University of Groningen

Serious Games Application for Memory Training Using Egocentric Images

de Oliveira, Gabriel; Bolanos, Marc; Talavera Martínez, Estefanía; Gelonch, Olga ; Garolera, Maite

Published in:
Workshop on Social Signal Processing and Beyond

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Early version, also known as pre-print

Publication date:
2017

Serious Games Application for Memory Training Using Egocentric Images

Gabriel Oliveira-Barra[1], Marc Bolaños[1], Estefania Talavera[1,2], Adrián Dueñas[1], Olga Gelonch[3], and Maite Garolera[3]

[1] Universitat de Barcelona
[2] University of Groningen
[3] Consorci Sanitari de Terrassa

Abstract. Mild cognitive impairment is the early stage of several neurodegenerative diseases, such as Alzheimer's. In this work, we address the use of lifelogging as a tool to obtain pictures from a patient's daily life from an egocentric point of view. We propose to use them in combination with serious games as a way to provide a non-pharmacological treatment to improve their quality of life. To do so, we introduce a novel computer vision technique that classifies rich and non rich egocentric images and uses them in serious games. We present results over a dataset composed by 10,997 images, recorded by 7 different users, achieving 79% of F1-score. Our model presents the first method used for automatic egocentric images selection applicable to serious games.

Keywords: lifelogging, serious games, egocentric vision, mild cognitive impairment, machine learning, computer vision

1 Introduction

Dementia can result from different causes, the most common being Alzheimers disease (AD) [10], and it is often preceded by a pre-dementia stage, known as Mild Cognitive Impairment (MCI), characterized by a cognitive decline greater than expected by an individual's age, but which does not interfere notably with their daily life activities [19, 11]. Currently, medical specialists design and apply special activities that could serve as a treatment tool for cognitive capabilities enhancement. Even though, these activities are not specially designed for the patients, which limits their engagement in some cases.

A possible alternative to the application of generic exercises would be the use of personalized images of the daily life of the patients
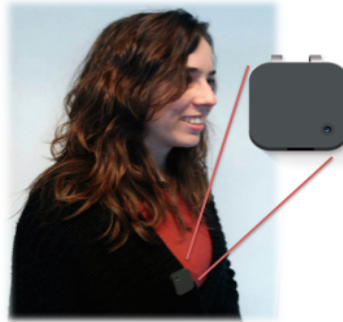

Fig. 1: Person using the Narrative Clip camera.

acquired by lifelogging devices.

Fig. 2: Examples of egocentric images recorded by the Narrative Clip camera.

Lifelogging consists of a user continuously recording their everyday experiences, typically via wearable sensors including accelerometers and cameras, among others. When the visual signal is the only one recorded, typically by a wearable camera, it is referred to as visual lifelogging [4]. This is a trend that is rapidly increasing thanks to advances in wearable technologies over recent years. Nowadays, wearable cameras are very small devices that can be worn all-day long and automatically record the everyday activities of the wearer in a passive fashion, from a first-person point of view. As an example, Fig. 2 shows pictures taken by a person wearing such a camera.

Recent studies have described wearable cameras or lifelogging technologies as useful devices for memory support for people with episodic memory impairment, such as the one present in MCI [15, 8]. The design of new technologies to be applied on this field requires to take into account people capabilities, limitations, needs and the acceptance of the wearable devices, since it can directly affect the treatment. So far, some studies have deeply focus into the factors associated to the use of these devices [24, 13].

Lifelogging and privacy: In terms of privacy, in 2011, the European Union agency ENISA evaluated the risks, threats and vulnerabilities of lifelogging applications with respect to central topics as privacy and trust issues. In their final report, they highlighted that lifelogging itself is still in its infancy but nevertheless will play an important role in the near future [3]. Therefore, they recommended further and extensive research in order to influence its evolution to be better prepared to mitigate the risks and maximize the benefits of these technologies. In addition, other researchers have also evaluated the possible ethical risks involved on using lifelogging devices on medical studies [7].

Serious games for MCI: Serious games (also known as games with a purpose) are digital applications specialized for purposes other than simply entertaining, such as informing, educating or enhancing physical and cognitive functions. Nowadays they are widely recognized as promising non-pharmacological tools to help assess and evaluate functional impairments of patients, as well as to aid with their treatment, stimulation, and rehabilitation [21]. Boosted by the publication of a Nature letter showing that video game training can enhance cognitive control in older adults [2], there is now a growing interest in developing serious games specifically adapted to people with AD and related disorders. Preliminary evidence shows that serious games can successfully be employed to train

physical and cognitive abilities in people with AD, MCI, and related disorders [17]. [18] performed a literature review of the experimental studies conducted to date on the use of serious games in neurodegenerative disorders and [21] studied recommendations for the use of serious games in people with AD and related disorders, reporting positive effects on several health-related capabilities of MCI patients such as voluntary motor control, cognitive functions like attention and memory or social and emotional functions. For instance they can improve their mood and increase their sociability, as well as reduce their depression.

Our contribution: Different studies have proven the benefits of directly stimulating the working memory. Our contribution in this paper consists in using as stimuli the autobiographical images of the MCI patients that was acquired by the wearable cameras. By doing this, we intend to accomplish the goal of enhancing their motivation and at the same time treat them in a more functional and multimodal manner [9, 16, 1]. The application, which will allow the user to exercise either at the sanitary center or at home, will be composed by serious games where the patient has to observe a series of images and interact with them.

Although the stimuli provided by egocentric images can be of greater importance than non-personal images, it is important to note both, that egocentric images are captured in an uncontrolled environment, and that wearable cameras usually have free motion that might cause most images to be blurry, dark or empty of semantic content. Considering this important limitations together with the limited capabilities of MCI patients, we propose the development of an egocentric rich images detection system intended to select only images with semantic and relevant content. Our hypothesis is that, by using personal daily life rich images, the motivation of the patient will increase, and as a consequence, the health-related benefits provided by the treatment.

This paper is organized as follows. We describe the proposed serious game and model for rich images selection in Section 2 and Section 3, respectively. In Section 4, we describes the experimental setup and show quantitative and qualitative evaluation. Finally, Section 5 draws conclusions and outlines future works.

2 Proposed Serious Game: "Position Recall"

MCI patients experiment problems in their working memory [23], herefore, it is of high importance to do exercises for stimulating it. All this under the neuroplasticity paradigm, which has proven that it is possible to modify the brain capabilities and the hypothesis of "use it or lose it", which are the basis of the studies related to the cognitive stimulation of elderly people [22]. Thus, in this work, we introduce a serious game that we name as "Position Recall", which was designed by neuropsychologist of Consorci Sanitari de Terrassa for improving the working memory. The mechanics of this game follow this scheme:

The first screen explains to the patient the instructions of the game and in the second the patient is informed that, before starting the game, there will be some practice examples that will serve to understand its logic. To start, the patient must select his preferred level of difficulty (Level 1, 2 or 3).

- **Level 1** shows 3 images of the patients' day during 8 seconds and they are asked to remember their positions. Immediately after they disappear, a single "target" image is shown and they are asked to select in what position it was placed. After some trials the number of images displayed are increased to 4 and then to 5.
- **Level 2** follows the same procedure as the 1st level, but the timespan between the moment where the images disappear and the target image is shown is increased. During this timespan, called latency time, a black screen is shown.
- **Level 3** follows the same procedure as the 2nd level, but now a distractor image is shown instead of a black screen during the latency time. The distractor image is also an image belonging to the patients' day.

The reward system of the game are points that are given after each level, and are calculated as $100 x number\ of\ correct\ answers$. There are 10 trials per level translating into a maximum of 1000 points per level and maximum of 3000 points per game. Figures 3a and 3b show the mechanics of the developed game.
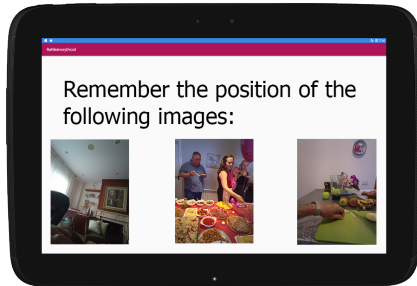
(a) A predefined number of pictures of the patient is shown to him during few seconds at random positions in the screen.
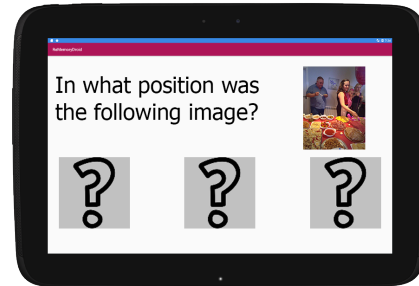
(b) After a certain time passed, the patient is asked to recall in what position one of the pictures, picked up randomly, was placed before.

The images to be shown during the serious games should be significant for the patient. We propose to use images that represent past moments of the user's life, i.e. from the egocentric photostreams recorded by the patient. On the following section, we describe the proposed model for rich images selection.

3 What did I see? Rich images detection

The main factor for providing a meaningful image selection algorithm is the fact that the proposed serious games intend to work on cognitive and sentiment enhancement. Considering the free-motion and non-intentionality of the pictures taken by wearable cameras [4], it is very important to provide a robust method for images selection.

Two of the most important and basic factors that determine the memorability of an image [14, 5] can be described as 1) the appearance of human faces, and 2) the appearance of characteristic and recognizable objects. In this paper, we focus on satisfying the second criterion by proposing an algorithm based on computer vision. Our proposal consists in a rich images detection algorithm, which intends to detect images with a high number of objects and variability and at the same time avoids images with low semantical content, understanding as rich any image that is neither blur, nor dark and that contains clearly visible non-occluded objects. In Fig. 4 we show the general pipeline of our proposal.
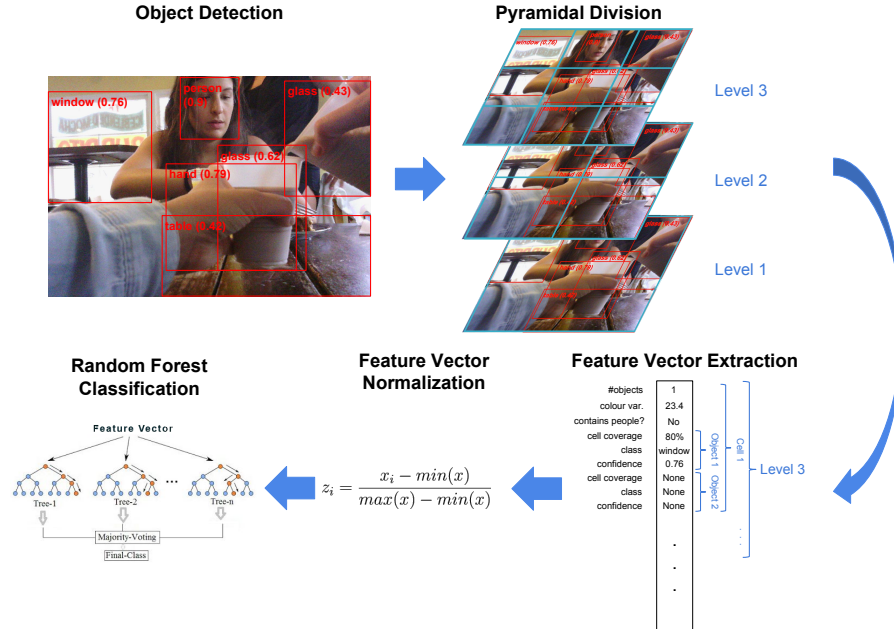

Fig. 4: Scheme of the proposed rich images detection model.

Our algorithm for rich images detection (consists in 1) objects detection: where the neural network named YOLO9000 [20] is applied in order to detect any existent object in the images and their associated confidences c_i. 2) the image is divided in a pyramidal structure of cells, 3) a set of richness-related features

are extracted, 4) the extracted features are normalized and 5) a Random Forest Classifier (RFC) [6] is trained to distinguish the differences between rich or non-rich images. When extracting features, the image is divided in a pyramidal structure of cells with different sizes at each level. The set of extracted features are:

- **Numbers of objects the cell contains.**
- **Variance of color in the cell.**
- **Does the cell contain people?**
- **Object Scale.** Real number between 0 and 1.
- **Object Class.** Class identifier that varies between 1 and 9418.
- **Object Confidence** c_i.

where all features are repeated for each cell and the last three kinds of features are repeated for each object appearing in the cells. The image cell divisions applied are $1x1$, $2x2$ and $3x3$, the maximum of objects selected per cell are 5, 3 and 2, respectively and all objects are sorted by their confidence c_i before selection. If the number of objects is less than the maximum number are found, the feature value in that specific position is set to 0.

The pyramidal division of the images helps us consider smaller objects at higher levels (more cells) and bigger objects at lower levels (less cells). Thus, both small and big objects will be considered for the final prediction.

In order to define the feature "Does the cell contain people?" we manually selected a set of person-related objects detected by the employed object detection method. The concepts representing people that we selected are "person", "worker", "workman", "employee", "consumer", "groom" and "bride".

4 Results

This section describes the results obtained in a quantitative and qualitative form. We compare the results obtained by variations of the proposed method on a self-made dataset of rich images.

Dataset: The dataset used for evaluating our model was acquired by the wearable camera Narrative Clip 2[4], which takes a picture every thirty seconds automatically. The camera was worn during 15 days by 7 different people. Considering that on average the camera takes 1,500 images per day, our dataset consists of 10,997 photographs.

The resulting data was labeled by neuropsychologist experts on MCI cognition following the criteria that any rich image has to be 1) properly illuminated, 2) not blurry and 3) contain one or more objects that are not occluded. After this manual selection the acquired images where split in 6,399 rich images and 4,598 non-rich images.

---

[4] www.getnarrative.com

In Fig. 5a we can see some examples of egocentric rich images and in Fig. 5b non-rich images. We observe that rich images show people or recognizable places. However, non-rich images are meaningless or dark images (that can hardly be seen), including pictures of the sky, ceilings or floor.

(a) Rich images

(b) Non-rich images

The resulting data was divided in training, validation, and test. Considering the pictures taken during the same day can be very similar, we proceeded to randomly separate the different days into the three different sets. First, the training set consists of 60% of the days, in this case 9. Second, 20% of the days, in this case 3, were defined as the validation set. Finally, the remaining 20% was used for the test set.

Evaluation Metrics: In order to evaluate the different results and compare them to get the best one, we make use of the F1-score (or F-measure) metric:

$$F1 = 2 * \frac{1}{\frac{1}{precision} + \frac{1}{recall}} = 2 * \frac{precision * recall}{precision + recall}$$

where *precision* is the quotient between the number of True Positives objects and the number of predicted positive elements; and *recall* is the quotient between the number of True Positives objects and the number of real positive elements.

Quantitative Results: Currently, there are no previous works addressing the challenge we introduce in this work. Thus, in order to compare the performance of our proposed model, we have defined and compared several variations to our main pipeline (see results in Table 1).

As an alternative to our proposed approach (1), we tested an alternative feature vector representation by means of using the (2) Word2Vec word embedding [12]. This word characterization is a 300-dimensional vector representation created by Google that represents words in space depending on their semantic meaning (i.e. words with similar definitions will be represented close in space). The Word2Vec representation was used in two ways. On the one hand was used

for defining the set of concepts related to "person" in the feature described as "Does the cell contain people?". Thus, we computed the similarity between the word "person" and any other concept detected in the image by the object detection and the maximum similarity achieved was used as an alternative to a 0/1 representation. On the other hand, the feature described as "Object Class" was replaced by the 300-dimensions Word2Vec representation.

In the test setting (3) we additionally applied a PCA dimensionality reduction to the Word2Vec representation. Finally, in (4) we used a Support Vector Machine (SVM) classifier instead of a Random Forest Classifier. We applied a Grid Search on the variables C and $gamma$ for parameter selection over the validation set.

	Precision	Recall	F1-score
(1) RFC	0.79	0.79	0.79
(2) RFC + Word2Vec	0.78	0.78	0.78
(3) RFC + Word2Vec + PCA	0.74	0.75	0.75
(4) SVM	0.68	0.67	0.68

Table 1: Comparison of the results

In conclusion we can see that using an RFC classifier (1) obtains better results than SVM (4) and at the same time none of the Word2Vec representations (2) and (3) helped improving the base results.
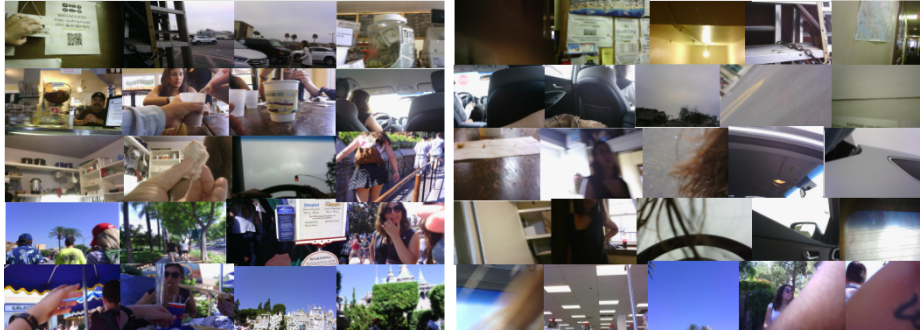

Fig. 6: Example of rich (left) images selection, vs non-rich images rejection. From an egocentric photostream composed by 972 images, 221 were considered rich.

Qualitative Results: Examples of the selected images by the proposed algorithm are shown in Fig. 6. On one hand, we can observe that rich images (left) are clearer, without shadows and with people or focused objects, which allows the user to infer what is happening in the scene. On the other hand, non-rich

images (right) are discarded since they are not illustrative and make difficult the scene interpretation.

Images selected by the proposed model are rich in information and memory trigger. We can foresee that the proposed model cannot only be used for serious games images selection, but also as a tool for images selection for autobiographical memories creation.

5   Conclusions

In this work, we have introduced a novel type of wearable computing application, aiming to provide non pharmacological treatment for MCI patients and to improve their life quality. We discussed lifelogging pictures obtained from wearable cameras combined with serious games as a channel for personalized treatments. We also introduced and tested a novel computer vision technique to classify rich and non rich images obtained from first-person point of view. We obtain 79% F1-score, promising results that will be further studied.

As future work, we will implement more serious games to be included in the application tool. Specialists will use it for MCI patients, aiming to prove the the memory reinforcement hypothesis introduced in this work, as well as the motivation experienced by the subjects increase when using personalized rich images and serious games. Furthermore, in [25], positiveness from egocentric images was addressed. Moreover, we will go deeper on the analysis of users acceptance over the proposed technology, their willingness to use it, and the factors that determine their acceptance toward it. Further improvements of the methodology will be developed in order to obtain more accurate results.

Acknowledgements

References

1. Alves, J., Alves-Costa, F., Magalhães, R., Gonçalves, Ó.F., Sampaio, A.: Cognitive stimulation for portuguese older adults with cognitive impairment: a randomized controlled trial of efficacy, comparative duration, feasibility, and experiential relevance. American Journal of Alzheimer's Disease & Other Dementias® 29(6), 503–512 (2014)
2. Anguera, J.A., Boccanfuso, J., Rintoul, J.L., Al-Hashimi, O., Faraji, F., Janowich, J., Kong, E., Larraburo, Y., Rolle, C., Johnston, E., et al.: Video game training enhances cognitive control in older adults. Nature 501(7465), 97–101 (2013)

3. Askoxylakis, I., Brown, I., Dickman, P., Friedewald, M., Irion, K., Kosta, E., Langheinrich, M., McCarthy, P., Osimo, D., Papiotis, S., et al.: To log or not to log?-risks and benefits of emerging life-logging applications (2011)
4. Bolaños, M., Dimiccoli, M., Radeva, P.: Toward storytelling from visual lifelogging: An overview. IEEE Transactions on Human-Machine Systems 47(1), 77–90 (2017)
5. Carné, M., Giro-i Nieto, X., Radeva, P., Gurrin, C.: Egomemnet: Visual memorability adaptation to egocentric images
6. Criminisi, A., Shotton, J., Konukoglu, E., et al.: Decision forests: A unified framework for classification, regression, density estimation, manifold learning and semi-supervised learning. Foundations and Trends® in Computer Graphics and Vision 7(2–3), 81–227 (2012)
7. Doherty, A.R., Hodges, S.E., King, A.C., Smeaton, A.F., Berry, E., Moulin, C.J., Lindley, S., Kelly, P., Foster, C.: Wearable cameras in health. American journal of preventive medicine 44(3), 320–323 (2013)
8. Doherty, A.R., Pauly-Takacs, K., Caprani, N., Gurrin, C., Moulin, C.J., O'Connor, N.E., Smeaton, A.F.: Experiences of aiding autobiographical memory using the sensecam. Human–Computer Interaction 27(1-2), 151–174 (2012)
9. Flak, M.M., Hernes, S.S., Skranes, J., Løhaugen, G.C.: The memory aid study: protocol for a randomized controlled clinical trial evaluating the effect of computer-based working memory training in elderly patients with mild cognitive impairment (mci). Trials 15(1), 156 (2014)
10. Fratiglioni, L., Grut, M., Forsell, Y., Viitanen, M., Grafström, M., Holmen, K., Ericsson, K., Bäckman, L., Ahlbom, A., Winblad, B.: Prevalence of alzheimer's disease and other dementias in an elderly urban population relationship with age, sex, and education. Neurology 41(12), 1886–1886 (1991)
11. Gauthier, S., Reisberg, B., Zaudig, M., Petersen, R.C., Ritchie, K., Broich, K., Belleville, S., Brodaty, H., Bennett, D., Chertkow, H., et al.: Mild cognitive impairment. The Lancet 367(9518), 1262–1270 (2006)
12. Goldberg, Y., Levy, O.: word2vec explained: deriving mikolov et al.'s negative-sampling word-embedding method. arXiv preprint arXiv:1402.3722 (2014)
13. Gurrin, C., Smeaton, A.F., Doherty, A.R., et al.: Lifelogging: Personal big data. Foundations and Trends® in Information Retrieval 8(1), 1–125 (2014)
14. Khosla, A., Raju, A.S., Torralba, A., Oliva, A.: Understanding and predicting image memorability at a large scale. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 2390–2398 (2015)
15. Lee, M.L., Dey, A.K.: Lifelogging memory appliance for people with episodic memory impairment. In: Proceedings of the 10th international conference on Ubiquitous computing. pp. 44–53. ACM (2008)
16. Li, H., Li, J., Li, N., Li, B., Wang, P., Zhou, T.: Cognitive intervention for persons with mild cognitive impairment: A meta-analysis. Ageing Research Reviews 10(2), 285–296 (2011)
17. Manera, V., Petit, P.D., Derreumaux, A., Orvieto, I., Romagnoli, M., Lyttle, G., David, R., Robert, P.H.: kitchen and cooking,a serious game for mild cognitive impairment and alzheimers disease: a pilot study. Frontiers in aging neuroscience 7 (2015)
18. McCallum, S., Boletsis, C.: Dementia games: A literature review of dementia-related serious games. In: International Conference on Serious Games Development and Applications. pp. 15–27. Springer (2013)
19. Petersen, R.C., Smith, G.E., Waring, S.C., Ivnik, R.J., Tangalos, E.G., Kokmen, E.: Mild cognitive impairment: clinical characterization and outcome. Archives of neurology 56(3), 303–308 (1999)

20. Redmon, J., Farhadi, A.: Yolo9000: better, faster, stronger. arXiv preprint arXiv:1612.08242 (2016)
21. Robert, P.H., König, A., Amieva, H., Andrieu, S., Bremond, F., Bullock, R., Ceccaldi, M., Dubois, B., Gauthier, S., Kenigsberg, P.A., et al.: Recommendations for the use of serious games in people with alzheimer's disease, related disorders and frailty. Frontiers in aging neuroscience 6 (2014)
22. Salthouse, T.A.: Mental exercise and mental aging: Evaluating the validity of the use it or lose it hypothesis. Perspectives on Psychological Science 1(1), 68–87 (2006)
23. Saunders, N.L., Summers, M.J.: Attention and working memory deficits in mild cognitive impairment. Journal of Clinical and Experimental Neuropsychology 32(4), 350–357 (2010)
24. Sellen, A.J., Whittaker, S.: Beyond total capture: a constructive critique of lifelogging. Communications of the ACM 53(5), 70–77 (2010)
25. Talavera, E., Strisciuglio, N., Petkov, N., Radeva, P.: Sentiment Recognition in Egocentric Photostreams (2017)