental results obtained from this basis confirm the right choice of the histobin distance for the DTW algorithm.

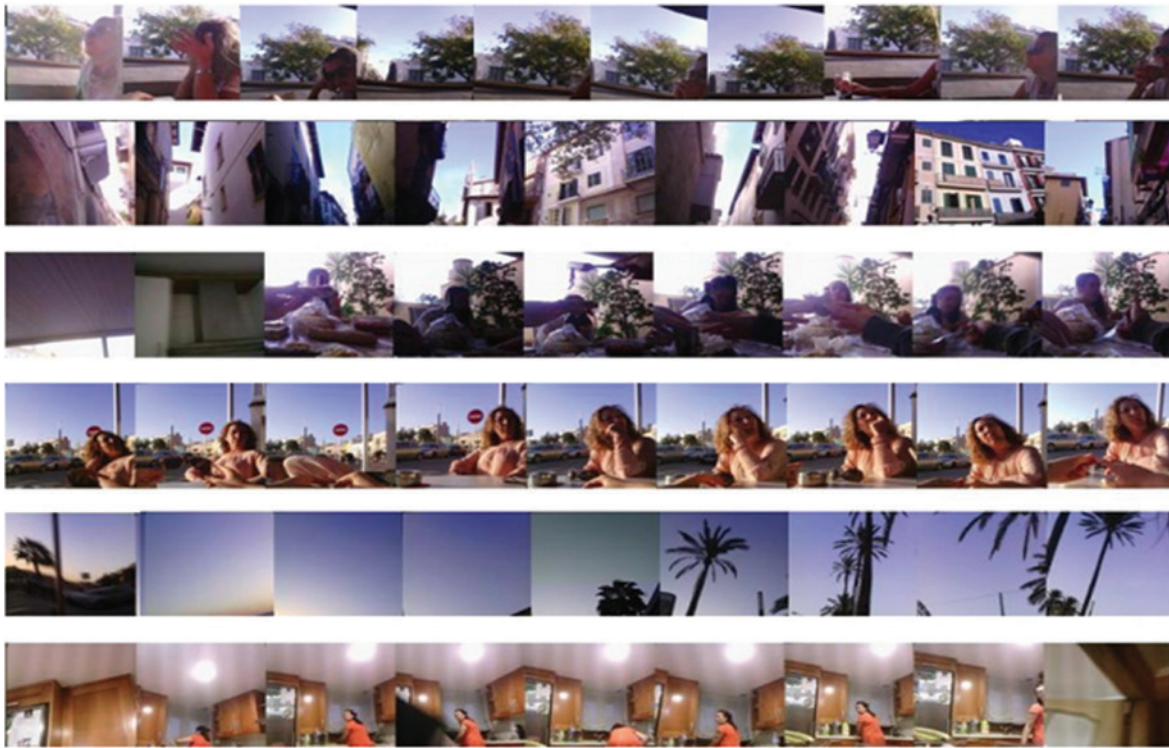


Figure 4: Example of images of the EDUB dataset acquired.

Table 1: Correspondence between subject and days.

Name of day on EDUB	Refer to
Subject 1_1	Day 1
Subject 1_2	Day 2
Subject 2_1	Day 3
Subject 2_2	Day 4
Subject 3_1	Day 5
Subject 3_2	Day 6
Subject 4_1	Day 7
Subject 4_2	Day 8

7.3 Tests for Visual Lifelog Similarity Estimation

To evaluate the performance of the proposed method, we use the accuracy as the statistical comparison parameter. As activities and environments for different people can vary, we test if our algorithm can detect and match events from outdoor to outdoor and indoor to indoor events as a base for day similarity assessment. The accuracy of the day similarity is computed by the following formula:

$$A = \frac{\#(\text{in} / \text{in}) + \#(\text{out} / \text{out})}{\#(\text{in} / \text{in}) + \#(\text{out} / \text{out}) + \#(\text{in} / \text{out}) + \#(\text{out} / \text{in})} \quad (5)$$

where in (out) refers to indoor (resp outdoor) images, and $\#(\cdot/\cdot)$ gives the number of corresponding indoor/indoor, outdoor/outdoor, indoor/outdoor, and outdoor/indoor images, respectively.



Figure 5: The most similar images of a query image at the top left corner using the histobin distance.

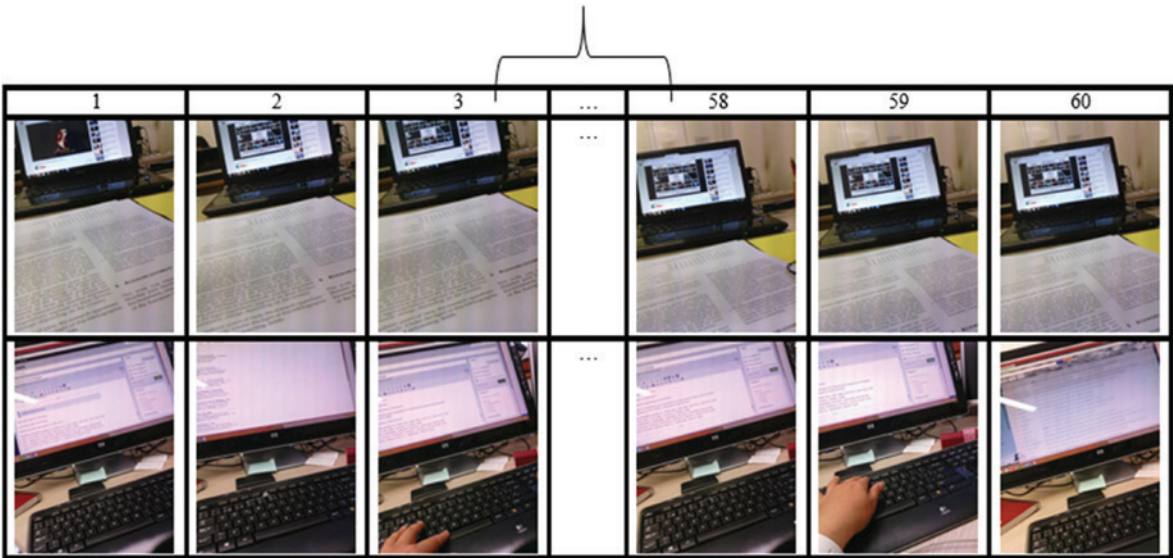


Figure 6: Short example of similarity between days 6 and 8.

Figure 6 illustrates an example of the results obtained between day 6 and day 8. Lines 1 and 2 contain the 60 corresponding images from both sequences detected by our algorithm applied on these 2 days. To make the figure clearer, we just display the first three and the last three images from both sequences detected by applying our proposed algorithm on these 2 days (See Figures 6 and 7).

To check the day similarity using the proposed method, several experiments are conducted on the EDUB dataset. We compute the accuracy [see equation (5)] for each day compared to the rest of the days in the dataset as shown in Table 2. In order to evaluate the similarity algorithm, we compute the total mean accuracy and the mean variance of all pairs.

In order to evaluate the similarity algorithm, we compute the total mean accuracy and the mean variance of all pairs using Table 2 as follows:

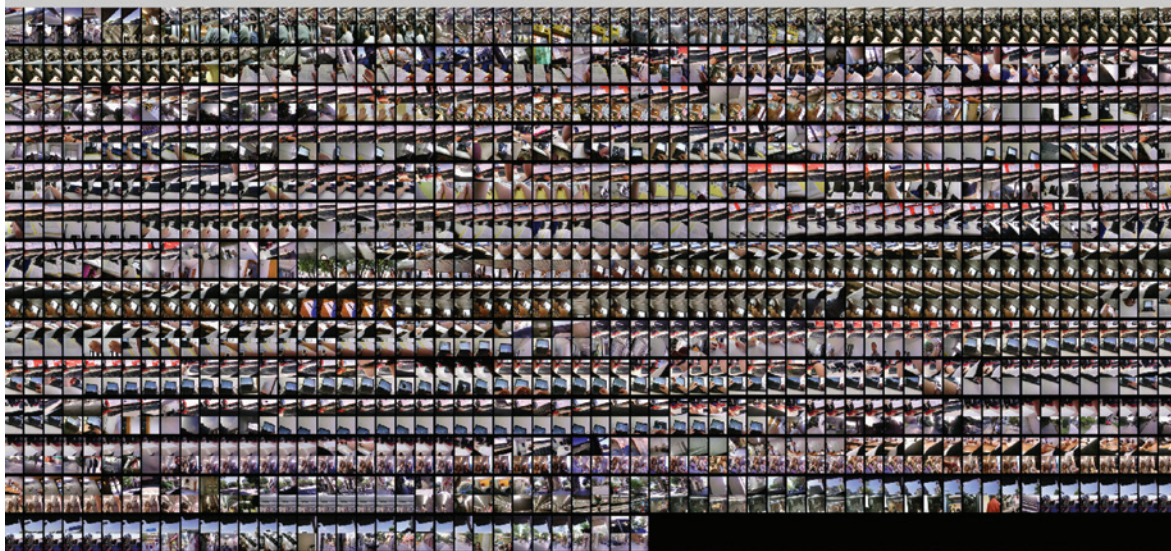


Figure 7: Full example of similarity between days 6 and 8.

Table 2: Accuracy results, mean accuracy, and variation on EDUB.

	Subject1_2	Subject2_1	Subject2_2	Subject3_1	Subject3_2	Subject4_1	Subject4_2
Subject1_1	93	39	52	41	38	42	26
Subject1_2	93	90	82	81	84	69	76
Subject2_1	39	90	75	94	95	92	84
Subject2_2	52	82	75	85	99	95	84
Subject3_1	41	81	94	85	94	99	95
Subject3_2	38	84	95	99	94	89	92
Subject4_1	42	69	92	95	99	89	88
Subject4_2	26	76	84	84	95	92	88
Mean Accuracy	53	76.37	81.12	83	87.25	83.37	79.12
Variance SD	25.67	16.62	14.18	18.11	20.43	18.93	22.21

$$\text{Total mean accuracy} = \frac{\sum \text{Mean Accuracy}}{7}$$

$$\text{Total mean accuracy} = \frac{53 + 76.37 + 81.12 + 83 + 87.25 + 83.37 + 79.12}{7}$$

$$\text{Total mean accuracy} = \frac{543.23}{7} = 77.6\%$$

$$\text{Mean Variance SD} = \frac{\sum \text{Variance SD}}{7}$$

$$\text{Mean Variance SD} = \frac{25.67 + 16.62 + 14.18 + 18.11 + 20.43 + 18.93 + 22.21}{7}$$

$$\text{Mean Variance SD} = \frac{136.15}{7} = 19.45\%$$

The number 7 refers to the number of day in the EDUB dataset.

7.4 Interpreting the Results

This study sets up a new algorithm for the day similarity estimation that must be considered as an application of the concept of visual lifelogs to automatically characterize days using lifelogging data. With our preliminary experimental analysis, observation, and evaluation, the proposed method achieves promising performance and provides a good accuracy for similarity estimate between lifelogging day data. As Table 2 shows, the total mean accuracy achieves up to 77.6%, and the mean variance SD is 19.45% in automatic characterizing corresponding to between days in an egocentric dataset. These first results imply the potential of our proposed algorithm.

Some of the main benefits in using the DTW are: it allows comparing two signal sequences with different lengths. It is more robust against noise and offers scaling along the time axis. Moreover, the use of the DTW is simple; it does not require complex mathematical models and is a fast and efficient algorithm in measuring the similarity between two sequences. Another significant issue with the DTW is to achieve a high accuracy for lifelog matching.

In our method we used the Swain's distance as a distance between the images in the DTW algorithm because it provides satisfactory results and better performance than the classically used Euclidean distance [13–16]. The same tests were carried out using other color spaces with distance measurements such as Manhattan, Chebyshev, Minkowski, and we finally found that the best choice that provides optimal results is to use the proposed method with the Swain's distance.

8 Conclusions

In this paper, we have addressed the following problem: how to compare two visual logs corresponding to 2 days captured by a wearable camera. We presented a new approach that is able to automatically estimate the similarity between a pair of days. Although the work presented a preliminary validation, we believe it demonstrates the potential of lifelogging techniques to characterize the corresponding similarity between days for further routine assessment.

Future work covers a deeper validation using not only indoor and outdoor images but also applying scene classification and/or physical activities of the wearer.

Acknowledgments: This work was partially founded by the Ministerio de Ciencia e Innovación of the Gobierno de España, through the research project TIN2015-66951-C2. SGR 1219, CERCA, ICREA Academia 2014, and Grant 20141510 (Marató TV3). The funders had no role in the study design, data collection, analysis, and preparation of the manuscript.

Authors' contributions: Khalid El Asnaoui carried out the experiments and the programming stage under the supervision of Petia Radeva. The two authors contributed equally to this work. All authors wrote the paper, and all approve this submission.

Bibliography

- [1] S. Alletto, G. Serra, S. Calderara and R. Cucchiara, Head pose estimation in first-person camera views, in: *Pattern Recognition (ICPR), 22nd International Conference on IEEE*, Stockholm, Sweden, pp. 4188–4193, 2014.
- [2] C. Bahlmann and H. Burkhardt, The writer independent online handwriting recognition system frog on hand and cluster generative statistical dynamic time warping, *IEEE Trans. Pattern Anal. Mach. Intell.* **26** (2004), 299–310.
- [3] R. Bellman and R. Kalaba, On adaptive control processes, *IRE Automat. Contr.* **4** (1959), 1–9.
- [4] J. Biagioni and J. Krumm, *Days of our lives: assessing day similarity from location traces*, ADFA, p. 1, Springer-Verlag, Berlin Heidelberg, 2013.
- [5] M. Bolaños, M. Dimiccoli and P. Radeva, Towards storytelling from visual lifelogging: an overview, *J. Trans. Hum. Mach. Syst.* **47** (2017), 77–90.
- [6] D. Byrne, A. R. Doherty, C. G. M. Snoek, G. J. F. Jones and A. F. Smeaton, Everyday concept detection in visual lifelogs: validation, relationships and trends, *Multimed. Tools Appl.* **49** (2010), 119–144.
- [7] V. Chandrasekhar, C. Tan, W. Min, L. Liyuan, L. Xiaoli and L. J. Hwee, Incremental graph clustering for efficient retrieval from streaming egocentric video data, in: *Pattern Recognition (ICPR), 22nd International Conference on IEEE*, Stockholm, Sweden, pp. 2631–2636, 2014.
- [8] A. Corradini, Dynamic time warping for o-line recognition of a small gesture vocabulary, in: *RATFG-RTS'01: Proceedings of the IEEE ICCV Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems (RATFG-RTS'01)*, Washington, DC, USA, IEEE Computer Society, 2001.
- [9] A. R. Doherty and A. F. Smeaton, Combining face detection and novelty to identify important events in a visual lifelog, in: *IEEE International Conference on Computer and Information Technology Workshops*, Sydney, Australia, pp. 348–353, 2008.
- [10] A. R. Doherty, K. Pauly-Takacs, N. Caprani, C. Gurrin, C. J. A. Moulin, N. E. O'Connor and A. F. Smeaton, Experiences of aiding autobiographical memory using the sensecam, *Hum. Comput. Interact.* **27** (2012), 151–174.
- [11] A. R. Doherty, E. S. Hodges, A. C. King, A. F. Smeaton, E. Berry, J. C. Moulin, P. K. Lindley and C. Foster, Wearable cameras in health. *Am. J. Prev. Med.* **44** (2013), 320–323.
- [12] A. Efrat, Q. Fan and S. Venkatasubramanian, Curve matching, time warping, and light fields: new algorithms for computing similarity between curves, *J. Math. Imaging Vis.* **27** (April 2007), 203–216.
- [13] K. El Asnaoui, B. Aksasse and M. Ouanan, Content-based color image retrieval based on the 2D histogram and statistical moments, *World Acad. Sci. Eng. Technol. Comput. Inf. Eng.* **2** (2015), 603–607.
- [14] K. El Asnaoui, B. Aksasse and M. Ouanan, Color image retrieval based on a two-dimensional histogram, *Int. J. Math. Comput.* **26** (2015), 10–18.
- [15] K. El Asnaoui, Y. Chawki, B. Aksasse and M. Ouanan, A content based image retrieval approach based on color and shape, *Int. J. Tomogr. Simul.* **29** (2016), 37–49.
- [16] K. El Asnaoui, Y. Chawki, B. Aksasse and M. Ouanan, Efficient use of texture and color features in content based image retrieval (CBIR), *Int. J. Appl. Math. Stat.* **54** (2016), 54–65.
- [17] W. Euachongprasit and C. Ratanamahatana, Efficient multimedia time series data retrieval under uniform scaling and normalization, in: *ECIR 2008*, LNCS, vol. 4956, pp. 506–513, Springer, Heidelberg, 2008.
- [18] A. Fathi, A. Farhadi and J. M. Rehg, Understanding egocentric activities, in: *IEEE International Conference on Computer Vision (ICCV)*, Barcelona, Spain, pp. 407–414, 2011.
- [19] A. Fathi, Y. Li and J. M. Rehg, Learning to recognize daily actions using gaze, in: *European Conference on Computer Vision*, pp. 314–327, Springer, 2012.
- [20] M. S. Ferdous, S. Chowdhury and J. M. Jose, Analysing privacy in visual lifelogging, *Pervasive Mob. Comput.* (2017). DOI: 10.1016/j.pmcj.2017.03.003.
- [21] J. Gu and X. Jin, A simple approximation for dynamic time warping search in large time series database, in: *Proceedings of the 7th International Conference on Intelligent Data Engineering and Automated Learning*, Burgos, Spain, pp. 841–848, 2006.
- [22] 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: *UbiComp: Ubiquitous Computing*, pp. 177–193, Springer, Heidelberg, 2006.
- [23] A. Jinda-Apiraksa, J. Machajdik and R. Sablatnig, *A Keyframe Selection of Lifelog Image Sequences*, Erasmus Mundus M.Sc. In Visions and Robotics thesis, Vienna University of Technology, 2012.
- [24] T. Kahveci and A. Singh, Variable length queries for time series data, in: *IEEE Proceedings of the 17th International Conference on Data Engineering*, Heidelberg, Germany, pp. 273–282, 2001.
- [25] T. Kahveci, A. Singh and A. Gurel, Similarity searching for multiattribute sequences, in: *IEEE Proceedings of the 14th International Conference on Scientific and Statistical Database Management, 2002*, Edinburgh, Scotland, pp. 175–184, 2002.
- [26] B. Kikhia, A. Y. Boytsov, J. Hallberg, H. Jonsson and K. Synnes, Structuring and presenting lifelogs based on location data, in: *Pervasive Computing Paradigms for Mental Health*, pp. 133–144, Springer, Cham, Switzerland, 2014.
- [27] K. M. Kitani, T. Okabe, Y. Sato and A. Sugimoto, Fast unsupervised ego-action learning for first-person sports videos, in: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, USA, pp. 3241–3248, 2011.

- [28] A. Kuzmanic and V. Zanchi, Hand shape classification using dtw and lcsm as similarity measures for vision-based gesture recognition system, in: *IEEE EUROCON, The International Conference on "Computer as a Tool"*, Warsaw, Poland, pp. 264–269, 2007.
- [29] M. L. Lee and A. K. Dey, Lifelogging memory appliance for people with episodic memory impairment, in: *Proceedings of the 10th International Conference on Ubiquitous Computing*, Seoul, South Korea, pp. 44–53, ACM, 2008.
- [30] A. Lidon, M. Bolaños, M. Dimiccoli, P. Radeva, M. Garolera and X. Giró Nieto, Semantic summarization of egocentric photo stream events, arXiv preprint arXiv:1511.00438, 2015.
- [31] M. Ma, H. Fan and K. M. Kitani, Going deeper into first-person activity recognition, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Colorado Springs, CO, USA, pp. 1894–1903, June 2016.
- [32] S. Majed, Robust face localization using dynamic time warping algorithm, reviews, refinements and new ideas in face recognition, Dr. Peter Corcoran (Ed.), ISBN: 978-953-307-368-2, InTech., 2011.
- [33] M. Muller, Dtw-based motion comparison and retrieval, in: *Information Retrieval for Music and Motion Part II*, pp. 211–226, Springer, New York City, 2007.
- [34] M. Muller, H. Mattes and F. Kurth, An efficient multiscale approach to audio synchronization, in: *Proc. ISMIR*, Victoria, Canada, pp. 192–197, 2006.
- [35] C. Myers, L. Rabiner and A. Rosenberg, Performance tradeoffs in dynamic time warping algorithms for isolated word recognition, *IEEE Trans. Acoust. Speech Signal Process.* [see also *IEEE Trans. Signal Process.*], **28** (1980), 623–635.
- [36] V. Niennattrakul and C. A. Atanamahatana, On clustering multimedia time series data using k-means and dynamic time warping, in: *IEEE International Conference on Multimedia and Ubiquitous Engineering, MUE'07*, Seoul, South Korea, pp. 733–738, 2007.
- [37] H. Pirsiavash and D. Ramanan. Parsing videos of actions with segmental grammars, in: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Columbus, OH, USA, pp. 612–619, 2014.
- [38] A. Ratanamahatana and E. Keogh. Making time-series classification more accurate using learned constraints, in: *The SIAM Intl. Conf. on Data Mining*, pp. 11–22, Lake Buena Vista, Florida, 2004.
- [39] H. Sakoe and S. Chiba, Dynamic programming algorithm optimization for spoken word recognition, *IEEE Trans. Acoust. Speech Signal Process.* **26** (1978), 43–49.
- [40] A. F. Smeaton, P. Over and A. R. Doherty, Video shot boundary detection: seven years of TRECVID activity, *Comput. Vis. Image Underst.* **114** (2010), 411–418.
- [41] S. Sundaram and W. W. Mayol-Cuevas, Egocentric visual event classification with location-based priors, in: *Advances in Visual Computing*, pp. 596–605, Springer, 2010.
- [42] M. J. Swain and D. H. Ballard, Color indexing, *Int. J. Comput. Vis.* **7** (1991), 11–22.
- [43] C. C. Tappert, C. Y. Suen and T. Wakahara, The state of the art in online handwriting recognition, *IEEE Trans. Pattern Anal. Mach. Intell.* **12** (1990), 787–808.
- [44] J. Vial, H. Nocairi, P. Sassi, S. Mallipatu, G. Cognon, D. Thiebaut, B. Teillet and D. Rutledge, Combination of dynamic time warping and multivariate analysis for the comparison of comprehensive two-dimensional gas chromatograms application to plant extracts, *J Chromatogr. A* **1216** (2009), 2866–2872.
- [45] Z. Wang, M. D. Hoffman, P. R. Cook and K. Li, Vferret: content-based similarity search tool for continuous archived video, in: *Proceedings of the 3rd ACM workshop on Continuous archival and retrieval of personal experiences*, Santa Barbara, CA, USA, pp. 19–26, 2006.
- [46] B. Xiong and K. Grauman. Detecting snap points in egocentric video with a web photo prior, in: *European Conference on Computer Vision*, pp. 282–298, Springer, Zurich, Switzerland, 2014.
- [47] Z. Zhang, K. Huang and T. Tan, Comparison of similarity measures for trajectory clustering in outdoor surveillance scenes, in: *ICPR'06: Proceedings of the 18th International Conference on Pattern Recognition (ICPR'06)*, Washington, DC, USA, IEEE Computer Society, pp. 1135–1138, 2006.

Bionote



Khalid El Asnaoui

Moulay Ismail University, Faculty of Sciences
and Techniques, Department of Computer
Sciences, BP 509 Boutalamine, 52000
Errachidia, Morocco, <https://orcid.org/0000-0003-4260-6898>
khalid.elasnaoui@gmail.com.