Video Activity Recognition Framework for assessing motor behavioural disorders in Alzheimer Disease Patients

Véronique Joumier, Rim Romdhane, Francois Bremond, Monique Thonnat, Emmanuel Mulin, P. H. Robert, A. Derreumaux, Julie Piano, J. Lee

Video Activity Recognition Framework for assessing motor behavioural disorders in Alzheimer Disease Patients

Joumier V.(1,3), Romdhane R.(1), Brémond F.(1,3), Thonnat M.(1), Mulin E.(2), Robert PH.(2,3), Derreumaux A.(3), Piano J.(2,3), Lee J.(2,3)

(1) INRIA-Pulsar Team, Sophia Antipolis, France
(2) Centre Mémoire de Ressources et de Recherche, CHU Nice, France
(3) EA CoBTek, University of Sophia Antipolis, France
Veronique.Joumier@inria.fr,Rim.Romdhane@inria.fr

Abstract. Patients with Alzheimers disease show cognitive decline commonly associated with psycho-behavioural disorders like depression, apathy and motor behaviour disturbances. However current evaluations of psycho-behavioural disorders are based on interviews and battery of neuropsychological tests with the presence of a clinician. So these evaluations show limits of subjectivity (e.g., subjective interpretation of clinician at a date t). In this work, we study the ability of a proposed automatic video activity recognition system to detect activity changes between elderly subjects with and without dementia during a clinical experimentation. A total of 28 volunteers (11 healthy elderly subjects, 17 Alzheimer's disease patients (AD)) participate to the experimentation. The proposed study shows that we could differentiate the two profiles of participants based on motor activity parameters, such as the walking speed, computed from the proposed automatic video activity recognition system. These primary results are promising and validating the interest of automatic analysis of video as an objective evaluation tool providing comparative results between participants and over the time.

Keywords: automatic video behavioural disorders; monitoring older people; event recognition; gerontechnology.

1 Introduction

Aging disorders represent a major challenge for health care systems. Many efforts are currently undertaken to investigate on psycho-behavioural disorders. Over the last decades, research has focussed on developing and using various sensors to monitor activities in elderly as well as the AD patients including cameras, microphones, or embedded sensors on the body. These evaluation methods allow providing quantitative and clinical relevant information in real-time on the patient, and also to establish objective criteria. The main objective of this work is to show the ability of the proposed automatic video monitoring system to detect motor disturbances in Alzheimers Disease patients (AD patients) compared with the healthy control subjects during a clinical experimentation. This work differs from prior studies in some points: (i) we build an automatic video monitoring system which is able to describe and recognize events in formal models that can be easily used by clinicians and recognize automatically complex activities and posture, (ii) we do not use embedded sensors which may disturb elderly and be not accepted.

2 Related work

Over the last several years much research has addressed developing and employing various sensors to monitor activity in the elderly people, including cameras, microphones [1], [2], or embedded sensors on the body [3], [4]. Boger et al [5], [6] have developed the system named COACH, a cognitive aide for Alzheimers patients that monitors a user attempting a handwashing task and offers assistance in the form of task guidance (e.g. prompts or reminders) when it is most appropriate. Bouchard et al. [7] address the problem of recognizing the behaviour of person suffering from AD at early-intermediate stages by using plan recognition model based action description logic.

3 Materials and Methods

3.1 Video Activity Recognition Framework

The proposed automatic video monitoring system takes as input video streams and a priori knowledge for 3D scene modelling and events to recognize. The system contains a vision component (e.g. detection, classification and tracking processes) and an event recognition component. The vision component allows detecting people and tracking their different movements over the time in the scene. The event recognition component allows recognising temporal and spatial events, and posture associated with the tracked persons (Fig. 1).

3.2 Experimental site and clinical evaluation

Experimentations description for this study was described in [9]. Briefly, clinical evaluation was conducted in an observation room, located in the Nice Memory Center of the Nice University Hospital. Two fixed monocular-video camera (8 frames/seconde) were positioned in two opposite places in order to capture activities of participants during experimentation. Furthermore, the observation room was equipped with everyday object for use in ADLs, such as a phone, a TV, a coffee corner, a library, an armchair. To assess cognitive and psycho-behavioural disturbances in the AD patients, three experimentation parts are proposed to participants:
- Experimentation 1: called guided activities consists to execute various physical

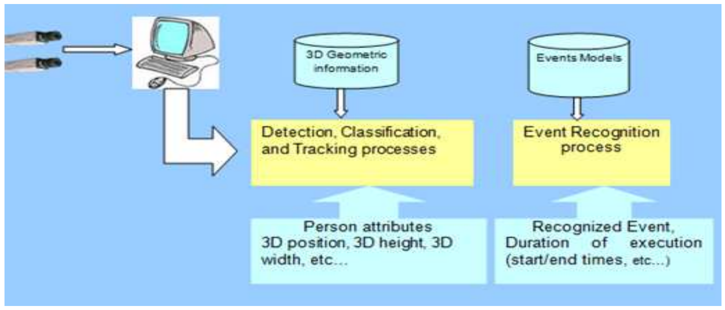

Fig. 1. Architecture of the proposed Video Event Recognition Framework.

exercises timed with the presence of a clinician such as walking exercise, sit-to-stand exercise, etc.
- Experimentation 2: called semi guided activities describes a set of ordered activities that volunteers have to perform alone without the presence of a clinician.
- Experimentation 3: called free activities, volunteers are free to do whatever they want (books reading, watching TV, playing cards).
For this work, we only use video records of the experimentation 1.

3.3 Demographic and clinical characteristics of volunteers

We use video records of 28 volunteers, composed of 11 healthy elderly subjects (healthy control group, G1) and 17 AD patients (AD group, G2).

	Control Group (G1)	AD patients (G2)
Sample size (N)	11	17
Age, years (Mean SD)	74.8 6.67	77 7.43
Sex Ratio (F/M)	6/5	12/5

Table 1. Demographic and clinical characteristics of volunteers.

3.4 Parameters - Judgement criteria

For this work, we focus on the automatic detection of motor disturbances. For that, we analyse different sub-activities composing three main guided physical activities proposed in the first experimentation part (experimentation 1).

Walking exercise For this exercise, participants have to walk on 4 meters (two exercises: go exercise and go-back exercise). Walking speed from automatic video system was computed in two ways. We define a zone called "zone Exercise Walking" which is included into the zone bounded by two lines defining the start and the end points for the exercise (distance between start-end lines = 4m). We model a spatial event e_1: "Inside zone Exercise Walking" (Fig. 2) identifying the time interval Δt_{e_1} when the tracked person is inside the "zone Exercise Walking".

We define an event called e_2: "Is making displacement" identifying the time interval Δt_{e_2} when the tracked person makes a displacement superior to 50 cm per second. At first, we compute the time interval Δt_{M1} used for computing the walking speed as follows: $\Delta t_{e_3(i)} = \max_k(\Delta t_{e_3(k)})$, where e_3 defined the composite event: "is making displacement inside zone Exercise Walking", and k refers to an e_3 occurrence during go exercise (respectively go-back exercise). Then we define the distance $d_{M1,e_3(i)}$ using the 3D coordinates of tracked person at the time points associated with the beginning and end of event $e_3(i)$ selected. We define 2 methods for computing the speed:
- For **M1** method, we define the walking speed as follows (Eq 1):

$$V_{M1,j} = d_{M1,e_3(i)}(j)/\Delta t_{e_3(i)}(j) \tag{1}$$

Where j refers to one walking exercise (j=1, 2, with 1 and 2 refer to the go and go-back exercise respectively). d_{M1,e_3} is the computed distance based on the 3D tracked coordinates of the person.
- For **M2** method, we define the walking speed as follows (Eq 2):

$$V_{M2,j} = d_{M1,e_1}(j)/\Delta t_{e_1}(j) \tag{2}$$

where $d_{M1,e_1}$is the constant length of the "zone Exercise Walking" ($3m < d_{M1,e_1} < 3.5m, d_{M1,e_1}$ is specific to each participant) and $\Delta t_{e_1}(j)$ refers to the time interval when e_1 occurred during the j exercise.
In order to correct the impact of calibration problem on the computation of distance from video processing, we define a reference measure on the ground ($d_{referencevalueGT}$ = 4m = constant value), which is a constant virtual landmark along the direction of the walking exercise). We measure the 3D value for each scene computed by the calibration tool. Then for each scene l, we correct the distance computed by the video processing algorithms using the multiplicative factor $d_{referencevalueGT}/d_{calibrationvalueVideo}(l)$ specific to each scene.

Transfer position exercise For this exercise, participants have to execute consecutively several transfers of position from sitting to standing position (at least 5 transfer positions). We perform the duration of this exercise from the first position transfer to the last position transfer executed using two primitive states related to the posture, e_4:"is standing" and e_5: "is sitting" defined from the 3D height of participant. We build an operator o_1 that computes the number of transfer position n_{Video} (from sitting to standing position, and, from standing to sitting position) based on the historic of postures detected for the tracked participant. Then we model the event e_6: "Transfer position exercise" from temporal constraints (minimal duration of e_4 and e_5, for detecting the start and

```
PrimitiveState (Inside zone Exercise Walking)
PhysicalObjects ((p : Person), (z : Zone))
Constraints ((p in z)
             (z'Name = zone Exercise Walking))
Action (Priority "Normal")
```

Fig. 2. Formal description of the event model e_1 "Inside zone Exercise Walking", the physical objects used for the description of the model are p, the person and z, the zone. The spatial constraint is the person inside the zone named "zone Exercise Walking".

end of composite event e_6, and the minimal duration to execute a consecutive sequence of transfer position). Then we define like gait parameters the execution duration Δt_{e_6} by the number of position transfer n_{GT,e_6} executed when e_6 was recognised by the system (i.e. $\mu_{e_6} = \Delta t_{e_6}/n_{GT,e_6}$,). In other hand we compare the system performance to detect the number of position transfer n_{Video,e_6} when e_6 was recognised by the system (Table. 2).

Up and Go exercise For this exercise, participants start from the sitting position. At the start signal given by the clinician, participant wake up, walk on 3 meters, make a U-turn in the center of the room, go-back on 3 meters, make a U-turn, and sit on the chair. We compute 4 durations of execution: (1) D1 as the duration for walking on 3 meters from the standing position to the time when participant reach the zone where they undertake the U-turn, (2) D2 as the duration for making the U-turn, event modelled by a spatial constraint e_7: "Inside zone U-turn", (3) D3 as the duration for going back on 3 meters from the zone U-turn to the zone where the chair is, (4) D4 as the duration of the exercise (from the sitting position at the start point to the last position transfer at the end of exercise).

3.5 Performance system evaluation

For each video record, events modelled were annotated by the same expert in blind test. Then execution duration of activities and gait parameters associated with event of interest were performed. To compute the walking speed by GT method V_{GT}, we use as start time the instant when participant cross the start line, and, as end time the instant when they cross the end line or make a halt in front of the end line. Reference value used as walked distance is the same for all participants, such as $d_{referencevalueGT} = 4m$. Mean of speed parameters is used to compare performance between estimations provided by the proposed automatic video system and the ones provided by GT (i.e., position and event annotations). To compare the ability of the proposed system to detect differences between patients profiles compared to GT, statistics analyses are conducted with SPSS release 19.0 software using the non parametric Mann-Withney test and their associated p-value determined from Monte-Carlo simulation in order to have more meaningfull conclusion given the small sample size of two groups.

3.6 Results

Parameter values Differences in the estimation of walking speed (Fig. 3) are (i) higher with M2 method than with M1 method, (ii) except for the walking speed for AD patients for the go-back exercise. These differences can be explained as follows: (i) in most cases, the participant l walks a distance inferior to the virtual landmark distance $d_{M1,e_1}(l)$ when the e_1 event (Inside Exercise Walking) is recognized, so the walking speed computed with M2 method is overestimated compared with the one computed by M1 method, (ii) for example, on go-back exercise, some AD patients stop before the end line and are detected inside the zone Exercise Walking (compared to other participants who cross the end line), so the walking speed of these participants is underestimated. This problem is limited with the constraint e_2 used by the M1 method. With the GT method, we underestimate the walking speed (see Fig. 3) because most of participant have reduced their speed before reaching the end line so we have higher values in using the spatial constraint e_1 "Inside Exercise Walking" which defines a zone included between the start and end lines in order to avoid the effect of the walking speed reduction at the end of one walking exercise.
For most of parameters (e.g., execution duration), the proposed automatic video system provides smaller values than GT methods, which is essentially due to the necessary delay of the proposed system to detect the beginning of event. For the duration necessary to perform the U-turn, we have higher values with the proposed automatic video system, which is due to the difference between the way GT has been defined and the models used for event detection.

Ability of system to highlight differences between two profiles of participants Differences in motor disturbances between the two profiles of participants are concordant between the one identified by GT and the one identified by the proposed automatic video system. We have similar results in terms of statistics differences detected from automatic video system and ground truth results performed: walking speed, the duration to execute one position transfer during the transfer position exercise (μ_{e_6}), and D3 (i.e., parameter related to the walking speed in a complex activity).

TP	FP	FN	S+	FPR	PPV
169	1	5	97%	1%	99%

Table 2. Performance system to compute n_{Video,e_6} during transfer position exercise. TP: True Positive; FP: False Positive number; FN: False Negative number; S+: Sensitivity (=TP/(TP+FN)); FPR: False Positive Rate ($FP/(FP+TP)$; PPV: Positive Predictive Value ($= TP/(TP+FP)$)

Mean, ±sd	G1	G2	p-value, [CI(95%)] (&)
Walking Exercise			
Walking speed (cm/s)			
1^{st} exercise (go)			
- *Video, $V_{M1,1}$*	95.8, (21.7)	74.9, (22.3)	0.014, [0.012, 0.016] (*)
- *Video, $V_{M2,1}$*	100.1, (23.7)	75.7, (24.5)	0.013, [0.011, 0.015] (*)
- *GT $V_{GT,1}$*	85.5, (22.3)	66.3, (21.4)	0.034, [0.030, 0.037] (*)
2^{nd} exercise (back)			
- *Video, $V_{M1,2}$*	109.9, (30.7)	88.6, (23.7)	0.089, [0.083, 0.094]
- *Video, $V_{M2,2}$*	113.9, (36.4)	85.8, (25.9)	0.021, [0.018, 0.024] (*)
- *GT $V_{GT,2}$*	94.1 (23.0)	74.4, (19.2)	0.012, [0.010, 0.014] (*)
Mean on the exercise			
- *Video, mean($V_{M1,1}$, $V_{M1,2}$)*	102.9, (24.7)	81.7, (20.5)	0.013, [0.011, 0.015] (*)
- *Video, mean($V_{M2,1}$, $V_{M2,2}$)*	107.0, (28.5)	80.7, (21.5)	0.003, [0.002, 0.004] (*)
- *GT, mean($V_{GT,1}$, $V_{GT,2}$)*	89.8, (21.2)	70.4, (18.1)	0.005, [0.004, 0.006] (*)
Transfer position Exercise			
Execution duration Δt_{e6} (s)			
- Video	13.60, (5.70)	17.43, (6.85)	0.170, [0.162, 0.177]
- GT	15.72, (6.20)	19.96, (7.31)	0.141, [0.135, 0.148]
$\mu_{e6} = \Delta t_{e6} / n_{GT,e6}$			
- Video	1.26, (0.57)	1.71, (0.61)	0.019, [0.016, 0.021] (*)
- GT	1.46, (0.61)	1.95, (0.66)	0.012, [0.010, 0.014] (*)
Up and Go Exercise			
Execution duration (s)			
For walking 3m, D1 (go)			
- *Video*	2.31, (0.96)	2.81, (1.22)	0.058, [0.054, 0.063]
- *GT*	2.67, (1.16)	3.37, (1.43)	0.051, [0.046, 0.055]
For doing the U-turn, D2			
- *Video*	2.38, (1.13)	3.77, (2.65)	0.099, [0.093, 0.104]
- *GT*	1.68, (0.65)	2.56, (1.79)	0.053, [0.049, 0.057]
For walking 3m, D3 (back)			
- *Video*	2.38, (0.81)	3.01, (1.27)	0.020, [0.017, 0.023] (*)
- *GT*	2.95, (0.84)	4.19, (1.90)	0.002, [0.001, 0.003] (*)
Between posture changes, D4			
- *Video*	10.13, (3.95)	12.52, (5.83)	0.130, [0.123, 0.137]
- *GT*	10.19, (3.85)	12.64, (5.97)	0.071, [0.066, 0.076]

Fig. 3. Parameters estimation for healthy elderly participants and AD patients. (&) Non parametric Mann-Withney test was used to compare the results between two groups G1 vs G2. Bilateral p-value associated with the Mann-Whitney test and its 95% confidence interval [CI (95%) were estimated using Monte-Carlo simulation based on a sample size of $10,000$. (*) Intergroup comparisons: differences between healthy elderly participants and AD patients, using a significance level of $.05 (p - value < .05)$.

3.7 Discussion

With the proposed automatic video recognition system, we are able to highlight differences in motor activities between participant profiles. Results show that the walking speed is a sensitive parameter in view of the different results obtained from three computation methods, and by the fact that for a same participant the walking speed differs between the go and go-back exercises (see Fig. 3). Thus, the reliability on this parameter may be discussed. At first for having a more robust estimation, it would be necessary to reproduce the walking exercise on a longer distance, what should limit impact of a short time of execution associated with the short distance to walk on the walking speed computation. Secondly, the automatic video recognition system used 3D spatial information characterizing the tracked participant for computing distance and/or duration when a spatial constraint is satisfied. These spatial attributes on tracked participant are sensitive to illumination changes, segmentation and occlusions issues. These contextual problems may affect the results on walking speed but also on posture recognition and execution duration of activities which are modelled from events represented by a set of spatial and temporal constraints. So it would be interesting to compare the presented results with these one provided by another approach of event recognition using probabilistic reasoning for handling uncertainty in order to validate and/or provide more robust results [8].

4 Conclusion and Future work

We present an automatic video event recognition system that is able to track people and recognize a set of predefined activities. A set of parameters relevant of motor abilities were extracted to compare participant profiles. Results show that we could differentiate between the healthy elderly subjects and AD patients from our automatic video system. Other experiments are planned to deal with other parameters relevant of cognitive disturbances that could enhance the discrimination between healthy and Alzheimer profiles. We plan also to validate the walking speed results with data from an embedded sensor used for this experimentation, and to investigate with a larger number of volunteers for having more robust results.

References

1. Clarkson B, Sawhney N, Pentland A.: Auditory context awareness via wearable computing. In Proceedings of PUI98: The Perceptual User Interfaces Workshop; 1998. p. 37-42.
2. Moore D, Essa I, Hayes M.: Exploiting human actions and object context for recognition tasks. In Proceedings of ICCV99: The 7th IEEE International Conference on Computer Vision; 1999. p. 80-86.
3. Biswas J, et al.: Sensor based micro context for mild dementia assistance. Proceedings of the 3rd International Conference on PErvasive Technologies Related to Assistive Environments (PETRA), 2010.
4. Wang S, et al.: Common sense based joint training of human activity recognizers. n Proceedings of IJCAI 2007: The 20th IEEE International Joint Conference on Artificial Intelligence; 2007. p. 2237-42.
5. Boger J., Poupart P., Hoey J., Boutilier C., Fernie G. and Mihailidis A.: A Decision-Theoretic Approach to Task Assistance for Persons with Dementia. In Proc. of the International Joint Conference on Artificial Intelligence (IJCAI05), Edinburgh, Scotland, (2005), 1293-1299.

6. Mihailidis A, Boger J., Poupart P., Hoey J., Marcelle Canido: The use of an intelligent Prompting System for People with Dementia. Journal interaction, vol 14, no. 4, pp34-37, 2007.
7. Bouchard B., Bouzouane A., and Giroux S.: A Keyhole Plan Recognition Model for Alzheimers Patients: First Results. Jounal Applied Artificial Intelligence, Volume 21 Issue 7, August 2007.
8. Romdhane R., Bremond F., Thonnat M.: A Framework dealing with Uncertainty for Complex Event Recognition. AVSS, 2010.
9. Romdhane R., Mulin E., Derreumeaux A., Zouba N., Piano J., Lee J., Leroi I., Mallea P., Thonnat M., Bremond F., Robert P.H. : Automatic Video Monitoring system for assessment of Alzheimer's Disease symptoms", The Journal of Nutrition, Health and Aging (JNHA), 2011.