A Computer system to monitor older adults at home: Preliminary results

Nadia Zouba, François Brémond, Monique Thonnat, Alain Anfosso, Eric
Pascual, Patrick Mallea, Veronique Mailland, Olivier Guerin

A computer system to monitor older adults at home: Preliminary results

Nadia Zouba MSc
François Bremond PhD
Monique Thonnat PhD
INRIA Sophia-Antipolis, 2004 route des Lucioles, 06902 Sophia Antipolis Cedex, France
E: nadia.zouba@sophia.inria.fr
E: francois.bremond@sophia.inria.fr
E: monique.thonnat@sophia.inria.fr

Alain Anfosso MSc
Éric Pascual MSc
CSTB Sophia Antipolis, 290 route des Lucioles, 06904 Sophia Antipolis Cedex, France
E: alain.anfosso@cstb.fr
E: eric.pascual@cstb.fr

Patrick Mallea MD
Véronique Mailland MD
Olivier Guerin MD
CHU Nice, Hôpital Cimiez, 4 avenue Reine Victoria, 06003 Nice Cedex 1, France
E: mallea.p@chu-nice.fr
E: mailland-putegnat.v@chu-nice.fr
E: guerin.o@chu-nice.fr

N. Zouba, F. Bremond, M. Thonnat, A. Anfosso, É. Pascual, P. Mallea, V. Mailland, O. Guerin, A computer system to monitor older adults at home: Preliminary results. Gerontechnology 2009;8(3):129-139, doi: 10.4017/gt.2009.08.03.011.00 Determining the individual transition from the 3rd to the 4th or frailty phase of life is important for both the safety of the older person and to support the care provider. We developed an automatic monitoring system consisting of cameras and different sensors that analyze human behaviors and looks for changes in activities by detecting the presence of people, their movements, and automatically recognizing events and Activities of Daily Living (ADLs). Assessment took place in a laboratory environment (GERHOME) comprised of four rooms (kitchen, living-room, bedroom, and bathroom). Data from 2 volunteers (64 and 85 years old) were analyzed. Precision in recognizing postures and events ranged from 62-94%, while sensitivity fell in the range of 62-87%. The system could differentiate ADL levels for the 64 and 85 year old subjects. These results are promising and merit replication and extension. Considerable work remains before the complete transition from 3rd to 4th life phase can be reliably detected. The GERHOME system is promising in this respect.
Keywords: monitoring system, ADLs, activity recognition, cognitive assessment

Observing Activities of Daily Living (ADLs) can provide clues to emerging physical and mental health problems in the passage between the 3rd (active retirement) to the

4th (frailty) phase of life[18]. Over the last several years much research has addressed developing and employing various sensors to monitor home activities, including camera networks for people tracking[4], cameras and microphones for activity recognition[5,6], embedded sensors[7], or sensors placed on the body[8]. Sensor monitoring of ADLs has also been used to assist older adults with and without dementia symptoms [9-13]. The ADLs are monitored in a number of integrated endeavors, such as the Aware Home Project[16], the Assisted Cognition Project[17], and the House_n project[18]. Simulated physiological data have been employed to generate health alarms[14,15]. A combination of body-worn and environmental sensors may automatically recognize ADLs, such as cooking, making tea, walking or changing posture[19].

Our work differs from prior studies in two ways: (i) we combine video cameras with a small number of sensors embedded in the home infrastructure, and (ii) we describe activities in formal models with the aid of a simple descriptive language[2]. We built a framework for modeling and evaluating ADLs at home on the basis of multi-sensor information and algorithms to assess behavioral trends, with an assessment platform[1] that includes an event recognition algorithm[2]. Our aim is twofold: (i) increased security for older persons, and (ii) support of the care provider.

This paper describes the system and presents an initial evaluation.

Methodology

Experimental Site

The experimental laboratory (GERHOME) was constructed in the CSTB (Scientific Centre of Building Techniques) at Sophia Antipolis in France. It simulates a typical apartment of an older person: 41m^2 with entrance, living-room, bedroom, bathroom, and kitchen. The kitchen includes an electric stove, a microwave oven, a refrigerator, cupboards, and drawers.

GERHOME is equipped with different sensors to evaluate ADL scenarios predefined by investigating gerontologists (Patrick Mallea, Véronique Mailland and Olivier Guerin). Four video cameras are installed: one in the kitchen, two in the living-room and the last one in the bedroom; however this paper addresses data obtained from the single living-room video camera.

Volunteers

Volunteers, 6 females and 8 males aged 60-85 years, were recruited by advertisements for a study of ways to make sensing technologies easier to use in the home. Volunteers were encouraged to maintain as normal as possible routines in preparing and taking meals, washing dishes, cleaning the kitchen, and watching TV while staying at GERHOME, and each was observed for a 4 hour interval. In total 14 video scenes were acquired by 4 video cameras (at a rate of ten frames per second). Each volunteer was alone in the laboratory during the observation period. Currently only the results of 2 of these volunteers (a male of 64 years, and a female of 85 years) have been analyzed.

Cameras and sensors

Commercially available sensing devices were used for data gathering including video cameras, and environmental sensors embedded in the home infrastructure: 12 contact sensors in the kitchen for detecting the opening and closing of cupboard

doors, drawers, etc., 3 proximity detectors to detect the presence of people near sinks, cooking stoves and washbowls, 4 pressure sensors located beneath 2 chairs, 1 pressure sensor located in the armchair and the bed (to detect when a person is sitting or not), 2 power consumption sensors (for detecting electrical appliance use) and 3 hot and cold water consumption sensors in the kitchen and bathroom. To reduce cost, environmental sensors with video sensors were combined (*Figure 1*). Environmental sensors are robust and precise but costs are high due to the number required. Cameras are less precise but typically a single camera in a room suffices.

An a priori knowledge base was used containing: (i) a 3D model of a person (i.e. height, width, depth); (ii) models of events of interest, (iii) 3D geometric description of the static environment, (iv) Semantic information (for example that people are expected to sit on a chair but not on a table), and (v) Calibration information for the camera.

The assessment system consists of (i) a video analysis component that detects and tracks those observed, recognizes their posture and a set of events detected by one video camera, (ii) an environmental sensor analysis component that collects information about interpersonal interactions and contextual objects and recognizes a set of simple environmental events (e.g., the refrigerator door is open), and (iii) a multimodal activity recognition component that combines video and environmental events to recognize complex activities (e.g., the person prepares a meal).

The output of the system is a set of recognized events as represented by xml files, by alarms text or by 3D visualization.

Object detection and tracking

Detecting and tracking a moving object is accomplished with a set of vision algorithms within the video interpretation platform[1]. A background subtraction method[20] segments the picture and compares intensity and color with a periodically updated reference background image not containing the moving object[21].

3D information is obtained by using a calibration step which computes the transformation of a 2D image referential point to a 3D scene referential point. The 3D position of the moving object is estimated from the detected blob and the calibration matrix associated with the video camera by supposing that the bottom of the 3D moving object is on floor level. When people's legs are occluded (not visible by the camera).

Internal parameters of the camera determine the matrix (image center, focal length and distortion coefficients) which are combined with external parameters (position and orientation relative to a world coordinate system). In the Tsai camera calibration method[22], the 3D world coordinates of a point in the image are computed under the assumption that the world point belongs to a particular plane; in our case the floor plane.

Next, a classification task uses the obtained 2D blobs, the calibration matrix of the camera and predefined 3D parallelepiped models (described by their width, height, length, position, and orientation) of the expected objects on the scene, to define the most likely 3D model for each object. Finally, a merging task is performed to improve the classification performance by assembling 2D blobs showing a better 3D object likelihood.
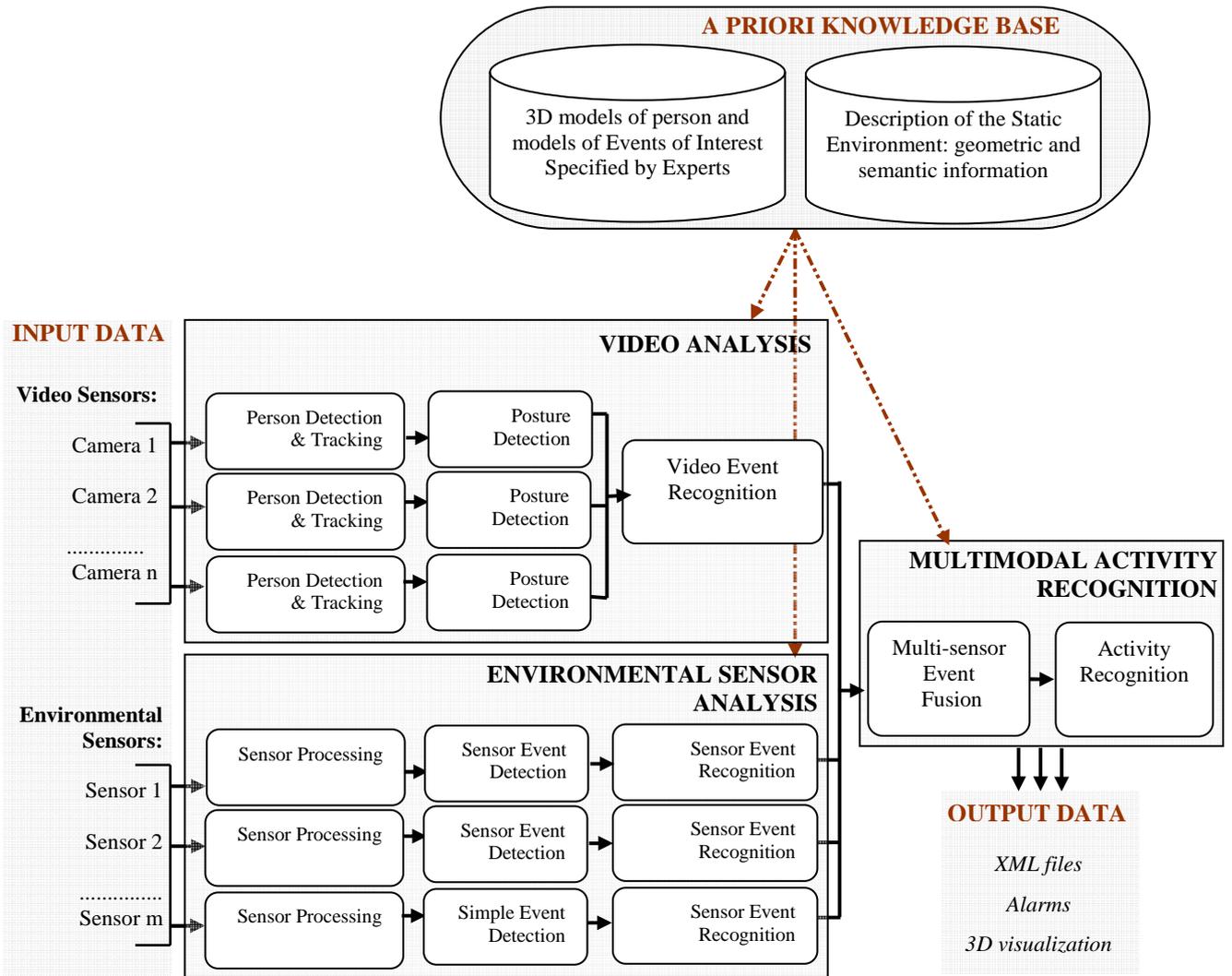
Figure1. Architecture of the GERHOME monitoring system

For each moving region, a 3D classifier adds an object class label (e.g. person, vehicle)[23]. After that, the tracking task adds a unique identifier to each new classified blob, and maintains it globally throughout the whole video (*Figure* 2)[1].

(a) Original image   **(b) Detection**   **(c) Classification**   **(d) Tracking**

Figure 2. Detection, classification and tracking of a person. (a) The original image; (b) Moving pixels highlighted in white and clustered into a mobile object; (c) Classification of the object as a person and a 3D parallelepiped indicates the position and orientation of that person; (d) Tracking at 3 of the same person (IND 0) at 3 occasions

Posture detection

The human posture recognition algorithm[24] determines the posture of the detected person using its silhouette and 3D position and bases its det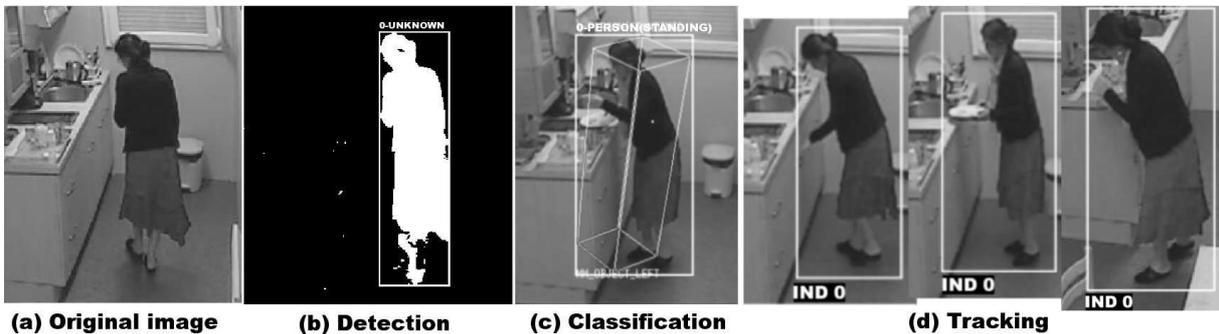ection on the combined set of 3D human models versus their 2D model comparison. The 3D models are projected in a virtual scene observed by a virtual camera which has the same characteristics (position, orientation and field of view) than the real camera (*Figure 3*). The 3D silhouettes are then extracted and compared to the detected silhouette using a 2D technique which projects the silhouette pixels on the horizontal and vertical axes. The most similar extracted 3D silhouette is considered to most accurately correspond to the current posture of the observed person. The algorithm is real-time (about eight frames per second), and does not depend on camera position.
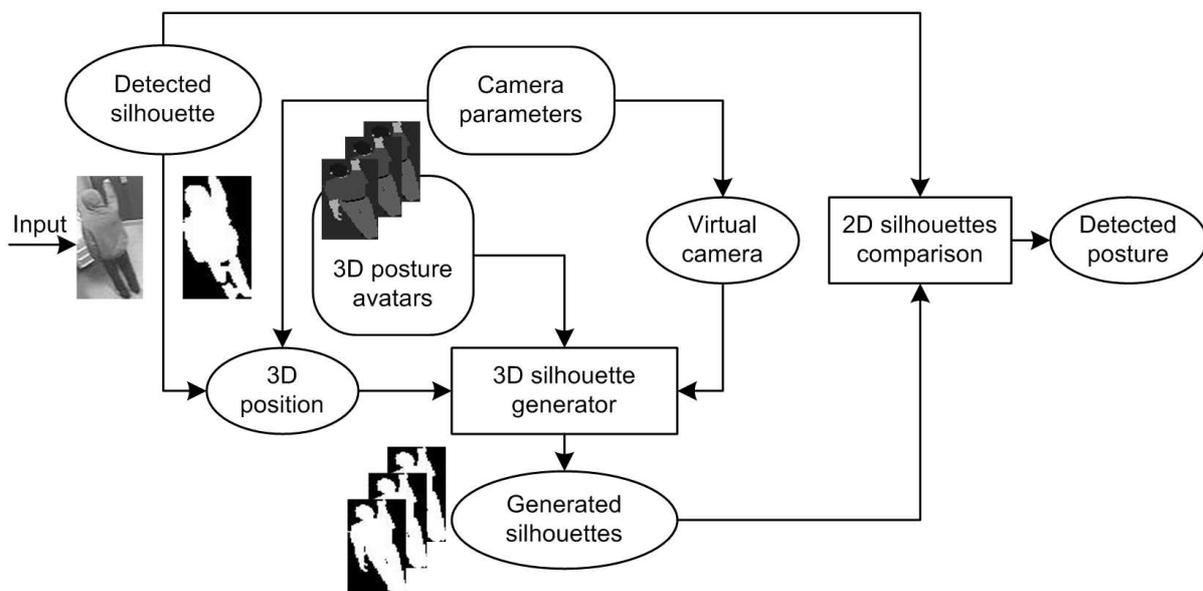

Figure 3. Simplified scheme showing the posture recognition approach

To adapt the human posture recognition algorithm for homecare applications, we have identified and modeled ten 3D key human postures that are useful to detect activities at home and critical situations for older persons: (i) standing with arms down, (ii) standing with arm up, (iii) standing with hands up, (iv) bending, (v) sitting on a chair, (vi) sitting on the floor with outstretched legs, (vii) sitting on the floor with flexed legs, (viii) slumping, (ix) lying on the side with flexed legs, and (x) lying on the back with outstretched legs. Each of these postures plays a significant role in the recognition of ADLs. For example, the posture 'standing with arm up' is used to detect reaching and opening kitchen cupboards. The posture 'standing with hands up' is used to detect carrying an object such as plates for the dinner table. These 10 human postures represent the key postures of ADLs[26].

Event assessment

To express the semantics of events of interest, we employ an event description language based on constraint resolution methods[2]. It uses a declarative representation of events that are defined as a set of spatio-temporal and logical constraints. The description is declarative and intuitive (in natural terms), so that the experts of the application domain can easily define and modify it.

The following concepts are defined in the context of events using the following event ontology[2]. A state is a spatio-temporal property valid at a given instant or stable on a time interval, and can characterize several mobile objects. An event is one or several state transitions at two successive time points or in a time interval. A *primitive state* is a spatio-temporal property valid at a given instant or stable over a time interval that is directly inferred from the visual attributes of physical objects computed by vision routines (e.g. a person is located inside a zone). A *primitive event* is a primitive state transition and represents the finest granularity of events (e.g. a person changes a zone). A *composite state* is a combination of primitive states. A *composite event* is a combination of primitive states and events. This is the coarsest granularity of events. Composite events are also known in video understanding literature as complex events, behaviors, and scenarios[2].

A definition of an event E consists of: (i) an event name, (ii) a list of physical objects (mobile or static) involved in the event, (iii) a list of components (variable values) representing sub-events that describe simple activities concerned, (iv) a list of forbidden components, a set of variables corresponding to all event instances that are not allowed to be recognized during the recognition of the event, (v) a list of constraints, a set of conditions among physical objects and/or the components to be verified for the recognition of the event, and (vi) a list of alerts (Not-Urgent, Urgent and Very-Urgent) as an optional part of the event model with a set of actions to be performed when the event recognized.

Constraints can be logical, spatial or temporal[25] depending on their meaning, and can have a symbolic or numeric form. For example, the spatial symbolic constraint 'object inside zone' is a spatial numeric constraint that is defined as follows: *distance (object1, object2) ≤ threshold*. A temporal constraint may also have a numeric form: *duration (event) ≤ 20 [secs]*.

In collaboration with gerontologists and geriatrics from the Nice hospital in France, we have modeled 26 events for homecare applications: 10 are related to the location of a person, another 10 relate to human postures, and 6 concern the transitions in human postures. '*Standing up*' represents a transition from sitting or slumping to standing, '*sitting down*' from standing or bending to sitting, '*sitting up*' from lying to sitting, '*lying down*' from standing or sitting to lying, '*feeling'* from standing or bending, to sitting on the floor with flexed legs and sitting on the floor with outstretched legs, and '*falling down*' from standing or bending, to sitting with flexed legs and lying with outstretched legs.

The event recognition process[2] uses the tracking of mobile objects, the a priori knowledge of the scene and predefined event models. The algorithm operates in 2 stages: (i) at each incoming frame, it computes all possible primitive states related to all mobile objects present in the scene, and (ii) it computes all possible events (i.e. primitive events, and then composite states and events) that may end with the previously recognized primitive states. The recognition algorithm also searches for previously recognized events to optimize the whole recognition.

Sensor data processing

The environmental sensor analysis task collects information about interactions between people and the contextual objects and processes them. Using transmission by radio frequency, a single USB receiver connected to a PC, acquires and logs data 24 hours a day, 7 days a week, from the environmental sensors. Currently, acquired

data is processed off-line. The following format was used for events: (i) *TimeStamp* (time the event occurred in YYMMDD-HHMMSS.MS), (ii) *SensorUnit* (class of information provided by the sensor, e.g., contact, presence), (iii) *SensorLocation* (location of the sensor, e.g., kitchen), (iv) *SensorValue* (value provided by the sensor, e.g., open/close); and (v) *SensorID* (a: single identifier of the transmitting sensor).

Ten environmental events have been modeled by using an event description language[2]: (i) 2 contact events (open/close) applied for kitchen cupboards, kitchen drawers, refrigerator, and bathroom cupboards, (ii) 2 usage events (on/off) applied for microware, stove, and TV, (iii) 2 presence events (present/not-present) applied near cooking stove and near washbowl; (iv) 2 pressure events (pressed/not-pressed) applied for chairs, armchair and bed, and (v) 2 water events (hot water consumed / cold water consumed) applied for water consumption in the kitchen and in the bathroom.

Multimodal activity recognition

A fusion process at the event level (i.e., decision level) combines video and environmental events. We extended the event description language[2] to address complex activity recognition involving several physical objects (e.g., person, chair) over an extended period of time to define a set of multimodal activities.

The multi-sensor event fusion algorithm takes the events (i.e. video and environmental events) and the models of events as input. An event model M is recognized at an instant **t** if its last (using the temporal order) component (i.e. sub-event) has been recognized at the same instant **t**. This sub-event can be a video event or an environmental event.

Together with gerontologists we modeled the 12 most relevant ADLs: using (i) the fridge, (ii) cupboards, (iii) drawers, (iv) the microwave, (v) the stove, (vi) TV, executing (vii) dish washing, (vii) taking a meal, and (viii) 4 variations of preparing a meal: breakfast, lunch, dinner, and warming up of a meal. Each activity is modeled with sub-activities relating to objects involved in that activity. For example, in the definition of the model of preparing lunch, the person should enter the kitchen, open the fridge to take foods, open food-cupboards to take ingredients, open a cupboard to take dishes, open the cutlery-drawer to take fork, knife and spoon, use the stove to cook the meal, and set up the table. Two modeled activities are detailed below

Falling down

'Falling down' has many forms. The event of a transition state from standing, sitting on the floor (with flexed or outstretched legs) and lying (with flexed or outstretched legs) is modeled as follows:

```
CompositeEvent (PersonFallingDown,
    PhysicalObjects ((p: Person))
        Components ((pStand: PrimitiveState Standing(p))
                    (pSit: PrimitiveState Sitting_Flexed_Legs(p))
                    (pLay: PrimitiveState Lying_Outstretched_Legs(p)))
        Constraints ((pSit before_meet pLay)
                    (pLay's Duration >=threshold))
        Alert (AText ("Person is Falling Down")
            AType ("VERYURGENT")) )
```

This 'falling down' model contains 1 physical object (the person), 3 components (human postures), 2 temporal constraints and 1 alert. When these components occurred and all the constraints are verified, the falling down event is recognized, and an alert is triggered.

Taking a meal

The 'taking a meal' model contains four physical objects (person, zone, equipment 1, equipment 2), four components, six constraints and an alert. The components are: location of the person in the living-room, close to table, the pressed state of the chair and the sitting posture of the person in the living-room. The constraints include 3 spatial constraints related to the zone and the equipments involved in the event, and include also 3 temporal constraints. When these components occurred and all the constraints are verified, the taking meal event is recognized and an alert is triggered.

> *CompositeEvent (**TakingMeal**,*
> *PhysicalObjects ((p : Person), (z : Zone), (eq1 : Equipment), (eq2 : Equipment))*
> *Components ((s_inz : PrimitiveState **inside_Livingroom**(p, z))*
> *(s_close : PrimitiveState **close_to_table**(p, eq1))*
> *(c : CompositeState **chair_pressed**(p, eq2))*
> *(s_sit : CompositeState **person_sitting_in_Livingroom**(p, z)))*
> *Constraints ((z->Name = Livingroom)*
> *(eq1->Name = table)*
> *(eq2->Name = chair1)*
> *(s_close's **Duration** >= threshold1)*
> *(c's **Duration** >= threshold2)*
> *(s_sit's **Duration** >= threshold3))*
> *Alert (AText ("**Person Takes a Meal**")*
> *AType("**NOTURGENT**")))*

If the event takes place at another location, the constraint *z->Name = Livingroom* will be lifted.

Model validation

Two validation experiments have been performed. The first one with one actor, and the second one with fourteen older persons.

The performance of the system is assessed as follows:

$$P = \frac{TP}{TP + FP} \quad\quad [1]$$

$$S = \frac{TP}{GT} \quad\quad [2]$$

$$GT = TP + FN \quad\quad [3]$$

With P being the precision , S the standard metrics, GT the ground truth, TP a true positive when the system correctly claims that an event occurs, FP a false positive when an incorrect event is claimed, and FN a false negative, if an event occurs and the system does not report it.

Behavior characterization

To compare the behavior of the 2 volunteers the Normalized Difference of mean durations of Activity (NDA) and the Normalized Difference of Instance number (NDI) were computed as follows:

$$NDA = \frac{|m1 - m2|}{(m1 + m2)} \qquad [4]$$

$$NDI = \frac{|n1 - n2|}{(n1 + n2)} \qquad [5]$$

Where m1 and m2 are the mean durations of a certain activity and n1 and n2 the number of instances it occurred during the observation period.

Results & discussion

Ten video sequences were acquired with one human actor (female, 33 years). The duration of each video is about ten minutes and each video contains about 4800 frames (about eight frames per second). We tested some normal activities such as: open and close kitchen cupboards, use microwave and warm up a meal. We have also tested two abnormal activities: 'feeling faint' and 'falling down'.

The preliminary results of the recognition of the different postures show a sensitivity of 63-79% and a precision of 62-85% (*Table 1*). When the system errs in the recognition of postures, it mixes postures such as bending and sitting due to segmentation errors (shadow, light change ...) and object occlusions.

Table 1. Recognition of the different postures of one human actor (female, age 33 years; GT=ground truth, TP= true positive, FN=false negative, FP=false positive, P=precision, S=sensitivity

Postures	Frequency				P=TP/(TP+FP)	S=TP/ GT
	GT	TP	FN	FP	[%]	[%]
Standing	120	95	25	20	82	79
Sitting	80	58	22	18	76	72
Slumping	35	25	10	15	62	71
Lying	6	4	2	2	66	66
Bending	92	66	26	30	68	71
Standing up	57	36	21	6	85	63
Sitting down	65	41	24	8	83	63
Sitting up	6	4	2	1	80	66

Recognition of states and events in case of older adult 1 (male, 64 years) showed a sensitivity of 62-87% and a precision of 71-94% (*Table 2*). This is comparable to the results of poster recognition. Usually the primitive states 'in the kitchen' and 'in the living-room' are well recognized. Errors occur mainly at the border between living-room and kitchen. Other errors are due to noise and shadow problems or sensor measurement errors.

Table 2. Recognition of states and events for volunteer 1 (male, 64 years); GT=ground truth, TP= true positive, FN=false negative, FP=false positive, P=precision, S=sensitivity

States and events	Frequency				P=TP/(TP+FP) [%]	S=TP/ GT [%]
	GT	TP	FN	FP		
Inside kitchen	8	5	3	2	71	62
Inside living-room	22	18	4	2	90	81
Using fridge	16	14	2	3	82	87
Using Stove	40	35	5	2	94	87
Preparing meal	1	1	0	0	100	100
Taking meal	1	1	0	0	100	100

Among the 22 activities for which the 2 older volunteers were compared (Table 3) 10 show differences. Of these 5 activities are considered meaningful and discriminative. Volunteer 1 of 64 years changed zones more often than the volunteer 2 of 85 years (for 'entering living-room' 20 vs. 13), and did this at a quicker pace (1:25 vs. 2:38), showing a greater ability to walk. Volunteer 1 was more often seen 'sitting on chair' (12 vs. 2, NDI=71%), but volunteer 2 was 'sitting on chair' for a longer duration (92:42 vs. 6:07, NDA=87%), showing also a greater ability for the volunteer 1 to move in the apartment.

Similarly volunteer 1 was 'bending' twice as much as volunteer 2 (30 vs. 15, NDI=33%), and in a quicker way (0:04 vs. 0:20, NDA=67%), showing greater dynamism for the younger volunteer. Volunteer 1 was also using more the 'upper-cupboard' than the volunteer 2 (25 vs. 9, NDI=47%), and in a quicker way (0:51 vs. 4:42, NDA=69%). She was also more able to use the *stove* (less trials for 'stove use' 35 vs. 102, NDI=49%).

All these measures show the greater ADL ability of the 64 years old adult as compared to those of the 85 years old.

Conclusion

We described a monitoring system that is able to automatically recognize a set of ADLs in two older adults, such as preparing and taking a meal, with a low false alarm rate. The obtained results demonstrate that the described method allows for detecting and recognizing of a set of ADLs. This is a first step to establish a behavioral profile of an observed person and to determine changes in this profile in time that are relevant for the transition from the 3rd to the 4th phase of life.

Acknowledgments

Table 3. Monitored activities, their frequencies (n1 & n2), mean and total duration of 2 volunteers staying in the GERHOME laboratory for 4 hours; NDA= Normalized Difference of mean durations of Activities==|mean 1-mean 2| / (mean 1 + mean 2); NDI=Normalized Difference of Instances number=|n1-n2| / (n1+n2); possible differences in behavior of the 2 volunteers and signified in bold

Activity	Sensor(s) used	Time spent [min:sec] and number of events (n)						NDA [%]	NDI [%]
		Male, 64 years			Female, 85 years				
		Mean 1	Total	n1	Mean 2	Total	n2		
Fridge use	Video, contact	0:12	2:50	**14**	0:13	1:09	**5**	4	**47**
Stove use	Video, power	**0:08**	4:52	**35**	**0:16**	27:57	**102**	**33**	**49**
Microwave use	Video, power	0:00	0:00	0	0:00	0:00	0	0	0
Kitchen hot-water tap	Water	0:19	12:40	40	0:09	3:31	22	4	29
Kitchen cold-water tap	Water	**0:28**	9:36	20	**0:03**	0:58	19	**81**	3
Upper cupboard use	Video, contact	**0:51**	21:34	**25**	**4:42**	42:24	**9**	**69**	**47**
Middle cupboard use	Video, contact	**0:10**	1:51	**11**	**0:10**	0:52	**5**	0	**38**
Lower cupboard use	Video, contact	**0:21**	3:09	**9**	**1:50**	7:23	**4**	**68**	**38**
Sitting on chair	Video, pressure	**6:07**	73:27	**12**	**92:42**	185:25	**2**	**87**	**71**
Sitting on armchair	Video, pressure	0:00	0:00	**0**	0:01	0:06	**6**	100	**100**
Toilet use	Water	**0:00**	0:00	**0**	**0:56**	0:56	**1**	**100**	**100**
Bathroom cupboard use	Contact	0:03	0:07	2	0:03	0:07	2	0	0
Bathroom hot-water tap	Water	0:10	0:21	2	0:12	0:36	3	9	20
Bathroom cold-water tap	Water	0:12	0:24	**2**	0:07	0:07	**1**	26	**33**
TV use	Power	**42:18**	169:12	4	**16:19**	65:18	4	**44**	0
Entering kitchen	Video	2:45	12:00	5	2:36	8:00	3	3	25
Entering living room	Video	**1:25**	25:00	**20**	**2:38**	35:00	**13**	**30**	21
Entering entrance	Video	2:00	8:00	4	1:43	11:00	6	8	20
Entering bedroom	Video	**1:25**	5:00	4	**3:28**	23:00	7	**42**	27
Entering bathroom	Video	1:00	2:00	**2**	1:00	5:00	**5**	0	**43**
Standing	Video	0:09	30:00	**200**	0:16	12:00	**45**	28	**63**
Bending	Video	**0:04**	2:00	**30**	**0:20**	5:00	**15**	**67**	**33**

11

References

1. Avanzi A, Bremond F, Tornieri C, Thonnat M. Design and assessment of an intelligent activity monitoring platform. EURASIP Journal on Applied Signal Processing, special issue in Advances in Intelligent Vision Systems: Methods and Applications, 2005;2005(14):2359-2374; doi:10.1155/ASP.2005.2359

2. Vu V, Bremond F, Thonnat M. Automatic video interpretation: A novel algorithm based for temporal scenario recognition. In Proceedings of IJCAI'03: The 15th IEEE International Joint Conference on Artificial Intelligence; 2003; pp.1295–1302

3. Bouma H. Gerontechnology: A framework on technology and aging. In Bouma H, Graafmans JAM, editors, Gerontechnology. Amsterdam: IOS; 1993; pp 1–6

4. Sidenbladh H, Black M. Learning image statistics for Bayesian tracking. In Proceedings of ICCV 2001: The 8th IEEE International Conference on Computer Vision; 2001; pp 709–716, doi:10.1109/ICCV.2001.937696

5. Clarkson B, Sawhney N, Pentland A. Auditory context awareness via wearable computing. In Proceedings of PUI'98: The Perceptual User Interfaces Workshop; 1998; pp 37-42

6. 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; pp 80-86; doi:10.1109/ICCV.1999.791201

7. Wang S, Pentney W, Popescu AM, Choudhury T, Philipose M. Common sense based joint training of human activity recognizers. In Proceedings of IJCAI 2007: The 20th IEEE International Joint Conference on Artificial Intelligence; 2007; pp 2237-2242

8. Foerster F, Fahrenberg J. Motion pattern and posture: correctly assessed by calibrated accelerometers. Behaviour Research Methods, Instruments, & Computers 2000;32(3):450–457

9. Hoey J, Bertoldi AV, Mihailidis A. Assisting persons with dementia during handwashing using a partially observable Markov Decision Process. In Proceedings of ICVS 2007: The 5th International Conference on Computer Vision Systems; 2007; doi:10.2390/biecoll-icvs2007-89

10. Wan D. Magic Medicine Cabinet: A Situated Portal for Consumer Healthcare. Lecture Notes in Computer Science 1999; Volume 1707; pp 352–355; doi:10.1007/3-540-48157-5-44

11. Barger T, Alwan M, Dalal S, Kell S, Turner B, Wood S, Naidu A. Objective Remote Assessment of Activities of Daily Living: Analysis of Meal Preparation Patterns. Poster presentation, Medical Automation Research Center, University of Virginia Health System; 2002

12. Tran Q, Truong K, Mynatt E. Cook's Collage: Recovering from Interruptions. Demo at 3rd International Conference on Ubiquitous Computing (UbiComp), Atlanta; 2001

13. Glascock A, Kutzik D. Behavioral Telemedicine: A New Approach to the Continuous Nonintrusive Monitoring of Activities of Daily Living. Telemedicine Journal 2000;6(1):33–44; doi:10.1089/107830200311833

14. Noury N, Rialle V, Demongeot J. AILISA : plateformes d'évaluation pour des technologies de télésurveillance médicale. Journal Gérontologie et Société 2005 ; 2(113) :97–119; doi:GS.113.0097

15. Rialle V, Duchene F, Noury N, Bajolle L, Demongeot J. Health 'smart' home: information technologie for patients at home. Telemedicine Journal E-Health 2002;8(4):395– 409; doi:10.1089/15305620260507530;

16. Mynatt E, Essa I, Rogers W. Increasing the opportunities for ageing in place. In Proceedings of the Association for Computing Machinery (ACM) conference on Universal Usability; 2000; pp 65–71; doi:10.1145/355460.355475

17. Kautz H, Arnstein L, Borriello G, Etzioni O, Fox D. An overview of the assisted cognition project. In Proceedings of the AAAI-2002 Workshop on Automation as Caregiver: The Role of Intelligent Technology in Elder Care, Edmonton; 2002; pp 60-65

18. Intille SS. The goal: smart people, not smart homes. In Proceedings of ICOST2006: The International Conference on Smart Homes and Health Telematics. Amsterdam: IOS Press; 2006; pp 3-6

19. Munguia-Tapia E, Intille SS, Larson K. Activity Recognition in the Home Setting Using Simple and Ubiquitous Sensors. Lecture Notes in Computer Science 2004; Volume 3001; pp 158-175; doi:10.1007/b96922

20. J. Heikkila J, and Silven O. A real-time system for monitoring of cyclists and pedestrians. In Proceedings of the Second IEEE Workshop on Visual Surveillance, pp 74–81, Fort Collins, Colorado, June 1999

21. McIvor A. Background subtraction techniques. In Proceedings of the Conference on Image and Vision Computing (IVC 2000), Hamilton, New Zealand, November 27-29 2000

22. Tsai R.Y, An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach, FL, pp. 364-374, 1986.

23. Zúñiga M, Brémond F, Thonnat M, Fast and reliable object classification in video based on a 3D generic model. The 3rd International Conference on Visual Information Engineering (VIE 2006), pp 433–440, Bangalore, India, September 26-28, 2006

24. Boulay B, Bremond F, Thonnat M. Applying 3d human model in a posture recognition system. Pattern Recognition Letter 2006;27(15):1788-1796; doi:10.1016/j.patrec.2006.02.008

25. Allen JF. Maintaining knowledge about temporal intervals. Communications of the Association for Computing Machinery (ACM) 1983;26(11):832-843; doi:10.1145/182.358434

26. Zouba N, Boulay B, Brémond F, Thonnat M, Monitoring Activities of Daily Living (ADLs) of Elderly Based on 3D Key Human Postures. The 4th International Cognitive Vision Workshop (ICVW 2008), pp 37-50, Santorin, Greece, May 2008.