Journal of Alzheimer's Disease xx (20xx) x–xx
DOI 10.3233/JAD-141767
IOS Press

Validation of an Automatic Video Monitoring System for the Detection of Instrumental Activities of Daily Living in Dementia Patients

Alexandra König[a,b,*], Carlos Fernando Crispim Junior[d], Alexandre Derreumaux[a], Gregory Bensadoun[a], Pierre-David Petit[a], François Bremond[a,d], Renaud David[a,c], Frans Verhey[b], Pauline Aalten[b] and Philippe Robert[a,c]

[a]*EA CoBTeK, University of Nice Sophia Antipolis, France*
[b]*School for Mental Health and Neuroscience, Alzheimer Center Limburg, Maastricht University Medical Center, Maastricht, The Netherlands*
[c]*Centre Mémoire de Ressources et de Recherche, CHU de Nice, Nice, France*
[d]*INRIA - STARS team - Sophia Antipolis, France*

Accepted 17 September 2014

Abstract. Over the last few years, the use of new technologies for the support of elderly people and in particular dementia patients received increasing interest. We investigated the use of a video monitoring system for automatic event recognition for the assessment of instrumental activities of daily living (IADL) in dementia patients. Participants (19 healthy subjects (HC) and 19 mild cognitive impairment (MCI) patients) had to carry out a standardized scenario consisting of several IADLs such as making a phone call while they were recorded by 2D video cameras. After the recording session, data was processed by a platform of video signal analysis in order to extract kinematic parameters detecting activities undertaken by the participant. We compared our automated activity quality prediction as well as cognitive health prediction with direct observation annotation and neuropsychological assessment scores. With a sensitivity of 85.31% and a precision of 75.90%, the overall activities were correctly automatically detected. Activity frequency differed significantly between MCI and HC participants ($p < 0.05$). In all activities, differences in the execution time could be identified in the manually and automatically extracted data. We obtained statistically significant correlations between manually as automatically extracted parameters and neuropsychological test scores ($p < 0.05$). However, no significant differences were found between the groups according to the IADL scale. The results suggest that it is possible to assess IADL functioning with the help of an automatic video monitoring system and that even based on the extracted data, significant group differences can be obtained.

Keywords: Alzheimer's disease, assessment, autonomy, dementia, mild cognitive impairment, information and communication technologies, instrumental activities of daily living, video analyses

INTRODUCTION

The increase of persons with dementia is accompanied by the need to identify methods that allow for an easy and affordable detection of decline in functionality in the disorder's early stages. Consequently, the development of computerized assessment systems for

*Correspondence to: Alexandra König, School for Mental Health and Neuroscience, Alzheimer Center Limburg, Maastricht, EA CoBTek - Centre Mémoire de Ressources et de Recherche, Institut Claude Pompidou, 10 Rue Molière, 06100 Nice, France. Tel.: +33 0 4 92 03 47 70; Fax: +33 0 4 92 03 47 72; E-mail: a.konig@maastrichtuniversity.nl.

ISSN 1387-2877/14/$27.50

the elderly is of high interest, and represents a promising new research domain that aims to provide clinicians with assessment results of higher ecological validity.

Dementia is one of the major challenges affecting the quality of life of the elderly and their caregivers. Progressive decline in cognitive function represents a key symptom and results often in the inability to perform activities of daily living (ADL) and instrumental activities of daily living (IADL) [1] such as managing finances or cooking.

Many efforts are currently being undertaken to investigate dementia pathology and develop efficient treatment strategies considering its rapidly increasing prevalence. Mild cognitive impairment (MCI) [2–4] is considered as a pre-dementia stage for Alzheimer's disease (AD), as many MCI patients convert to AD over time [5]. Studies show that impairment in complex functional tasks, notably due to slower speed of execution [6], may already be detectable in the early stages of cognitive decline and therefore gradually becomes an important target in clinical assessments [7, 8]. Rating scales and questionnaires constitute the essential tools for the assessment and monitoring of symptoms, treatment effects, as well as (I)ADL functioning.

Nevertheless, changes in (I)ADL functioning observed in MCI may be too subtle to be detected by traditional measures assessing global ADLs [9, 10]. Thus, standard tools are limited to some extent in ecological validity, reproducibility, and objectivity [11]. They do not fully capture the complexity of a patient's cognitive, behavioral, and functional statuses, which do not always evolve in parallel but rather idiosyncratically.

To overcome these problems, Schmitter-Edgecombe et al. developed a naturalistic task in a real world setting to examine everyday functioning in individuals with MCI using direct observation methods [12]. However, this method can also suffer from possible observation biases and difficulties in reproducibility.

For this reason, information and communication technology (ICT) involving imaging and video processing could be of interest by adding more objectively measured data to the diagnostic procedure. Functionality in (I)ADL, which is very closely linked to executive functions [13, 14], may be reflected in activity patterns measurable through computerized systems such as automatic video detection of activities.

Dawadi et al. showed that it is possible to automatically quantify the task quality of daily activities and to perform limited assessment of the cognitive functioning of individuals in a 'smart' home environment (equipped with various sensors) as long as the activities are properly chosen and the learning algorithms are appropriately trained [15]. Sablier and colleagues developed a technological solution designed for people with difficulties managing ADL, providing a schedule manager as well as the possibility to report occurrences of experiences of symptoms such as depression and agitation [16]. However, indicators of cognitive functioning and autonomy were measured using a test battery and scales [16]. Okahashi et al. created a Virtual Shopping Test—using virtual reality technology to assess cognitive functions in brain-injured patients—correlating variables on the virtual test with scores of conventional assessments of attention and memory [17]. Similar work has been done by Werner et al. using a virtual action planning Supermarket game for the diagnosis of MCI patients [18].

Along this line, a project was launched under the name Sweet-HOME (2012), defining a standardized scenario where patients are asked to carry out a list of autonomy relevant (I)ADLs, such as preparing tea, making a phone call, or writing a check, in an experimental room equipped with video sensors. Within this project, Sacco et al. performed a functional assessment with the help of visual analyses by computing a DAS (Daily Activity Scenario) score able to differentiate MCI from healthy control (HC) subjects [19]. However, analysis was based purely on annotations made by a direct observer, and therefore still risked lack of objectivity and reliability. Automatic, computer-based video analysis, which allows for the recognition of certain events and patients' behavioral patterns, may offer a new solution to the aforementioned assessment problems.

To date, automatic video event recognition has been employed in clinical practice simply for feasibility studies with small samples [20–22]. Banerjee et al. presented video-monitoring for fall detection in hospital rooms by extracting features from depth information provided by a camera [23]. Wang et al. used automatic vision analyses for gait assessment using two cameras to differentiate between the gait patterns of residents participating in realistic scenarios [22].

In order to further evaluate the potential contribution of such technologies for clinical practice, this study aims to validate the use of automatic video analyses for the detection of IADL performance within a larger group of MCI patients and HC subjects carrying out a predefined set of activities. More specifically, the objectives of the study are (1) to compare IADL performances of elderly HC subjects and patients with MCI in a predefined scenario; (2) to compare automatically extracted video data with so-called 'ground-truth'

(GT) annotations made manually by a human observer; and (3) to assess the importance of automatic video analyses data for the differentiation between the two populations. As a secondary objective, we investigate the relationship between the participants' performance in the scenario and the results of classical neuropsychological testing, in order to verify whether or not the performance in the created scenario is associated with the status of cognitive functioning.

We expect automatically extracted video detection to achieve results as GT annotations when differentiating between the MCI group and the HC group. We also hypothesize that individuals with MCI will perform poorer in the predefined IADL scenario than HC subjects and that difficulties in executive functioning will be related to the amount of completed activities. Further, we expect a significant relationship between the video captured performance in the scenario and the classical neuropsychological test results such as the Frontal Assessment Battery (FAB) [24] or the Mini-Mental State Examination (MMSE) [25] and IADL scales [26].

METHODS

Participants

The study was approved by the local Nice ethics committee and only participants with the capacity to consent to the study were included. Each participant gave informed consent before the first assessment. Participants aged 65 or older were recruited at the memory center in Nice located at the Geriatric Department of the University Hospital. For the MCI group, patients with a MMSE score higher than 24 were included using the Petersen clinical criteria [4]. Participants were excluded if they had any history of head trauma, loss of consciousness, psychotic aberrant motor behavior, or a score higher than 0 on the Unified Parkinson's Disease Rating scale (UPDRS) [27] in order to control for any possible motor disorders influencing the ability to carry out IADLs.

Assessments

Participants were administered a cognitive and behavioral examination prior to completing the video monitoring session. General cognitive status was assessed using neuropsychological tests including: MMSE [25], Frontal Assessment Battery (FAB) [24], Instrumental Activities of Daily Living scale (IADL-E) [28], Montgometry-Asberg Depression Rating Scale (MADRS) [29], and Geriatric Depression Scale (GDS) to assess depression levels [30]. Additionally, neuropsychiatric symptoms were assessed using the Neuropsychiatric Inventory Scale (NPI) [31].

Clinical scenario: The ecological assessment

The ecological assessment of IADLs was conducted in an observation room located in the Nice Research Memory Center. This room was equipped with everyday objects for use in ADLs and IADLs, e.g., an armchair, a table, a tea corner, a television, a personal computer, and a library. Two fixed monocular video cameras (eight frames per second) were installed to capture the activity of the participants during the experiment. Using an instruction sheet, participants had to carry out 10 daily-living-like activities, such as making a phone call or preparing a pillbox, in a particular order within a timeframe of 15 min (Table 1). The aim of this ecological assessment of autonomy was to determine to which extent the participant could undertake a list of daily activities with respect of some constraints after being given a set of instructions. After each participant carried out the scenario, a clinician verified the amount of activities initiated and carried out completely and correctly, as well as repetitions and omissions. The information was manually annotated and entered into the database via a tablet. The scenario was recorded using a 2D-RGB video camera (AXIS, Model P1346, 8 frames per second) and a RGB-D camera (Kinect, Microsoft).

Table 1
List of the activities proposed to the patient during the ecological assessment

	Daily Living scenario associated with the protocol
Activities	« Your task is to perform this list of 10 activities in a logical manner within 15 minutes. These 15 minutes represent a typical morning period of everyday life. »
	– Read the newspaper
	– Water the plant
	– Answer the phone
	– Call the taxi
	– Prepare today's medication
	– Make the check for the Electricity Company
	– Leave the room when you have finished all activities
	– Watch TV
Constraints	– Prepare a hot tea
	– Write a shopping list for lunch
	1. Watch TV before the phone call
	2. Water the plant just before leaving the room
	3. Call the taxi which will take 10 min to arrive and ask the driver to bring you to the market

For a more detailed analysis, the main focus was placed particularly on three IADLs, namely preparing a pillbox, making a phone call, and preparing tea, because they fall within the commonly used IADL-Lawton scale, and are the most challenging activities for appropriately representing a patient's general autonomy level. However, all other activities were included in the overall IADL assessment procedure and analyses.

Automatic video monitoring system and event recognition

In the first step, after each assessment, a clinician manually gathered data of the amount of activities carried out by the participants. This included parameters such as activity occurrence, activity initiation, and the number of activities carried out completely and correctly. In the next step, a computer vision algorithm was used to automatically extract different parameters representing movement patterns of the participants during the ecological assessment period.

The Automatic Video Monitoring System (AVMS) herein used has been fully described [32]. It is composed of two main modules: the vision and the event recognition. The vision module is responsible for detecting and tracking people on the scene. The event recognition module uses the generic constraint-based ontology language proposed by Zouba et al. [33] for event modeling and the reasoning algorithm proposed by Vu and colleagues [34] to describe and detect the activities of daily living of interest in this study.

The vision module detects people in the scene using an extension of the Gaussian Mixture Model algorithm for background subtraction proposed by Nghiem et al. [35]. People tracking over time is performed by a multi-feature algorithm proposed by Chau et al. using features such as 2D size, 3D displacement, color histogram, and dominant color. The detected people and their tracking information (their current and previous positions in the scene) are then passed to the event recognition module [36].

The event recognition module is composed of a framework for event modeling and a temporal scenario recognition algorithm which assess whether the constraints defined in the event models are satisfied [34]. Event models are built taking into account *a priori* knowledge of the experimental scene and attributes dynamically obtained by the vision module. Event modeling follows a declarative and intuitive ontology-based language that uses natural terminology to allow end users (e.g., medical experts) to easily add and modify the models. The *a priori* knowledge consists of a decomposition of a 3D projection of the room's floor plan into a set of spatial zones that have semantic information regarding the events of interest (e.g., TV position, armchair position, desk position, tea preparation). The ontology employed by the system hierarchically categorizes event models according to their complexity, described here in ascending order:

- **Primitive State** models an instantaneous value of a property of a person (posture or position inside a certain zone.
- **Composite State** refers to a composition of two or more primitive states.
- **Primitive Event** models a change in a value of person's property (e.g., change in posture to model whether or not a person changes from a Sitting to a Standing state).
- **Composite Event** refers to the composition of two of the previous event model types in terms of a temporal relationship (e.g., Person changes from Sitting to Standing posture before Person in Corridor).

IADL modeling

The semantic information of the observation room where patients conducted the activities of daily living was defined. Contextual or Semantic Elements were defined at the locations where the activities of daily living would be carried out (e.g., telephone zone at top-left corner, tea and plant zones at top-right corner, and pharmacy zone at bottom-left corner).

The activity modeling was performed with the support of domain experts. The models were mostly made taking into account one or more of the following constraints: the presence of the person in a specific zone, their posture, and their proximity to the object of daily living (when static, e.g., the telephone). These constraints were defined as primitive state models. The combination of these models, along with their temporal order, was defined as a composite event. Duration constraints were also used to establish a minimum time of execution for the whole or sub-components of the composite event.

Statistical analysis

Spearman's correlations were performed to determine the association between the extracted video parameters and the established assessment tools in particular for executive functioning, e.g., the FAB.

Comparison between the two groups (i.e., MCI patients and HC subjects) was performed with a Mann-Whitney test for each outcome variable of the automatic video analyses. Differences were reported as significant if $p < 0.05$.

Automatic activity recognition evaluation

The evaluation compared the performance of the AVMS at automatically detecting IADL with respect to the annotations manually made by human experts. The AVMS performance was measured based on the indices of recall and precision, described in Equations 1 and 2, respectively. Recall index measures the percentage of how many of the targeted activities have been detected compared to how many existed. Precision index evaluates the performance of the system at discriminating a targeted activity type from others.

1. Recall = TP/(TP+FN) 2. Precision = TP/(TP+FP)

TP: True Positive rate, FP: False Positive rate, FN: False Negative rate.

RESULTS

Population

19 MCI patients (age = 75.2 ± 4.25) and 19 HC (age = 71.7 ± 5.4) were included. Table 2 shows the clinical and demographic data of the participants. Significant intergroup differences in demographic factors (gender and age) were not seen. However, significant differences were found between for the MMSE score, with a mean of 25.8 (±2.2) for the MCI group and 28.8 (±1.0) for the HC group (p, 0.001), as well as for the FAB score with a mean of 14.16 (±1.92) for the MCI group and 16.2 (±1.44) for the HC group. The mean IADL-E scores did not differ between groups, with a mean IADL-E score of 9.9 (±1.7) for the MCI group and 9.6 (±1.1) for the HC group.

Automatic video monitoring results versus ground-truth annotation

The participants performed differently on the IADL scenario according to their diagnostic group; in all three activities (preparing the pillbox, preparing tea, and making/receiving a phone call), the obtained parameters (manually as automatic) showed variations. All results are presented in detail in Table 3. The total frequency of activities as well as the number of correctly completed activities according to manual annotations differed significantly between MCI and HC groups ($p < 0.05$). Two activities, namely preparing the pillbox and making/receiving the phone call, generally took the MCI participants a longer time to carry out. In turn, for the activity of preparing tea, HC participants took a longer time. The same trends, even if not significant, were detected as well by the automatic video analyses; a significant difference was found between MCI and HC groups ($p < 0.05$) in the phone call time. Furthermore, MCI and HC participants differed in the total amount of detected activities carried out; the same activities, preparing the pillbox and making/receiving a phone call took longer for MCI

Table 2
Characteristics of the participants

Characteristics	HC group $n = 19$	MCI group $n = 19$	p
Female, n (%)	15 (78.9%)	9 (47.4%)	0.091
Age, years mean ST	71.7 ± 5.37	75.2 ± 4.25	0.07
Level of Education, n (%)			
Unknown	2 (10.5%)	2 (10.5%)	1
No formal education	0 (0%)	0 (0%)	–
Elementary school	1 (5.3%)	5 (26.3%)	0.405
Middle school	4 (21.0%)	7 (36.8%)	0.269
High school	4 (21.0%)	3 (15.8%)	1
Post-secondary education	8 (42.1%)	2 (10.5%)	0.062
MMSE, mean ± SD	28.8 ± 1.03	25.8 ± 2.22	**0.001****
FAB, mean ± SD	16.2 ± 1.44	14.16 ± 1.92	**0.002***
IADL-E, mean ± SD	9.6 ± 1.12	9.9 ± 1.73	0.488
NPI total, mean ± SD	0.42 ± 1.43	6.16 ± 6.73	**0.00***

Data shown as mean ± SD. Bold characters represent significant p-values <0.05. Scores on the Mini Mental State Examination (MMSE) range from 0 to 30, with higher scores indicating better cognitive function; Scores on the Instrumental Activities of Daily Living for Elderly (IADL-E) range from 0 to 36, with lower score indicating a better functional independency; Scores on the Montgomery Asberg Depression Rating Scale (MADRS) range from 0 to 60 (10 items range from 0 to 6), with higher scores indicating depressive state; Scores on the Geriatric Depression Scale (GDS) range from 0 to 30, with higher scores indicating depressive state. HC, healthy control; MCI, mild cognitive impairment.

Table 3
Comparison of parameters from video analyses between groups

Video analyses data	HC $n = 19$	MCI $n = 19$	p
Manually annotated:			
Activities carried out completely and correctly†	9.68 ± 0.48	8.21 ± 1.48	**0.00***
Activity frequency total‡ (activities initiated)‡	11.74 ± 2.62	9.58 ± 1.89	**0.007***
Preparing Pillbox (f)	1.05 ± 0.23	0.89 ± 0.32	0.086
Preparing Pillbox time	41.17 ± 17.04	46.17 ± 31.18	0.609
Making tea (f)	2.68 ± 0.82	2 ± 1	0.068
Making tea time	41.21 ± 30.60	32.16 ± 35.3	0.175
Phone call (f)	2 ± 0.47	2.21 ± 0.53	0.198
Phone call time	66.61 ± 21.75	83.30 ± 30.96	0.118
Automatically extracted: Activity frequency total	13.26 ± 3.89	10.95 ± 3.15	0.056
Preparing Pillbox (f)	1.05 ± 0.23	1.17 ± 0.38	0.271
Preparing Pillbox time	47.64 ± 22.28	70.26 ± 38.01	0.204
Making tea (f)	2.74 ± 1.33	2.12 ± 1.22	0.136
Making tea time	102.3 ± 77.3	79.57 ± 40.92	0.531
Phone call (f)	1.95 ± 0.52	2.17 ± 0.79	0.38
Phone call time	60.32 ± 21.52	112.61 ± 46.31	**0.000***

Mann-Whitney test: $^{*}p < 0.05$, $^{**}p < 0.01$ HC, healthy control; MCI, mild cognitive impairment; (f), mean frequency of detected event; †Represents the total amount of completely carried out activities without a mistake, ‡Represents the total of simply initiated activities which are not always necessarily accomplished completely and without mistakes.

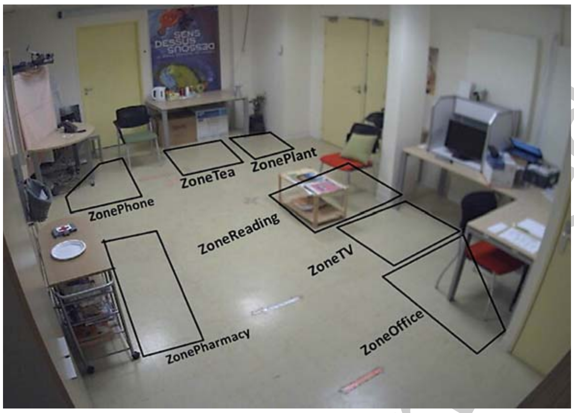

Fig. 1. The experimental room for the IADL assessment. For the automatic activity detection, the room was divided in different zones according to the designated IADL.

participants whereas making tea took longer for the HC group.

According to the amount of carried out activities and rapidity, the best and worst performers were determined in each group. Next, we investigated if participants that performed well showed a difference in the parameters extracted from the automated video analyses compared to participants that did not perform as well on the assessment. This, in turn, could help establish diagnostic-specific profiles of IADL functioning. The results are presented in Fig. 2.

Moreover, the manually and automatically extracted video data parameter ‘activity frequency’ correlated significantly with neuropsychological test results namely the MMSE ($p < 0.01$) and FAB score ($p < 0.05$). The obtained correlation analyses results are presented in Table 4. Particularly, from the manually annotated parameters, the time spent to prepare the pillbox correlated significantly negatively with the MMSE scores ($p < 0.01$), whereas the time spent to make a phone call correlated significantly negatively with the FAB scores ($p < 0.05$). The mean frequency of the activity ‘making tea’ correlated significantly positively with the FAB scores ($p < 0.05$). From the automatically extracted parameters, the detected time spent to prepare the pillbox ($p < 0.01$) and to make the phone call ($p < 0.05$) correlated significantly negatively with the MMSE scores. None of the extracted parameters correlated with the IADL-E scores.

Automatic video monitoring results: Experimental results

Table 5 presents the results of the evaluation of the AVMS with respect to its accuracy at detecting the number of activities of daily living annotated by domain experts while watching the experiment video.

From all 10 proposed activities, ‘Reading’ was detected automatically with the highest precision of 91.30%, followed by ‘Preparing pillbox’ with 90.24%, and ‘Making phone call’ with 89.47%.

DISCUSSION

The presented study demonstrates the additional value of employing new technologies such as automatic video monitoring system in clinical practice for

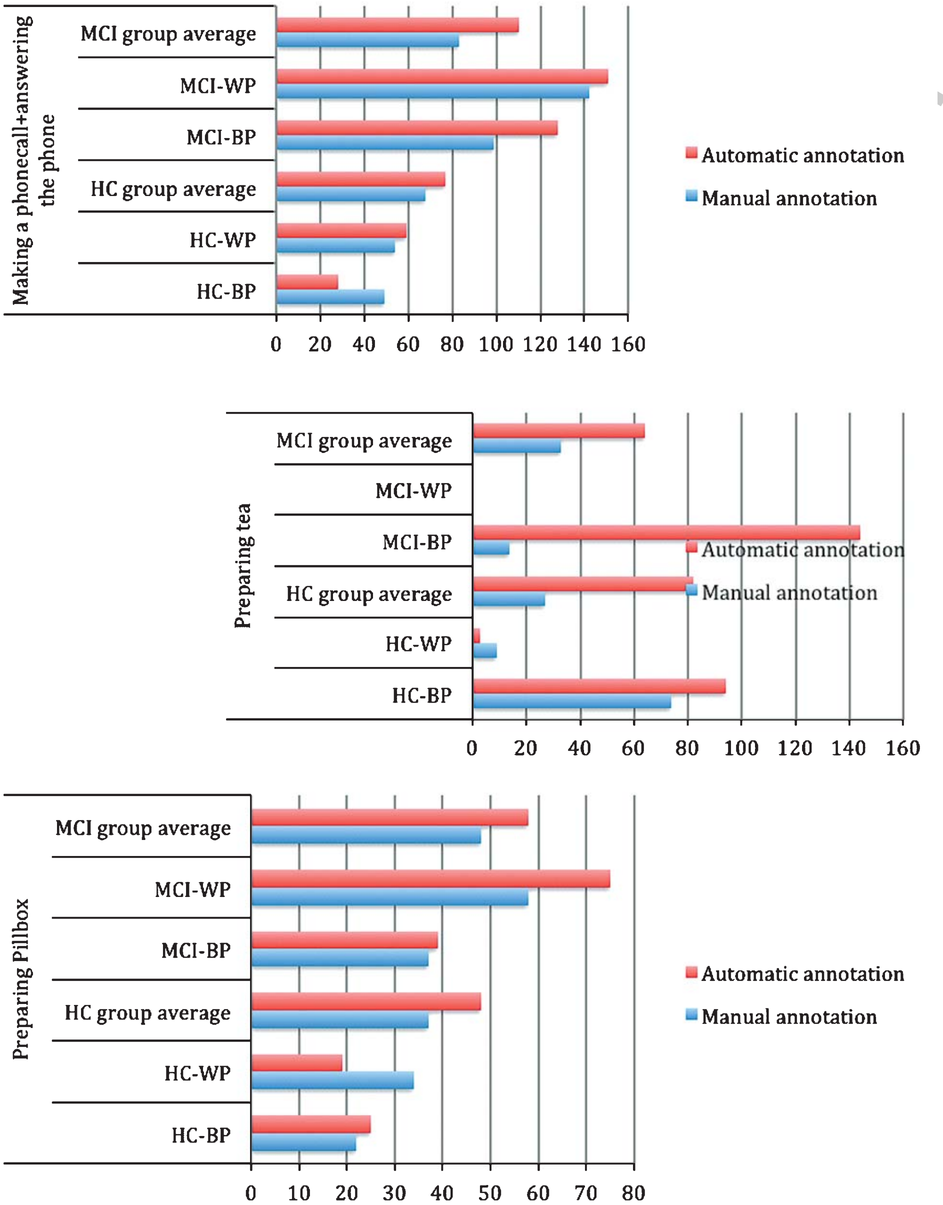

Fig. 2. The average execution times for each activity in blue annotated manually and in red detected automatically. MCI, mild cognitive impairment; WP, worst performer; BP, best performer; HC, healthy control.

the assessment of (I)ADL in dementia patients. The two main goals of the study were (1) to investigate if differences in IADL functioning can be detected between MCI and HC and (2) to compare between manual and automated assessments of IADL performances in contrast to standard paper scales.

The obtained results demonstrate that significant group differences between MCI and HC participants (even with just a small sample size) can be detected when using such techniques, and this when regular assessment tools such as the IADL-E questionnaire lack sensitivity to detect these group differences. A detection accuracy of up to 90% for the ‘Preparing pillbox’ activity has been achieved validating clearly the use of AVMS for evaluation and monitoring purposes. Furthermore, the correlation analyses demonstrated that extracted parameters, particularly execution times of activities, correlated significantly with neuropsy-

Table 4
Correlation between automatic video parameters, manually annotated parameters and conventional cognitive assessments (Spearman's correlation coefficient)

Video analyses data Spearman correlation coefficient (r) / p-values	MMSE	FAB	IADL-E
Manually annotated	*(r)*	*(r)*	*(r)*
Activities frequency	0.491**	0.394*	–0.035
	$p = 0.002$	$p = 0.014$	$p = 0.834$
Activities completed correctly	0.819**	0.660**	–0.107
	$p = 0.000$	$p = 0.000$	$p = 0.522$
Automatically extracted			
Activity frequency	0.415**	0.273*	−0.071
	$p = 0.005$	$p = 0.048$	$p = 0.337$
Manually annotated			
Preparing Pillbox (f)	0.055	0.299	−0.149
	$p = 239$	$p = 0.063$	$p = 0.127$
Preparing Pillbox time	−0.468**	−0.114	−0.179
	$p = 0.001$	$p = 0.409$	$p = 0.211$
Making tea (f)	0.27	0.363*	−0.5
	$p = 0.083$	$p = 0.042$	$p = 0.391$
Making tea time	–0.143	0.053	−0.002
	$p = 0.222$	$p = 0.343$	$p = 0.396$
Phone call (f)	−0.123	−0.235	0.002
	$p = 0.128$	$p = 0.084$	$p = 0.465$
Phone call time	−0.280*	−0.332*	−0.145
	$p = 0.044$	$p = 0.041$	$p = 0.291$
Automatically extracted			
Preparing Pillbox (f)	–0.287*	−0.073	0.125
	$p = 0.043$	$p = 0.295$	$p = 0.222$
Preparing Pillbox time	−0.618**	−0.241	−0.05
	$p = 0.001$	$p = 0.340$	$p = 0.128$
Making tea (f)	0.223	0.221	−0.264
	$p = 0.60$	$p = 0.083$	$p = 0.051$
Making tea time	0.016	−0.101	−0.114
	$p = 0.392$	$p = 0.261$	$p = 0.197$
Phone call (f)	−0.248	0.077	0.158
	$p = 0.095$	$p = 0.330$	$p = 0.223$
Phone call time	−0.373*	−0.277*	−0.054
	$P = 0.002$	$p = 0.049$	$p = 0.451$

*$p < 0.05$, **$p < 0.01$.

Table 5
Activity/Event detection performance

Activity	Recall	Precision
Phone call	85	89.47
Watching TV	83.33	73.77
Making tea	80.9	80
Preparing Pillbox	100	90.24
Watering Plant	75	61.22
Reading	75	91.3
Average Recognition	85.31	75.9

n: 38, MCI: 19 / HC: 19.

chological tests results, namely the MMSE and FAB scores.

The study's results were consistent with those previously presented in [32], where a recall of 88.30 and a precision of 71.23 were demonstrated. Although our evaluation results were obtained from different patients and from a larger cohort, small differences were observed in precision index which is higher by ∼5%, and in the recall index which is lower by 3%. These differences are a result of a trade-off between AVMS precision and recall performance due to a refinement of the event-modeling step. By opting for more strict constraints in such models, we make the system less prone to errors such as misleading evidence. For instance, instead of patients walking toward the plant to water it, they just stretch from the tea table to do so, as this table is just beside the plant.

Activities where the AVMS presented lower precision refer to at least one of two factors: participants performing the activity far from the camera and/or noise from low-level vision components of the AVMS. For example, a few patients stopped close by or inside the activity zones for long periods to read the instructions sheet, which caused false-positive detections of the zone-related activities. In addition, noisy data from low-level vision components sometimes shifted the estimation of the position of participants from their actual location to an activity zone close by, mostly when the participants were far from the camera. For the described problems, possible solutions include the adoption of a probabilistic framework to handle noise and event modeling uncertainty, and a multi-sensor approach for cases where the activities are mis-detected by a lack of view of the participants.

If we try to interpret the results, it is not surprising that MCI participants carried out fewer activities in general and took more time, especially for preparing the pillbox and the phone call, which was detected by the observer as well as by the automatic video analysis. Recent studies demonstrated that even in MCI patients, difficulties in the execution of complex IADL tasks, could be observed and linked to possible early impairment of executive function [8]. This is further in line with our finding of significant group differences in the studied population (see Table 2) on the FAB, a test that specifically measures levels of executive functioning.

Interestingly, the preparing tea activity took longer for HC participants and can be explained by the fact that, for the most part, they correctly completed this activity (which takes at least a minimum of 60 s), whereas MCI patients initiated this activity but did not always finish it completely. Therefore, their execution time was shorter but may serve as an indicator of poor task performance.

One major drawback of this study was that healthy control subjects were recruited through the Memory Clinic and therefore suffered in most cases from subjective memory complaints. However, according to classical assessment tools and diagnostic manual they were cognitively healthy. Thus it is debatable whether or not to classify them as healthy controls, as the MMSE and FAB mean scores for that group were relatively low. Furthermore, the study was only based on a small population size. This does not mean that the chosen parameters were not helpful indicators, and they should be validated with a larger population in the future, potentially combined with other ICT data such as actigraphy [37] or automatic speech analyses [38], given the fact that certain significant group differences could be observed.

It can be further argued that the experiment was conducted in an artificial laboratory environment and not in a complete natural setting such as a patient's home. This could have had increased the stress level of the participants and consequently an impact on their IADL performance. It is therefore desirable in the future to conduct this type of assessment in more naturalistic settings, but that may also represent a less controlled environment and therefore a bigger challenge from a technical point of view. Finally, the current study placed less emphasis on multi-tasking in IADL performances, but rather focused more on the simple execution of tasks sequentially. However, in real life, multi-tasking is of great importance and represents complex cognitive processing required for functional ability.

It is important to mention that in the field of automatic video analysis, it is almost impossible to achieve 100% accuracy in the activity recognition, often caused as well by inaccurate manual annotations. The challenge is to define, for example, the beginning and the end of an activity, which represents a common problem in video analyses. Nevertheless, the activity detection by video analyses might be actually a much closer representation of the reality and the real events happening than annotations of a human observer because the latest can be influenced by various confounding factors such as fatigue, distraction, lack of concentration, etc.

The advantages of using such techniques are that the application in daily practice is easy and reproducible, and add an objective measure to the assessment of autonomy. Furthermore, this evaluation provides quicker results than manual annotations and could be even used as an outcome measure in clinical trials in order to evaluate the effect of certain treatments (pharmacological and non-pharmacological) on the functioning of IADLs of patients.

Overall, the study showed in particular that manually annotated data gives a more accurate picture of a patient's status to date, and is better validated by traditional diagnostic and neuropsychological assessment tools. This means that qualitative assessments still seem to better correlate with conventional scoring than quantitative video extracted parameters. Until now, the obtained data still needs interpretation of an experienced clinician regarding the quality of the carried out activities. It should be emphasized that this cannot be replaced by technology and is not the objective of this research.

However, in future studies, we aim for improvement in the activity detection with a larger group sample, in particular to improve the detection of the quality of activity execution, i.e., if an activity was carried out successfully and completely.

ACKNOWLEDGMENTS

This study was supported by grants from the ANR-09-TECS-016-01 – TecSan – SWEET HOME, the

FP7 Dem@care project, by the Innovation Alzheimer associations, by the CoBTek (Cognition – Behaviour – Technology) Research Unit from the Nice Sophia-Antipolis University (UNS), the CMRR Nice team and by the platform patients of the Nice CHU member of the CIU-S.

Authors' disclosures available online (http://www.j-alz.com/disclosures/view.php?id=2560).

REFERENCES

[1] Reppermund S, Brodaty H, Crawford JD, Kochan NA, Draper B, Slavin MJ, Trollor JN, Sachdev PS (2013) Impairment in instrumental activities of daily living with high cognitive demand is an early marker of mild cognitive impairment: The Sydney memory and ageing study. *Psychol Med* **43**, 2437-2445.

[2] Albert MS, DeKosky ST, Dickson D, Dubois B, Feldman HH, Fox NC, Gamst A, Holtzman DM, Jagust WJ, Petersen RC, Snyder PJ, Carrillo MC, Thies B, Phelps CH (2011) The diagnosis of mild cognitive impairment due to Alzheimer's disease: Recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* **7**, 270-279.

[3] Artero S, Petersen R, Touchon J, Ritchie K (2006) Revised criteria for mild cognitive impairment: Validation within a longitudinal population study. *Dement Geriatr Cogn Disord* **22**, 465-470.

[4] Petersen RC, Smith GE, Waring SC, Ivnik RJ, Tangalos EG, Kokmen E (1999) Mild cognitive impairment: Clinical characterization and outcome. *Arch Neurol* **56**, 303-308.

[5] Morris JC, Cummings J (2005) Mild cognitive impairment (MCI) represents early-stage Alzheimer's disease. *J Alzheimers Dis* **7**, 235-239; discussion 255-262.

[6] Wadley VG, Okonkwo O, Crowe M, Ross-Meadows LA (2008) Mild cognitive impairment and everyday function: Evidence of reduced speed in performing instrumental activities of daily living. *Am J Geriatr Psychiatry* **16**, 416-424.

[7] Gold DA (2012) An examination of instrumental activities of daily living assessment in older adults and mild cognitive impairment. *J Clin Exp Neuropsychol* **34**, 11-34.

[8] Marshall GA, Rentz DM, Frey MT, Locascio JJ, Johnson KA, Sperling RA, Alzheimer's Disease Neuroimaging I (2011) Executive function and instrumental activities of daily living in mild cognitive impairment and Alzheimer's disease. *Alzheimers Dement* **7**, 300-308.

[9] Burton CL, Strauss E, Bunce D, Hunter MA, Hultsch DF (2009) Functional abilities in older adults with mild cognitive impairment. *Gerontology* **55**, 570-581.

[10] Jefferson AL, Byerly LK, Vanderhill S, Lambe S, Wong S, Ozonoff A, Karlawish JH (2008) Characterization of activities of daily living in individuals with mild cognitive impairment. *Am J Geriatr Psychiatry* **16**, 375-383.

[11] Sikkes SA, de Lange-de Klerk ES, Pijnenburg YA, Scheltens P, Uitdehaag BM (2009) A systematic review of Instrumental Activities of Daily Living scales in dementia: Room for improvement. *J Neurol Neurosurg Psychiatry* **80**, 7-12.

[12] Schmitter-Edgecombe M, McAlister C, Weakley A (2012) Naturalistic assessment of everyday functioning in individuals with mild cognitive impairment: The day-out task. *Neuropsychology* **26**, 631-641.

[13] Nelson AP, O'Connor MG (2008) Mild cognitive impairment: A neuropsychological perspective. *CNS Spectr* **13**, 56-64.

[14] Razani J, Casas R, Wong JT, Lu P, Alessi C, Josephson K (2007) Relationship between executive functioning and activities of daily living in patients with relatively mild dementia. *Appl Neuropsychol* **14**, 208-214.

[15] Dawadi PN, Cook DJ, Schmitter-Edgecombe M, Parsey C (2013) Automated assessment of cognitive health using smart home technologies. *Technol Health Care* **21**, 323-343.

[16] Sablier J, Stip E, Jacquet P, Giroux S, Pigot H, Franck N, Mobus G (2012) Ecological assessments of activities of daily living and personal experiences with Mobus, an assistive technology for cognition: A pilot study in schizophrenia. *Assist Technol* **24**, 67-77.

[17] Okahashi S, Seki K, Nagano A, Luo Z, Kojima M, Futaki T (2013) A virtual shopping test for realistic assessment of cognitive function. *J Neuroeng Rehabil* **10**, 59.

[18] Werner P, Rabinowitz S, Klinger E, Korczyn AD, Josman N (2009) Use of the virtual action planning supermarket for the diagnosis of mild cognitive impairment: A preliminary study. *Dement Geriatr Cogn Disord* **27**, 301-309.

[19] Sacco G, Joumier V, Darmon N, Dechamps A, Derreumaux A, Lee JH, Piano J, Bordone N, Konig A, Teboul B, David R, Guerin O, Bremond F, Robert P (2012) Detection of activities of daily living impairment in Alzheimer's disease and mild cognitive impairment using information and communication technology. *Clin Interv Aging* **7**, 539-549.

[20] Romdhane R, Mulin E, Derreumeaux A, Zouba N, Piano J, Lee L, Leroi I, Mallea P, David R, Thonnat M, Bremond F, Robert PH (2012) Automatic video monitoring system for assessment of Alzheimer's disease symptoms. *J Nutr Health Aging* **16**, 213-218.

[21] Stone EE, Skubic M (2012) Capturing habitual, in-home gait parameter trends using an inexpensive depth camera. *Conf Proc IEEE Eng Med Biol Soc* **2012**, 5106-5109.

[22] Wang F, Stone E, Dai W, Banerjee T, Giger J, Krampe J, Rantz M, Skubic M (2009) Testing an in-home gait assessment tool for older adults. *Conf Proc IEEE Eng Med Biol Soc* **2009**, 6147-6150.

[23] Banerjee T, Keller JM, Skubic M (2012) Resident identification using kinect depth image data and fuzzy clustering techniques. *Conf Proc IEEE Eng Med Biol Soc* **2012**, 5102-5105.

[24] Dubois B, Slachevsky A, Litvan I, Pillon B (2000) The FAB: A Frontal Assessment Battery at bedside. *Neurology* **55**, 1621-1626.

[25] Folstein MF, Folstein SE, McHugh PR (1975) Mini-mental state. A practical method for grading the cognitive state of patients for the clinician. *J Psychiatr Res* **12**, 189-198.

[26] Lawton MP, Brody EM (1969) Assessment of older people: Self-maintaining and instrumental activities of daily living. *Gerontologist* **9**, 179-186.

[27] Fahn S, Elton RL (1987) UPDRS program members. Unified Parkinson's Disease Rating Scale. In *Recent developments in Parkinson's disease*, Fahn S MC, Goldstein M, Calne DB, ed. Macmillan Healthcare Information, Florham Park, NJ, pp. 153-163.

[28] Mathuranath PS, George A, Cherian PJ, Mathew R, Sarma PS (2005) Instrumental activities of daily living scale for dementia screening in elderly people. *Int Psychogeriatr* **17**, 461-474.

[29] Montgomery SA, Asberg M (1979) A new depression scale designed to be sensitive to change. *Br J Psychiatry* **134**, 382-389.

[30] Yesavage JA, Brink TL, Rose TL, Lum O, Huang V, Adey M, Leirer VO (1982) Development and validation of a geriatric depression screening scale: A preliminary report. *J Psychiatr Res* **17**, 37-49.

[31] D'Arcy S (2008) Speech as a means of monitoring cognitive function of elderly subjects. *Interspeech* (Brisbane, Australia).

[32] Crispim-Junior CF, Bathrinarayanan V, Fosty B, Konig A, Romdhane R, Thonnat M, Bremond F (2013) Evaluation of a monitoring system for event recognition of older people. *10th IEEE International Conference on Advanced Video and Signal-Based Surveillance* (Krakow, Poland).

[33] Zouba N, Bremond F, Thonnat M (2010) An activity monitoring system for real elderly at home: Validation study. *7th IEEE International Conference on Advanced Video and Signal-Based Surveillance* (Boston, USA).

[34] Vu T, Brémond F, Thonnat M (2003) Automatic video interpretation: A novel algorithm for temporal scenario recognition. *The Eighteenth International Joint Conference on Artificial Intelligence (IJCAI'03)* (Acapulco, Mexico).

[35] Nghiem AT, Bremond F, Thonnat M (2009) Controlling background subtraction algorithms for robust object detection. *3rd International Conference on Imaging for Crime Detection and Prevention* (London, UK), pp. 1-6.

[36] Chau DP, Bremond F, Thonnat M (2011) A multi-feature tracking algorithm enabling adaptation to context variations. *International Conference on Imaging for Crime Detection and Prevention*.

[37] Yakhia M, Konig A, van der Flier WM, Friedman L, Robert PH, David R (2014) Actigraphic motor activity in mild cognitive impairment patients carrying out short functional activity tasks: Comparison between mild cognitive impairment with and without depressive symptoms. *J Alzheimers Dis* **40**, 869-875.

[38] Satt A, Sorin A, Toledo-Ronen O, Barkan O, Kompatsiaris I, Kokonozi A, Tsolaki M (2013) Evaluation of speech-based protocol for detection of early-stage dementia. *Interspeech* (Lyon, France).