Validation of an Automatic Video Monitoring System for the Detection of Instrumental Activities of Daily Living in Dementia Patients
Alexandra König, Carlos Fernando Crispim-Junior, Alexandre Derreumaux, Gregory Bensadoun, Pierre-David Petit, François Bremond, Renaud David, Frans Verhey, Pauline Aalten, Philippe Robert

To cite this version:
Alexandra König, Carlos Fernando Crispim-Junior, Alexandre Derreumaux, Gregory Bensadoun, Pierre-David Petit, et al.. Validation of an Automatic Video Monitoring System for the Detection of Instrumental Activities of Daily Living in Dementia Patients. 2014. <hal-01094093>

HAL Id: hal-01094093
https://hal.inria.fr/hal-01094093
Submitted on 11 Dec 2014
HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.
Validation of an Automatic Video Monitoring System for the Detection of Instrumental Activities of Daily Living in Dementia Patients

Alexandra König1,2,∗, Carlos Fernando Crispin Junior3, Alexandre Darreumaux4, Gregory Bensadoun5, Pierre-David Petit4, François Bremond4,5, Renaud Davat6,7, Frans Verhey6, Pauline Aalten6,7, Pierre-David Petita, François Bremonda, Nicola De Pauw1,2, Renaud Davida, Frans Verheyb, Pauline Aaltenb and Philippe Roberta,c

1 EA CoBiTeK, University of Nice Sophia Antipolis, France
2 School for Mental Health and Neuroscience, Alzheimer Center Limburg, Maastricht University Medical Center, Maastricht, The Netherlands
3 Centre Memorie de Ressources et de Recherche, CHU de Nice, Nice, France
4 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...
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 IADL functioning.

Nevertheless, changes in IADL 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 states, 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 IADL, 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 ADLs, 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 IADLs, 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], Montgomery-Asberg Depression Rating Scale (MADRS) [29], and Geriatric Depression Scale (GDS) [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

<table>
<thead>
<tr>
<th>Activities</th>
<th>Constraints</th>
</tr>
</thead>
<tbody>
<tr>
<td>– 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.</td>
<td>– Prepare a hot tea</td>
</tr>
<tr>
<td>– Read the newspaper</td>
<td>– Write a shopping list for lunch</td>
</tr>
<tr>
<td>– Watch TV</td>
<td>1. Watch TV before the phone call</td>
</tr>
<tr>
<td>– Call the taxi</td>
<td>2. Water the plant just before leaving the room</td>
</tr>
<tr>
<td>– Prepare today’s medication</td>
<td>3. Call the taxi which will take 10 min to arrive and ask the driver to bring you to the market</td>
</tr>
<tr>
<td>– Make the check for the Electricity Company</td>
<td>4. Watch TV before the phone call</td>
</tr>
<tr>
<td>– Leave the room when you have finished all activities</td>
<td>5. Answer the phone</td>
</tr>
<tr>
<td>– Answer the phone</td>
<td>– Prepare a hot tea</td>
</tr>
</tbody>
</table>

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).
For a more detailed analysis, the main focus was placed particularly on three IADLs, namely preparing a pillow case, 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 Ngheim 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 semantometric 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 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 ($age = 75.2 \pm 4.25$) and 19 HC ($age = 71.7 \pm 5.4$) 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 the MMSE scores, with a mean of 25.8 ($\pm 2.2$) for the MCI group and 28.8 ($\pm 1.0$) for the HC group ($p = 0.001$), as well as for the FAB score with a mean of 14.16 ($\pm 1.92$) for the MCI group and 16.2 ($\pm 1.44$) for the HC group. The mean IADL-E scores did not differ between groups, with a mean IADL-E score of 9.9 ($\pm 1.7$) for the MCI group and 9.6 ($\pm 1.1$) 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 (manual 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 a 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.

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 scores indicating a better functional independence; 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>
<thead>
<tr>
<th>Characteristics</th>
<th>HC group $n = 19$</th>
<th>MCI group $n = 19$</th>
<th>$p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female, n (%)</td>
<td>15 (78.9%)</td>
<td>9 (47.4%)</td>
<td>0.091</td>
</tr>
<tr>
<td>Age, years mean ST</td>
<td>71.7 $\pm$ 5.57</td>
<td>75.2 $\pm$ 4.25</td>
<td>0.07</td>
</tr>
<tr>
<td>Level of Education, n (%)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>2 (10.5%)</td>
<td>2 (10.5%)</td>
<td>1</td>
</tr>
<tr>
<td>No formal education</td>
<td>0 (0%)</td>
<td>0 (0%)</td>
<td></td>
</tr>
<tr>
<td>Elementary school</td>
<td>1 (5.3%)</td>
<td>5 (26.3%)</td>
<td>0.405</td>
</tr>
<tr>
<td>Middle school</td>
<td>4 (21.0%)</td>
<td>7 (36.8%)</td>
<td>0.269</td>
</tr>
<tr>
<td>High school</td>
<td>4 (21.0%)</td>
<td>3 (15.8%)</td>
<td>1</td>
</tr>
<tr>
<td>Post-secondary education</td>
<td>8 (42.1%)</td>
<td>10 (52.6%)</td>
<td>0.062</td>
</tr>
<tr>
<td>MMSE, mean $\pm$ SD</td>
<td>25.8 $\pm$ 1.03</td>
<td>25.8 $\pm$ 2.22</td>
<td>0.001**</td>
</tr>
<tr>
<td>FAB, mean $\pm$ SD</td>
<td>16.2 $\pm$ 1.44</td>
<td>14.16 $\pm$ 1.92</td>
<td>0.002*</td>
</tr>
<tr>
<td>IADL-E, mean $\pm$ SD</td>
<td>9.6 $\pm$ 1.12</td>
<td>9.9 $\pm$ 1.73</td>
<td>0.488</td>
</tr>
<tr>
<td>NPI total, mean $\pm$ SD</td>
<td>0.42 $\pm$ 1.45</td>
<td>6.36 $\pm$ 7.37</td>
<td>0.001**</td>
</tr>
</tbody>
</table>

Significantly different characteristics were in bold text. 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 scores indicating a better functional independence; 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

<table>
<thead>
<tr>
<th>Video analyses data</th>
<th>HC n = 19</th>
<th>MCI n = 19</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manually annotated:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activities carried out completely and correctly†</td>
<td>9.68 ± 0.48</td>
<td>8.21 ± 0.48</td>
<td>0.00*</td>
</tr>
<tr>
<td>Activity frequency total (activities initiated)‡</td>
<td>11.74 ± 1.62</td>
<td>9.58 ± 1.68</td>
<td>0.00*</td>
</tr>
<tr>
<td>Preparing Pillbox time</td>
<td>41.17 ± 17.04</td>
<td>46.17 ± 11.18</td>
<td>0.609</td>
</tr>
<tr>
<td>Making tea (f)</td>
<td>2.64 ± 0.82</td>
<td>2.3 ± 0.83</td>
<td>0.68</td>
</tr>
<tr>
<td>Preparing Pillbox (f)</td>
<td>1.05 ± 0.23</td>
<td>0.89 ± 0.32</td>
<td>0.086</td>
</tr>
<tr>
<td>Making tea</td>
<td>41.21 ± 10.60</td>
<td>32.16 ± 15.3</td>
<td>0.175</td>
</tr>
<tr>
<td>* Phone call</td>
<td>2 ± 0.47</td>
<td>2 ± 0.37</td>
<td>0.198</td>
</tr>
<tr>
<td>* Phone call time</td>
<td>66.61 ± 21.75</td>
<td>43.30 ± 30.96</td>
<td>0.118</td>
</tr>
<tr>
<td>Automatically extracted: Activity frequency total</td>
<td>13.26 ± 3.89</td>
<td>10.57 ± 3.15</td>
<td>0.056</td>
</tr>
<tr>
<td>Preparing Pillbox time</td>
<td>1.05 ± 0.23</td>
<td>1.97 ± 0.38</td>
<td>0.271</td>
</tr>
<tr>
<td>Preparing Pillbox (f)</td>
<td>47.64 ± 22.28</td>
<td>70.26 ± 38.01</td>
<td>0.204</td>
</tr>
<tr>
<td>Making tea</td>
<td>2.74 ± 1.33</td>
<td>2.12 ± 1.22</td>
<td>0.136</td>
</tr>
<tr>
<td>Making tea time</td>
<td>102.3 ± 77.3</td>
<td>75.57 ± 40.92</td>
<td>0.531</td>
</tr>
<tr>
<td>* Phone call</td>
<td>1.95 ± 0.52</td>
<td>1.17 ± 0.79</td>
<td>0.38</td>
</tr>
<tr>
<td>* Phone call time</td>
<td>60.32 ± 21.32</td>
<td>112.64 ± 46.31</td>
<td>0.005*</td>
</tr>
</tbody>
</table>

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.

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...
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 pill-box’ 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)

<table>
<thead>
<tr>
<th>Video analysis data</th>
<th>MMSE</th>
<th>FAB</th>
<th>IADL-E</th>
</tr>
</thead>
<tbody>
<tr>
<td>Spearman correlation coefficient ($r$)</td>
<td>($r$)</td>
<td>($r$)</td>
<td>($r$)</td>
</tr>
<tr>
<td>p-values</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Manually annotated</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activities frequency</td>
<td>0.491*</td>
<td>0.394*</td>
<td>-0.035</td>
</tr>
<tr>
<td>($p=0.002$)</td>
<td>($p=0.014$)</td>
<td>($p=0.834$)</td>
<td></td>
</tr>
<tr>
<td>Activities completed correctly</td>
<td>0.819**</td>
<td>0.667**</td>
<td>-0.107</td>
</tr>
<tr>
<td>($p=0.000$)</td>
<td>($p=0.000$)</td>
<td>($p=0.522$)</td>
<td></td>
</tr>
<tr>
<td>Automatically extracted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Activity frequency</td>
<td>0.415*</td>
<td>0.273*</td>
<td>-0.071</td>
</tr>
<tr>
<td>($p=0.005$)</td>
<td>($p=0.048$)</td>
<td>($p=0.337$)</td>
<td></td>
</tr>
<tr>
<td>Preparing Pillbox (f)</td>
<td>0.055</td>
<td>0.299</td>
<td>-0.149</td>
</tr>
<tr>
<td>($p=0.239$)</td>
<td>($p=0.063$)</td>
<td>($p=0.427$)</td>
<td></td>
</tr>
<tr>
<td>Preparing Pillbox time</td>
<td>-0.468**</td>
<td>-0.114</td>
<td>-0.179</td>
</tr>
<tr>
<td>($p=0.001$)</td>
<td>($p=0.409$)</td>
<td>($p=0.211$)</td>
<td></td>
</tr>
<tr>
<td>Making tea (f)</td>
<td>0.27</td>
<td>0.363*</td>
<td>-0.15</td>
</tr>
<tr>
<td>($p=0.038$)</td>
<td>($p=0.042$)</td>
<td>($p=0.391$)</td>
<td></td>
</tr>
<tr>
<td>Making tea time</td>
<td>-0.413</td>
<td>0.053</td>
<td>-0.002</td>
</tr>
<tr>
<td>($p=0.322$)</td>
<td>($p=0.434$)</td>
<td>($p=0.396$)</td>
<td></td>
</tr>
<tr>
<td>Phone call (f)</td>
<td>-0.423</td>
<td>-0.235</td>
<td>0.082</td>
</tr>
<tr>
<td>($p=0.128$)</td>
<td>($p=0.084$)</td>
<td>($p=0.465$)</td>
<td></td>
</tr>
<tr>
<td>Phone call time</td>
<td>-0.123</td>
<td>-0.332*</td>
<td>-0.145</td>
</tr>
<tr>
<td>($p=0.644$)</td>
<td>($p=0.041$)</td>
<td>($p=0.291$)</td>
<td></td>
</tr>
<tr>
<td>Automatically extracted</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Preparing Pillbox (f)</td>
<td>-0.287*</td>
<td>-0.073</td>
<td>0.125</td>
</tr>
<tr>
<td>($p=0.043$)</td>
<td>($p=0.205$)</td>
<td>($p=0.222$)</td>
<td></td>
</tr>
<tr>
<td>Preparing Pillbox time</td>
<td>-0.618**</td>
<td>-0.244</td>
<td>-0.05</td>
</tr>
<tr>
<td>($p=0.001$)</td>
<td>($p=0.340$)</td>
<td>($p=0.128$)</td>
<td></td>
</tr>
<tr>
<td>Making tea (f)</td>
<td>0.223</td>
<td>0.221</td>
<td>-0.264</td>
</tr>
<tr>
<td>($p=0.60$)</td>
<td>($p=0.063$)</td>
<td>($p=0.051$)</td>
<td></td>
</tr>
<tr>
<td>Making tea time</td>
<td>0.016</td>
<td>-0.101</td>
<td>-0.114</td>
</tr>
<tr>
<td>($p=0.392$)</td>
<td>($p=0.261$)</td>
<td>($p=0.197$)</td>
<td></td>
</tr>
<tr>
<td>Phone call (f)</td>
<td>-0.248</td>
<td>0.077</td>
<td>0.158</td>
</tr>
<tr>
<td>($p=0.095$)</td>
<td>($p=0.330$)</td>
<td>($p=0.223$)</td>
<td></td>
</tr>
<tr>
<td>Phone call time</td>
<td>-0.273*</td>
<td>-0.277**</td>
<td>-0.054</td>
</tr>
<tr>
<td>($p=0.002$)</td>
<td>($p=0.049$)</td>
<td>($p=0.451$)</td>
<td></td>
</tr>
</tbody>
</table>

*p < 0.05, **p < 0.01.

Table 5
Activity/Event detection performance

<table>
<thead>
<tr>
<th>Activity</th>
<th>Recall</th>
<th>Precision</th>
</tr>
</thead>
<tbody>
<tr>
<td>Phone call</td>
<td>85</td>
<td>89.47</td>
</tr>
<tr>
<td>Watching TV</td>
<td>88.31</td>
<td>73.77</td>
</tr>
<tr>
<td>Making tea</td>
<td>80</td>
<td>80</td>
</tr>
<tr>
<td>Preparing Pillbox</td>
<td>780</td>
<td>90.24</td>
</tr>
<tr>
<td>Watering Plant</td>
<td>75</td>
<td>61.22</td>
</tr>
<tr>
<td>Reading</td>
<td>75</td>
<td>91.3</td>
</tr>
<tr>
<td>Average Recognition</td>
<td>85.31</td>
<td>73.9</td>
</tr>
</tbody>
</table>

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

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 psychological 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 surpris-
ing 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 impair-
ment 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 from most cases from sub-
jective 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 rel-
avively 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 differ-
ces could be observed.

It can be further argued that the experiment was con-
ducted 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 set-
tings, but that may also represent a less controlled envi-
ronment and therefore a bigger challenge from a tech-
nical 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 cogni-
tive processing required for functional ability.

It is important to mention that in the field of auto-
matic video analysis, it is almost impossible to achieve
100% accuracy in the activity recognition, often caused
as well by inaccurate manual annotations. The chal-
lenge 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 analysis might be actually a much closer rep-
resentation of the reality and the real events happening
than annotations of a human observer because the lat-
est 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 manu-
ally annotated data gives a more accurate picture of
a patient’s status to date, and is better validated by
traditional diagnostic and neuropsychological assess-
ment tools. This means that qualitative assessments
still seem to better correlate with conventional scor-
ing than quantitative video extracted parameters. Until
now, the obtained data still needs interpretation of an
experienced clinician regarding the quality of the car-
ried out activities. It should be emphasized that this
cannot be replaced by technology and is not the objec-
tive 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
REFERENCES


