Multisensor Fusion for Monitoring Elderly Activities at Home

Nadia Zouba Ep Valentin

UNIVERSITÉ DE NICE - SOPHIA ANTIPOLIS
UFR Sciences

École Doctorale STIC

THÈSE
pour obtenir le titre de
Docteur en Sciences
de l'Université de Nice - Sophia Antipolis

Spécialité : INFORMATIQUE

présentée et soutenue par

Nadia ZOUBA VALENTIN

Multisensor Fusion for Monitoring Elderly Activities at Home

Thèse dirigée par Monique THONNAT
et co-dirigée par François BRÉMOND
Équipe d'accueil : PULSAR – INRIA Sophia-Antipolis

Soutenue publiquement le
devant le jury composé de :

Président :	Philippe	ROBERT	Pr. UNSA, CHU Nice, France
Directeur :	Monique	THONNAT	DR, INRIA Sophia Antipolis, France
Co-Directeur :	François	BRÉMOND	CR, INRIA Sophia Antipolis, France
Rapporteurs :	Alessandro	SAFFIOTTI	Pr. AASS Mobile Robotics Orebro, Sweeden
	James	CROWLEY	Pr. INP Grenoble, France
Examinateur :	Jenny	BENOIS-PINEAU	Pr. LaBRI Bordeaux, France
Invités :	Alain	ANFOSSO	CSTB Sophia Antipolis, France
	Olivier	GUERIN	CHU Nice, France

Université Nice SOPHIA ANTIPOLIS

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA

centre de recherche SOPHIA ANTIPOLIS - MÉDITERRANÉE

UNIVERSITÉ DE NICE - SOPHIA ANTIPOLIS
UFR Sciences

École Doctorale STIC

THÈSE
pour obtenir le titre de
Docteur en Sciences
de l'Université de Nice - Sophia Antipolis

Spécialité : INFORMATIQUE

présentée et soutenue par

Nadia ZOUBA VALENTIN

Fusion Multicapteurs pour la Reconnaissance d'Activités des Personnes Agées à Domicile

Thèse dirigée par Monique THONNAT
et co-dirigée par François BRÉMOND
Équipe d'accueil : PULSAR – INRIA Sophia-Antipolis

Soutenue publiquement le
devant le jury composé de :

Président :	Philippe	ROBERT	Pr. UNSA, CHU Nice, France
Directeur :	Monique	THONNAT	DR, INRIA Sophia Antipolis, France
Co-Directeur :	François	BRÉMOND	CR, INRIA Sophia Antipolis, France
Rapporteurs :	Alessandro	SAFFIOTTI	Pr. AASS Mobile Robotics Orebro, Sweeden
	James	CROWLEY	Pr. INP Grenoble, France
Examinateur :	Jenny	BENOIS-PINEAU	Pr. LaBRI Bordeaux, France
Invités :	Alain	ANFOSSO	CSTB Sophia Antipolis, France
	Olivier	GUERIN	CHU Nice, France

Université Nice SOPHIA ANTIPOLIS

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

INRIA

centre de recherche SOPHIA ANTIPOLIS - MÉDITERRANÉE

Acknowledgments

I would like to thank Pr. Alessandro Saffiotti and Pr. James Crowley for accepting to review this manuscript. I want to thanks them for their very pertinent advices and remarks.

Merci à Jenny Benois-Pineau d'avoir accepté d'être membre du jury de thèse. Merci à Olivier Guerin et Alain Anfosso pour leur participation au jury. Je tiens aussi à remercier tout particulièrement le Pr. Phillippe Robert pour avoir accepté de présider ce jury.

Merci à Monique Thonnat et à François Brémond de m'avoir donné la chance de faire cette thèse dans leur équipe. Merci à tous les deux pour toutes leurs remarques pertinentes. Merci à tous les deux également pour avoir été disponible et m'avoir guidée tout au long de la thèse.

Un grand merci à Catherine Martin, pour toute l'aide administrative qu'elle m'a apporté tout au long de ces trois années et pour sa grande gentillesse. Merci aussi à Chantal Joncour qui a remplacé efficacement Catherine dans la dernière ligne droite de ma thèse.

Un grand merci à Bernard, Valéry, Guillaume, Etienne, Luis, Phu, Jean-Yves pour leur aide et pour toutes leurs remarques pertinentes.

Un remerciement plus général à tous les membres de l'équipe Pulsar: Valéry, Lan, Marcos, Bernard, Vincent, Etienne, Guillaume, Luis, Phu, Mohammed, Anh-Tuan, Rim, Guido, Daniel, Anja, Julien, Ikhlef, Zouhaier, Slawek, Erwan, Jean-Paul, Annie, Sabine, Jean-Yves, Christophe, pour avoir su y faire régner une ambiance propice au travail et à la détente.

Un grand merci à mon très cher mari Valéry qui m'a supportée pendant la phase de rédaction de ce manuscrit et qui a su m'épauler et encourager tout le long de la thèse, et un trés grand merci à mon adorable fille Lisa qui est restée sage tout au long de la rédaction de ce manuscrit (aussi bien avant qu'après l'accouchement).

Enfin un grand merci à toute ma famille, et plus particulièrement à mes parents Said et Tassadit, à mes frères et mes soeurs en France et en Kabylie et leur familles respectives, à mes beaux parents Charles et Gabrielle, et à ma belle soeur Véronique pour leurs soutiens et leurs encouragements.

ABSTRACT

In this thesis, an approach combining heterogeneous sensor data for recognizing elderly activities at home is proposed. This approach consists in combining data provided by video cameras with data provided by environmental sensors to monitor the interaction of people with the environment.

The first contribution is a new sensor model able to give a coherent and efficient representation of the information provided by various types of physical sensors. This sensor model includes an uncertainty in sensor measurement.
The second contribution is a multisensor based activity recognition approach. This approach consists in detecting people, tracking people as they move, recognizing human postures and recognizing activities of interest based on multisensor analysis and human activity recognition. To address the problem of heterogeneous sensor system, we choose to perform fusion at the high-level (i.e. event level) by combining video events with environmental events.
The third contribution is the extension of a description language which lets users (i.e. medical staff) to describe the activities of interest into formal models.
The results of this approach are shown for the recognition of ADLs of real elderly people evolving in an experimental apartment called Gerhome equipped with video sensors and environmental sensors. The obtained results of the recognition of the different ADLs are encouraging.

Keywords: Activities of Daily Living (ADLs), sensor model, probability density function (PDF), video events, environmental events, multimodal events, multisensor activity recognition, Dempster Schäfer Theory (DST).

Table of Contents

Abstract i

Table of Contents iii

List of Tables viii

List of Figures xii

1 Introduction 1
- 1.1 Motivations . . . 2
- 1.2 Objectives . . . 3
- 1.3 Context of the Study . . . 3
- 1.4 Hypotheses . . . 4
- 1.5 Thesis Contributions . . . 5
- 1.6 Thesis Layout . . . 6

2 State of the Art 9
- 2.1 Elderly Care Monitoring at Home . . . 9
  - 2.1.1 Technologies for Monitoring Human Activities at Home . . . 10
    - 2.1.1.1 Sensing Modalities . . . 10
    - 2.1.1.2 Industrial and Research Projects for Monitoring ADLs . . . 13
  - 2.1.2 Acceptance of Technologies . . . 16
- 2.2 Human Activity Recognition Approaches . . . 19
  - 2.2.1 Vision-Based Activity Recognition Approaches . . . 19
    - 2.2.1.1 Probabilistic and Stochastic Approaches . . . 20
    - 2.2.1.2 Constraint-Based Approaches . . . 21
  - 2.2.2 Sensor-Based Activity Recognition Approaches . . . 22
- 2.3 Multisensor-Based Activity Recognition Approaches . . . 23
  - 2.3.1 Definition of Sensor Fusion . . . 23
  - 2.3.2 Potential Advantages in Fusion of Multiple Sensors . . . 24
  - 2.3.3 Possible Problems in Multisensor Fusion . . . 25
  - 2.3.4 Sensor Fusion Levels . . . 26
  - 2.3.5 Sensor Fusion Approaches . . . 29
    - 2.3.5.1 Inference Methods . . . 29

2.3.5.2 Estimation Methods . . . 31
2.3.6 Sensor Fusion Work for Healthcare . . . 32
2.4 Conclusion . . . 34

3 Activity Recognition Approach Overview 35
3.1 Objectives . . . 35
3.1.1 A Framework for Activity Recognition . . . 36
3.1.2 Challenges in Activity Recognition . . . 36
3.1.3 Monitoring Goals . . . 37
3.2 Proposed Activity Recognition Approach . . . 37
3.2.1 Architecture of the Proposed Approach . . . 37
3.2.2 Video Analysis . . . 40
3.2.2.1 Person Detection and Person Tracking . . . 40
3.2.2.2 Posture Detection . . . 42
3.2.3 Sensor Analysis . . . 45
3.2.3.1 Sensor Processing and Modeling . . . 45
3.2.4 Event Recognition . . . 46
3.2.4.1 Event Modeling . . . 46
3.2.4.2 Event Recognition Algorithm . . . 48
3.2.5 Multisensor Event Fusion . . . 49
3.2.5.1 Video & Environmental Event Fusion . . . 49
3.2.5.2 Activity Recognition . . . 49
3.3 Conclusion . . . 50

4 Sensor Modeling 53
4.1 Introduction . . . 53
4.2 Smart Sensor . . . 54
4.2.1 Physical Sensors . . . 54
4.2.1.1 Physical Sensor Characteristics . . . 54
4.2.1.2 Physical Sensor Observation . . . 55
4.2.2 Logical Sensor . . . 56
4.3 Logical Sensor Modeling (LSM) . . . 57
4.3.1 Sensor Model with Uncertainty . . . 57
4.3.2 Binary Sensors . . . 58
4.3.3 Sensor Model . . . 58
4.4 Conclusion . . . 59

5 Multisensor Activity Recognition 61
5.1 Introduction . . . 61
5.2 Instrumentation of the Home . . . 61
5.2.1 Sensor Choice and Placement . . . 61
5.2.2 Sensor Mode . . . 62
5.3 Sensor Fusion . . . 63
5.3.1 Multisensor Properties . . . 63
5.3.2 High-Level Sensor-Fusion . . . 64

5.4 Activity Modeling . . . 65
5.4.1 Event Modeling Approach . . . 67
5.4.1.1 Event Description Language (EDL) . . . 67
5.4.1.2 Event Models . . . 68
5.4.2 Ontology for Daily Activities . . . 69
5.4.3 Ontology Hierarchy of Activities . . . 74
5.4.3.1 Ontology Concepts . . . 74
5.4.3.2 Examples . . . 74
5.4.4 The Proposed Event Models for Daily Activities . . . 75
5.4.4.1 Video Event Models . . . 77
5.4.4.2 Environmental Events . . . 86
5.4.4.3 Multimodal Event Models . . . 89
5.5 Activity Recognition . . . 93
5.5.1 Event Recognition Process . . . 94
5.5.1.1 Video Event Recognition Process . . . 95
5.5.1.2 Environmental Event Recognition Process . . . 97
5.5.2 Multisensor Event Recognition Process . . . 99
5.5.2.1 Multisensor Event Fusion . . . 99
5.5.2.2 Multimodal Event Recognition . . . 100
5.5.2.3 Algorithm for Multimodal Event Recognition . . . 101
5.6 Behavioral Profile . . . 101
5.7 Discussion . . . 102
5.8 Handling Uncertainty in Sensor Measurements . . . 102
5.8.1 Applying Dempster-Shafer Theory of Evidence for Fusing Sensors . . . 103
5.8.2 Evidential Network for Activity Recognition . . . 105
5.8.3 Evidential inference of activities . . . 106
5.8.4 Evidential network representation . . . 107
5.8.5 Activity Inference on Evidential Network . . . 108
5.9 Conclusion . . . 114

6 Evaluation and Results of the Proposed Approach 115
6.1 Experimental Site . . . 116
6.1.1 Gerhome Laboratory . . . 116
6.1.2 Video Cameras and Environmental Sensors . . . 117
6.2 Evaluation Metrics . . . 118
6.3 Performed Experiments . . . 122
6.3.1 Predefined Scenarios and Data Collection . . . 122
6.3.2 With One Human Actor . . . 122
6.3.3 With Fourteen Elderly Volunteers . . . 123
6.4 Performance Evaluation . . . 124
6.4.1 Evaluation of the Vision-Based Framework . . . 124
6.4.2 Evaluation of the Sensor-Based Framework . . . 130
6.4.3 Evaluation of the Multisensor-Based Fusion Framework . . . 131

6.4.3.1 Results of Recognition . . . 131
6.4.3.2 Results on Behavior for 2 Elderly Volunteers . . . 134
6.4.3.3 Results of the Recognition using DS Uncertainty . 140
6.4.3.4 Discussion . . . 140
6.5 Medical Evaluation . . . 141
6.5.1 Events Durations for 9 Elderly Persons . . . 141
6.5.2 Leave-One-Out Cross Validation . . . 143
6.5.3 Discussion . . . 152
6.6 Conclusion . . . 153

7 Conclusion and Future Work **155**
7.1 Overview of the Contributions . . . 155
7.2 Discussion . . . 156
7.3 Future Work . . . 157
7.3.1 Short-Term Perspectives . . . 157
7.3.1.1 Improving Object Detection . . . 157
7.3.1.2 Learning Event Models and Learning Temporal Information of Events . . . 158
7.3.1.3 Incorporate Another Uncertainty . . . 158
7.3.2 Long Term Perspectives . . . 158
7.3.2.1 Activity Monitoring in Other Environment . . . 159
7.3.2.2 Improve Activity Recognition Algorithms . . . 159
7.3.2.3 Embedding the sensors into common architectural components . . . 160

A Publications of the Author **161**

B Predefined Scenario **163**

Bibliography **167**

List of Tables

2.1 Sensor types, and considerations of their use 12
2.2 Multisensor Fusion Methods 29

5.1 Physical objects for monitoring activities at home. Pets object is useful but not yet implemented 70
5.2 List of body postures. Body postures presented in normal font are already implemented and used, and body postures presented in italic font are useful but not yet implemented 71
5.3 List of daily activities. Activities presented in normal font are already implemented and used, and activities presented in italic font are useful but not yet implemented 72
5.4 List of daily activities (table 5.3 continued). Activities presented in normal font are already implemented and used, and activities presented in italic font are useful but not yet implemented 73
5.5 Ground truth mean durations of 4 daily activities for 5 observed elderly persons. Time unit is hh:mm:ss 77
5.6 Mean durations $\mu_{Ei,Pk}$ of 4 daily activities Ei for 4 persons. Time unit is hh:mm:ss 77
5.7 Examples of frames of discernment 108
5.8 Examples of multivalued mappings; SFrid: represents a fridge sensor (a contact sensor associated to the contextual object "Fridge"), SCupb: represents a cupboard sensor (a contact sensor associated to the contextual object "Cupboard"), and SWatr: represents a water sensor (a water sensor associated to the contextual object "Water Pipe"), Frid: represents the contextual object "Fridge", Cupb: represents the contextual object "Cupboard", and Watr: represents the contextual object "Water Pipe" 109
5.9 Examples of evidence mappings 111

6.1 List of installed sensors per room 120
6.2 Results for recognition of a set of states and events by using video camera; Recognition of person location in the kitchen and in the livingroom. Recognition of the different human postures. 125
6.3 Results for recognition of a set of states and events by using environmental sensors 130

6.4 Results for recognition of a set of multimodal events of one volunteer among the 5 volunteers with ground truth 131
6.5 Results of recognition of a set of daily activities for 2 observed elderly persons 133
6.6 Results of recognition of a set of daily activities for 2 observed elderly persons 133
6.7 Monitored activities, their frequencies (n1 and n2), mean (m1 and m2) and total duration of 2 volunteers staying in the GERHOME laboratory for 4 hours; NDA =Normalized Difference of mean durations of Activities= $|mean1 - mean2|/(mean1 + mean2)$; NDI =Normalized Difference of Instances number= $|n1 - n2|/(n1 + n2)$. Time unit is hh:mm:ss 138
6.8 Monitored activities, their frequencies (n1 and n2), mean (m1 and m2) and total duration of 2 volunteers staying in the GERHOME laboratory for 4 hours; NDA =Normalized Difference of mean durations of Activities= $|mean1 - mean2|/(mean1 + mean2)$; NDI =Normalized Difference of Instances number= $|n1 - n2|/(n1 + n2)$. Time unit is hh:mm:ss 139
6.9 Results of recognition (using uncertainty) of a set of daily activities for 2 observed elderly persons 141
6.10 Durations of six daily activities for the 9 observed elderly persons. Time unit is hh:mm:ss 142
6.11 The mean duration of each activity. $MD_{Ei,P1}$ represents the mean duration using the data for 8 persons (i.e. by removing data from the person P1), and so on 144
6.12 Standard deviations $\sigma_{Ei,Pj}$ of each event Ei for each person Pj . . 146
6.13 Intervals $I_{Ei,Pj} = [MD_{Ei,Pj} - \sigma_{Ei,Pj}; MD_{Ei,Pj} + \sigma_{Ei,Pj}]$ 148
6.14 Intervals $I_{Ei,Pj} = [MD_{Ei,Pj} - \sigma_{Ei,Pj}; MD_{Ei,Pj} + \sigma_{Ei,Pj}]$ 148
6.15 Comparison between the duration of the event "Using Fridge" and the interval $I_{E1,Pj}$ 149
6.16 Comparison between the duration of the event "Using Stove" and the interval $I_{E2,Pj}$ 150
6.17 Comparison between the duration of the event "Sitting on a Chair" and the interval $I_{E3,Pj}$ 150
6.18 Comparison between the duration of the event "Sitting on an Armchair" and the interval $I_{E4,Pj}$ 151
6.19 Comparison between the duration of the event "Using TV" and the interval $I_{E5,Pj}$ 151
6.20 Comparison between the duration of the event "Using Upper Cupboard" and the interval $I_{E6,Pj}$ 152

List of Figures

1.1 Example of person falling down 2

1.2 Detection, classification, tracking and recognition of activities of a person in an experimental laboratory; (a) Represents the original image acquired by video camera, (b) the moving pixels are highlighted in white and clustered into a mobile object enclosed in an orange bounding box, (c) the mobile object is classified as a person, (d) shows the individual identifier (IND 0) and a colored box associated to the tracked person, (e) shows the 3D visualization of activity recognition. 7

2.1 Sensors for Activity Monitoring 11

2.2 The Japanese i-pot system 14

2.3 The Vivago system 15

2.4 RFID glove 15

2.5 RFID reader bracelet (left), RFID tagged toothbrush and toothpaste (right), tags circled 16

2.6 Fusion Levels: a) Data Level Fusion; b) Feature Level Fusion; c) Decision Level Fusion; Adapted from Yang et al. [Yang, 2006] . . . 28

2.7 Difference between the two concepts of Probability versus the concept of Dempster-Shafer 32

3.1 General Architecture of the Approach 39

3.2 The video analysis architecture. Our major contributions are represented in bold lines with white background. Our minor contributions are represented with dashed background and the existing methods are represented with gray background. 40

3.3 Video Analysis. (a) Represents the original image, (b) the detection of moving pixels which are highlighted in white and clustered into a mobile object, (c) the mobile object is classified as a person, (d) shows the tracking at 2 different times of the same person (IND 0), (e) shows the corresponding 3D posture of the tracked person in the 3D environment. 41

3.4 (a) Classification of the object as a person with standing posture and a 3D parallelepiped indicates the position and orientation of that person; (b) Tracking at 2 different times of the same person (IND 0) . . . 43
3.5 Model of human posture described by a set of 23 parameters . . . 43
3.6 Simplified scheme showing the posture recognition approach . . . 44
3.7 The proposed 3D human postures. . . . 45
3.8 The sensor analysis architecture. Our major contributions are represented in bold lines with white background. Our minor contributions are represented with dashed background. . . . 46
3.9 The event recognition architecture. Our major contributions are represented in bold lines with white background. Our minor contributions are represented with dashed background. . . . 47
3.10 Model of Events. . . . 48
3.11 The multisensor fusion event architecture. Our major contributions are represented in bold lines with white background. Our minor contributions are represented with dashed background. . . . 50

4.1 A smart sensor with a physical sensor and the encapsulated data processing functions and the encapsulated Interface File System (IFS). . . . 55

5.1 Overview of a typically instrumented home . . . 63
5.2 Data level, feature level and decision level fusion . . . 65
5.3 Model of composite event . . . 70
5.4 Graphical notation of hierarchical ontology . . . 75
5.5 A general ontology network of Activities . . . 75
5.6 An ontology network of preparing a meal activity . . . 76
5.7 A representation of the "Inside Kitchen" primitive state to model the status of a person p being geometrically inside a zone z which name is Kitchen . . . 79
5.8 A description of a "Person Enters Bedroom" primitive event. . . . 79
5.9 View and 3D visualization of a "hands-up" posture . . . 80
5.10 View and 3D visualization of an "arm-up" posture . . . 80
5.11 Two defined avatars . . . 81
5.12 Primitive posture-based state representing the model of "standing" posture . . . 81
5.13 View and 3D visualization of "standing" posture . . . 82
5.14 Example of "standing-up from the armchair" activity. . . . 82
5.15 "Standing-up from the armchair" model . . . 83
5.16 Example of "sitting-down" activity. . . . 83
5.17 Example of "sitting-up" activity. . . . 84
5.18 Example of "lying-down" activity. . . . 84
5.19 Different forms of "fainting" . . . 84
5.20 Example of "Fainting" model . . . 85

5.21 "Falling-down1" model 86
5.22 Illustration of elderly falls; (a) Example of "falling-down2" of elderly women; (b) Example of "falling-down3" of elderly women . . 86
5.23 Primitive environmental states of events provided by contact sensor. 88
5.24 Primitive environmental states of events provided by pressure sensor. 88
5.25 Primitive environmental states of events provided by an electrical sensor. 88
5.26 Primitive environmental event related to the status of a fridge. . . 89
5.27 Primitive environmental event related to the status of a microwave. 89
5.28 Illustration of "using microwave" activity 90
5.29 An illustration of "slumping in an armchair" activity and a corresponding model 91
5.30 Using microwave model 92
5.31 Temporal constraints between states and events constituting a composite multimodal event "using microwave" 92
5.32 Example of "preparing lunch" model 93
5.33 Taking a meal model 93
5.34 Taking a meal model with a logical location 94
5.35 Processing of video event models 95
5.36 Processing of environmental event models 98
5.37 The multisensor (video-environmental) event recognition process at each instant 100
5.38 Examples of evidential networks of activity type. (The graphical notations are summarized in figure 5.4) 105
5.39 Examples of evidential networks of sensor type; (a) Prepare Cold Meal, (b) Prepare Hot Meal. Sensor abbreviations: SFrid: fridge sensor, SCupb: cupboard sensor, SStov: stove sensor, SMicr: microwave sensor, SWatr: water sensor, SVideo: video sensor. (The graphical notations are summarized in figure 5.4) 107

6.1 External views of the Gerhome laboratory. 116
6.2 Internal views of the Gerhome laboratory. 117
6.3 3D visualization of the Gerhome laboratory. 118
6.4 Position of the sensors in the Gerhome laboratory. 119
6.5 Views from the installed video cameras in the Gerhome laboratory. 119
6.6 Views of some environmental sensors installed in the Gerhome laboratory. Sensors are circled. (a) Contact sensor on cupboard door in the kitchen; (b) Electrical sensor on electrical outlet in the kitchen; (c) Presence sensor in front of the washbowl in the bathroom; (d) Water sensor on water pipe in the kitchen; (e) Pressure sensor under the armchair in the livingroom. 120
6.7 An illustration of a 3D visualization. 122
6.8 A person is slumping in the armchair when he/she warms up a meal in the microwave oven. 123

6.9 (a) Recognition of "bending in the kitchen" activity and (b) the 3D visualization of this recognition. 126
6.10 Visualization of the recognition of "slumping in the armchair" activity in the Gerhome laboratory. 127
6.11 Recognition and the 3D visualization of the recognition of "fainting" situation. 128
6.12 Recognition and the 3D visualization of the recognition of "falling down" situation. 129
6.13 Visualization of the recognized events in the Gerhome laboratory. . 132
6.14 The drawer is still open when the person does not use it 135
6.15 The recognition and the 3D visualization of the recognition of "preparing a meal" activity. 136
6.16 The recognition and the 3D visualization of the recognition of "taking a meal" activity. 137
6.17 Duration of each activity for the 9 observed elderly persons 143

B.1 A predefined scenario (step 1) for the fourteen volunteers. 164
B.2 A predefined scenario (step 2) for the fourteen volunteers. 165
B.3 A predefined scenario (step 3) for the fourteen volunteers. 166

Chapter 1

Introduction

Human activity recognition is an important part of cognitive vision systems because it provides accurate information about the behavior of the observed people. A major goal of current computer vision research is to recognize and understand human motion, short-term activities and long-term activities. The application areas for these vision systems are mostly surveillance and safety. Activity recognition is becoming also important in the application area of healthcare.
Demographic changes associated with the aging population and the increasing numbers of elderly people living alone are leading to a significant change in the social and economic structure of our society. The elderly population is expected to grow dramatically over the next 50 years. The proportion of people aged 60-plus around the world is expected to be doubled from the current 10% to 22% [Jones, 2006]. The number of people requiring care will grow accordingly, while the number of people able to provide this care will decrease. Without receiving sufficient care, elderly are at risk of loosing their independence. It is well known that even subtle changes in the behavior of the elderly can give important signs of progression of certain diseases. Disturbed sleeping patterns could be caused, for example, by heart failure and chronic disease. Changes in gait, on the other hand, can be associated with early signs of neurological abnormalities linked to several types of dementias. These examples highlight the importance of continuous observation of behavioral changes in the elderly in order to detect health deterioration before it becomes critical. Thus a system permitting to analyse the elderly behaviours and looking for changes in their activities is more than needed. With the increasingly accessible sensor technology, automatic activity recognition is becoming a reality. By attaching different types of sensors on various objects, locations and on the human body, activities of a person can be tracked and continuously monitored.

The following sections describe the motivations, the objectives of this thesis, the context of the study, my hypotheses, my contributions and the thesis layout.

1.1 Motivations

This work was greatly motivated by research done in understanding human activity. Over the last several years much effort has been put into developing and employing a variety of sensors to monitor activities at home. Most systems that have been built to recognize home activities have been limited in the variety of activities they recognize. In particular, most previous work on activity recognition has used sensors that provide only a very coarse idea of what is going on. For example, by detecting only movement in a room, it is not possible to detect which activity occurs in the room. In this work we propose an approach to activity recognition that addresses these problems by combining the use of video cameras with environmental sensors to determine when a person uses the household equipment and to detect most of the activities at home. This approach consists in analyzing human behaviors and looking for changes in their activities. In particular, the goal is to collect and combine multisensor information to detect activities and assess behavioral trends to provide different services.
Our approach aims to provide several services for elderly people in order to help them to retain their independence and to live safely longer at home.
In particular, elderly people are prone to accidents and falls in the home and can often lie injured and undiscovered for long periods of time. The most important provided service is concerned with medical monitoring. Medical monitoring includes handling emergencies (e.g. people falling, gas leakage or taking overdose of medication) and the evaluation of frailty evolution of elderly people to prevent, for instance, fall (see figure 1.1) or depression. This type of services should be designed by physicians who have specified risky situations.

Figure 1.1: Example of person falling down

1.2 Objectives

The main objective of this thesis is to propose a new cognitive approach based on using ambient sensors technologies to recognize interesting activities at home. This approach includes an algorithm for real-time recognition of primitive and complex activities that have occurred in the observed scene by video cameras and sensors attached to house furnishings. The proposed approach consists in detecting people, tracking people as they move (see figure 1.2), and recognizing activities of interest based on multisensor analysis and human activity recognition.

This approach involves a complete framework for event recognition including video frame segmentation, object classification, object tracking, and event recognition tasks:

1. First, at each video frame, a segmentation task detects the moving regions, represented by bounding boxes enclosing them (see figure 1.2(b)).

2. Second, to each moving region, a 3D classifier associates an object class label (e.g. person, vehicle) and a 3D parallelepiped described by its width, height, length, position, and orientation (see figure 1.2(c)).

3. Third, a tracking algorithm associates to each new classified object a unique identifier and maintains it globally throughout the whole video (see figure 1.2(d)).

4. Finally, an adapted event recognition algorithm recognizes events occurring in the observed scene (see figure 1.2(e)).

1.3 Context of the Study

Healthcare technology for the elderly is a popular area of research. This technology represents a sub-discipline of "gerontechnology" [Bouma and Graafmans, 1993]. Automatic monitoring of Activities of Daily Living (ADLs) has been a popular focus in gerontechnology. Activities of Daily Living (ADLs) are routine activities that people tend to do everyday, such as eating, bathing and toileting. These activities are used by physicians to benchmark the physical and cognitive abilities of patients. According to gerontologists, identifying changes in daily living activities (ADLs) is often more important than biometric information for the early detection of emerging physical and mental health problems, particularly for the elderly [Manabe et al., 2000]. Typical ADLs include preparing meals, eating, getting in and out of bed, using the toilet, bathing or showering, dressing, using the telephone, housekeeping, doing laundry, and managing medications. Detection of these activities would enable systems to monitor and recognize changes in patterns of behavior that might be indicators of developing physical or mental medical conditions. Similarly, it could help to determine the level of independence of elderly. If it is possible to develop systems that recognize such activities, the medical experts may be able to automatically

detect changes in patterns of behavior of people at home that indicate declines in health.
This PhD work has been conducted in the Pulsar team at INRIA Sophia Antipolis in France. Pulsar is a multi-disciplinary team at the frontier of computer vision, artificial intelligence and software engineering. Pulsar work focuses on two main application areas: safety/security and healthcare. This work takes place in this context and aims at recognizing human activities for healthcare applications. In this study, we collaborate with gerontologists from Nice hospital to determine which elder activities are most important to monitor.
Sensor technology plays a fundamental role in human activity analysis. In order to test new sensors and new activity recognition techniques, we have set-up an experimental laboratory at Sophia Antipolis together with CSTB (the French scientific and technical center for building). This laboratory looks like a typical apartment for elderly people and is equipped with many sensors such as video cameras, contact sensors, pressure sensors, water sensors. We instrumented this laboratory in order to conduct experiments using real data.

1.4 Hypotheses

This thesis assumes the following hypotheses:

- **Fixed Video Camera:** In this work we assume that the used video cameras are fixed on a wall and without pan, tilt or zoom. In plus, we suppose the availability of a model for transforming 2D image referential points to 3D scene referential points. The 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 by supposing that the bottom of the 3D mobile object is on the ground floor. There are no restrictions on video cameras orientation.
  The quality of the analyzed video sequence must be sufficient for detecting the objects moving in the scene with an acceptable level of reliability. Excessive video noise, too low video frame rate, or a big lack of contrast between the objects and the background of the scene, among others, can be the factors which prevent the right detection of an object. This constraint does not mean that the interest is only centered in video sequences of high definition and of high quality.

- **Tracking one Individual:** For medical reasons and for reason of an increasing numbers of elderly people living alone at home, in this thesis we assume that we track only one individual living alone in his/her apartment. This hypothesis implies that the tracked person has only one identifier during the period of tracking. This identifier changes if we loose the person (e.g. when a person enters in a zone which is located outside of the field of view of the video camera).

1.5 Thesis Contributions

The global contributions of this work are the following:

- **My first main contribution consists in a new sensor model which is necessary for multisensor fusion systems**. It includes uncertainty in sensor measurement. This sensor model is able to give a coherent and efficient representation of the information provided by various types of sensors. This representation provides means for recovery from sensor failure and also facilitates reconfiguration of the sensor system when adding or replacing sensors. In the proposed sensor model we define the type of information (e.g. pressure, image, motion) and the measurement $\mathbf{y}$ which is the value of the physical property measured by the sensor. We also define the uncertainty Δy of measurement $\mathbf{y}$.

- **The second main contribution consists in a new cognitive approach for activity recognition based on multisensor fusion data**. This multisensor based activity recognition approach uses video cameras and environmental sensors in order to recognize interesting human activities at home. The input of the approach is the data provided by the different sensors. We use video cameras to detect and track mobile objects (mostly people) moving in the scene and environmental sensors (e.g. contact sensors, pressure sensors, water sensors) attached to house furnishings to collect information about the interactions with the objects in the scene. The output of the approach is a set of XML files, alarms and a 3D visualization of the recognized events. The proposed approach consists in a 4D (3D + time) analysis of multisensor data. It exploits three major sources of knowledge: the 3D information of the scene, the 3D model of mobile objects (e.g. person), and the models of activities predefined in collaboration with gerontologists.

- **The third main contribution consists in a new set of 3D human postures useful to recognize important activities at home.** We propose ten 3D key human postures to detect typical body configurations (e.g. sitting position) and critical situations for elderly (e.g. falling down).

- **The fourth main contribution consists in a new set of computational models of interesting activities at home**. We propose to represent the interesting activities in a formal model that satisfies a number of constraints by using the event description language developed in the Pulsar team [Vu et al., 2003]. We improved this language by adding information provided by non vision algorithms. We propose a knowledge base of models of interesting activities at home. These models of activities can be used in other applications in different environments. This proposed knowledge base contains 100 events including 16 ADLs.

1.6 Thesis Layout

This thesis is organized as follows:

- In **chapter 2** we discuss related work in the area of ADLs monitoring. We present different technologies for monitoring human activities at home and different types of sensors and sensed data in healthcare monitoring are briefly introduced. After that we describe different techniques for activity recognition and for multisensor fusion data.

- In **chapter 3** we present an overview of the proposed cognitive vision approach. We give a general architecture of the proposed multisensor based activity recognition approach. We describe the inputs, the outputs and the major sources of knowledge of our approach. We define the activity recognition problem as a key component of automatic health monitoring.

- In **chapter 4** we describe the proposed sensor model which is used to perform the multisensor system. We present firstly the physical sensors and their characteristics. After that, we present the logical sensor modeling with uncertainty.

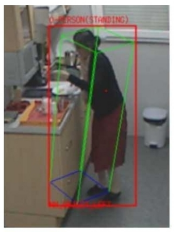
- In **chapter 5** we present the proposed multisensor activity recognition approach. This approach is based on using the combination of video cameras and environmental sensors to collect data about people activities and probabilistic models that are used to transform the raw sensor data into higher-level descriptions of people behaviors. We demonstrate that by the use of video sensors and environmental sensors, it is possible to provide rich information that can be used for analyzing most types of human activities at home. We present the activity recognition modeling. We present the event modeling approach and the proposed knowledge base of activity models. We propose also to define a behavioral profile for each person and we also propose to compare these behavioral profiles.

- In **chapter 6** we evaluate our approach and we test our proposed activity models in a set of scenarios performed in a realistic experimental laboratory. We present separately the obtained results by the vision algorithm, the obtained results by the environmental sensors and the obtained results by the multisensor fusion algorithm, and we compare the results. Evaluations are made using our datasets which contain sensors data of one human actor (aged of 33 years) and also of fourteen elderly volunteers (aged from 60 to 85 years) observed in an experimental laboratory, each one during 4 hours. The volunteers were given a sequence of activities to perform, like preparing a meal, and taking a meal.

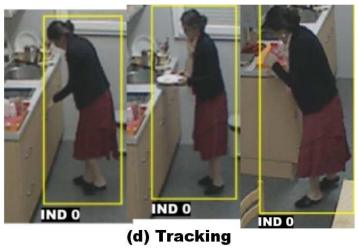
- Finally, in **chapter 7** we conclude this work, by summarizing the contributions of this thesis, and by presenting short-term and long-term perspectives.

(a) Original image

(b) Detection



(c) Classification



(d) Tracking

(e) Activity Recognition

Figure 1.2: Detection, classification, tracking and recognition of activities of a person in an experimental laboratory; (a) Represents the original image acquired by video camera, (b) the moving pixels are highlighted in white and clustered into a mobile object enclosed in an orange bounding box, (c) the mobile object is classified as a person, (d) shows the individual identifier (IND 0) and a colored box associated to the tracked person, (e) shows the 3D visualization of activity recognition.

Chapter 2

State of the Art

As seen in the previous chapter, human activity recognition is an important part of cognitive vision systems. In this chapter, previous work on monitoring elderly activities at home is described in section 2.1.
Related work on human activity recognition techniques using video sensors is described in section 2.2.1. Techniques using non-video sensors are described in section 2.2.2. Finally, section 2.3 describes fusion techniques between multiple and different sensors to recognize human activities.

2.1 Elderly Care Monitoring at Home

Healthcare technology for the elderly is a popular area of research. This technology represents a sub-discipline of "gerontechnology" [Bouma and Graafmans, 1993]. Automatic monitoring of elderly activities at home has been a common focus in gerontechnology.
In France, the proportion of people aged 75 and over in the population (approximately 7% in 2000) should reach nearly 10% in 2020 [Colin and Coutton, 2000]. In future years, the difference between the needs of the dependent elderly and the number of places available in hospitals and in specialized centers will become even more important than it is currently [Mesrine, 2003].
Healthcare technologies to maintain elderly at home allow the concerned person to live in a familiar environment and to benefit from a maximal independence. If these technologies generally enable to delay the loss of autonomy, they present however some risks at short-term (e.g. falls) and longer-term (e.g. bad feeding, insufficient hygiene, dementia).
Dependence of a person is defined as partial or total impossibility for a person to perform, without technical or human assistance, one or many daily activities [CNEG, 2000]. It is the consequence of one or many incapacities, deficiencies, or diseases, leading to limitations of activity or restrictions of participation. Autonomy can be defined as the absence of dependence.
Many scales were proposed to measure dependence of a person and among them some are particularly used in geriatrics. We enumerate here three scales usable

to measure the degree of dependence of elderly: the Katz ADL index, the Lawton IADL scale and the AGGIR grid.

The Katz ADL index (Index of Independence in Activities of Daily Living) [Katz et al., 1963], [Katz et al., 1970], [Katz, 1983] and the Lawton IADL scale (Instrumental Activities Daily Living Scale) [Lawton and Brody, 1969] were referred in the international literature as tools for assessment of the autonomy centered on the person.
The AGGIR grid (Autonomie Gérontologique Groupes Iso Ressources) [Benaim et al., 2005] is dedicated to evaluate cost of the dependence, the load in care, and was registered in 1997 in the French law as a tool for assessment of the dependence in order to determine if a person could have a specific allocation of money.

2.1.1 Technologies for Monitoring Human Activities at Home

Tracking and identification of daily physical activities are key factors to evaluate the quality of life and health status of a person. Research on this field is well recognized in rehabilitation, assessment of physical treatment [Pentland, 2004], [Aggarwal and Cai, 1999] and is shown to have significant impacts on healthcare of elderly persons and patients [Najafi et al., 2003].

Monitoring activities at home by using ambient sensor technologies can provide some proactive and situation aware assistance to sustain the autonomy of the elderly. It also can be helpful in reducing costs for public health systems and in providing advantages for older person by increasing his/her quality of life.

Over the last several years much effort has been put into developing and employing a variety of sensors to monitoring activities at home. These sensors include camera networks for people tracking [Sidenbladh and Black, 2001], cameras and microphones for activity recognition [Clarkson et al., 1998], [Moore et al., 1999], and embedded sensors for activity detection [Moeslund et al., 2000], [Wang et al., 2007].

2.1.1.1 Sensing Modalities

Sensors are devices which can be used to detect the interaction between a person and his/her environment. They are ultimately the source of all the input data in a multisensor data fusion system [Fowler and Schmalzel, 2004]. The physical sensor may be any device which is able of perceiving a physical property, or environmental attribute, such as light, sound, pressure, motion, image. To be useful, the sensor must transform the value of the property or attribute to a quantitative measurement.
A sensor system that is able to automatically recognize activities at home would allow many potential applications in healthcare area. The various sensor

technologies differ from each other in terms of price, ease of installation and the type of data they output [Fogarty et al., 2006].

Figure 2.1 illustrates the range of sensor technologies that are being investigated for activity monitoring. As shown there, researchers are exploring both environmental sensors and biosensors. The environmental sensors include different types of sensors such as motion and video sensors that determine the location of the person, contact sensors on cabinets and refrigerator doors that indicate whether they have been opened, pressure sensors that indicate whether a person is sitting in a bed or a chair, and electrical sensors that indicate whether a stove has been turned on. Biosensors are generally worn by a person to measure vital signs such as heart rate and body temperature.

This range of sensors can be used to determine where a person is and what household objects he/she has used, as well as to get a general sense of his/her activity level. This information can be used to infer specific daily activities performed, and in turn, that knowledge, perhaps combined with biometric information, leads to a general assessment of health and wellbeing.

Table 2.1 illustrates a set of sensor types and some considerations of

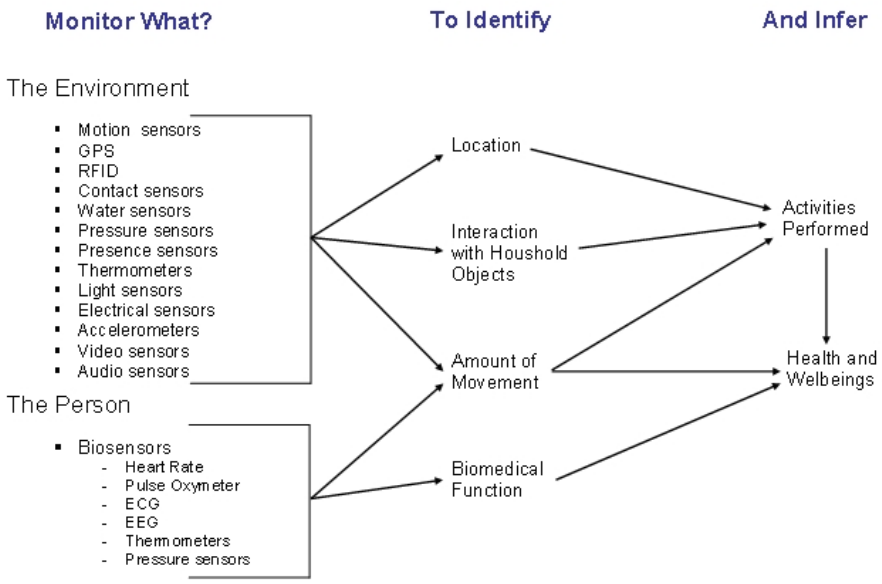


Figure 2.1: Sensors for Activity Monitoring

their use. Each of the sensors described in table 2.1 has been assigned to one or more of three types: activity, context and biomedical. Activity sensor types may be used to infer activity or behaviors. This may be movement around the

Sensor	Considerations	Type
Accelerometers	Must be worn	Activity
Motion	Inability to distinguish between subjects	Activity
RFID	Requires reader be worn and tags installed	Activity
GPS	Privacy	Activity
Contact	Inability to distinguish between subjects	Activity
Water	Inability to distinguish between subjects	Context
Light	Inability to distinguish between subjects	Context
Pressure	Inability to distinguish between subjects	Activity, Biomedical
Presence	Inability to distinguish between subjects	Activity
Temperature	Inability to distinguish between subjects	Context, Biomedical
Video	Privacy	Context, Activity
Audio	Privacy	Context, Activity
Heart Rate	Must be worn	Biomedical
Pulse Oxymeter	Must be worn	Biomedical

Table 2.1: Sensor types, and considerations of their use

home, activities such as meal preparation or leisure activities like watching TV or reading. Context sensors consist in sensors attached to house furnishings in order to collect context information about the scene. These include light sensors, water sensors, temperature sensors.

Biomedical sensors are designed to provide continuous monitoring of vital signs and patient attributes. As we can see from the above table, biomedical sensors include heart rate and pulse oxymeter as well as weight determined using a pressure sensor. In healthcare applications, biomedical sensors play an important role to obtain information of an elderly person. About the presented biomedical sensors, they could indicate a decreasing for physical health of the person. For example, decreased activity coupled with increasing weight of a person may signal a physical health problem of a person with congestive heart failure.
Accelerometers, motion sensors, video sensors, audio sensors, GPS, and RFID can be used in a home healthcare environment. All of these sensors can give us an indication of where a person is and what the person is doing. If motion sensors are placed in each zone in the home environment of the person, it is then easy to see the movement of the person around his/her home. A similar argument is true

for a GPS receiver worn about the person, although prior knowledge of the scene are required. A carried RFID sensor could perform a similar job using RFID tags fixed to objects that interact with the person. Video sensors can be used to detect a change in posture, audio sensors to detect sound waves and body worn accelerometers to detect a rapid change in acceleration.

2.1.1.2 Industrial and Research Projects for Monitoring ADLs

Medical professionals believe that one of the best ways to detect emerging physical and mental health problems (before it becomes critical - particularly for the elderly) is to look for changes in the activities of daily living (ADLs). Typical ADLs are sleeping, food preparation, eating, housekeeping, bathing or showering, dressing, using the toilet, doing laundry, and managing medications.
There are different commercial systems available for monitoring elderly activities at home. The best-known projects include the QuietCare system [QuietCare, 2002], and the Japanese "i-pot" system [i pot, 2005]. QuietCare system was created in 2002 by Living Independently, a next generation health and eldercare company that has been helping seniors live with greater safety in their own homes. This system is the result of 12 years of dedicated research, design, and testing by Professors Anthony P. Glascock and David M. Kutzik [Glascock and Kutzik, 2006] of Drexel University. Research was partially funded by grants from the USA National Institute of Health and Aging. QuietCare system uses wireless motion sensors to monitor the person in their own home. These sensors are installed in the bedroom, the bathroom, the kitchen, and in medication area in order to measure bathroom stays, use of medications, and the number of times a person gets out of bed at night. The main limitation of this commercial systems is that it provides a limited analysis of activity. For example, by detecting only movement in a room, it is not possible to detect which activity occurs in the room.
The Japanese "i-pot" system (information pot, see figure 2.2) consists in an electric kettle that keeps track of when it is used and sends a signal to a server with the data. The idea is to detect a sudden change in an elderly person's tea habits, in order to act as an early warning system in case of emergency. The i-pot system is in use in Japan, where an increasing number of the elderly are living, and dying, alone. Seniors who use the i-pot system report feeling less alone, knowing that somebody else is able to monitor them via the data sent by the kettle. The main limitation of the i-pot system is that it detect only one activity.

Among the most important reasons for the transfer from home to institutional care are the security concerns. Hence, all means for improving security for an independently living elderly person are essential. Among successful and widely adopted methods to respond to this need are social alarm systems. Traditional social alarm systems are based on a panic button, which is usually worn on person's wrist or as a necklace. Vivago system (see figure 2.3) is an active

Figure 2.2: The Japanese i-pot system

social alarm system, which combines intelligent social alarms with continuous remote monitoring of the user's activity profile [Sarela et al., 2003]. The system can provide long-term monitoring of the user's circadian rhythm, which, in turn, may be used to monitor changes in the wellbeing. The system has been especially designed to fit the needs of elderly homecare and institutional care settings.
These social alarm systems are mostly closed, stand-alone systems with a limited ability to describe the actual situation, often just too difficult for the elderly people to operate and useless in emergencies. The main problem with these alarm systems is that a significant portion, even 27-40% of the users, do not wear the alarm device on daily basis [Porteus and Brownsell, 2002], in case of an emergency the alarm is hence not possible. Furthermore, if the user is unable to push the button (e.g. has loss his/her consciousness) no alarm is generated.
Another important aspect in supporting independent living is remote monitoring of elderly health status in order to allow early intervention and monitoring of changes in their general wellbeing. For example, the incidence of dementia is increasing in the elderly population. Sleeping disorders are common in demented person, and sleep/wake rhythm in Alzheimer's disease is extremely disturbed. Reports of poor sleep correlate strongly with health complaints and depression [Phillips and Ancoli-Israel, 2001].

There has also been a significant amount of research work in the area of recognition of Activities of Daily Living (ADLs). Recognition of ADLs can be split into three subcomponents; feature detection, feature extraction and models for recognition.
A currently popular technique for detecting features of ADLs collects a wide range of sensor data. In [Philipose et al., 2004] from university of Washington, for example, a set of household objects such as microwave and cupboards are tagged with wireless sensors and transponders that transmit information via

Figure 2.3: The Vivago system

an RFID (Radio Frequency Identification) reader, mounted on hand glove (see figure 2.4), when the object is being used or touched.

Another technique for feature detection is the use of wearable sensors such as

Figure 2.4: RFID glove

wearable accelerometers that provide data about body motion and the surroundings where the data has been collected from. Previous work [Lester et al., 2005] has shown that a variety of activities like climbing stairs and running can be determined using this technique. The authors in [Wang et al., 2007] used accelerometers to detect fine-grained arm actions like "drink", "chop with knife". These were then combined with object-use data to achieve accurate activity recognition. The accurate recognition was based on a joint probabilistic model of object-use activities, which showed that it was possible to combine the data from both for accurate activity recognition.

Several projects have investigated the use of different sensors to provide a "smart" home for the observation of activities of daily living (ADLs). Examples include Georgia Tech's "Aware Home" [Abowd et al., 2002], Imperial College's UbiMon system [Yang et al., 2004], SAPHE project [Saphe, 2006], the Welfare-Techno house in Japan [Tamura, 2005] and MIT's PlaceLab [Cook and Das, 2007].

However, the use of heterogeneous sensors, including both wearable and ambient sensors, in such large deployment projects poses a number of interesting challenges. These include dealing with energy constraints, memory and processing power restrictions, as well as privacy and security issues.
Recent advances in miniaturization and wireless communication have seen the emergence of a third approach to sensing. In this approach, sensors are directly attached to many objects of interest. These sensors are either battery-free wireless stickers called Radio Frequency Identification (RFID) tags [Wyatt et al., 2005]. The sensors transmit the usage of the objects they are attached to by detecting either motion or hand-proximity to the object.
At the University of Washington [Wang et al., 2007] an RFID reader bracelet (see figure 2.5) records information about objects being manipulated by a person. A model of activities is obtained through web data mining techniques. While the authors report positive results, there is one main disadvantage to this approach: the inconvenience of wearing a bracelet. Ogawa, et al. [Ogawa et al., 2002] used

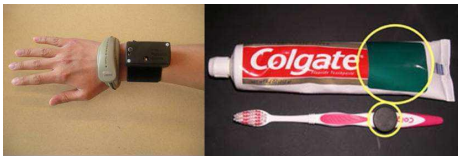


Figure 2.5: RFID reader bracelet (left), RFID tagged toothbrush and toothpaste (right), tags circled

sensors to detect movement, use of appliances, and presence in a room and from this information were able to analyze behavior patterns of two elderly ladies living alone. Nambu, et al. [Nambu et al., 2005] found that analyzing TV watching patterns alone was effective at identifying and analyzing behavior patterns, without the need for additional customized sensors.
The Ailisa project [Noury et al., 2001] (Intelligent Apartments for effective longevity) is an experimental platform to evaluate remote care and assistive technologies in gerontology. This ambitious project regroups specialists of smart home, networks and computing, electronics, and signal processing.
Overall the systems presented in this section lack one or more features to infer a large number of users activities. More importantly the majority of these systems rely on a single technology that effectively decreases the richness of information generated as a result of user actions and behavior, which limits the number of activities that can be recognized.

2.1.2 Acceptance of Technologies

A smart home is an environment equipped with technology that enhances safety of patients at home and monitors their health conditions. Smart home technology was initially developed for the elderly in order to give them a more independent lifestyle. Many of the elderly are "well aware about their problems resulting from

the age and the handicaps" and, "willing to accept the extra technical support and costs, above all, if the only remaining option is to leave their home and their well known environment in order to change to an old age home" [Brey, 2005]. Smart homes are greatly beneficial to the elderly because it gives them much more control over their environment and a "quality of life that they might not have otherwise had".
There are two main ways in which smart home technology can benefit the elderly and disabled. One way is by providing a monitoring system which can alert a healthcare system or an emergency system in case that the person has an accident or needs medical help. The second way is by giving the person access to a system in which they can control the devices in their home and be alerted as to actions so that they may control it. Monitoring the elderly in their home is a way to provide them an extra measure of safety and care. For example, location sensor technology can tell where the person is located in the house at all times. If the monitoring system detects that the person has not moved from the same position for a predetermined period of time, an alert is sent and the person is either called or visited to ensure that all is okay.
The devices and sensors chosen to be installed and maintained in the elderly homes need to address functional limitations and social and healthcare needs. Several pilot projects have introduced "smart home" technologies both in the US and Europe. One such pilot project, the SmartBo project in Sweden [Elger and Furugren, 1998], was created in a two-room ground floor demonstration apartment operated by the Swedish Handicap Institute. The project utilizes solutions for elderly with mobility impairments and/or cognitive disabilities (such as dementia). Devices and sensors control lighting, windows, doors, locks, water pipes, and electrical outlets. A similar project for elderly was introduced in the Netherlands [Berlo, 1998] using devices for control of lighting, sensors for optimal processing of temperature and heating, and remote control of several other functions. The project Prosafe in France [Chan et al., 1999] identified abnormal behavior of a monitored patient that can be interpreted as an accident, by collecting representative data on a patient's nocturnal and daily activity.

Little evaluation research exists on user acceptance of smart home technologies. There are only a few studies that investigate elderly perceptions of smart home technologies or other home-based technological applications. One of these studies that address this concept is by Vincent et al. [Vincent et al., 2002] who examined the application of environmental control systems in the homes of users and caregivers and concluded that the use of remote control by people with moderate cognitive impairments was difficult, while verbal reminders were greatly appreciated.
A further study by Demiris et al. [Demiris et al., 2001] investigated elderly perceptions of videophone and monitoring technology that can be installed in their homes and found that the respondents had an overall positive attitude

toward the use of home based technology. The findings from this study indicate that privacy can be a barrier for elderly to adopt smart home technologies; however their perception of their need for the technology may override their own privacy concerns. Privacy was considered important by the elderly, but they stated that when you need help it becomes less important;
Use and acceptance of technologies of monitoring human activities at home, and technical devices depend on various factors: adequate design, financial resources, the housing situation, which functions shall be compensated or strengthened by technologies and which skills and competences still exist. Although it is quite true that use and acceptance of innovations can only be roughly estimated today we nevertheless can list a set of important aspects that should be considered when speaking about user requirements and acceptance:

- People do not accept everything that is technologically possible and available
- Ambient Assisted Living concerns a heterogeneous group, where solutions therefore are accordingly multifaceted. There is no such thing as a typical, standard user or use rather a diversity of users and uses
- Acceptance by a user depends on the obvious advantages, functionality, utility, usability, price/financial resources, (data)security and adequate design of the device as well as on her biographical and technological experiences
- New products should consider "old" habits of the users
- The systems should stay user-determined. At any time user intervention must be possible
- Information, training for usage, support, error diagnosis and error removal has to be appropriate for the target group
- Technologies should provide an additional aid to improve social life conditions; they can never replace social interaction
- The new living environment/ambiance should not generate new risks
- Integration into existing infrastructure should be easily accomplished
- Possibility of easy expansion/upgrades of products or integration of new devices according to (changing) user requirements and financial boundary conditions should be given.

According to the elderly, the cost of such a system was an important variable in deciding whether they were going to use it or not. If they live alone and if they can afford it, they would like to use and buy such a system. However, some of them stated that security was more important than cost. Other elderly persons did not care much about their privacy. As someone said: "What does privacy matters at our age" [Brey, 2005].

2.2 Human Activity Recognition Approaches

The ability to recognize human activities is a key factor if computing systems are to interact seamlessly with the persons environment. Research into enabling computer systems to recognize human activities has emerged as an application domain of computer vision research [Gavrila, 1999]. However, the more recent trends in human activity recognition have witnessed the appearance of another strand in this domain. Technological advancements have enabled instrumentation of our living environments with a large variety of multimodal sensors. Such environments possess the ability to monitor person behavior and provide information pertaining to persons, which is then filtered and processed in order to infer persons activities.
Human activity recognition can be divided into three major approaches, namely vision-based activity recognition, sensor-based activity recognition and multisensor-based activity recognition. Many approaches for human activity recognition have been proposed in the literature [Moeslund et al., 2006], [Gavrila, 1999]. Most of the work on activity recognition has focused on either identifying single activities in a particular scenario, or on analyzing sequences of activities.
Recognition of human activity has many important applications that rely on linking observed behavior with particular actions. However, activity recognition systems are usually built for specific applications, and the used architectures and solutions are often not applicable in other domains.
In this section we present related work on human activity recognition approaches. Firstly, we present the vision-based approaches and secondly, we present the sensor-based approaches. In the next section, we present the multisensor-based approaches .

2.2.1 Vision-Based Activity Recognition Approaches

The recognition of human activities from video sequences is a very important and active area of research for applications in video surveillance, multimedia communications, and medical diagnosis. Video surveillance is of increasing importance to many applications, such as security and healthcare of elderly [Harmo et al., 2005]. Automatic activity recognition plays an important part for video surveillance applications. It has become an important research topic in computer vision in recent years.
The problem of activity detection and recognition in the context of visual surveillance has received considerable attention [Aggarwal and Cai, 1999]. There been significant work that span across techniques for low level event detection [Zelnik-Manor and Irani, 2001], [Cohen and Medioni, 1999] and for activity modeling [Ivanov and Bobick, 2000], [Chowdhury and Chellappa, 2003].
A large number of different approaches have been developed, whose complexity and underlying models depend on the goals of the particular application which is targeted.

Many approaches to human activity recognition rely on background subtraction for extracting the location and shape of people in video sequences. The largest body on activity recognition is carried out using cameras and computer vision techniques [Gavrila, 1999], [Moeslund et al., 2006].

2.2.1.1 Probabilistic and Stochastic Approaches

There is a vast amount of literature in the area of computer vision, where the aim is to determine different types of human activity, mostly motion, from video images [Moeslund and Granum, 2001]. Usual modern methods applied include variations of neural networks (NNs) and hidden Markov models (HMMs). They are represented by graphs.
HMMs have been a popular tool for activity modeling, motivated primarily by its successful use in speech recognition. An HMM is a stochastic finite state machine which models an activity pattern by learning transition probabilities among its non-observable states such that the likelihood of observation of a temporal sequence of symbols representing the activity is maximized. HMMs have been used to model simple and more complex hand gestures [Oliver et al., 1999] and layered HMMs [Oliver et al., 2002] have been proposed to model events such as interaction between multiple mobile objects.
Chomat and Crowley [Chomat and Crowley, 1999] proposed a probabilistic method for recognizing activities from local spatio-temporal appearance. In [Yamato et al., 1992] the authors use HMM techniques to model human activities and to perform behavior recognition, but they are only based on representation of data.
Another popular approach for activity recognition is though the use of Bayesian networks. The authors in [Carter et al., 2006] combined Bayesian networks and Markov chains to recognize human behavior in airport apron scenes (AVITRACK project). In [Kumar et al., 2005] the authors proposed a framework for behavior understanding from traffic. Recently, in [Hoey et al., 2007] the authors successfully used only cameras to assist person with dementia during hand-washing. The system uses only video inputs, and combines a Bayesian sequential estimation framework for tracking hands and towel, with a decision using a partially observable Markov decision process.
Most of these methods mainly focus on a specific human activity and their description are not declarative and it is often difficult to understand how they work (especially for NNs). In consequence, it is relatively difficult to modify them or to add a priori knowledge.
The main advantage of Bayesian classifier and HMM approaches is that they are capable to model uncertainty by using probabilities. In the Bayesian classifier approaches the a priori probability needs to be learned and the learning stage is often tiresome. The Bayesian approaches are not adapted to model the temporal relations, because the time when the visual features have to be computed needs to be explicitly indicated.

Another advantage for the HMM approaches consists in their ability to recognize sequences of events, but they are limited when the recognition involves several mobile objects.

2.2.1.2 Constraint-Based Approaches

Constraint-based approaches have also been largely used to recognize activities for few decades. The main trend consists in designing symbolic networks whose nodes or predicates correspond to the Boolean recognition of simpler events. The first constraint-based approaches have been developed in the 70s and include plan recognition [Kautz and Allen, 1986] and event calculus [Kowalski and Sergot, 1986]. However, these approaches have not been applied to scene understanding based on real-world perceptual observations. Other approaches including Petri Net [C. Castel and Tessier, 1996] and [Lesire and Tessier, 2005], logic programming [L. Davis and Shet, 2005], script-based language, constraint resolution [Rota and Thonnat, 2000] and [C. J. Needham and Cohn, 2005] and chronicle recognition [Ghallab, 1996] and [C. Dousson and Ghallab, 1993], etc. have been adapted for recognizing activities through videos. For instance, Lesire and Tessier [Lesire and Tessier, 2005] have designed a Petri Net to recognize a given activity, whose nodes correspond to typical situations and the tokens to the mobile objects involved in the activity. But, this approach uses just one Petri Net to recognize one activity type and cannot recognize all the occurrences of the same activity. Stochastic grammar has been proposed to parse simple actions recognized by vision modules [Ivanov and Bobick, 2000]. Logic and Prolog programming have also been used to recognize activities defined as predicates [L. Davis and Shet, 2005]. Constraint Satisfaction Problem (CSP) has been applied to model activities as constraint networks [Rota and Thonnat, 2000].
These last three approaches are interesting and have successfully recognized complex activities. However, they do not have specific mechanisms to handle temporal constraints so they have to explore all possible temporal combinations of events and to store all totally recognized events to be used to recognize other more complex events. In practice, these approaches can recognize in real-time only activities involving a small number of physical objects.
Other techniques for the recognition of human activities have been proposed to reduce this combinatorial explosion by propagating the temporal constraints inside the constraint network. Then, the recognition is limited to only the sub-networks (complying with the satisfied temporal constraints) that can lead to a possible activity. These approaches store all partially recognized events and envisage all combinations that can occur and store only these predictions to recognize complete events in the future. For instance, an efficient version was proposed by Pinhanez and Bobick who described a temporal constraint network (called PNF for Past, Now and Future) to recognize activities. However, this network cannot represent event duration and is mainly dedicated to the recognition

of event sequences. More generally, the notion of chronicle was first introduced in [Kumar and Mukerjee, 1987] (and called dynamic situation) and then extended by Dousson and Ghallab [Ghallab, 1996] and [C. Dousson and Ghallab, 1993]. A chronicle is represented as a set of events (detected by specific routines) and sub-chronicles (recognized by the recognition process) linked by temporal constraints. The temporal aspects are the starting/ending time points of a chronicle and also the delay between two chronicles.
This approach recognizes correctly predefined chronicles and makes the recognition of chronicles possible in real-time. This approach has been applied to the video surveillance of metro stations [Chleq and Thonnat, 1996]. However, this algorithm was designed to recognize mono-physical-object events (i.e. chronicles), so, it contains a number of drawbacks for multi-physical-object events. For a multi-physical-object events, the algorithm has to create all predictions corresponding to all combinations of potential physical objects.

Techniques which are based on constraint resolution are among the most sophisticated event recognition techniques to date. They are able to recognize complex events involving multiple actors having complex temporal relationships. These techniques are used in [Vu et al., 2003] where the authors use a declarative representation of events which are defined as a set of spatio-temporal and logical constraints. These techniques have the advantage of being easily since they are based on constraints which are defined in a declarative way.

2.2.2 Sensor-Based Activity Recognition Approaches

An increasingly popular alternative approach is to use personalized sensors such as accelerometers to get precise information about a particular small set of features related to the person, such as limb-movement and person location. The majority of research using wearable devices has concentrated on using multiple sensors of a single modality, typically accelerometers on several locations on the body [Kern et al., 2003]. The placement of sensors in multiple predefined locations can be quite obtrusive and is one of the limitations of such an approach. The authors in [Guralnik and Haigh, 2002] describe the approach of collected data from a set of motion sensors installed in living environments. They used sequential patterns learning algorithms to extract the behavior patterns of the person (e.g. bathroom motion sensor fires between 7:00 am and 8:00 am after bedroom motion sensor which fires between 6:45 am and 7:45 am at 75% of the time). However, using only motion sensors is insufficient to deduce activities with high accuracy and also makes it very difficult to understand specific user behaviors. In [Kern et al., 2003], the authors describe a hardware platform equipped with three-dimensional accelerometers. However, results reported show only a small number of simple activities that are recognized including sitting, standing, walking, which may be attributed to using only one type of sensors. Bao and Intille [Bao and Intille, 2004] also propose recognizing human activities based on accelerometers. Authors report recognition accuracy up to

95%. However, their approach limits the number of activities the system can recognize. Another initiative in activity inference comes from University of Aarhus in Denmark [Bardram and Christensen, 2004]. Although the authors describe the issues that surround the activity inference, with a special focus on healthcare, inferring users activity based on the set of artifacts and other context information was found to be difficult, since activities are triggered by sources that are too complex to capture.

2.3 Multisensor-Based Activity Recognition Approaches

2.3.1 Definition of Sensor Fusion

Sensor Fusion is the combining of sensory data or data derived from sensory data such that the resulting information is in some sense better than would be possible when these sources were used individually. The main issue in sensor fusion is to provide higher accuracy and improved robustness against uncertainty and unreliable integration. The definition of sensor fusion does not say that input from more than one sensor is required; it only says that sensor data have to be combined in some sense. The definition also includes systems with a single sensor that takes multiple measurements that later on are fused [Elmenreich et al., 2001].

A non sensor fusion system may have to manage with a lot of different sensor types and ambiguous and incomplete data from these. If the input is fused prior it is sent to an application, the input interface of the application can be standardized and the application does not have to consider which sensor types that are used and by this reduce the complexity of the system. In [Elmenreich et al., 2001] the authors list a number of problems that physical sensor measurement can suffer from.

- **Sensor loss:** The loss of a sensor can cause a faulty observation of the object.
- **Limited spatial coverage:** A sensor covers usually only a restricted area.
- **Limited temporal coverage:** Limitation in the frequency in the production of measurements.
- **Imprecision:** The sensor may suffer from lack of precision.
- **Uncertainty:** May arise when the sensor fails to measure relevant attributes. Uncertainty, in contrast to imprecision, depends on the object being observed rather than the observing device. Uncertainty arises when features are missing (e.g. occlusions), when the sensor cannot measure all

relevant attributes of the percept, or when the observation is ambiguous. A single sensor system is unable to reduce uncertainty in its perception because of its limited view of the object.

One solution to the listed problems is to use sensor fusion.

2.3.2 Potential Advantages in Fusion of Multiple Sensors

The purpose of external sensors is to provide a system with useful information concerning some features of interest in the system's environment. The potential advantages in fusing information from multiple sensors are that the information can be obtained more accurately, concerning features that are impossible to perceive with individual sensors, in less time, and at a lesser cost. The following advantages can be expected from the fusion of sensor data from a set of heterogeneous sensors [Grossmann, 1998]:

- **Redundant information** is provided from a group of sensors (or a single sensor over time) when each sensor is perceiving, possibly with a different fidelity, the same features in the environment. The integration or fusion of redundant information can reduce overall uncertainty and thus serve to increase the accuracy with which the features are perceived by the system. Multiple sensors providing redundant information can also serve to increase reliability in the case of sensor error or failure.

- **Robustness and reliability:** Despite partial system failure the system can produce information depending on the redundancy in a system with multiple sensors.

- **Complementary information** from multiple sensors allows features in the environment to be perceived that are impossible to perceive using just the information from each individual sensor operating separately. If the features to be perceived are considered dimensions in a space of features, then complementary information is provided when each sensor is only able to provide information concerning a subset of features that form a subspace in the feature space, i.e., each sensor can be said to perceive features that are independent of the features perceived by the other sensors; conversely, the dependent features perceived by sensors providing redundant information would form a basis in the feature space.

- **Extended spatial and temporal coverage**, the combination of data gives the system a better overview of the surroundings. As compared to the speed at which it could be provided by a single sensor, may be provided by multiple sensors due to either the actual speed of operation of each sensor, or the processing parallelism that may be possible to achieve as part of the fusion process.

- **Less costly information**, in the context of a system with multiple sensors, is information obtained at a lesser cost when compared to the equivalent information that could be obtained from a single sensor. Unless the information provided by the single sensor is being used for additional functions in the system, the total cost of the single sensor should be compared to the total cost of the integrated multisensor system.

- **Increased confidence:** Information from more than one sensor covering the same object can support each others observations.

- **Reduced ambiguity and uncertainty:** The fused information decreases the ambiguity of the collected values.

A further advantage of sensor fusion is the possibility to reduce system complexity. In a traditionally designed system the sensor measurements are fed into the application, which has to cope with a big number of imprecise, ambiguous and incomplete data streams. In a system where sensor data is preprocessed by fusion methods, the input to the controlling application can be standardized independently of the employed sensor types, thus facilitating application implementation and providing the possibility of modifications in the sensor system regarding number and type of employed sensors without modifications of the application software [Elmenreich and Pitzek, 2001].

The role of multisensor fusion in the overall operation of a system can be defined as the degree to which each of these seven aspects is present in the information provided by the sensors to the system. Redundant information can usually be fused at a lower level of representation compared to complementary information because it can more easily be made commensurate. Complementary information is usually either fused at a symbolic level of representation, or provided directly to different parts of the system without being fused.

2.3.3 Possible Problems in Multisensor Fusion

Many of the possible problems associated with creating a general methodology for multisensor fusion, as well as developing systems that use multiple sensors, center around the methods used for modeling the error or uncertainty in the fusion process, the sensory information, and the operation of the overall system including the sensors. For the potential advantages in integrating multiple sensors to be realized, solutions to these problems will have to be found that are practical.

- **Error in the Fusion Process**: The major problem in fusing redundant information from multiple sensors is the determination that the information from each sensor is referring to the same features in the environment. This problem is termed the correspondence and data association problem in stereo vision and multitarget tracking research, respectively. Barniv and Casasent [Bamiv and Casasent, 1981] have used the correlation coefficient

between pixels in the gray level of images as a measure of the degree of recording of objects in the images from multiple sensors. Hsiao [Hsiao, 1988] has detailed the different geometric transformations needed for recording.

- **Error in Sensory Information**: The error in sensory information is usually assumed to be caused by a random noise process that can be adequately modeled as a probability distribution. The noise is usually assumed not to be correlated in space or time (i.e., white), and Gaussian. The major reasons that these assumptions are made is that they enable a variety of fusion techniques to be used that have tractable mathematics and yield useful results in many applications. If the noise is correlated in time (e.g., gyroscope error) it is still sometimes possible to retain the whiteness assumption through the use of a shaping filter [Maybeck, 1982].
  The Gaussian assumption can only be justified if the noise is caused by a number of small independent sources. In many fusion techniques the consistency of the sensor measurements is increased by first eliminating spurious sensor measurements so that they are not included in the fusion process. Many of the techniques of robust statistics can be used to eliminated spurious measurements.

- **Error in System Operation**: When error occurs during operation due to possible coupling effects between components of a system, it may still be possible to make the assumption that the sensor measurements are independent if the error, after calibration, is incorporated into the system model through the addition of an extra state variable. In well-known environments the calibration of multiple sensors will usually not be a difficult problem, but when multisensor systems are used in unknown environments, it may not be possible to calibrate the sensors. Possible solutions to this problem may require the creation of detailed knowledge bases for each type of sensor so that a system can autonomously calibrate itself. One other important feature required of any intelligent multisensor system is the ability to recognize and recover from sensor failure [T.E. Bullock and Boudreau, 1988].

2.3.4 Sensor Fusion Levels

Sensor fusion can be classified into different levels according to the input and output data types [Dasarathy, 1996], [Dasarathy, 1997]. The fusion may take place at the data level also called signal level, feature level also called symbolic level and decision level related to task level.
In data level fusion, each sensor observes an object and the raw output data of sensors are combined [Luo and Kay, 1992]. Varieties of the methods are developed in this level, and were applied in image processing [Goodridge and Kay, 1996] and in visual and speech recognition [Kabre, 1995].
In feature level fusion, each sensor provides observational data from which a feature vector is extracted. These vectors are then concatenated together into

a single feature vector. Because most features have well-defined structures, the fusion methods in this level can be based on statistical approaches and pattern analysis approaches [Bajcsy et al., 1996], [MacLeod and Summerfield, 1987].
Decision level fusion involves combination of sensor high level output data (e.g. event). Decision fusion is a common problem in many research areas, such as decision theory and artificial intelligence.
Different levels of multisensor fusion can be used to provide information to a system that can be used for a variety of purposes; e.g. data-level fusion can be used in real-time applications and can be considered as just an additional step in the overall processing of the signals, feature-level fusion can be used to improve the performance of many image processing tasks like segmentation, and decision-level fusion can be used to provide an object recognition system with additional features that can be used to increase its recognition capabilities [Luo and Kay, 1990]. Figure 2.6 summarizes these three sensor fusion levels. Each of these fusion levels has distinct advantages and disadvantages in our scenario.

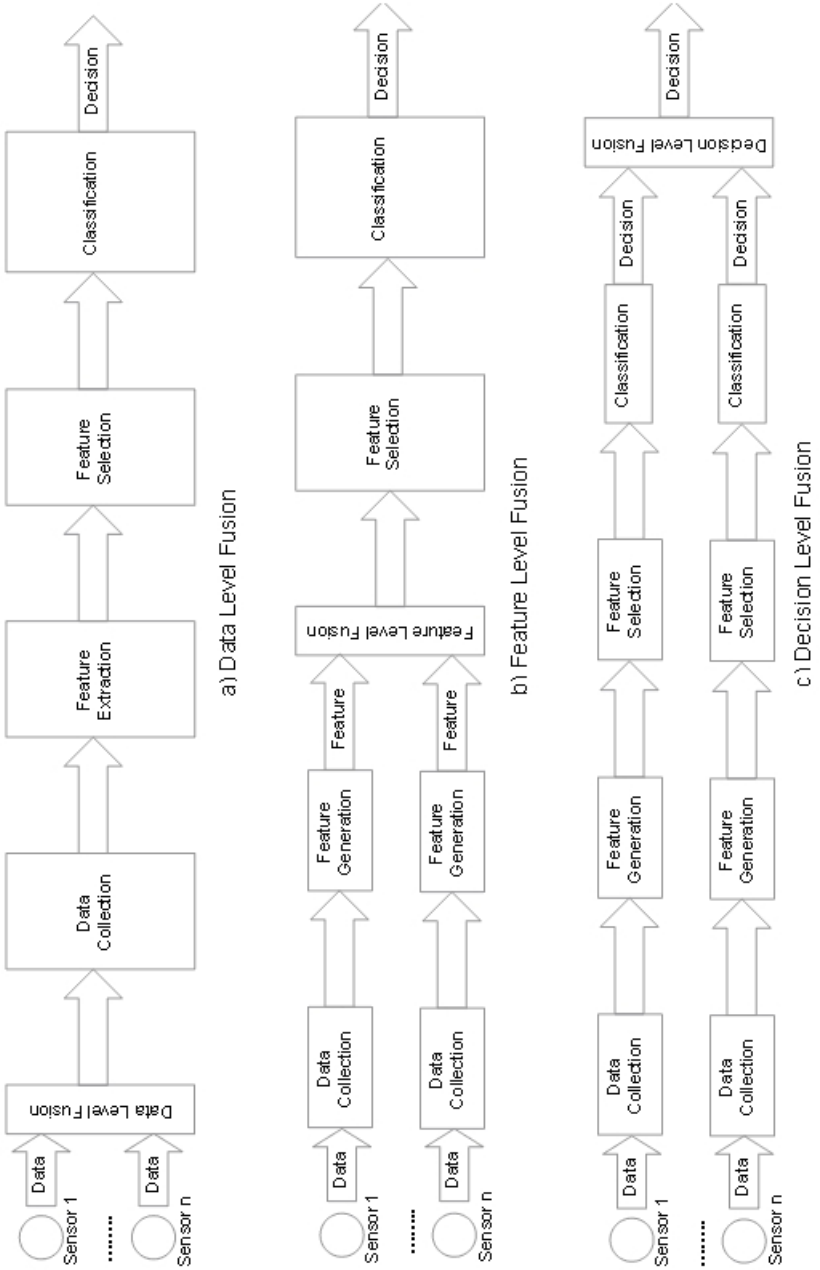


Figure 2.6: Fusion Levels: a) Data Level Fusion; b) Feature Level Fusion; c) Decision Level Fusion; Adapted from Yang et al. [Yang, 2006]

Inference Methods	**-Bayesian Inference** **-Dempster-Shafer Method** **-Evidence Processing**
Estimation Methods	**-Maximum Likelihood** **-Kalman Filter** **-Particle Filter** **-Bayesian Estimation**
Classification Methods	**-Cluster Analysis** **-K-means Clustering**

Table 2.2: Multisensor Fusion Methods

In data level fusion approach, sensors transmit all collected data without, or with minimal processing to a centralized processing system for analysis. Since reduction may now occur with all collected data available it is less likely that patterns observable across multiple sensors will be missed. Also as minimal processing is required by the sensors, these may be manufactured cheaply. This scheme may be problematic for wireless sensors however, as the high volume of communication may quickly diminish battery life.
The converse is true for decision level fusion, battery life may be traded for accuracy in inference by transmitting data from sensors only at the decision level. Unfortunately this burdens the sensors with a level of computation that may be unfeasible depending on the nature of the inference algorithms implemented by these sensor networks.
Feature level fusion stands in the middle ground between these two extremes. Features are generated that are representative of individual signals and transmitted onwards. Features are then composed, further reduced then used to classify the phenomenon under observation. If features are sufficiently descriptive of their signal then the loss of patterns across multiple sensors should not be a problem. Communication overhead is reduced compared to data level fusion as are the computation requirements over decision level fusion.

2.3.5 Sensor Fusion Approaches

As shown in Table 2.2, multisensor fusion algorithms can be broadly classified as follows: inference methods, estimation methods and classification methods.

2.3.5.1 Inference Methods

Inference methods are often applied in decision fusion. In this case, a decision is taken based on the knowledge of the observed situation. Here, inference refers to the transition from one likely true proposition to another, whose truth is believed to result from the previous one. Classical inference methods are based on Bayesian inference and Dempster-Shafer Belief Accumulation theory.

- **Bayesian Inference:** Information fusion based on Bayesian Inference offers a formalism to combine evidence according to rules of probability theory. The uncertainty is represented in terms of conditional probabilities describing the belief, and it can assume values in the [0, 1] interval, where 0 is absolute disbelief and 1 is absolute belief. Bayesian inference is based on the rather old Bayes rule, which states that:

$$Pr(Y|X) = \frac{Pr(X|Y)Pr(Y)}{Pr(X)} \tag{2.1}$$

  where the posterior probability $Pr(Y|X)$ represents the belief of hypothesis Y given the information X. This probability is obtained by multiplying $Pr(Y)$, the prior probability of the hypothesis Y , by $Pr(X|Y)$, the probability of receiving X , given that Y is true; $Pr(X)$ can be treated as a normalizing constant. The main issue regarding Bayesian Inference is that the probabilities $Pr(X)$ and $Pr(X|Y)$ have to be estimated or guessed beforehand since they are unknown.
  Coue et al. [Coue et al., 2002] use Bayesian programming, a general approach based on an implementation of Bayesian theory, to fuse data from different sensors (e.g. laser, radar, and video) to achieve better accuracy and robustness of the information required for high-level driving assistance. Work in event detection for wireless sensor networks is proposed by Krishnamachari and Iyengar [Krishnamachari and Iyengar, 2004] who explicitly consider measurement faults and develop a distributed and localized Bayesian algorithm for detecting and correcting such faults. This work is further extended by Luo et al. [Luo et al., 2006] who consider both measurement errors and sensor faults in the detection task.

- **Dempster-Shafer Inference:** Dempster-Shafer Inference is based on the Dempster-Shafer Belief Accumulation (also referred to as Theory of Evidence or Dempster-Shafer Evidential Reasoning), which is a mathematical theory introduced by Dempster [Dempster, 1968] and Shafer [Shafer, 1976] that generalizes the Bayesian theory. It deals with beliefs or mass functions just as Bayes rule does with probabilities. The Dempster-Shafer theory provides a formalism that can be used for incomplete knowledge representation, belief updates, and evidence combination [Provan, 1992]. The theory is based on a number of key propositions which are summarized as follows:
  Frame of discernment: A sensor can have either a value of one (active) or zero (inactive). The two values comprise the exhaustive set of mutually exclusive values that the sensor can hold. In DS theory, the set is called the frame of discernment of the sensor, denoted by Θ.
  For example, swatr, ¬swatr is the frame of discernment for the water flow sensor, in which swatr means the sensor is active and ¬swatr means an inactive sensor.
  Mass function: Many factors surrounding the sensor have an impact

on the quality of the sensor observation. For example, the person which drops his bag on the chair, may activate the chair sensor (sensor installed under the chair) and giving a false result. The observation of the sensor is inherently evidential. DS theory uses a number in the range [0, 1] to represent the degree of belief in the observation. The distribution of a unit of belief over the frame of discernment is called evidence. A function $m : 2^{\Theta} \rightarrow [0, 1]$ is called a mass function, representing the distribution of belief and satisfying the following two conditions:

$$m(\phi) = 0 \tag{2.2}$$

$$\sum_{A \subseteq \Theta} m(A) = 1 \tag{2.3}$$

Where: ϕ: is the empty set and A: is a sub-set of Θ.
Belief and plausibility: Dempster used a range of probability rather than a single probabilistic number to represent uncertainty. The lower and upper bounds of the probability are called the belief and plausibility respectively, which can be defined by mass functions as follows:

$$Bel(A) = \sum_{B \subseteq A} m(B) \tag{2.4}$$

$$Pls(A) = \sum_{B \supseteq A} m(B) \tag{2.5}$$

Bel represents the degree of belief to which the evidence supports A.
Pls describes the degree of belief to which the evidence fails to refute A, that is, the degree of belief to which it remains plausible.

The difference $Pls(A) - Bel(A)$ describes the uncertainty concerning the hypothesis A represented by the evidential interval, see figure 2.7.

$$\delta(A) = Pls(A) - Bel(A) \tag{2.6}$$

2.3.5.2 Estimation Methods

Estimation methods were inherited from control theory and use the laws of probability to compute a process state vector from a measurement vector or a sequence of measurement vectors [Bracio et al., 1997]. In this section, we present the estimation methods known as: Maximum Likelihood, Kalman filter, and Particle filter.

- **Maximum Likelihood (ML):** Estimation methods based on Likelihood are suitable when the state being estimated is not the outcome of a random variable [Brown et al., 1992].

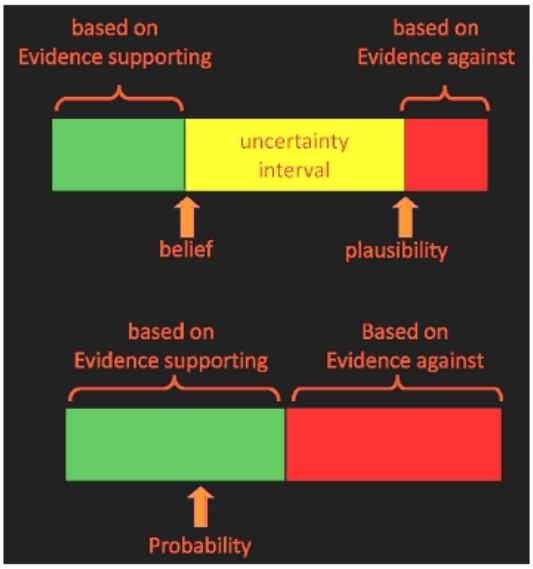


Figure 2.7: Difference between the two concepts of Probability versus the concept of Dempster-Shafer

In the context of information fusion, given x, the state being estimated, and $z = (z(1), ..., z(k))$, a sequence of k observations of x, the likelihood function $\lambda(x)$ is defined as the probability density function (pdf) of the observation sequence z given the true value of the state x:

$$\lambda(x) = P(z|x) \tag{2.7}$$

The Maximum Likelihood (ML) searches for the value of x that maximizes the likelihood function.

- **Kalman Filter:** The Kalman filter is a very popular fusion method. It was originally proposed in 1960 by Kalman [Kalman, 1960] and it has been extensively studied since then [Luo and Kay, 1992].
  The Kalman filter is used to fuse low-level redundant data. If a linear model can describe the system and the error can be modeled as Gaussian noise, the Kalman filter recursively retrieves statistically optimal estimates.

2.3.6 Sensor Fusion Work for Healthcare

This section presents some systems that perform elderly activity recognition with the aid of sensor fusion techniques.

In [Mehboob et al., 1997], a method entitled Robust Sensor Fusion (RSF) is used to fuse data from multiple, redundant sensors in order to obtain the

most accurate estimate of heart rate. In addition to consistency of data between multiple sensors RSF also utilizes temporal consistency at individual sensors during operation. RSF allows the combination of heart rate signals from multiple sensors such that the combined heart rate estimate is closer to the "true" value. RSF also provides a confidence value with every estimate indicating the likelihood of its correctness.
The heart rate sensors considered are the electrocardiogram (ECG) and the pulse oxymeter (SpO2). By fusing heart rate signals from each, the combined accuracy would be above the accuracy of any single sensor alone. Furthermore, the authors wished to examine whether this improved estimation would reduce the frequency of false heart rate alarms.
The motivation for this work is that each of the above individual sensors have independent causes of artifact. Heart rate estimate is calculated as a weighted average of individual signals, taking into account also past estimates using a Kalman filter. To obtain higher accuracy in estimation, erroneous sensor data are identified and excluded from the weighted averaging process using the consensus between measurements, the similarity of sensor data to an estimate based upon past estimates only and also upon the physiological consistency of these estimates.
French researchers Virone et al. [Virone et al., 2003], have experimented with the fusion of audio and contact sensors for home healthcare applications in their Smart Home Information System (HIS). The HIS consists of a multitude of sensors (such as door contacts and tensiometers) and is augmented through the use of 8 microphones linked to form a single smart audio sensor. The system is capable of generating both short term alerts, those which are instantaneously triggered on reception of a message from the HIS or audio sub-system, and long term alerts, triggered after data analysis of more long term data is performed. Minimal results are provided to show the advantage of sensor fusion for the detection of pathological disease in a home healthcare scenario.
Different types of Markov models have been used to carry out task identification from a sequence of sensor events. One such approach was by Wilson et al [Wilson et al., 2005], where episode recovery experiments were carried out and analyzed by a Hidden Markov Model (HMM) using the Viterbi algorithm which was responsible for determining which task is active from the sequence of sensor events. Although this approach enabled unsupervised task identification it was not as efficient when the tasks were carried out in a random order. Other approaches that have been developed in order to carry out reliable activity recognition and solve the incomplete sensor problem involve ontologies [Munguia-Tapia et al., 2006] and data mining techniques [Wyatt et al., 2005]. Ontologies have been utilized to construct reliable activity models that are able to match an unknown sensor reading with a word in an ontology which is related to the sensor event. For example, a Mug sensor event could be substituted by a Cup event in the task identification model "Make Tea" as it uses Cup.

2.4 Conclusion

In this chapter, previous work on human activity recognition has been presented. Most of the presented systems that have been built to recognize home activities have been limited in the variety of activities they recognize. In particular, most previous work on activity recognition has used sensors that provide only a very coarse idea of what is going on. For example, by detecting only movement in a room, it is not possible to detect which activity occurs in the room.
The accuracy of the techniques using a single type of sensor (video and non-video) has been shown but they are limited for application. In contrary, multisensor techniques are well adapted to healthcare applications and they are more generic than approaches using single sensor.
As previously introduced, our objective is to propose an approach based on multisensor data fusion. This approach combines the advantages of the visions techniques and the non-vision techniques and aims to determine when a person uses the household equipment and to detect most of the activities at home.
In the next chapter, an overview of the proposed approach is given.

Chapter 3

Activity Recognition Approach Overview

The goal of human activity recognition is to provide accurate information about the behavior of a person observed in a scene. As seen in chapter 2, the activity recognition problem has been treated with probabilistic approaches and constraint resolution approaches. Our goal is to propose a framework that takes the advantages of each approach.
The objectives are presented in section 3.1, an overview of the proposed cognitive vision approach for activity recognition is described in section 3.2 and finally, a conclusion is presented in section 3.3.

3.1 Objectives

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. By being able to recognize and monitor activities of daily living such as preparing a meal, eating, bathing, etc, automatic detection of changes in patterns of behavior is possible. This information can reveal a decline in health, risks in the environment, and emergency situations that may require the assistance of caregivers.
The goal of this work is to propose an approach based on using ambient sensor technologies to recognize interesting activities at home. This approach combines data from video cameras with data from environmental sensors to 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). It includes an algorithm for real-time recognition of primitive and complex activities that have occurred in the scene observed by video cameras and sensors attached to house furnishings. The proposed approach consists in detecting people, tracking people as they move, and recognizing activities of interest based on multisensor analysis and human activity recognition.

3.1.1 A Framework for Activity Recognition

In this section, we describe firstly the challenges in activity recognition and after that we describe our monitoring goals.

3.1.2 Challenges in Activity Recognition

To create algorithms that detect activities, computational models that capture the structure of activities must be developed. The behavior of an individual can be characterized by the temporal distribution of his/her activities such as patterns in timing, duration, frequency, sequential order, and other factors such as location, cultural habits, and age.
Based on the state of the art already presented in chapter 2, below are human behavior attributes that present challenges for recognition:

- **Multitasking:** Individuals often perform several activities at the same time when they do any kind of work that does not fully engage their attention.

- **Periodic variations:** Everyday activities are subject to periodic daily, weekly, monthly, annual, and even seasonal variations. For example, a person might typically prepare breakfast in 15 minutes on weekdays and for one hour during weekends.

- **Time scale:** Human activities also occur at different time scales. For example, cooking lunch can take 25 minutes, while toileting may only take a few minutes.

- **Sequential order complexity:** Sequential order, the position in time that an activity has in relation to those activities preceding and following it, is particularly important. The choice of what to do next as well as how that activity is performed is strongly influenced by what one has already done and what one will do after. For example, preparing lunch is very likely followed by eating.

- **False starts:** A person may start an activity, and then suddenly begin a new activity because something more important has caught his/her attention or because he/she simply forgot about the original activity.

- **Location:** Human behavior is also affected by location. For example, cleaning the kitchen involves a different sequence of actions than cleaning the bathroom.

- **Cultural habits:** Some cultural habits may be expressed by individuals in typical sequences of activities. For example, in some cultures, people take a nap after lunch while others have a cup of tea before having breakfast, lunch or dinner.

3.1.3 Monitoring Goals

As seen previously in chapter 1, monitoring activities at home is predominantly composed of location and activity information. Below is a list of exactly what we wish to automatically recognize:

- **Presence:** Determine whether one or several individuals are present in the environment.
- **People Tracking:** Determine the location of each person (e.g. in the kitchen).
- **Posture:** Recognize body configuration such as standing, bending, sitting.
- **Interactions:** Recognize how a person interacts with the environment (e.g. opens the fridge, sitting on a chair).
- **Activities of Daily Living (ADLs):** Recognize daily activities such as cooking, eating, bathing, toileting [Katz et al., 1963], [Lawton, 1990].

3.2 Proposed Activity Recognition Approach

In this thesis, an approach for recognizing activities at home is proposed (e.g. an elderly person living alone has taken a meal). The approach combines data provided by video cameras with data provided by environmental sensors to monitor the interaction of people with the environment. The environmental sensors we used are attached to house furnishings. They are easy to install in home environment and removable without damage to the cabinets or furniture. The proposed sensors require no major modifications to existing homes and can be easily retrofitted in real home environments.

3.2.1 Architecture of the Proposed Approach

The proposed approach consists in collecting multisensor data of the person in order to build up a "normal" profile of his/her daily activity patterns (e.g. use the refrigerator, prepare a meal, sitting on a chair, go to bed). Large deviations from this profile should alert a human operator. The proposed multisensor based activity recognition approach uses video cameras and environmental sensors.

As described in Figure 3.1, the input of the proposed approach consists in the data provided by the different sensors. Its output is a set of XML files and alarms and also a 3D visualization of the recognized events. The proposed approach consists in a 4D (3D + time) analysis of multisensor data. It exploits three major sources of knowledge: 3D models of person (e.g. 3D size of a person), the models of events predefined in collaboration with gerontologists and the 3D information of the scene (e.g. position and size of furniture, zones of interest).

The proposed multisensor based activity recognition approach is composed of four components:

1. **Video Analysis:** detects, tracks people moving in the scene, and also detects the body configuration of the person.

2. **Sensor Analysis:** collects information about interactions between people and the contextual objects and process them.

3. **Event Recognition:** recognizes a set of simple video events (e.g. a person leaves the kitchen) and also recognizes a set of simple environmental events (e.g. the fridge is open).

4. **Multisensor Event Fusion:** recognizes complex (multimodal) events by combining video events with environmental events (e.g. a person prepares a meal).

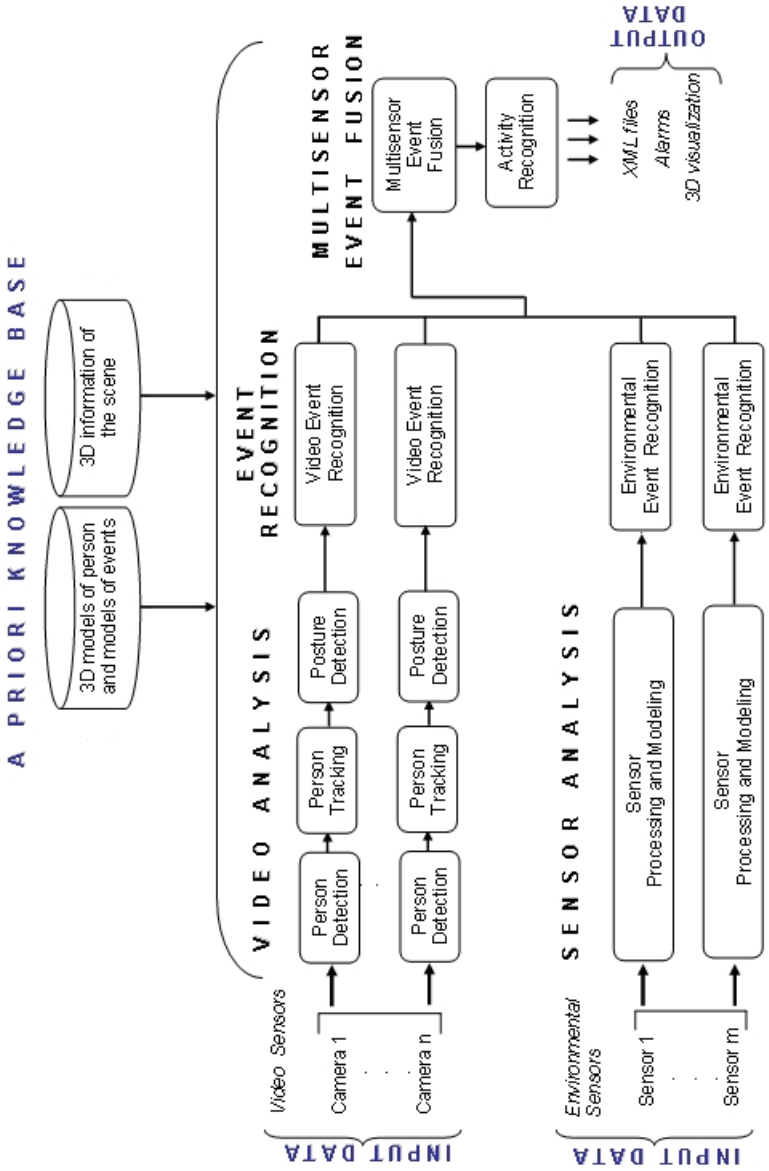


Figure 3.1: General Architecture of the Approach

3.2.2 Video Analysis

In this section we firstly describe person detection and person tracking methods. After that we describe posture detection method. This method includes human posture recognition algorithm, and a 3D human posture. Figure 3.2 illustrates the video analysis component with our contributions.

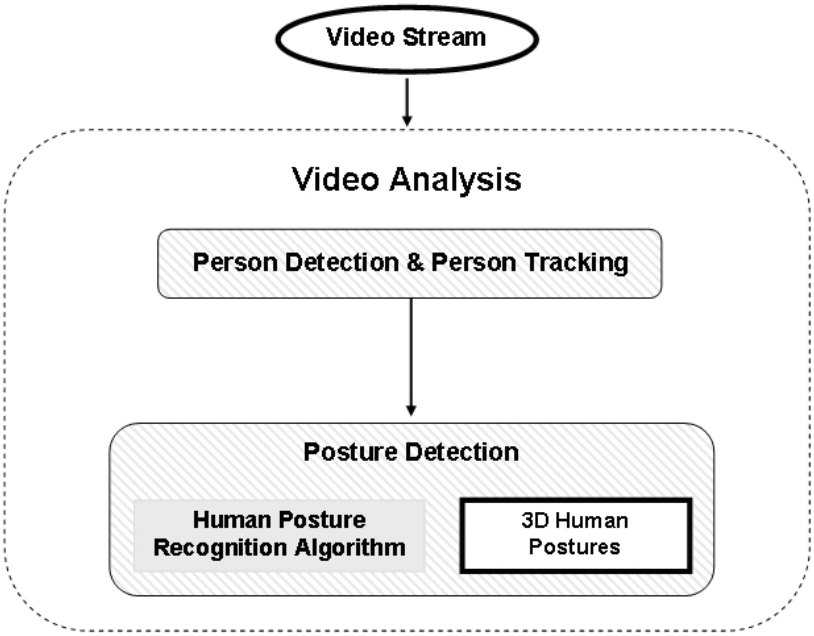


Figure 3.2: The video analysis architecture. Our major contributions are represented in bold lines with white background. Our minor contributions are represented with dashed background and the existing methods are represented with gray background.

3.2.2.1 Person Detection and Person Tracking

Video analysis aims at detecting and tracking people moving in the scene. To achieve this task, we have used a set of vision algorithms coming from a video interpretation platform described in [Avanzi et al., 2005].
A first algorithm segments moving pixels in the video into a binary image by subtracting the current image with the reference image. A background subtraction method [Heikkila and Silven, 1999] segments the picture and compares intensity and color with a periodically updated reference background image not containing the moving object [McIvor, 2000]. The reference image is updated along the time to take into account changes in the scene (e.g. light, object displacement, shadows).

A 3D information is obtained by using a calibration step which computes

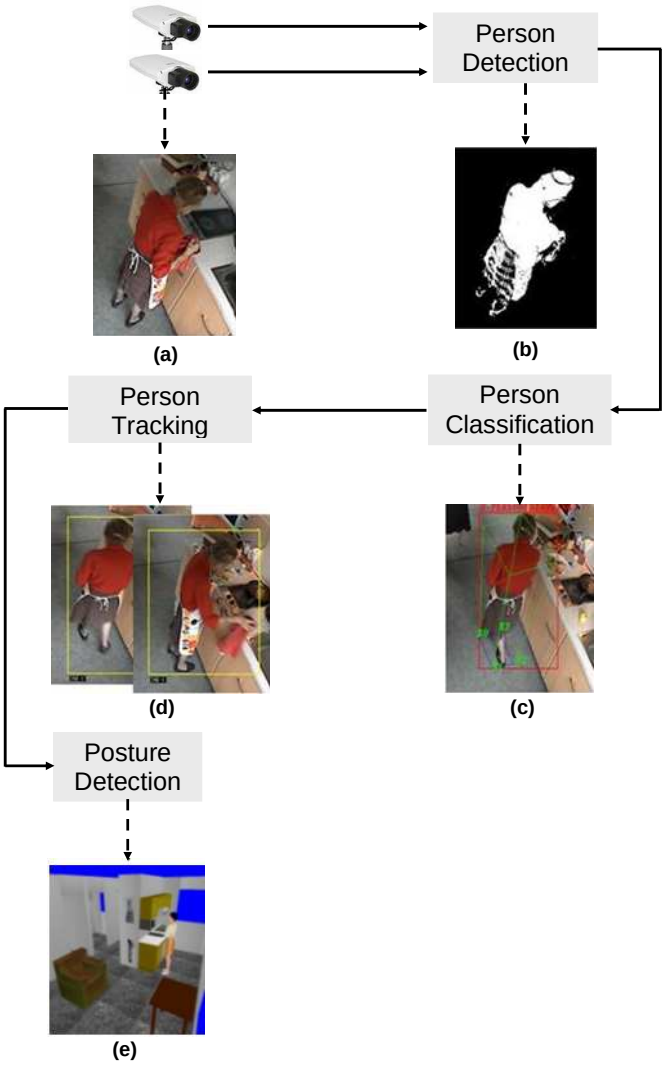


Figure 3.3: Video Analysis. (a) Represents the original image, (b) the detection of moving pixels which are highlighted in white and clustered into a mobile object, (c) the mobile object is classified as a person, (d) shows the tracking at 2 different times of the same person (IND 0), (e) shows the corresponding 3D posture of the tracked person in the 3D environment.

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 the legs of a person are occluded by a specified contextual object and therefore not visible by the camera, the person is supposed to be just behind the object.

Internal parameters of the camera (image center, focal length and distortion coefficients) are combined with external parameters (position and orientation relative to a world coordinate system) to compute the calibration matrix. In the Tsai camera calibration method [Tsai, 1986], 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.

The moving pixels are then grouped into connected regions, called blobs. A set of 3D features such as 3D position, width and height are computed for each blob. Then, 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.

For each moving region, a 3D classifier adds an object class label (e.g. person, vehicle) [M. Zúñiga, 2006]. After that, the tracking task adds a unique identifier to each new classified blob, and maintains it globally throughout the whole video (see Figure 3.4).

3.2.2.2 Posture Detection

In [Boulay et al., 2006] a very precise 3D model of human is utilized to detect postures. Human posture is described by a set of 23 parameters. This human model enables to generate 2D silhouettes to be compared with the one detected for a person in the scene (see Figure 3.5).

- **Human Posture Recognition Algorithm** We have used a human posture recognition algorithm [Boulay et al., 2006] in order to recognize in real time a set of human postures once the person moving in the scene is correctly detected. This algorithm determines the posture of the detected person using the detected silhouette and its 3D position. The human posture recognition algorithm is based on the combination between a set of 3D human models with a 2D approach. These 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 (see figure 3.6). 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

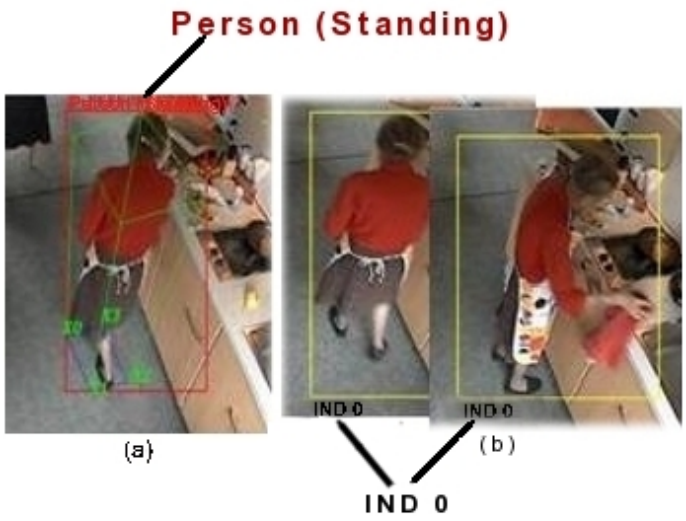


Figure 3.4: (a) Classification of the object as a person with standing posture and a 3D parallelepiped indicates the position and orientation of that person; (b) Tracking at 2 different times of the same person (IND 0)

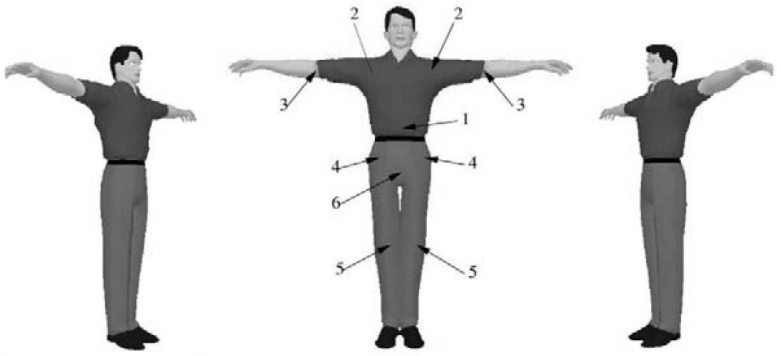


Figure 3.5: Model of human posture described by a set of 23 parameters

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.

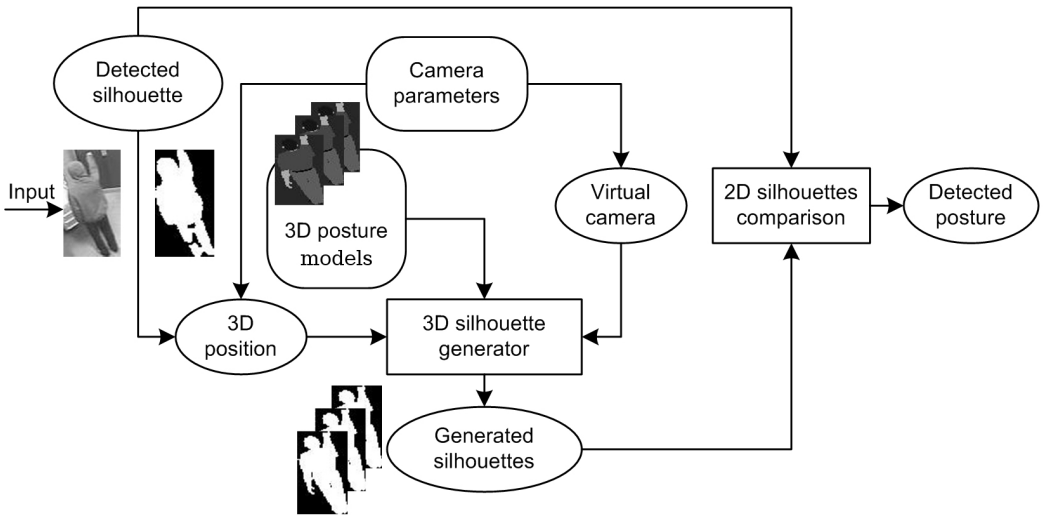


Figure 3.6: Simplified scheme showing the posture recognition approach

- **3D Human Posture** In this thesis, in collaboration with gerontologists and geriatrics from the Nice hospital in France, we have proposed a set of 3D human postures. These 3D human postures are based on a 3D geometrical human model. For homecare applications we propose ten 3D key human postures which are useful to recognize activities of interest at home. These postures are displayed in figure 3.7: standing (a), standing with arm up (b), standing with hands up (c), bending (d), sitting on a chair (e), sitting on the floor with outstretched legs (f), sitting on the floor with flexed legs (g), slumping (h), lying on the side with flexed legs (i), and lying on the back with outstretched legs (j).
  Each of the proposed 3D human postures plays a significant role in the recognition of the targeted activities of daily living or of abnormal activities. For example, the posture "standing with hands up" (see figure 5.9) is used to detect when a person is carrying an object such as plates. The posture "standing with arm up" (see figure 5.10) is used to detect when a person reaches and opens kitchen cupboard and his/her ability to do it. These proposed human postures are not an exhaustive list but represent the key human postures taking part in everyday activities.

Figure 3.3 illustrates the detection, classification, tracking and posture detection of a person in an experimental laboratory.

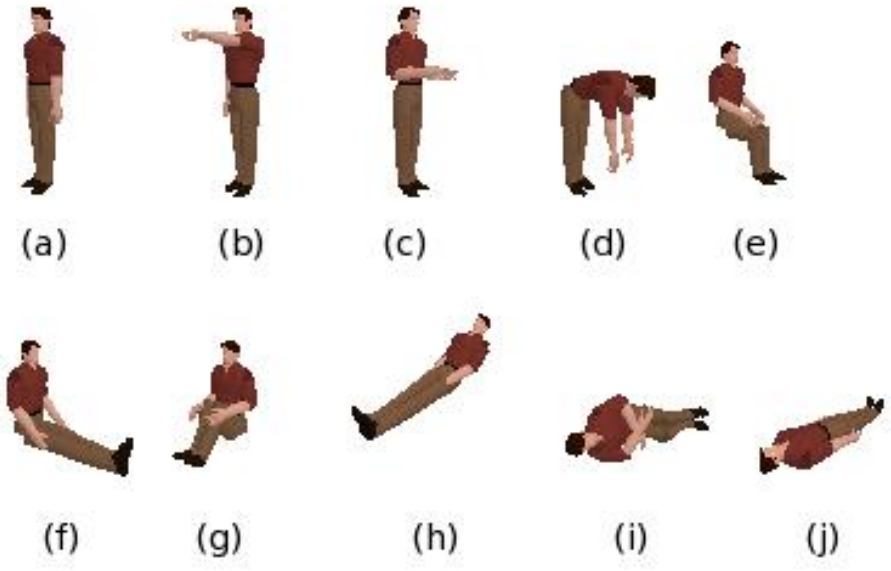


Figure 3.7: The proposed 3D human postures.

3.2.3 Sensor Analysis

In this section, we describe the sensor processing and modeling method which include the proposed sensor model which is necessary to fuse multisensor systems. This sensor model includes an uncertainty in sensor measurements.
Figure 3.8 illustrates the sensor analysis component with our contributions.

3.2.3.1 Sensor Processing and Modeling

The physical sensor (e.g. electrical sensor) produces a response to the surrounding environment. For instance the electrical sensor triggers a signal when an appliance is used. The raw data collected by the physical sensors is processed to produce high-level representations of sensed object. This process converts the physical sensor response into a representative value of the raw environmental characteristics, such as electrical current.

Handling Uncertainty in Sensor Measurement

Because each sensor type has different characteristics and functional description, it is necessary to find a general model that is independent from the physical sensors, and that enables comparison of the performance and robustness of such sensors. For solving this issue we propose a generic sensor model in order to develop a coherent and efficient representation of the information provided by sensors of different types. This sensor model is able to give a coherent and

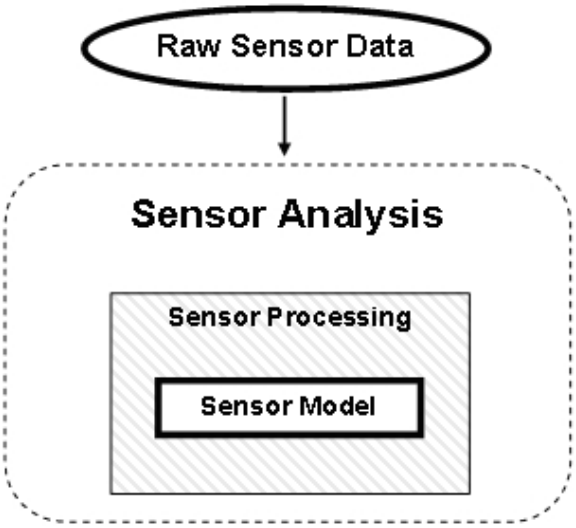


Figure 3.8: The sensor analysis architecture. Our major contributions are represented in bold lines with white background. Our minor contributions are represented with dashed background.

efficient representation of the information provided by various types of sensors. This representation provides means for recovery from sensor failure and also facilitates reconfiguration of the sensor system when adding or replacing sensors. In the proposed sensor model we define the type of information (e.g. pressure, image, motion) and the measurement $\mathbf{y}$ which is the value of the physical property measured by the sensor. We also define the uncertainty Δy of measurement $\mathbf{y}$. It includes errors in $\mathbf{y}$, such as measurement errors. More details about sensor modeling are described in chapter 4.

3.2.4 Event Recognition

In this work, we propose to represent the activities of interest into a formal model that satisfies a number of constraints by using the event description language proposed by Vu et al. [Vu et al., 2003]. We have extended this language to address complex activity recognition involving several physical objects of different types (e.g. person, chair) in a scene observed by video cameras and environmental sensors and over an extended period of time.
In this section, we firstly describe the event modeling. After that we describe the event recognition algorithm. Figure 3.9 illustrates the event recognition component with our contributions.

3.2.4.1 Event Modeling

The event models correspond to the modeling of all the knowledge used by the system to detect events occurring in the scene. The description of this

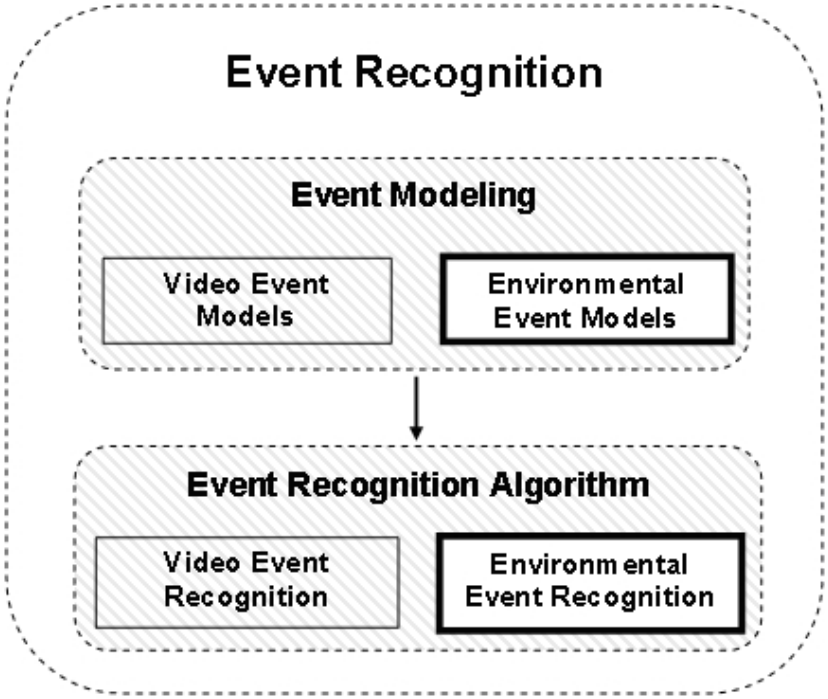


Figure 3.9: The event recognition architecture. Our major contributions are represented in bold lines with white background. Our minor contributions are represented with dashed background.

knowledge is declarative and intuitive (in natural terms), so that the experts of the application domain can easily define and modify it. Four types of event can be defined: primitive state, composite state, primitive event and composite event. 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** (e.g. a person is located inside a zone) corresponds to a perceptual property characterizing one or several physical objects. A **composite state** is a combination of primitive states. A **primitive event** corresponds to a change of primitive state values (e.g. a person changes a zone). A **composite event** is a combination of primitive states and/or primitive events.

An event model M of an event E is composed of five parts (see figure 3.10):

- "**Physical objects**" which are a set of variables whose values correspond to the physical objects involved in E,
- "**Components**" which are a set of variables whose values correspond to the event instances composing E,

- "**Forbidden components**" which are a set of variables corresponding to all event instances that are not allowed to be recognized during the recognition of E,
- "**Constraints**" which are a set of conditions between the physical objects and/or the components to be verified for the recognition of E, they include symbolic, logical, spatial and temporal constraints (Allens interval algebra operators [Allen, 1983]),
- "**Alerts**" which are an optional part of an event model which correspond to a set of actions to be performed when E is recognized.

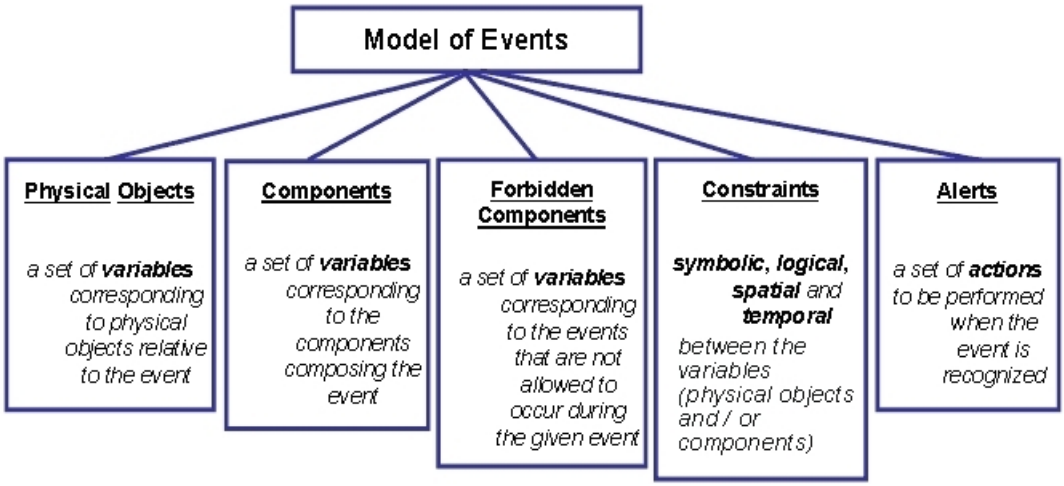


Figure 3.10: Model of Events.

A primitive state must contain at least, one physical object and one constraint. A primitive and composite events must contain at least, one physical object, one component and one constraint. Forbidden components and alerts are optional.

3.2.4.2 Event Recognition Algorithm

The event recognition process we used [Vu et al., 2003] is able to recognize which events are occurring in the scene at each instant. To benefit from all the knowledge, the event recognition process uses the coherent tracked mobile objects, the a priori knowledge of the scene and the predefined event models. To be efficient, the recognition algorithm processes in specific ways events depending on their type. Moreover, this algorithm has also a specific process to search previously recognized events to optimize the whole recognition. The algorithm is composed of two main stages. First, at each step, it computes all possible primitive states related to all mobile objects present in the scene. Second, it computes all possible events (i.e. primitive events then composite states and events) that may end with the recognized primitive states.

3.2.5 Multisensor Event Fusion

By using only vision sensors, we can detect some simple activities of the observed person such as the location and posture of the person in the apartment. Monitoring activities at home is predominantly composed of locations, postures and interactions with equipments. For this we choose to use video cameras combined with environmental sensors to determine when a person uses the household equipment and to detect most of the activities at home.
The environmental sensors are more robust and precise but need to be installed everywhere resulting on a prohibiting price (the cost of system is usually due to the number of sensors, wiring and maintenance). The cameras are less precise but more global and usually one camera can be enough to monitor one room.

In this section, we describe how to combine the video events with the environmental events and the activity recognition method.
Figure 3.11 illustrates the multisensor fusion event with our contribution. More details of this multisensor fusion approach are described in chapter 5.

3.2.5.1 Video & Environmental Event Fusion

As described in chapter 2, sensor fusion can be classified into different levels according to the input and output data types. Fusion may take place at the data level, feature level and decision level. In data level fusion, raw output data of sensors are combined. In feature level fusion, each sensor provides observational data from which a feature vector is extracted. These vectors are then concatenated together into a single feature vector. The decision level fusion involves combination of sensor high level output data (e.g. event).
The use of sensor fusion at the decision level facilitates an extensible sensor system, because the number and types of sensors are not limited.
In our approach, we use a fusion process at the decision level to address the problem of heterogeneous sensor system. For this, we combine the video events with the environmental events in order to detect rich and complex events (i.e. multimodal events). The environmental sensor data and the video sensor data are fused at the level of event recognition. The multimodal events can include video event and / or environmental event. Therefore, when the video and the environmental events are recognized, then the global multimodal event is also recognized.

3.2.5.2 Activity Recognition

The multisensor event recognition algorithm takes as input sensor events (i.e. video and environmental) and a priori knowledge of composite events to be recognized. An event model M should be recognized at an instant t if all its components have been recognized, its last (using the temporal order) component being recognized at the given instant t.

The use of an heterogeneous sensor system involves a synchronization task

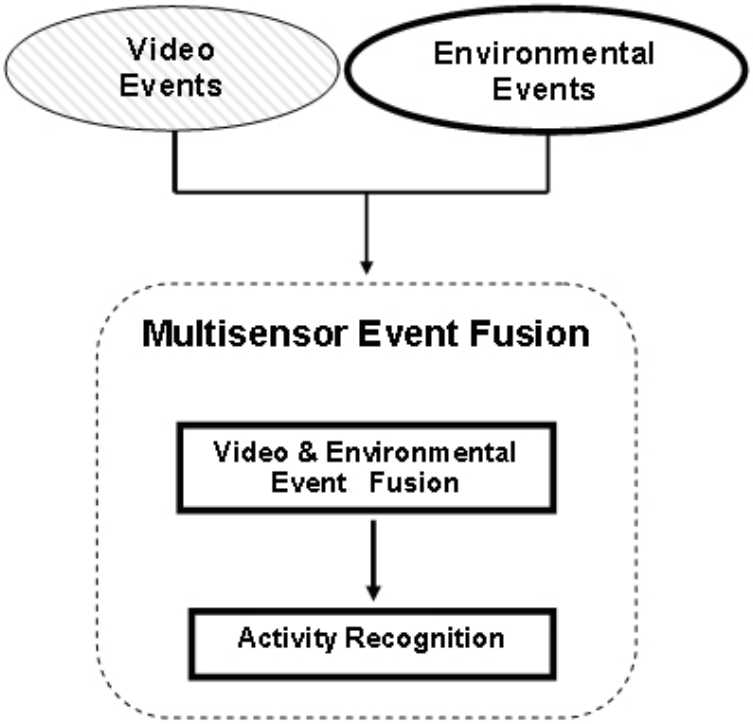


Figure 3.11: The multisensor fusion event architecture. Our major contributions are represented in bold lines with white background. Our minor contributions are represented with dashed background.

to cope with the different output data frequencies of the sensors. To solve this issue, we currently use different configurations of delays between components composing a multimodal event. More precisely, we define different event models corresponding to variations of delays between environmental and video sensor outputs.

3.3 Conclusion

We have presented in this chapter an overview of the proposed approach to recognize human activities at home. Human activity recognition is an important part of cognitive vision systems as seen in chapter 1. Our approach consists in combining data provided by video sensors with data provided by environmental sensors. A video analysis part consists in detecting people, and tracking people as they move and detecting primitive activities related to the person location. The sensor analysis part consists in processing raw sensor data in order to provide high-level representations of sensed objects. The multisensor fusion part combines the video event models with the environmental event models in order to recognize composite activities.

In the next chapters, the proposed approach for activity recognition is described in details.
In chapter 4, the proposed sensor modeling is described in more details. In chapter 5, the multisensor activity recognition approach is described and the proposed activity modeling is presented. The approach is evaluated in chapter 6.

Chapter 4

Sensor Modeling

4.1 Introduction

Sensors are devices which can be used to detect the interaction between a person and his/her environment. They are ultimately the source of all the input data in a multisensor data fusion system [Fowler and Schmalzel, 2004]. The sensor device used to detect this interaction is known as the physical sensor and may be any device which is capable of perceiving a physical property, or environmental attribute, such as light, sound, pressure or motion.

The sensing technologies provide a means to acquire data about the person movement and interactions within the home environment [Loke, 2007]. This data is then processed through an intelligent system which makes recommendations as to how the environment should be adapted to support the needs of the user [Pollack, 2005]. As such, sensors provide the fundamental low level data which forms the basis of how the smart home (see section 2.1.2) is able to provide assistive living conditions and improved levels of independence for the persons. The main concern is therefore that the data obtained from sensors within the home environment may not be totally reliable and may present different degrees of uncertainty in the measurements they report [Ranganathan et al., 2004]. This uncertainty may arise for a number of reasons. For example, it may be the case that the sensor is faulty or malfunctioning, it may be that it can never be 100% accurate due to the nature of what it is measuring.

Sensors must not only measure a physical property, but must also perform additional functions. These functions can be described in terms of compensation, data processing, communication and integration:

- Compensation. This refers to the ability of a sensor to detect and respond to changes in the environment through self-diagnostic tests, self-calibration and adaption.

- Data processing. This refers to processes such as signal conditioning, data reduction, event detection and decision-making, which enhance the information content of the raw sensor measurements.
- Communication. This refers to the use of a standardized interface and a standardized communication protocol for the transmission of information between the sensor and the outside world.
- Integration. This refers to the coupling of the sensing and computation processes on the same silicon chip. Often this is implemented using microelectro-mechanical systems (MEMS) technology.

A practical implementation of such a sensor is known as a smart, or intelligent, sensor [W. Elmenreich, 2003].

In this chapter we describe firstly a smart sensor, after that we describe the proposed sensor model.

4.2 Smart Sensor

A smart sensor (see figure 4.1) is a hardware/software device that comprises in a compact small unit a physical sensor and the associated software for data processing, calibration, and communication. The smart sensor transforms the raw sensor signal to a standardized digital representation, checks and calibrates the sensor, and transmits digital signal to the outside world via a standardized interface using a standardized communication protocol.

The transfer of information between a smart sensor and the outside world is achieved by reading (writing) the information from (to) an interface-file system (IFS) which is encapsulated in the smart sensor.
The IFS provides a structured (name) space which is used for communicating information between a smart sensor and the outside world [Elmenreich et al., 2001].

In this section we firstly describe the physical sensor and their characteristics. After that we describe the logical sensor.

4.2.1 Physical Sensors

The most frequently used types of sensors are physical ones. These hardware sensors (e.g. video camera, audio sensor, light sensors, temperature sensors) can detect almost any raw data, such as motion, audio, light, temperature.

4.2.1.1 Physical Sensor Characteristics

In selecting an appropriate sensor for a single sensor application, we need to consider the individual sensor characteristics. These characteristics are grouped

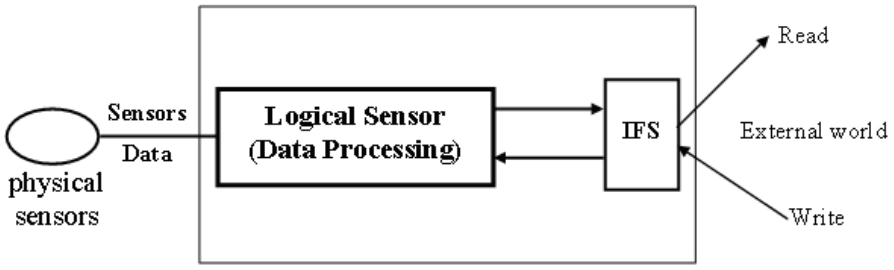


Figure 4.1: A smart sensor with a physical sensor and the encapsulated data processing functions and the encapsulated Interface File System (IFS).

into the following four categories:

- **Type:** The sensors are classified as being fixed on the person (i.e. wearable sensors) or on the environment (i.e. environmental sensors). Wearable sensors are devices used to measure "internal" parameters of a person such as pulse and circadian rhythm. Examples of such sensors include potentiometers, ECG, etc. Environmental sensors are devices which are used to monitor the interaction between a person and his/her environment. Examples of such sensors include video sensors, contact sensors, etc.
- **Function:** The sensors are classified in terms of their functions, i.e. in terms of the parameters, or measurements, which they measure. For example, the measurement include velocity, acceleration, motion, etc.
- **Performance:** The sensors are classified according to their performance measures. These performance include accuracy, sensitivity, resolution, reliability and range.
- **Output:** The sensors are classified according to the nature of their output signal: analog (a continuous output signal), digital (digital representation of measurement) and frequency (use of output signal frequency).

4.2.1.2 Physical Sensor Observation

A physical sensor is characterized by various parameters such as the zone it covers, the precision of its measurement through this zone, its placement and the perturbations to which it is sensitive. The covered zone can be very variable depending on the sensors. For a video camera, this zone is the field of view and for a contact sensor this zone is reduced to a point.

In this work, we consider seven attributes associated with each sensor observation:

- **Sensor ID** Id**:** Single sensor identifier which is transmitting the data;
- **Sensor Class** c**:** This includes the name of the physical property (e.g. temperature, light, pressure) which is measured by the sensor and the units in which it is measured (e.g. Celsius).
- **Sensor Location** x**:** This is the 3D position of the physical sensor in the scene referential.
- **Time** t**:** This is the time when the physical property is measured. In real-time systems the timestamps of a measurement is often as important as the value itself.
- **Sensor Mode** m**:** It represents the different modes allowing the sensors to provide their data (e.g. continuous, by event, on request).
- **Measurement** y**:** This is the value of the physical property as measured by the sensor. The physical property may have more than one dimension and this is the reason we represent it as a vector y.
- **Uncertainty** Δy**:** This is a generic term and includes errors relatively to y, such as measurement errors, and sensor failure errors.

Symbolically we represent a sensor observation using the following 7-tuples:

$$O =< Id, c, x, t, m, y, \Delta y > \tag{4.1}$$

Sometimes not all the attributes are present. In this case we represent the missing attributes by an asterix (*). For example, if the spatial location x is missing from the physical sensor, then we write the corresponding sensor observation as:

$$O =< Id, c, *, t, y, m, \Delta y > \tag{4.2}$$

4.2.2 Logical Sensor

Logical sensor detects raw data through events occurred in the system rather than by physical sensors. For example, a logical sensor can be constructed to detect the current position of a person by analyzing their movement and location.

Multisensor systems require a coherent and efficient treatment of the information provided by the various physical sensors. For this we propose a framework, the Logical Sensor Modeling (LSM), in which the sensors can be defined abstractly.
Modelling the sensor characteristics to an appropriate level of detail has the advantage of giving more accurate and robust mapping between the physical and logical sensor, as well as a better understanding of environmental dependency and its limitations.

4.3 Logical Sensor Modeling (LSM)

As we explained in section 4.2, the smart sensor checks and calibrates the sensor before it is transmitted to the outside world. In order to perform these functions, we require a sufficiently rich sensor model which will provide us with a coherent description of the sensors ability to extract information from its surroundings.
For this, we need to develop a model which can handle different physical sensors but provides a common interface to the multisensor fusion system. We do this by quantifying the uncertainty through probabilistic models of the sensors, taking into account their physical characteristics and interaction with the expected environment.

With binary sensor observations, the probability of making a specific observation is governed by the probabilities of detection and false alarm for the sensor making the observation. When non-binary sensor observations are considered, however, a probability density function (*pdf*) is used to describe the observation results.

In this section, we describe the proposed sensor model and how to model uncertainty in sensor measurements.

4.3.1 Sensor Model with Uncertainty

Sensor data is usually prone to noise and sensing errors [Henricksen and Indulska, 2006]. In many situations sensors can provide uncertain measurements. A malfunctioning sensor gives invalid output data that incorrectly reflect the status of the equipment for example which it is associated with. For instance the contact sensor installed on the door of a fridge may have a technical problem. As such the zero data does not necessarily mean that the person has not opened the fridge as it would whenever it is functioning correctly. The main concern is therefore that the data obtained from sensors within the home environment may not be totally reliable and may present different degrees of uncertainty in the measurements they report [Ranganathan et al., 2004].

Some sensors give information about contexts only at an abstract level. For example, a contact sensor is installed on the door of the fridge. There are many items contained in the fridge such as milk, juice, butter etc. When the fridge sensor is triggered, the state of the fridge context is changed which indicates the person interacts with the fridge (opening the fridge and getting food out of the fridge).
However, it is not possible to infer what food is removed from the fridge by simply considering the current state of the fridge door. Mapping from the sensed fridge to the item removed from the fridge is dynamic and uncertain. For example, if the person wants to make a cold drink, it is more likely that the juice is removed from the fridge. If the person wants to make a hot drink, then it is

more likely that he will remove milk from the fridge.

4.3.2 Binary Sensors

In the case on using binary sensors, the sensor framework presents a certain P_d (probability of detection) and P_f (probability of false alarm). Assume we have M binary sensors which give the state S for N physical objects and have binary states: $s = 1$ or $s = 0$, representing "sensor active" and "sensor not active" respectively. Sensor observations O are likewise binary, either "fridge is open" or "fridge in not open (i.e. closed)". The probability that an observation is made is determined by the P_d or P_f of the sensor making the observation. Letting O_k be k sensor observations, the probabilities are:

$$\begin{aligned} P(O_k = 1|S_s = 1) &= P_d \\ P(O_k = 0|S_s = 1) &= 1 - P_d \end{aligned} \tag{4.3}$$

$$\begin{aligned} P(O_k = 1|S_s = 0) &= P_f \\ P(O_k = 0|S_s = 0) &= 1 - P_f \end{aligned} \tag{4.4}$$

4.3.3 Sensor Model

In this thesis we have defined the uncertainty Δy of measurement y. This uncertainty will also be required when we consider the fusion of multisensor input data. This uncertainty represents the probability that a measurement is erroneous due to the failure of the sensor.

Requisite output data from the intelligent (smart) sensor include estimates of the measurement, plus an estimate of the measurement uncertainty level, for use in processes such as data fusion of multiple sensors of different modalities. Intuitively, if the sensor data has low certainty, then its weighting in the data fusion procedure can be correspondingly reduced. Statistically, this information is completely described by the probability density function (pdf) for the measurement, where the pdf mean value and variance correspond to the measurement estimate and the measurement uncertainty respectively.
In the proposed sensor model, we distinguish between the variable Θ in which we are interested, and a sensor measurement y. We directly observe N raw sensor measurements y_i, i $\in$ {1,2,...,N}, while the variable of interest Θ is not directly observed and must be inferred. In mathematical terms we interpret the task of inferring θ as estimating the a posteriori probability, $P(\Theta = \theta|y)$, where θ represents the true value of the variable of interest Θ and $y = (y_1^T, y_2^T, ..., y_N^T)^T$ denotes the vector of N sensor measurements.

The proposed sensor model is evaluated in chapter 6 using the Gerhome data.

4.4 Conclusion

This chapter introduced a framework for processing sensor measurements.
The use of non-binary observations allows a more robust modeling capability for the sensor manager. Uncertainty modeling is also essential because uncertainty will be present in any real-world problem, and the modeling of that uncertainty is vital to maintaining good sensor manager performance.

Chapter 5

Multisensor Activity Recognition

5.1 Introduction

Sensors are scheduled to detect events which occur anytime and anywhere. The information generated by sensors can be used to identify the activity that the observed person performs.

Considerable research has been devoted towards activity recognition through the deployment of sensing technology to detect interactions with objects, from visual sensors like video cameras [Wu et al., 2002] to sensors which provide binary "on" or "off" outputs such as contact sensors that are used to detect for example a door being opened or closed [Wilson and Atkeson, 2005], [Tran et al., 2004].

In this chapter, we firstly describe the instrumentation of the home care environment. After that we describe the proposed multisensor fusion approach.

5.2 Instrumentation of the Home

In this section, we describe which sensors we use and why. We list the sensors we used in this work, their placement in a home and the selected mode to provide their data.

5.2.1 Sensor Choice and Placement

In chapter 2 (see section 2.1.2), we found that cost of sensors and sensor acceptance are pivotal issues, especially in the home. We found that people are often hesitant, they forget to wear a badge, set of markers, or RFID tag. In particular, elderly people are often very sensitive to small changes in their environment [Burgio et al., 2001].
In this work, we choose to use commonly available sensors that they do not have to wear or carry. These sensors include video sensors and environmental sensors. The selected sensors can easily and quickly be installed in home environments

and are removable without damage to the cabinets or furniture.
The used environmental sensors give at any given time binary value "on" and "off" ("on" if the sensor is activated and "off" if the sensor is not activated). Whenever the value of the sensor associated to a context (e.g. kitchen equipment) changes the status of the associated context (i.e. equipment) changes also.

The list of sensors which we have selected and which we already installed and plugged in the home care environment (i.e. experimental laboratory) includes:

- **Video cameras:** These sensors are used to detect and track people evolving in the scene. They are installed in all rooms but bathroom to locate people at each time.
- **Contact sensors:** These inexpensive magnetic contact sensors indicate a closed or open status. They are embedded on the kitchen furnitures and bedroom closets. These sensors are useful in determining, for example, the interaction with kitchen furnitures, such as cupboards, drawers, and fridge.
- **Pressure sensors:** These sensors are used to detect presence on chairs and beds. They are placed under chairs, armchairs, and bed.
- **Water flow sensors:** When placed in water pipes these sensors trigger a signal when flow exceeds some thresholds. They are placed on hot and cold water pipes and toilets.
- **Electrical sensors:** These sensors measure consumption of the current flow in a circuit, reporting when current exceeds some thresholds, e.g., whenever an appliance is used. They are placed on electrical outlets, to monitor the amount of current flowing to circuits.
- **Presence sensors:** These sensors are installed in front of the sink, the cooking stove and the washbowl to detect the presence of people nearby.

See figure 5.1 for an overview of a typically instrumented home.

5.2.2 Sensor Mode

As previously described in section 4.2.1.2, in the world of the sensors, we find various modes allowing the sensors to provide their data. These modes are listed bellow:

- **Continuous mode** The sensor provides data without any interruption and with a frequency that can be fixed or dynamically modified.

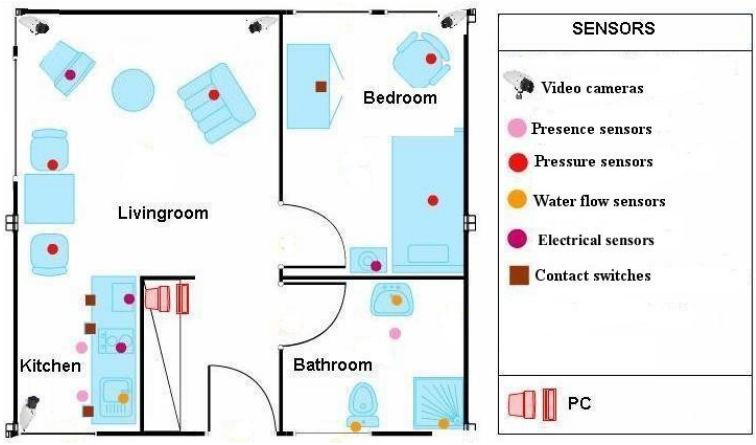


Figure 5.1: Overview of a typically instrumented home

- **By event mode** The sensor provides data when an event occurs. The event provides information with higher semantics than for the sensor with continuous transmission. For example we can think at a presence sensor which provides a binary data corresponding or not to a person presence.
- **On request mode** The sensor provides data in response to a request of an external entity. An external entity asks the sensor and is waiting for the sensor response.
- **By hybrid mode** Combines the three previous modes.

In this work, we use a "continuous" mode for the video sensors and "by event" mode for the environmental sensors. For instance, for the contact sensors installed on kitchen cupboards, a binary data (i.e. On or Off) is received every time a cupboard door is opened (i.e. a contact sensor is On) or closed (i.e. a contact sensor is Off).

5.3 Sensor Fusion

In the next sections, we describe multisensor properties and approaches for sensor fusion.

5.3.1 Multisensor Properties

In selecting a set of sensors for a multisensor application, we need to take into account not only the individual sensor characteristics (discussed in section 4.2.1.1 in chapter 4) but also the multisensor properties [Bellot et al., 2002]. We classify it in the following headings:

- **Distributed:** Sensors which give information on the same environment but from different points of view or from different subsets of the environment.

- **Complementary:** Sensors which together perceive the whole environment but which individually only perceive a subset of the environment.
- **Heterogeneous:** Sensors which give data with completely different characteristics and types.
- **Redundant:** Sensors which perceive the same environment or phenomenon, with little differences between them.
- **Synchronous/Asynchronous:** Sensors which provide data which are temporally concordant or not.

Example: Physiological Measurements [Bellot et al., 2002] We consider two physiological measurements made on a given patient: temperature and blood pressure. The measurements are provided by two sensors: a thermometer and a tensiometer. The two data sources are **distributed**, **complementary** and **heterogeneous** as defined on the physiological space of the patient.

5.3.2 High-Level Sensor-Fusion

Within data processing various different algorithm or special software are applied to obtain derived information from raw sensor data. So objects and their features can be derived from image data by segmentation algorithms, and the behaviour of these objects in the observed scene can be described. Based on this information, decisions can be made. Each processing step is equivalent to an increasing information extraction level. Fusion with other sensors is possible on each level. As described in chapter 2, basically, there are three possible levels on which to perform sensor fusion [Hall and Llinas, 2001]: on raw sensor data, on features extracted from raw data, and on the decision level (see figure 5.2):

- **Raw sensor data**: Fusion on raw sensor data is only possible if the domain of all sensors is the same, i.e. they are of the same type and measure the same quantity. In our approach, this would be not possible because we used different types of sensors.
- **Features**: Feature extraction is a technique that reduces the amount of data produced by a sensor and abstracts away all information that is irrelevant for the task at hand-in case of a positioning system, only information relevant to determining the current location is retained. Multiple sensors (working on different domains) can be combined after relevant features have been extracted from the raw data.
- **Decision**: Currently, sensor fusion is performed mainly on the decision level, i.e. each sensor module provides the system with a set of possible values (e.g. object locations) represented as a probability distribution. These distributions are combined to compute a new probability distribution that represents the most likely location of the object. This approach, sensor

fusion at the decision level, facilitates a modular and extensible system architecture. The number and types of sensors are not limited. Processing the sensor data can be performed remotely (i.e., not on the object itself) and pushed to the object in the form of an internal location event. When a single sensor fails, the quality of the localization is affected, but the system as a whole remains functional.
In this thesis we choose to make fusion at the decision level in order to address the problem of heterogeneous sensor system.

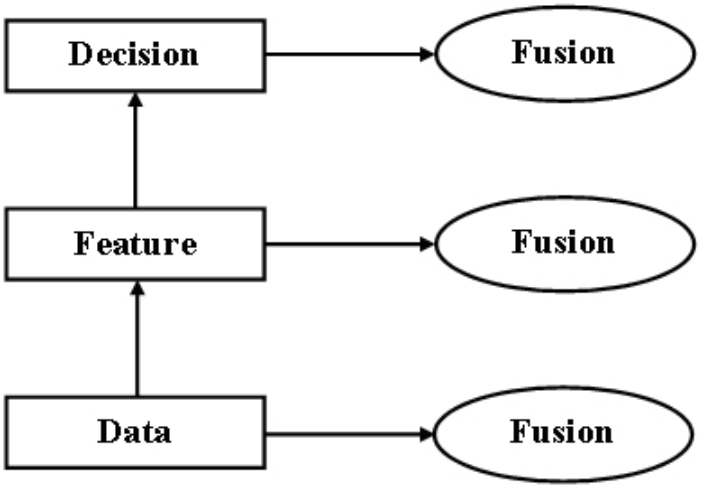


Figure 5.2: Data level, feature level and decision level fusion

5.4 Activity Modeling

The aim of activity recognition is to provide a high level interpretation of the tracked mobile objects in term of human behaviors. It consists in detecting events which have been predefined by application experts or learned through examples.
In order to express the semantics of the activities of interest of elderly at home a modeling effort is needed. The models correspond to the modeling of all the knowledge needed by the system to recognize events occurring in the scene.

To give the meaning of the activities of interest happening in the scene, we have defined a new 3D model of an apartment (without mobile objects) and a 3D model of the mobile objects present in the observed scene (e.g. a 3D model of a person).

- The defined 3D model of an apartment contains both geometric and semantic description of the specific zones, walls and the equipment located

in the observed apartment and contains also geometric information of the installed sensors.
In this 3D model, we have defined:
- A 3D referential which contains the calibration matrices and the position of the video cameras;
- A list of an environmental sensor positions. To define these positions, we have defined for each installed sensor the location of the associated equipment;
- A list of geometric areas corresponding to the different rooms (i.e. entrance, kitchen, livingroom, bedroom and bathroom) in the observed environment (i.e. an apartment);
- A list of geometric zones corresponding to the different zones of interest in the observed environment (i.e. entering zone, exiting zone, cooking zone, eating zone, sleeping zone and bathing zone);
- A list of walls to describe for instance home walls (e.g. kitchen north wall, bedroom west wall);
- A list of the different equipment present in the observed scene with its characteristics (e.g. table, fridge, microwave);

The geometric description of areas contains a polygon defined in a plane. The geometric description of equipment is defined by its size (i.e. height, width, length) and its coordinates in a plane. The semantic description of an area, of a zone, of a wall, and of an equipment contains two attributes: its type (area, zone, wall or equipment) and its name (e.g. kitchen, cooking zone, table).
The proposed 3D model of an apartment can be used in another environment, by redefining the geometric information of the observed environment.

- A 3D model of a mobile object is composed by a name of a model, and by a set of Gaussian functions which describe a 3D width, 3D height, and 3D depth of a mobile object. The availability of a 3D model of mobile objects allows us to have a more precise description of the mobile objects present in the scene (e.g. person. pets).

In the next sections, we firstly describe the proposed event modeling approach which includes the event description language. Secondly, we describe the proposed ontology for daily activities which we want to recognize and a graphical representation of this ontology. Thirdly, we describe the proposed event models for home care applications, which include the proposed video event models, the proposed environmental event models and the proposed multimodal event models. After that, we describe the proposed event recognition approach which includes singlesensor event recognition algorithm and multisensor event recognition algorithm. And finally, we describe the proposed approach to handle uncertainty in sensor measurements.

5.4.1 Event Modeling Approach

We have proposed a new representation formalism to help the experts to describe the events of interest occurring in the observed scene. This formalism contains a language called Video Event Description Language [Vu et al., 2003] which is both declarative and intuitive (in natural terms) so that the experts of the application domain can easily define and modify the event models. This language represents some significant drawbacks for modeling daily activities. His first drawback is that it is dedicated for data provided by only video cameras and does not take into account data provided by other types of sensors. His second drawback is that it does not allow to model complex activities by combining data from several different sensors.
For this, we have proposed 2 extensions of this language. The first extension concerns the adding of data provided by non-video sensors. The second extension allows the combination of several different sensors in order to address complex activity modeling in a scene observed by video cameras and environmental sensors and over an extended period of time.
We call the new proposed language "Event Description Language" instead of "Video Event Description Language".

5.4.1.1 Event Description Language (EDL)

The event description language uses a declarative representation of events that are defined as a set of spatio-temporal and logical constraints.

The following concepts are defined in the event ontology [Vu et al., 2003]. Four different types of events have been designed. The first distinction lies on the temporal aspect of events : we distinguish states and events. A state is a spatio-temporal property characterizing one or several mobile objects at time t or a stable situation over a time interval. An event is one or several state transitions at two successive time points or in a time interval. The second distinction lies on the complexity aspect : a state/event can be primitive or composite. 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 kitchen) or by other sensors (e.g. a fridge is open). A **primitive event** is a primitive state transition and represents the finest granularity of events (e.g. a person is staying close to table). A **composite event** is a combination of primitive states and events (e.g. a person is preparing a lunch). This is the coarsest granularity of events. Composite events are also known in video understanding literature as complex events, behaviors, and scenarios.
As described in section 3.2.4.1, a definition of an event E consists of: (i) an **event name**, (ii) a list of **physical objects** involved in the event such as contextual objects including static objects (i.e. equipment, wall and rooms) and mobile objects (e.g. person, pets), (iii) a list of **components** (variable

values) representing sub-events that describe simple activities concerned, (iv) a list of **forbidden components** which are variables corresponding to all event instances that are not allowed to be recognized during the recognition of the event, (v) a list of **constraints** which are 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 is recognized (e.g. activating an alarm or displaying a warning message). Constraints can be logical, spatial or temporal [Allen, 1983] depending on their meaning, and can have a symbolic or numeric form.

All these concepts describing mobile object interactions in a scene can involve one or several (at least one) mobile objects (e.g. person) and zero or several contextual objects (i.e. area, equipment).
The relations between the components and the physical objects indicate how the components are inferred from the physical objects. There are two types of relations: spatial and spatio-temporal relations. The spatial relations include distance and geometrical relations. Spatio-temporal relations characterize the evolution of spatial relations in time.
There are also two types of relations between the components: logical and temporal relations. Logical relations includes **and**, **or**, and conditional "**if .. then**". The temporal relations include **Allen's Algebra** operators and **quantitative relations** between the durations, **beginning** and **ending** of events. There is also relations between components which consists in a **sequential order** of components.

A spatial symbolic constraint "person is close to table" is a spatial numeric constraint that is defined as follows:

$$distance(person, table) <= 50[cm] \tag{5.1}$$

A temporal constraint may also have a numeric form:

$$duration(event) >= 20[secs] \tag{5.2}$$

5.4.1.2 Event Models

To model an event E, as described in section 5.4.1.1 we distinguish the set of physical objects (e.g. persons, tables) involved in E, a set of components (i.e. sub-events) composing E and a set of constraints on these physical objects and/or these components (see figure 5.3).

In this thesis, we have done a strong effort in event modeling. The result is 100 models which is our knowledge base of events:

- 58 customized video events for daily activities, among them 26 new posture-based events,
- 26 new environmental event models,
- 16 new multimodal event models.

In more details, we have modeled:

- 26 primitive video states which include 10 primitive posture-based states,
- 16 composite video states which include 10 composite posture-based states,
- 16 primitive video events which include 6 primitive posture-based events,
- 10 primitive environmental states,
- 16 primitive environmental events,
- 16 composite multimodal events.

In the next sections, we present firstly the proposed ontology (knowledge base) for daily activities. After that we describe the proposed event models for daily activities which include, a method to define event durations, the proposed video event models with the posture-based event models (see section 5.4.4.1), the proposed environmental event models (see section 5.4.4.2) and the proposed multimodal event models (see section 5.4.4.3).

5.4.2 Ontology for Daily Activities

In this thesis, we have proposed an ontology for daily activities. This ontology contains a set of physical objects (mobile objects and contextual objects) and a set of states and events (body postures and daily activities) which we are interesting to recognize.
Table 5.1 shows the proposed physical objects for home applications, including mobile objects and contextual objects.
Table 5.2 shows the proposed body postures interesting to recognize.
Table 5.3 and table 5.4 show the proposed daily activities interesting to recognize.
Physical objects, body postures and daily activities shown in the previous tables and presented in normal font are already implemented and used, and those presented in italic font are useful but not yet implemented.

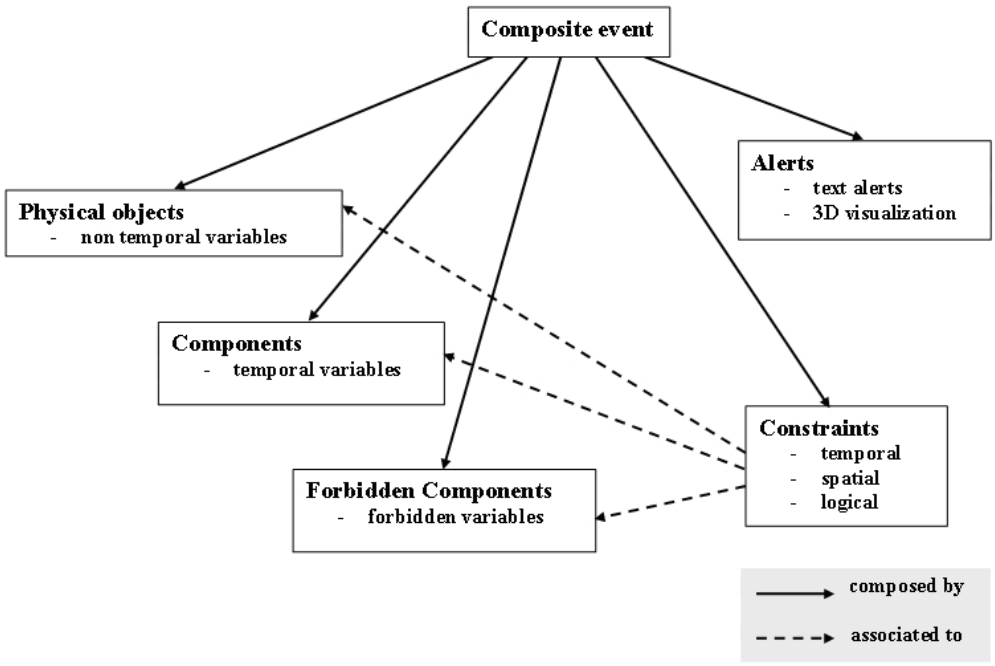


Figure 5.3: Model of composite event

Mobile Object	Person, *Pets*, Chair, and Armchair
Contextual Object	Kitchen, Livingroom, Bedroom, Bathroom, Entrance, Cooking Zone, Entering Zone, Exiting Zone, Eating Zone, Sleeping Zone and Bathing Zone, Fridge, Stove, Microwave, Sink, Countertop, Upper Right Cupboard, Upper Left Cupboard, Lower Right Cupboard, Lower Left Cupboard, Middle Cupboard, Right Drawer, Left Drawer, Chair, Armchair, Table, TV, Closet, Bed, Washbowl, Toilet, and Shower

Table 5.1: Physical objects for monitoring activities at home. Pets object is useful but not yet implemented

BODY POSTURE	**Description**
Standing	A person is standing for at least 2 seconds
Standing with Arms Up	A person is standing with arms up for at least 2 seconds
Standing with Hand Up	A person is standing with hand up for at least 2 seconds
Bending	A person is bending over at the waist for at least 2 seconds
Kneeling	A person is kneeling for at least 3 seconds
Squatting	A person assumes a sitting position in which the balls of his/her feet are in contact with the ground while the heel is lifted and in close proximity to the hindquarters, for at least 3 seconds
Sitting in a Chair	A person is sitting in a chair for at least 3 seconds
Sitting in an Armchair	A person is sitting in an armchair for at least 3 seconds
Sitting with Flexed Legs	A person is sitting (with flexed legs) on a floor or other flat surface, including the couch or bed, for at least 3 seconds
Sitting with Outstretched Legs	A person is sitting (with outstretched legs) on a floor or other flat surface, including the couch or bed, for at least 3 seconds
Slumping	A person is slumping on armchair for at least 3 seconds
Lying with Flexed Legs	A person is lying (with flexed legs) on a floor or other flat surface, including the couch or bed, for at least 3 seconds
Lying with Outstretched Legs	A person is lying (with outstretched legs) on a floor or other flat surface, including the couch or bed, for at least 3 seconds
Walking	A person is walking normally for at least 3 seconds
Standing Up	A person transitions from sitting, slumping, or kneeling to bending and/or standing
Sitting Up	A person transitions from lying to sitting
Sitting Down	A person transitions from standing and/or bending to sitting
Lying Down	A person transitions from standing and/or bending and/or sitting to lying
Turning/pivoting	A person is standing in place but rotates his or her body to face a different direction, turning or pivoting typically involves movement of the feet around a stationary point
Fainting	A person transitions from standing or bending to sitting on floor
Falling Down	A person transitions from standing or bending to sitting on floor and lying on floor

Table 5.2: List of body postures. Body postures presented in normal font are already implemented and used, and body postures presented in italic font are useful but not yet implemented

Kitchen Activities	**Activity**	**Description**
	Using Fridge	Using the fridge equipment for at least 3 seconds
	Using Stove	Using the stove equipment for at least 3 seconds
	Using Microwave	Using the microwave equipment for at least 3 seconds
	Using Middle Cupboard	Using the middle cupboard equipment for at least 3 seconds
	Using Upper Right Cupboard	Using the upper right cupboard equipment for at least 3 seconds
	Using Upper Left Cupboard	Using the upper left cupboard equipment for at least 3 seconds
	Using Lower Right Cupboard	Using the lower right cupboard equipment for at least 3 seconds
	Using Lower Left Cupboard	Using the lower left cupboard equipment for at least 3 seconds
	Using Right Drawer	Using the right drawer equipment for at least 3 seconds
	Using Left Drawer	Using the left drawer equipment for at least 3 seconds
	Preparing Breakfast	Gathering ingredients, utensil and cooking the breakfast
	Preparing Lunch	Gathering ingredients, utensil and cooking the meal
	Preparing Dinner	Gathering ingredients, utensil and cooking the meal
	Warming a Meal	Take a ready-made meal, and warm it
	Preparing Cold Meal	Gathering ingredients, utensil and preparing the cold meal
	Preparing Hot Meal	Gathering ingredients, utensil and cooking the meal
	Washing Ingredients	Using water from the sink to rinse ingredients before preparing them to be cooked
	Washing Dishes	A person is using soap and water to clean dishes in the sink for at least 3 seconds
	Taking Meal	A person is already preparing a meal, set up a table and sitting on a chair for at least 10 minutes.
Hygiene Activities		
	Washing hands or Face	A person is using water (and soap) to rinse hands or face for at least 2 seconds
	Bathing	Washing hair and body with water, soap, shampoo, while located in the shower or bathtub

Table 5.3: List of daily activities. Activities presented in normal font are already implemented and used, and activities presented in italic font are useful but not yet implemented

	Toileting	A person is using the toilet for at least 3 seconds
Leisure Activities		
	Watching TV	A person is sitting or standing in direct view and watching of the TV at least 3 seconds
	Using Telephone	Taking the phone and dialing a number (or answering a ringing phone), and then hanging up the phone when the call is finished
	Listening to music/radio	And audio device is producing output while a resident is at home
	Reading paper/book/magazine	A person is perusing or flipping through the pages of a paper/book/magazine for at least 3 seconds
	Relaxing/thinking	A person is sitting, slumping, or lying down (awake) for at least 3 seconds, while not engaging in any other activity
	Exercising	A person is engaging in particular cardiovascular or physical activity for an extended period of time (at least 10 minutes).
Bedroom Activities		
	Taking medication	Preparing the proper dosage of a medication and then consuming the medication
	Sleeping	Lying on a bed or couch (or possibly sitting in a chair), closing eyes, and remaining in this state for 3 hours or longer
	Napping	Same as sleeping, but the person is asleep under 3 hours
	Waking Up	Activity following sleeping, possibly including lying in bed, and standing up
Other Activities		
	Entering the house	Opening the door from the outside and then entering
	Leaving the house	Opening the door and then exiting to the outside

Table 5.4: List of daily activities (table 5.3 continued). Activities presented in normal font are already implemented and used, and activities presented in italic font are useful but not yet implemented

5.4.3 Ontology Hierarchy of Activities

5.4.3.1 Ontology Concepts

We refer to context as any information that can be used to characterize the activity of the person, including the zone that the person is in, contextual objects that the person interacts with, and the time of the day when an activity is being performed. The state change of the contextual object involved in an activity can be detected through low-level sensor data. When the value of a sensor changes, the state of the associated contextual object of that sensor changes also. This indicates that the person has just interacted with contextual objects related to an activity, which can then be used to infer the activity that the person is doing. The interaction with contextual objects involved in an activity are recorded by associated sensors which send signals to the central system for processing. The relationships between sensors, contextual objects and activities can be represented by a hierarchical network of concepts.

In the first instance it is possible to recognize that a particular activity can be performed or associated with a certain zone (e.g. the kitchen zone) in the home. As our first attempt of introducing the hierarchy we therefore group activities of daily living according to the spatial zone they can be performed in. Each ontology represents hierarchical relationships (e.g. contact sensor activates fridge door and fridge door is open) between sensors, related contextual objects and relevant activities within a zone location.
On a hierarchical ontology, from the point of view of graphical representation, a sensor is represented by a circular node. A rectangular node represents respectively contextual objects and activities (i.e. sub-activity and activity). A sensor node is directly connected to a contextual object node by an arrow. Figure 5.4 summarizes the graphical notations of hierarchical ontology.
Given that some contextual objects are related to several activities, they can also be connected to a set of activities (see figure 5.5).

5.4.3.2 Examples

If we consider the scenario of identifying the type of a meal a person is preparing it is possible to further expand on the concept of the ontology network. If for the sake of simplicity we reduce the possible activities to preparing a hot or cold meal, we begin to consider a simplified kitchen environment and a set of sensors which would be required to gather sufficient information to permit discrimination between these two activities.
Contact sensors are installed on the fridge door and the kitchen cupboards, water flow sensors are installed on the water pipes, and electrical sensors are installed on electrical outlets.

An ontology hierarchy for the activity "Prepare meal" being performed in the kitchen can be constructed as shown in figure 5.6.

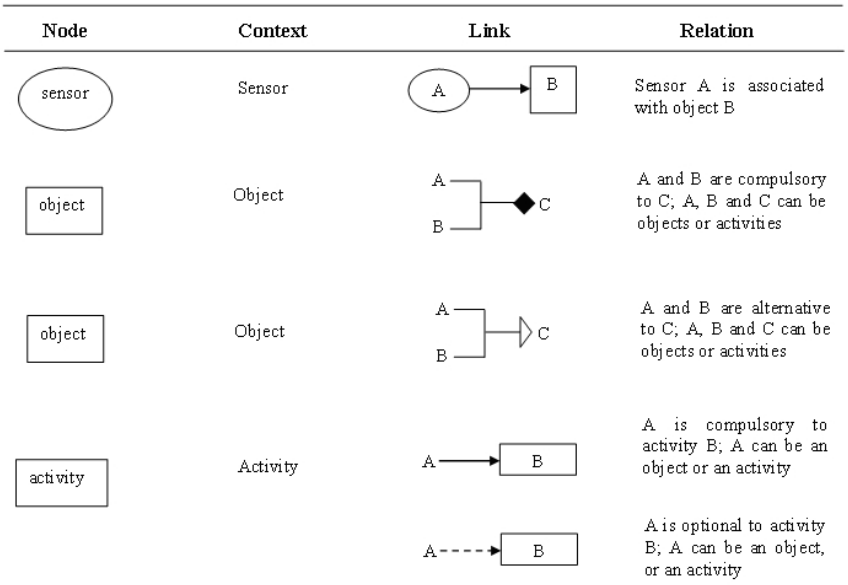


Figure 5.4: Graphical notation of hierarchical ontology

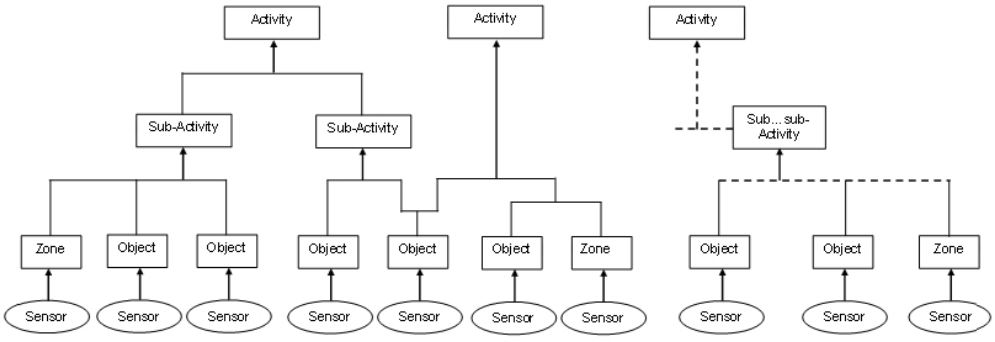


Figure 5.5: A general ontology network of Activities

5.4.4 The Proposed Event Models for Daily Activities

To estimate the performance of all the defined event models, and to define the values of all the following thresholds (i.e. the different thresholds introduced in the definition of event models in sections 5.4.4.1, 5.4.4.2, and 5.4.4.3), we have proposed an estimation of the duration of each event. This estimation is done

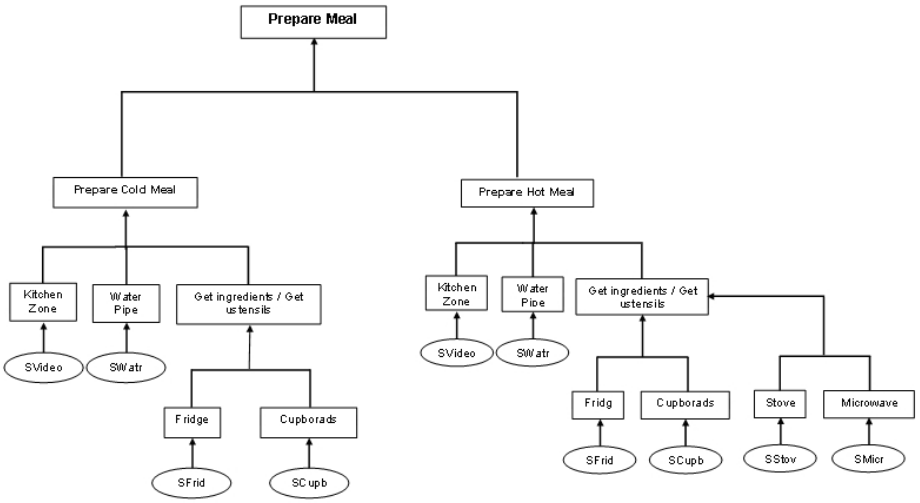


Figure 5.6: An ontology network of preparing a meal activity

by calculating the mean duration value of each event by using the leave-one-out cross-validation (LOOCV). This technique involves a single observation as the validation data, and the remaining observations as the training data. This is repeated such that each observation is used once as the validation data.

To estimate threshold values we have used the ground truth for 5 observed old persons among the experimental data, we calculate the mean duration of each activity for a training set of data (i.e. at each time, we remove 1 person (a testing set) and we used the 4 remaining as a training set). We calculate the mean duration values of each event by using the following equation:

$$\mu_{Ei,Pk} = \frac{\sum^{Pj \in P, Pj \neq Pk} D_{Ei,Pj}}{K-1}, \quad \forall Pk \in P \tag{5.3}$$

Where:

- $\mu_{Ei,Pk}$ represents the mean duration for a given event Ei for each person Pj without the person Pk;
- $P = \{P1, P2, P3, P4, P9\}$;
- $K-1$ represents the number of the training set (i.e. $K-1$=4 in this case);
- $D_{Ei,Pj}$ represents the mean duration (ground truth duration) of each event Ei for each person Pj.

Event (Ei)	**Ground truth mean durations**				
	$D_{Ei,P1}$	$D_{Ei,P2}$	$D_{Ei,P3}$	$D_{Ei,P4}$	$D_{Ei,P9}$
Using Fridge	00:00:15	00:00:16	00:00:10	00:00:29	00:00:18
Using Stove	00:00:18	00:00:11	00:00:25	00:00:11	00:00:15
Sitting on a Chair	00:04:33	00:06:03	00:22:16	00:05:42	00:51:27
Sitting on an Armchair	00:02:02	00:12:04	00:04:04	00:07:21	00:00:36

Table 5.5: Ground truth mean durations of 4 daily activities for 5 observed elderly persons. Time unit is hh:mm:ss

Event (Ei)	**Mean durations $\mu_{Ei,Pk}$ of each event Ei for 4 persons**				
	$\mu_{Ei,P1}$	$\mu_{Ei,P2}$	$\mu_{Ei,P3}$	$\mu_{Ei,P4}$	$\mu_{Ei,P9}$
Using Fridge	0:00:18	00:00:18	00:00:20	00:00:15	00:00:17
Using Stove	0:00:16	00:00:17	00:00:14	00:00:17	00:00:16
Sitting on a Chair	0:21:22	00:21:00	00:16:56	00:21:05	00:09:39
Sitting on an Armchair	0:06:01	00:03:31	00:05:31	00:04:41	00:06:23

Table 5.6: Mean durations $\mu_{Ei,Pk}$ of 4 daily activities Ei for 4 persons. Time unit is hh:mm:ss

Table 5.5 summarizes the ground truth mean durations of 4 daily activities for 5 observed elderly persons.

Table 5.6 summarizes the mean durations using the leave-one-out method of 4 daily activities for 5 observed elderly persons.

We have defined the $threshold_i$ of each event Ei as following:

$$min\left\{\mu_{Ei,Pk}\right\} <= threshold_i <= max\left\{\mu_{Ei,Pk}\right\} \tag{5.4}$$

For example (as shown in table 5.6) the $threshold_1$ of the event $E1$ (Using Fridge) is:

$$00:00:15 <= threshold_1 <= 00:00:20 \tag{5.5}$$

5.4.4.1 Video Event Models

We call video event each state and/or event detected by a video camera.
We have defined the following form for the provided video data:

- **SensorID** Id**:** Single sensor identifier which is transmitting the data;
- **SensorClass** c**:** Represents the class of information provided by the sensor (e.g. video);

- **SensorLocation** x**:** This is the location of the physical sensor in the scene referential;
- **Time** t**:** Represents the moment when the data was provided (YYMMDD-HHMMSS.MS);
- **SensorMode** m**:** It represents the different modes allowing the sensors to provide their data (i.e. "continuous" mode for the video cameras);
- **Measurement** y**:** This is the value of the physical property as measured by the sensor. The physical property may have more than one dimension and this is the reason we represent it as a vector y. For video cameras, the vector y represents the position of the person in the scene referential;
- **SensorUncertainty** Δy**:** This is a generic term and includes errors relatively to the measurement y.

Each provided video data is recorded with its date and time of occurrence. These data are stored in an XML file and transmitted via a parser to the event detection process.
In this thesis, we have modeled 58 video event models which include 26 posture-based event models.
We have modeled 16 primitive video states related to the location of the person in each zone (e.g. inside kitchen, inside livingroom, outside kitchen) and his/her location versus equipments in the observed scene (e.g. close to table, far from armchair). We have also modeled 6 composite video states related to a person staying in each zone (e.g. staying in the kitchen, staying in the bedroom) and 10 primitive video events related to time staying close to each equipment in the scene.

This section shows several examples of video event models using the presented event description language. Figure 5.7 shows the model of a primitive state called "Inside Kitchen" expressing the status of a person being inside a zone which name is kitchen . This video event involves two physical objects (a person p and a zone z), one spatial constraints and one symbolical constraint. The spatial constraint allows to verify whether p is geometrically inside the zone z and the symbolical constraint allows to verify the name of the zone z (i.e. kitchen). The operator "in" is a predefined spatial constraint involving two physical objects p and z to verify whether p is geometrically inside z. The evaluation of the spatial constraint is based on geometrical calculations.

Figure 5.8 shows an example of using spatial and temporal constraints to model an event. The modeled event "Person Enters Bedroom" expresses a primitive event where a person p is first located in livingroom (which is an adjacent zone to the bedroom), after that he/she enters bedroom. This event is composed of three physical objects (person p, zones z1 and z2), 2 components (i.e. first a person is located inside the livingroom, after that he/she is located inside the bedroom)

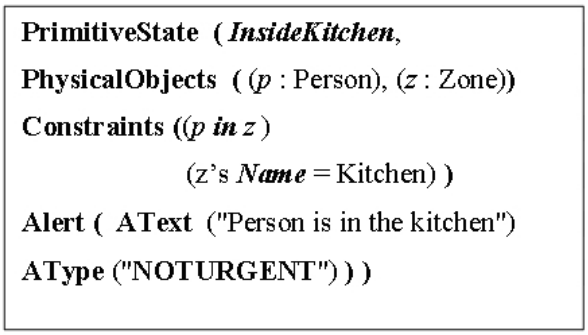


Figure 5.7: A representation of the "Inside Kitchen" primitive state to model the status of a person p being geometrically inside a zone z which name is Kitchen

and four constraints. Two symbolical constraints are related to the names of the two zones, one temporal constraint is related to the time staying inside a bedroom and the last constraint is related to a symbolical temporal constraint to express the sequence of c1 and c2: (c1 **before meet** c2).

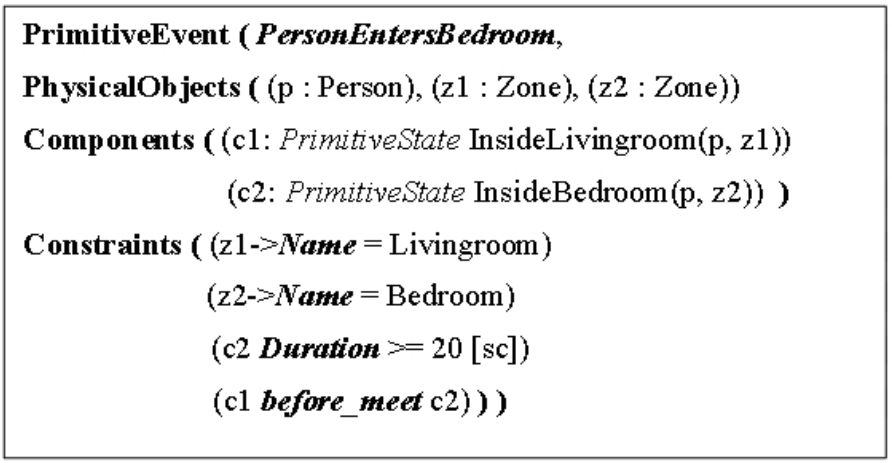


Figure 5.8: A description of a "Person Enters Bedroom" primitive event.

Posture-Based Event Models

Using the proposed 3D human postures already described in section 3.2.2.2 in chapter 3, we have modeled 26 posture-based events which are useful to recognize activities of interest at home.

Each of the proposed 3D human postures plays a significant role in the

recognition of the targeted activities of daily living or of abnormal activities. For example, the posture "standing with hands up" (see figure 5.9) is used to detect when a person is carrying an object such as plates. The posture "standing with arm up" (see figure 5.10) is used to detect when a person reaches and opens kitchen cupboard and his/her ability to do it. These proposed human postures are not an exhaustive list but represent the key human postures taking part in everyday activities.

Figure 5.9: View and 3D visualization of a "hands-up" posture

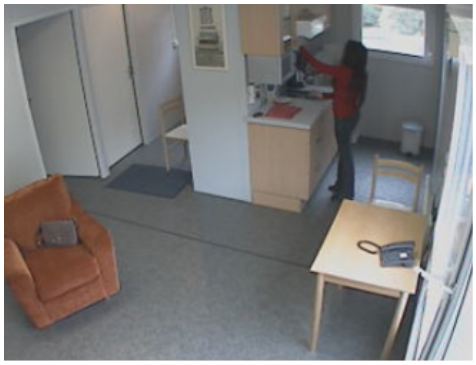


Figure 5.10: View and 3D visualization of an "arm-up" posture

We have defined two types of avatar: an avatar for the man and an avatar for the woman (see figure 5.11). These defined avatars do not take into account the shape of the person (i.e. a slim person, a fat person). But they take into account the height of the person.

For each 3D human posture displayed in figure 3.7 in chapter 3, we have associated a numeric value. For example we have associated a value "104" to the "standing" posture and a value "106" to the "bending" posture. These values are independent of the type of the defined avatar. For example man avatar and

(a) A man avatar (b) A woman avatar

Figure 5.11: Two defined avatars

women avatar have the same value for the "standing" posture.

In collaboration with gerontologists and geriatrics from the Nice hospital in France, for homecare applications, we have modeled 10 primitive posture-based states related to human postures (e.g. person is standing, person is bending), and 10 composite posture-based states related to the human posture with his/her location in the scene (e.g. person is standing in the kitchen).

Figure 5.12 shows the model of the "standing" posture and figure 5.13 shows an example of the "standing" posture.

PrimitiveState (***PersonStanding***,
PhysicalObjects ((p : Person))
Constraints ((p->***Posture3D*** = 104))
Alert (**AText** ("Person is Standing")
AType ("NOTURGENT")))

Figure 5.12: Primitive posture-based state representing the model of "standing" posture

1. **Normal Activities**: We have modeled 4 transitions in human postures related to normal activities using the proposed primitive posture-based states: standing-up, sitting-down, sitting-up and lying-down. "**Standing-up**" (see figure 5.14) represents a transition from slumping and/or sitting to bending

Figure 5.13: View and 3D visualization of "standing" posture

and/or standing, "**sitting-down**" (see figure 5.16) from standing and/or bending to sitting, "**sitting-up**" (see figure 5.17) from lying to sitting, "**lying-down**" (see figure 5.18) from standing and/or, bending, sitting on the floor, to lying on the floor.
Figure 5.15 shows "standing-up from the armchair" activity model which is modeled as a primitive event.

Figure 5.14: Example of "standing-up from the armchair" activity.

2. **Abnormal Activities**: Elderly persons are typically at higher risk of falls and other injuries. Elderly falling down have a high risk of injuring themselves. In some cases the resulting injury may involve broken bones and long recuperation times. Accidents are the fifth leading cause of death in older adults, with falls constituting two thirds of these accidents. As we age, we experience changes in vision, sensory processes, and hearing. Our reaction time slows, and we might lose our balance. An elderly people gait is often stiffer, less coordinated, and muscle strength and tone decline with age. Gait problems are a common cause for falls and a common cause of muscle weakness found in stroke, Parkinson, fractures, and arthritis.

PrimitiveEvent (***StandingUpArmchair***,
PhysicalObjects ((p : Person), (eq: Equipment))
Components ((c1: *PrimitiveEvent* stays_at (p, eq))
(c2: *PrimitiveState* PersonSlumping(p))
(c3: *PrimitiveState* PersonSitting(p))
(c4 : *PrimitiveState* PersonStanding(p)))
Constraints ((eq ->***Name*** = Armchair)
(c2 ***Duration*** >= threshold1)
(c3 ***Duration*** >= threshold2)
(c4 ***After*** c3))
Alert (AText("Person standing up from the armchair")
AType("NOTURGENT")))

Figure 5.15: "Standing-up from the armchair" model

Figure 5.16: Example of "sitting-down" activity.

For all these reasons, in this thesis we have modeled 2 abnormal activities of elderly living alone in his/her own home: fainting and falling-down. These abnormal activities can indicate the presence of health disorders (physical and/or mental) of elderly and can enable their early assistance.
"**Fainting**" which is the transition from standing and/or bending, to sitting with flexed legs and sitting with outstretched legs, and "**falling-down**" which is the transition from standing and/or bending, to sitting with flexed legs and lying with outstretched legs.

These modeled abnormal situations are detailed below:

- **Fainting.** This activity has many forms (see figure 5.19). In this thesis we have defined two types of fainting situation: fainting without

Figure 5.17: Example of "sitting-up" activity.

Figure 5.18: Example of "lying-down" activity.

loss of balance (see figures 5.19(a), 5.19(c)) and fainting with loss of balance (see figure 5.19(b)) which is composed of the transition states from standing to lying on the floor with outstretched legs. Fainting without loss of consciousness is composed of the transition states from standing, bending to sitting (with flexed and outstretched legs), which is modeled as described in figure 5.20.

(a) Fainting on the floor (b) Fainting with losing balance (c) Fainting on a chair

Figure 5.19: Different forms of "fainting"

PrimitiveEvent (*PersonFainting*,
PhysicalObjects ((*p* : Person))
Components ((c1: *PrimitiveState* Standing(p))
(c2: *PrimitiveState* Bending(p))
(c3: *PrimitiveState* SittingFlexedLegs(p))
(c4: *PrimitiveState* SittingOutstretchedLegs(p)) **)**
Constraints ((c1; c2; c3; c4)
(c4 ***Duration*** >= threshold) **)**
Alert (**AText** ("Person is Fainting ")
AType ("URGENT") **))**

Figure 5.20: Example of "Fainting" model

This "fainting" model contains 1 physical object (the person), 4 components (human postures), 2 constraints and 1 alert. The first constraint consists in **sequential order** between the components and the second constraint represents the **duration** of the sitting posture. When these components occurred and all the constraints are verified, the fainting event is recognized, and an alert is triggered.

- **Falling-down.** This activity has many forms. It is modeled by a transition between states: standing, bending, sitting on the floor (with flexed or outstretched legs) and lying (with flexed or outstretched legs). There are different visual definition for describing a person falling down. It depends on the wellness and the health of the person. Thus, we have modeled the event "falling-down" with three models:
  - **Falling-down 1**: Represents a change state from standing, sitting on the floor with flexed legs and lying with outstretched legs. Figure 5.21 shows a model of falling-down 1 situation.
    In this example, the "falling-down 1" model contains 1 physical object (the person), 3 components (human postures), 2 temporal constraints and 1 alert. The first constraint represents a symbolical temporal constraint to express the sequence of 2 components (pSit **before meet** pLay) and the second constraint represents the **duration** of the laying posture. When these components occurred and all the constraints are verified, the falling-down event is recognized, and an alert is triggered.
  - **Falling-down 2**: Represents a change state from standing, bending and lying with outstretched legs (see figure 5.22 (a)).
  - **Falling-down 3**: Represents a change state from standing, sit-

PrimitiveEvent (*PersonFallingDown1*,
PhysicalObjects ((p : Person) **)**
Components ((pStand: *PrimitiveState* Standing(p))
(pSit: *PrimitiveState* SittingFlexedLegs(p))
(pLay: *PrimitiveState* LyingOutstretchedLegs(p)) **)**
Constraints ((pSit ***before-meet*** pLay)
(pLay's ***Duration*** >= threshold) **)**
Alert (**AText** ("Person is Falling Down")
AType ("VERYURGENT ")) **)**

Figure 5.21: "Falling-down1" model

ting on the floor with flexed legs and lying with flexed legs (see figure 5.22 (b)).

Figure 5.22: Illustration of elderly falls; (a) Example of "falling-down2" of elderly women; (b) Example of "falling-down3" of elderly women

5.4.4.2 Environmental Events

We call environmental event each state and/or event detected by environmental sensors (i.e. contact sensors, pressure sensors, electrical sensors, presence sensors, and water sensors) except video cameras.
The environmental sensors provide data when its status changes. For instance the contact sensor determines an opening and closing states for various devices (i.e. kitchen cupboards, kitchen drawers, kitchen fridge, bedroom closets).

For the provided environmental data, we have used the same form as used for video cameras:

- **SensorID** Id**:** Single sensor identifier which is transmitting the data;
- **SensorClass** c**:** Represents the class of information provided by the sensor (e.g. pressure, electrical, contact);
- **SensorLocation** x**:** This is the location of the physical sensor in the scene referential;
- **Time** t**:** Represents the moment when the data was provided (YYMMDD-HHMMSS.MS);
- **SensorMode** m**:** It represents the different modes allowing the sensors to provide their data (i.e. "by event" for the environmental sensors);
- **Measurement** y**:** This is the value of the physical property as measured by the sensor. The physical property may have more than one dimension and this is the reason we represent it as a vector y. For environmental sensors, the value of y is 0 if the status of the sensor is OFF and is 1 if the status of the sensor is ON;
- **SensorUncertainty** Δy**:** This is a generic term and includes errors relatively to the measurement y.

Each provided environmental data is recorded with its date and time of occurrence. These data are stored in an XML file and transmitted via a parser to the event detection process. From these data, we infer the corresponding environmental event. For example, if the provided data is "On" and the sensor class is "contact" then we infer the contact event "Open". If the provided data is "Off" and the sensor class is "contact" then we infer the contact event "Closed".

We have modeled 10 primitive environmental states related to the data provided by the environmental sensors (i.e. a contact sensor provides "open/closed" events, a pressure sensor provides "pressed/not-pressed" events, an electrical sensor provides "used/not-used" events, a presence sensor provides "present/not-present" events, and a water sensor provides "water-consumed/water-not-consumed" events).
Using these primitive states we have modeled 16 primitive environmental events to describe the status of each equipment in the scene (e.g. microwave is used, microwave is not-used, fridge is open, fridge is closed).
Figure 5.23 shows the two primitive environmental states related to the events provided by a contact sensor.

Figure 5.24 shows the two primitive environmental states related to the events provided by a pressure sensor.

Figure 5.25 shows the two primitive environmental states related to the events

PrimitiveState (*IsOpen*,
PhysicalObjects ((eq: Equipment) **)**
Constraints ((eq's **State** = **OPEN**) **))**

PrimitiveState (*IsClosed*,
PhysicalObjects ((eq: Equipment) **)**
Constraints ((eq's **State** = **CLOSED**) **))**

Figure 5.23: Primitive environmental states of events provided by contact sensor.

PrimitiveState (*IsPressed*,
PhysicalObjects ((eq: Equipment) **)**
Constraints ((eq's **State** = **PRESSED**) **))**

PrimitiveState (*IsNotPressed*,
PhysicalObjects ((eq: Equipment) **)**
Constraints ((eq's **State** = **NOT-PRESSED**) **))**

Figure 5.24: Primitive environmental states of events provided by pressure sensor.

PrimitiveState (*IsUsed*,
PhysicalObjects ((eq: Equipment) **)**
Constraints ((eq's **State** = **USED**) **))**

PrimitiveState (*IsNotUsed*,
PhysicalObjects ((eq: Equipment) **)**
Constraints ((eq's **State** = **NOT-USED**) **))**

Figure 5.25: Primitive environmental states of events provided by an electrical sensor.

provided by an electrical sensor.

Figure 5.26 shows the primitive environmental event model related to the "Open" status of a fridge.

Figure 5.27 shows the primitive environmental event model related to the "Used" status of a microwave.

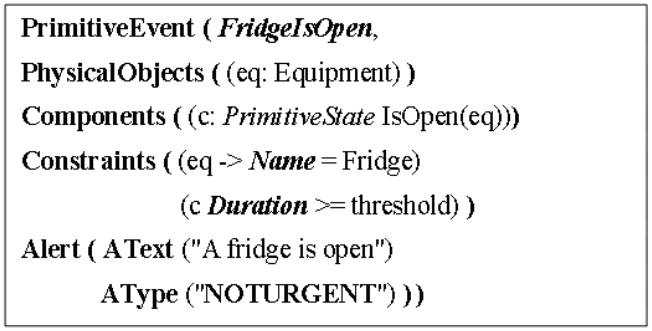


Figure 5.26: Primitive environmental event related to the status of a fridge.

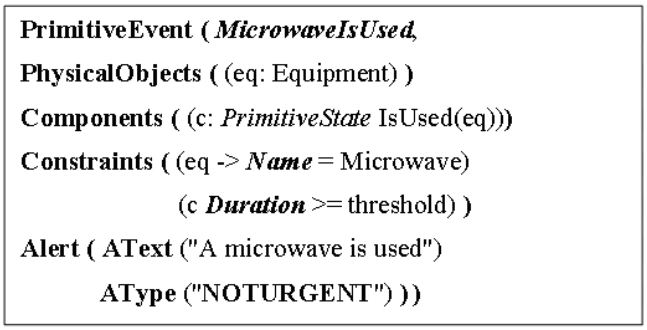


Figure 5.27: Primitive environmental event related to the status of a microwave.

5.4.4.3 Multimodal Event Models

We call multimodal event each event detected by both video cameras and environmental sensors.
In this thesis we have modeled 16 composite multimodal events: using (i) fridge, (ii) cupboards, (iii) drawers, (iv) microwave (see figure 5.28), (v) stove, (vi) telephone, (vi) watching TV, (viii) dish washing, (ix) slumping in armchair, (x) taking a meal, and (xi) 6 variations of preparing a meal: breakfast, lunch, dinner, warming a meal, cold meal and hot 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 be located close to the countertop in the kitchen and staying at this location for a while, the person opens cupboards to take ingredients and dishes (e.g. plates, fork, knife), opens the fridge to take foods, uses the stove to cook the meal, and set up the table. Figure 5.29 shows a composite multimodal event "Slumping in an Armchair" combining both pressure sensor installed under armchair to detect when a person

Figure 5.28: Illustration of "using microwave" activity

is sitting and video camera to detect a person approaching the armchair and to detect the slumping posture (i.e. by using posture model). This event model is currently used to detect when a person is slumping in an armchair which can indicate an abnormal activity if a stove is still running.

On figure 5.30 is depicted an example of a composite multimodal event, "Using Microwave": a person p is using a microwave equipment. This scenario is based on video and environmental events and will be recognized if a sequence of five sub-events are recognized and four constraints described on figure 5.31 has been verified.

- **Preparing a Meal**

  We have defined the "preparing meal" activity as follows:
  IF *the person is located close to the countertop in the kitchen* ***AND*** *(a person accesses to meal ingredients* ***AND*** *a person accesses plates or utensil cupboards)* ***AND*** *a person uses an appliance (e.g. microwave, stove)* ***THEN*** *a meal is prepared.*

  The location of a person close to the countertop in the kitchen lasting for a minimum period of time is detected by video camera. The use of meal ingredients is detected by the use of a food storage cupboard (contact sensor) and/or by the use of the fridge (contact sensor). The use of plates and/or utensil is detected by the use of dishes cupboard and/or drawer (contact sensor) and the use of appliance is detected by the use of stove or

(a) An illustration of slumping activity

CompositeEvent (***PersonSlumpingInArmchair*** **,**
PhysicalObjects ((p : Person), (z : Zone), (eq: Equipment))
Components ((c1: *PrimitiveState* CloseToArmchair(p, eq)) ◄····· *detected by video camera*
(c2: *PrimitiveEvent* ArmchairIsPressed(p, eq)) ◄····· *detected by pressure sensor*
(c3: *PrimitiveState* PersonSlumping(p)) ◄····· *detected by video camera (i.e. posture model)*
Constraints ((eq's ***Name*** = Armchair)
(c2 ***Duration*** >= threshold1)
(c3 ***After_meet*** c2)
(c3 ***Duration*** >= threshold2) **)**
Alert (AText ("Person is slumping in an armchair")
AType ("NOTURGENT") **))**

(b) Slumping model

Figure 5.29: An illustration of "slumping in an armchair" activity and a corresponding model

microwave (electrical sensor and presence sensor).

In this thesis, we have modeled 6 variations of preparing a meal: breakfast, lunch, dinner, warming a meal, cold meal and hot meal.
Figure 5.32 shows a model of preparing lunch. This model involves 5 physical objects (the person, and 4 equipments), 4 components: stays at countertop (video camera), and three composite multimodal events related to the using of the kitchen equipment (Cupboards, Fridge and Stove) and 2 temporal constraints.

- **Taking a Meal** An example of the modeling event "taking a meal" is presented in figure 5.33.
  This "taking a meal" model contains 5 physical objects (a person p, 2 zones: a kitchen and a livingroom, 2 equipments: a table and a chair), 4

CompositeEvent (*UsingMicrowave*,
PhysicalObjects ((*p* : Person), (*Microwave* : Equipment) **)**
Components ((p_stays: *PrimitiveEvent* **stays_at**(*p, Microwave*))
(m_used: *PrimitiveEvent* **MicrowaveIsUsed**(Microwave))
(m_not-used: *PrimitiveEvent* **MicrowaveIsNotUsed**(Microwave))
(p_far: *PrimitiveEvent* **far_from**(*p, Microwave*)) **)**
Constraints ((m_used ***During*** p_stays)
(m_used ***Duration*** >= threshold)
(m_not-used ***After*** m_used)
(p_far ***After*** m_not-used) **)**
Alert (**AText** ("Person is using microwave")
AType ("NOTURGENT") **))**

Figure 5.30: Using microwave model

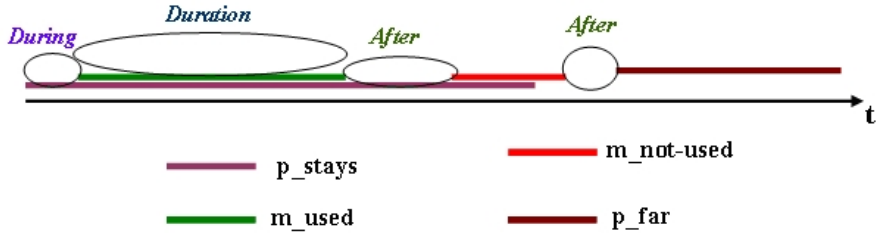


Figure 5.31: Temporal constraints between states and events constituting a composite multimodal event "using microwave"

components, 4 spatio-temporal constraints and 1 alert. The components are: a presence of preparing lunch in a kitchen, detection of a person enters in a livingroom, detection of a person close to table, and the sitting posture of the person in a chair. The spatio-temporal constraints are related to the **duration** and to **sequential order** between the components. When these components occurred and all the constraints are verified, the taking meal event is recognized and an alert is triggered.

We have defined another model "taking a meal 2" of activity with a logical location where the person is supposed to have his/her meal. This logical location can correspond to several real physical locations such as kitchen table, livingroom table, livingroom armchair, etc (see figure 5.34).

CompositeEvent (*PreparingLunch*,

PhysicalObjects ((p : Person), (Countertop : Equipment), (UpperCupboard: Equipment), (Fridge : Equipment), (Stove : Equipment))

Components ((c1: *PrimitiveEvent* Stays_at(p, Countertop))
(c2: *CompositeEvent* UsingCupboard(p, UpperCupboard))
(c3: *CompositeEvent* UsingFridge(p, Fridge))
(c4: *CompositeEvent* UsingStove(p, Stove)))

Constraints ((c1 ***Duration*** >= threshold1)
(c4 ***Duration*** >=threshold2))

Alert (**AText**("Person prepares lunch")
AType("NOTURGENT")))

Figure 5.32: Example of "preparing lunch" model

CompositeEvent (*TakingMeal*,

PhysicalObjects ((p : Person), (Kitchen : Zone), (Livingroom : Zone), (Table : Equipment), (Chair : Equipment))

Components ((c1 : *CompositeEvent* PreparingLunch(p, Kitchen))
(c2: *PrimitiveEvent* PersonEntersLivingroom(p, Livingroom))
(c3 : *PrimitiveState* CloseToTable(p, Table))
(c4 : *CompositeState* PersonSeatedOnChair(p, Chair)))

Constraints ((***Start*** of c4 ***after End*** of c1),
(c3 ***Duration*** >=threshold1),
(c4 ***after_meet*** c3)
(c4 ***Duration*** >= threshold2))

Alert (**AText** ("Person eats a meal in a livingroom")
AType ("NOTURGENT")))

Figure 5.33: Taking a meal model

5.5 Activity Recognition

The automatic recognition of activities is a real challenge for cognitive vision research because it addresses the recognition of complex activities. The challenge is to perform a real-time event recognition algorithm able to efficiently recognize all the events occurring in the scene at each instant.

CompositeEvent (***TakingMeal2***,
PhysicalObjects ((p : Person), (Kitchen: Zone), (z : Zone), (eq : Equipment))
Components ((c1 : *CompositeEvent* PreparingMeal(p, Kitchen))
(c2: *PrimitiveState* PersonInsideZone(p, z))
(c3 : *PrimitiveState* CloseTo (p, eq))
(c4 : *CompositeState* PersonSeated(p, eq)))
Constraints ((***Start*** of c4 ***after End*** of c1),
(c2 ***Duration*** >=threshold1),
(c4 ***Duration*** >= threshold2))
Alert (**AText** ("A person is inside a zone z and he/she is seated in an equipment after preparing a meal ")
AType ("NOTURGENT")))

Figure 5.34: Taking a meal model with a logical location

The event recognition process 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.

We have extended the existing event recognition algorithm [Vu et al., 2003] to address complex activity recognition involving several physical objects of different types (e.g. persons, chairs) in a scene observed by video cameras and environmental sensors over an extended period of time. We propose a method to recognize video and environmental events based on spatio-temporal reasoning taking full advantage of a priori knowledge about the observed environment and of video and environmental event models.

In the next sections we describe firstly the proposed algorithm for event recognition using only video sensors or environmental sensors, after that we describe the proposed algorithm for multisensor event recognition using the both sensors.

5.5.1 Event Recognition Process

The proposed event recognition algorithm is able to recognize which video or environmental events are occurring in a scene at each instant. To benefit from all the knowledge, the event recognition algorithm uses the coherent tracked mobile

objects, the a priori knowledge of the scene and the predefined video or environmental event models. To be efficient, the recognition algorithm processes in specific ways video or environmental events depending on their type. Moreover, this algorithm has also a specific process to search previously recognized video or environmental events to optimize the whole recognition.

5.5.1.1 Video Event Recognition Process

Video events are first represented by experts using the event description language (see section 5.4.1.1) by defining video event models. Then, video event models are automatically (off-line) parsed and analyzed to be used later during the recognition process. Finally, analyzed video event models are automatically (on-line) used with incoming low level video events to determinate which events are occurring in the observed scene (see figure 5.35).
In this section we describe firstly the video event recognition process. After that we describe the dedicated algorithm.

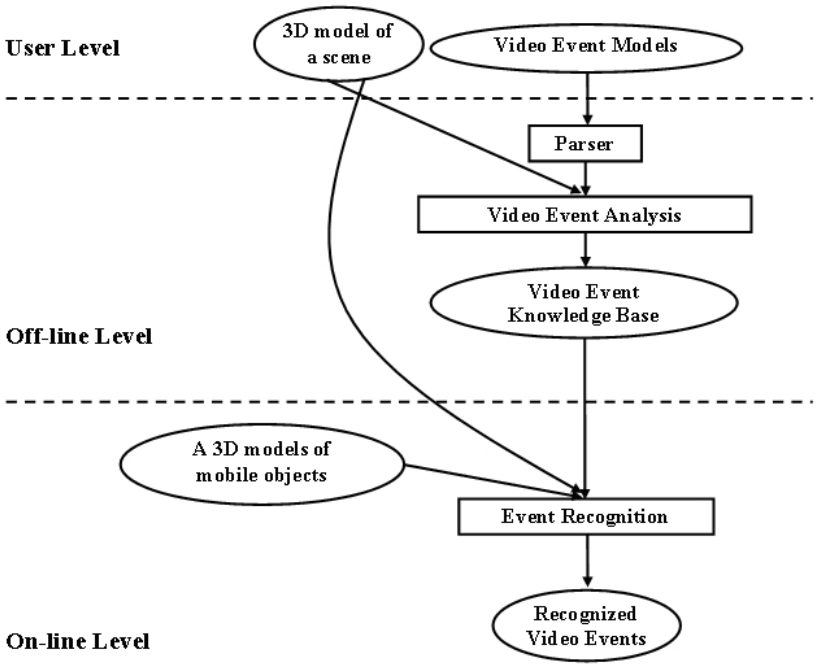


Figure 5.35: Processing of video event models

Recognition of Primitive Video States

To recognize a primitive video state, the recognition algorithm performs a loop of two operations:

1. The selection of a set of physical objects then

2. The verification of the corresponding atemporal constraints until all combinations of physical objects have been tested.

Once a set of physical objects satisfies all the atemporal constraints, the primitive state is recognized. To enhance primitive event recognition, after a primitive state has been recognized, event triggers are generated for each primitive event the last component of which corresponds to the recognized primitive state. The event trigger contains the list of the physical objects involved in the primitive state.

Recognition of Primitive Video Events

To recognize a primitive video event, given the event trigger partially instantiated, the recognition algorithm consists in looking backward in the past for a previously recognized primitive state matching the first component of the event model. If these two recognized components verify the event model constraints, the primitive event is recognized.
To enhance the composite video event recognition, after a primitive event has been recognized, event triggers are generated for all composite events the last component of which corresponds to the recognized primitive event.

Recognition of Composite Video States and Events

The recognition of composite video states and events usually implies a large space search composed of all the possible combinations of components and physical objects. All the composite states and events are decomposed into states and events composed at the most of two components. Then the recognition of composite states and events is performed similarly to the recognition of primitive events.

To recognize the predefined event models at each instant, we first select a set of event triggers that indicate which events can be recognized. These triggers correspond to a video event (primitive state or event) or to a composite video event that terminates with a component recognized at the previous or current instant.
For each of these event triggers, solutions are found by looking for component instances already recognized in the past to complete the recognition of event. A solution of an event model M is a set of physical objects that are involved in the recognized event and the list of corresponding component instances satisfying all the constraints of M.

Algorithm for Video Event Recognition

We define a trigger as a video event which can be potentially recognized. There are three types of triggers : the primitive video event models (type 1), the composite video states (type 2) and the composite video events already recognized at the previous instant (type 3) (see the algorithm 1).
We have initialized a list LT of triggers with all triggers of type 1 (i.e. primitive

Algorithm 1 *VideoEventRecognitionAlgorithm*

For each primitive video state model
Create a trigger T of *type 1* for the primitive video event model
For each solution ρ_e of T
If ρ_e is not extensible **Then**
Add ρ_e to the list of recognized video events
Add all triggers of *type 2* of ρ_e to the list LT (List of Triggers)
If ρ_e is extensible with $\rho_e^{'}$ recognized at previous instant **Then**
Merge ρ_e with event $\rho_e^{'}$
Add all triggers of *type 2 and 3* of $\rho_e^{'}$ to the list LT
While $LT \neq \oslash$
Order LT by the inclusive relation of video event models
For each trigger $T_0 \in LT$
For each solution ρ_0 of T_0
Add ρ_0 to the list of recognized video events
Add all triggers of *type 2 and 3* of ρ_0 to the list LT

video event models). Once we have recognized a primitive video events ρ_e, we try to extend ρ_e with a recognized video events $\rho_e^{'}$ at the previous instant. If ρ_e cannot be extended, we add the triggers of type 2 that terminate with ρ_e to the list LT. If ρ_e is extended with $\rho_e^{'}$, we add the triggers of type 2 and 3 that terminates with $\rho_e^{'}$. The triggers of type 2 are the instances of composite video states of $\rho_e^{'}$ and the triggers of type 3 are the composite video events $\rho_e^{'}$ already recognized at the previous instant and that terminates with $\rho_e^{'}$. After this step, there is a loop process first to order the list LT by the inclusive relation of event models contained in the triggers and second to solve the triggers of LT. If a trigger contains an instance of video event $\rho_0^{'}$ that can be solved (i.e. totally instantiated), we add the triggers of type 2 and 3 that terminate with $\rho_0^{'}$.

5.5.1.2 Environmental Event Recognition Process

Environmental events are firstly parsed (using an XML parser) to the event description language (see section 5.4.4.2) by defining environmental event models. Then, environmental event models are automatically (off-line) analyzed to be used later during the recognition process. Finally, these models are used (on-line) with incoming low level environmental events to determine which events are occurring

in the observed scene (see figure 5.36).
In this section we describe firstly the environmental event recognition process. After that we describe the dedicated algorithm.

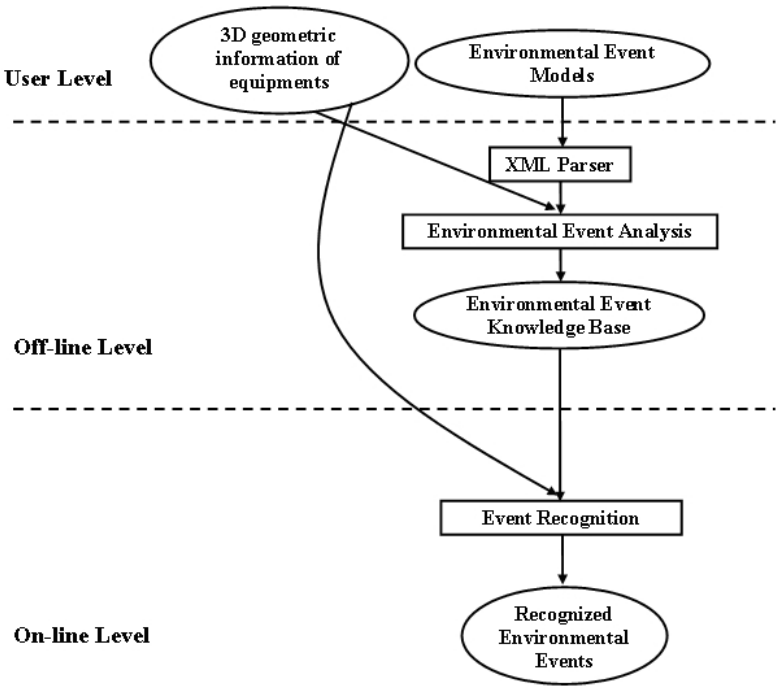


Figure 5.36: Processing of environmental event models

Recognition of Primitive Environmental States

After receiving data from the environmental sensors, we infer the occurred environmental events.
To define the primitive environmental states for each sensor, the recognition algorithm associates for each provided data a corresponding state. When a sensor is activated, a provided data is "On", then the recognition algorithm searches the associated sensor class to define which states are the current (e.g. open/closed states, used/not-used states).
Symbolical constraint is used to define each primitive environmental state. This constraint is related to the status of the equipment associated to the environmental sensor. Once this constraint is satisfied, the primitive environmental state is recognized.

Recognition of Primitive Environmental Events

To recognize a primitive environmental event, the recognition algorithm uses the previously recognized primitive environmental states matching with the associated equipments.

Algorithm for Environmental Event Recognition

We have defined two types of environmental events: primitive environmental states (type 1) and primitive environmental events (type 2) (see algorithm 2).

Algorithm 2 *EnvironmentalEventRecognitionAlgorithm*

Initialize all the sensors status with the value "Off"
If the status of the sensor change to value "On" **Then**
Create a primitive state model which represents the new status of the sensor
For each primitive environmental state model
Create a variable V of *type 1* for the primitive environmental event model
For each solution ρ_s of V
Add ρ_s to the list of recognized environmental events
Add all the variable of *type 1 and 2* of ρ_s to the list of recognized environmental events

---

5.5.2 Multisensor Event Recognition Process

This section presents the recognition of multisensor (i.e. multimodal) events (i.e. video-environmental events). The multisensor event recognition process is able to recognize which events are occurring in the scene at each instant. The recognition process takes as input video and environmental events and the a priori knowledge of multimodal events to be recognized. These events are first processed to synchronize them. Then, the event recognition process takes as input the synchronized events and tries to understand which events (i.e. video-environmental events or activities) are occurring.
Figure 5.37 shows the process of recognizing multimodal events at each instant. Our goal is to obtain an algorithm that is able to recognize in real-time (video rate) multimodal events (or complex events) that totally occurred through videos. Thus, there are two main issues to be focused on: (1) the fusion of environmental and video events 5.5.2.1 to obtain better synchronized events as input, and (2) the recognition of multimodal events 5.5.2.2.

5.5.2.1 Multisensor Event Fusion

The objective of event fusion is to synchronize dated events (i.e. time stamped events) received from different sources (e.g. video cameras and environmental

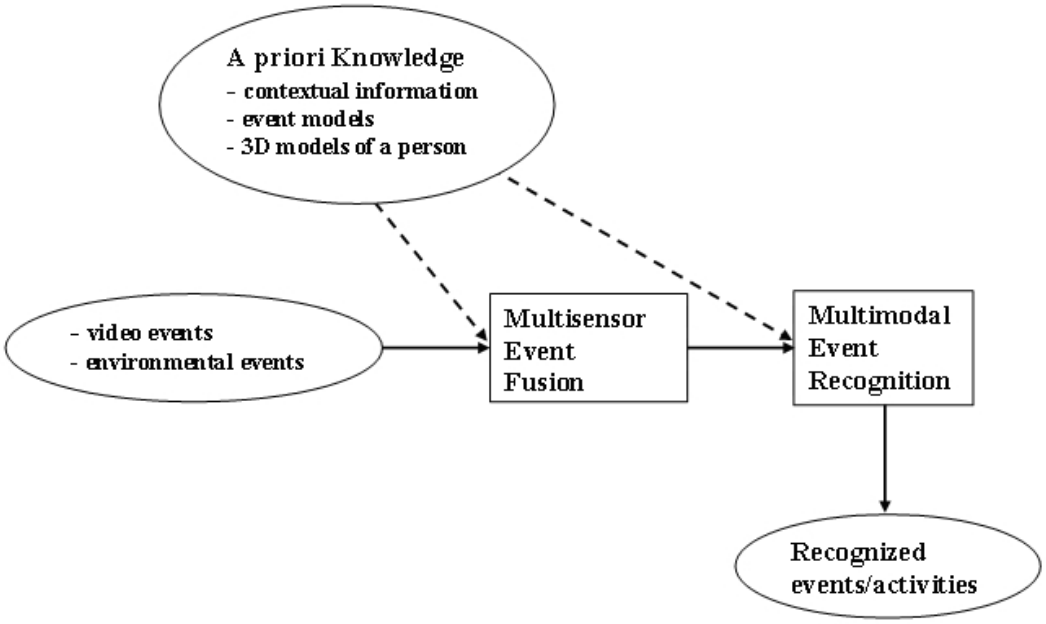


Figure 5.37: The multisensor (video-environmental) event recognition process at each instant

sensors). Synchronization of events received from different sources is an important step to have all received events in the temporal order they occurred. The need of event synchronization is due to delays because the information acquisition frequencies of different sources are different. For instance, video acquisition frequency is 25 frames/second at the maximum.
To cope with the synchronization issue, we use different configurations of delays between components composing a multimodal video-environmental event for the recognition algorithm to process temporal constraints with time tolerances. More precisely, we define different event models corresponding to variations of delays between non-video and video processing for modeling one multimodal event.
This method to cope with synchronization issue is not fully satisfactory, since the time delays between the occurrences of video-environmental events should not impose an order for the recognition of more complex events. However, experiments for healthcare applications (see section 6.4.3) show that this method can be used for the efficient recognition of multimodal events in limited conditions (e.g. environmental events are only considered as additional information to confirm the recognition of complex events based on visual information).

5.5.2.2 Multimodal Event Recognition

After an initial work on simple activity recognition (primitive states and events) to show the large diversity of events which can be addressed, the next challenge is to handle automatic recognition of complex activities (i.e. multimodal events) by

combining video events with environmental events already recognized. Complex activities refers to the activities detected by both video and environmental sensors during an extended time period.

The next section presents an algorithm for real-time recognition of complex events that totally occurred in the observed scene depicted by video-environmental sequences. We first specify the two hypotheses for the recognition.

- **Hypothesis 1 (good video processing):** all mobile objects are well detected by vision algorithms.
- **Hypothesis 2 (good sensor processing):** all non-video data are well detected by sensor processing algorithms.

These hypotheses sound strong and not realistic. Experimentally (see section 6.4.3), environmental events and video events can be missed (rarely wrong) due to segmentation errors and to sensor failures. Despite these errors, we have managed to get successful overall results as shown in the section 6.4.3.

5.5.2.3 Algorithm for Multimodal Event Recognition

We define variable V_1 as a video event and a trigger V_2 as an environmental event which can be potentially recognized. There are 3 types of variables : the primitive video event (type 1), the primitive environmental event (type 2), and the composite events already recognized at the previous instant (type 3) (see the algorithm 3).

Algorithm 3 *MultimodalEventRecognitionAlgorithm*

For each primitive video model and each primitive environmental model
Create variables V_1 of *type 1* and V_2 of *type 2* for the multimodal event model
For each solution δ_m of V_1 and V_2
Add δ_m to the list of recognized multimodal events
Create a variable V_3 of *type 3* for the multimodal event model Add all variables of *type 1,2, and 3* of δ_m to the list of recognized multimodal events

5.6 Behavioral Profile

The first step to establish a behavioral profile of an observed person is to determine his/her daily activities. This behavioral profile is defined as a set of the most frequent and interesting activities of an observed person. The basic goal of determining a behavioral profile is to measure variables from persons during their daily activities in order to capture deviations of activity and posture to facilitate timely intervention or provide automatic alert in emergency cases.

The obtained results of the behavioral profile for 9 elderly persons are described in chapter 6 in section 6.5.2.

5.7 Discussion

The proposed event recognition algorithm is able to recognize which events occur in the scene at each instant. The main problem of this algorithm is that it does not take into account uncertainty in sensor measurements. For example, for a pressure sensor, the person which drops his bag on the chair, may activate the chair sensor (sensor installed under the chair) and giving a false alarm. The false alarm is due when the person is located close to the chair (the same chair where the person has drop his bag) with poor detection for a person (with video camera). To reduce this kind of false alarm we propose to handle uncertainty in sensor measurements by using Dempster-Shafer theory.

5.8 Handling Uncertainty in Sensor Measurements

Advances in technology have provided the ability to equip the home environment with a large number of different sensors like the ones described in the previous section. These sensors may provide information about human activities. The main problem is that data obtained from sensors have different degrees of uncertainty [Ranganathan et al., 2004]. This uncertainty may arise for a number of reasons, as described in section 4.3.1. The question which is to be asked is if a sensor provides a value of "on" or "off" how sure can we be about this measurement and how can we accommodate for any uncertainty that may exist. Bayesian methods and Evidence Theory of which the Dempster-Shafer (DS) theory of evidence (DS theory) is a major constituent are commonly used to handle uncertainty.

In this section, we propose an evidential approach to reasoning under uncertainty in the sensor measurements. The proposed approach is based on the use of Dempster-Shafer theory through the fusion of contextual information inferred from uncertain sensor data.
As described in section 2.3.5.1 in chapter 2, Dempster-Shafer (DS) theory of evidence originated from Dempster work [Dempster, 1968] and further extended by Shafer [Shafer, 1976], is a generalization of traditional probability which allows us to better quantify uncertainty.
In plus to the basic concepts of Dempster-Shafer theory already described in section 2.3.5.1, the following evidential operations are involved when inferring activities along evidential networks:

- **Reliability discounting**: Some sensors are more vulnerable to misreading or malfunctioning due to their type and location and where they are

installed. The impact of evidence is discounted to reflect the sensor's credibility, in terms of discount rate r ($0 \leq r \leq 1$). The discounted mass function for each $A \subset \Theta$ is defined as follows:

$$m^r(A) = (1 - r)m(A) \tag{5.6}$$

Where:
- r=0; the source is absolutely reliable,
- 0<r<1; the source is reliable with a discount rate r,
- r=1; the source is completely unreliable.

- **Multivalued mapping**: Dempster used a multivalued mapping to reflect the relationship between two frames of discernment both representing evidence to the same problem but from different views.
  For two frames of discernment Θ_E and Θ_H, a multivalued mapping Γ describes a mapping function $\Gamma : \Theta_E \leftarrow 2^{\Theta_H}$, assigning to each element e_i of Θ_E a subset Γe_i of Θ_H.

- **Translation**: The evidential operation called translation can be used to determine the impact of evidence originally appearing on a frame of discernment upon elements of a compatibly related frame of discernment. Suppose the frame of discernment Θ_E carries a mass function m, the translated mass function over the compatibly related frame of discernment Θ_H is:

$$m^{'}(H_j) = \sum_{\Gamma e_i = H_j} m(e_i) \tag{5.7}$$

  Where:
  - $e_i \in \Theta_E, H_j \subseteq \Theta_H$, and $\Gamma : \Theta_E \leftarrow 2^{\Theta_H}$ is a multivalued mapping.

5.8.1 Applying Dempster-Shafer Theory of Evidence for Fusing Sensors

In this thesis we used environmental sensors which provide two binary values "On" if the sensor is activated and "Off" if the sensor is not activated.
The challenges posed with the use of binary sensor technology and the determination if a sensor provides a value of "On" or "Off" are huge. By applying Dempster-Shafer (DS) theory of evidence for the representation and management of sensor uncertainty will provide a possible solution to this problem.

The Dempster-Shafer Theory provides a means to numerically represent our belief on the value set of a variable in the form of a mass function.
The exhaustive set of mutually exclusive values that a variable can hold is represented by the frame of discernment (Θ). For instance, an environmental sensor denoted S can be in two states "On" and "Off". If we use "1" to represent "On" and "0" to represent "Off". Then $\Theta = \{1, 0\}$ is a frame of discernment for

the variable S.

A mass value can be committed to either a subset of Θ. This property makes DS theory more expressive than probability theory. When a mass value is committed to a subset that has more than one element, it is explicitly stating that there is not enough information to distribute this belief more precisely to each individual element in the subset. In particular, the total belief is assigned to the whole frame of discernment, $m(\Theta) = 1$, when there is no evidence about Θ at all. In contrast, probability theory lacks this ability by dividing the total belief equally among elements of Θ. If $m(A) > 0$, the subset A of Θ is called a focal element of the belief distribution.

The main difference between these definitions and conventional probability is that a mass value can be committed to either a subset of Θ. Mass functions can be used to define the lower and upper bounds of the probability. The lower bound called the belief (Bel) represents the degree of belief in supporting A. The upper bound called the plausibility (Pls) describes the degree of belief on failing to refute A.

Combination Rule: One of the main advantages of Evidence Theory is that Dempsters rule of combination allows us to accumulate evidence from distinct sources. In the case of imperfect data (uncertain, imprecise and incomplete), fusion is an interesting solution to obtain more relevant information. Evidence theory offers appropriate aggregation tools. From the basic belief assignment denoted m_i obtained for each information source, it is possible to use a combination rule in order to provide combined masses of the different sources. Let m_1 and m_2 be two mass functions on Θ. A new mass function m then is formed by combining m_1 and m_2 as:

$$m(C) = (m_1 \oplus m_2)(C) = \frac{\sum_{A \bigcap B=C \neq \phi} m_1(A) m_2(B)}{1 - \sum_{A \bigcap B=\phi} m_1(A) m_2(B)} \tag{5.8}$$

With $A, B, C \in \Theta$.
In the equation 5.8, the numerator represents the accumulated evidence for the sets A and B, which supports the hypothesis C, and the denominator sum quantifies the amount of conflict between the two sets.

Maximization: "Preparing Cold Meal" and "Preparing Hot Meal" are two alternative sub-activities of "Preparing Meal" activity.

We define the maximization operation to calculate the aggregated belief values on an activity contributed from its alternative sub-activities.

$$Bel(C) = max(Bel(A), Bel(B)), and Pls(C) = max(Pls(A), Pls(B))$$

Where C is the composite of alternatives A and B.

5.8.2 Evidential Network for Activity Recognition

Sensors, once activated, present contextual evidence such as which room the person is in, which objects the person is interacting with and whether or not a person is moving around the home. All of this information provides valuable evidence which in turn can be considered indicative as to what activities the person is performing.
Based on the proposed concept of ontology networks of activity as presented in the previous section, we propose evidential networks of activity inference. Lower level activities can be considered as evidence of higher level activities where the lowest level activities are inferred from sensed contexts. We propose two types of evidential networks: **activity** type and **sensor** type.

- An activity network contains only activities in a tree hierarchy. An activity can be composed by one or several sub-activities. An activity may also be a sub-activity to another activity.

  There are two types of connections between an activity and its sub-activities.

  - For the first type of connection, the activity is said to be carried out only when any of its sub-activities have been performed (i.e. an activity i is performed when his sub-activity i_1 **or** his sub-activity i_2 is performed). Such a network is drawn as a tree in which the connections between an activity and its sub-activities are represented by lines coming from the sub-activities which then merge into a single line ended by a triangle. For example, the network shown in figure 5.38-a indicates that preparing meal (activity) can be either the preparing cold meal sub-activity **or** the preparing hot meal sub-activity.

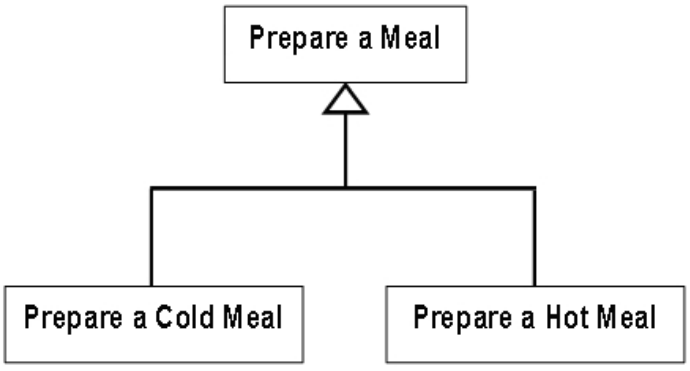


Figure 5.38: Examples of evidential networks of activity type. (The graphical notations are summarized in figure 5.4)

  - In the second type of connection, the activity is only considered complete when all his sub-activities have been performed (i.e. an activity

i is performed when his sub-activity i_1 **and** his sub-activity i_2 were performed). This type of connection is drawn by lines from the sub-activities which all merge into a single line ended by a square.

- A sensor network is also represented as a tree hierarchy in which sensors are represented by circle nodes, contextual objects and activities are represented by rectangular nodes. With different involvements of contextual objects in performing a given activity it is possible to divide them into two groups: necessary and accessory contextual objects. Necessary contextual objects are the compulsory contextual objects which must be interacted with when performing a certain activity. Accessory contextual objects can be considered as optional and may or may not be involved in the performance of a specific activity.
  The connections between the necessary contextual objects with the activity are presented by lines coming from the contextual objects which then merge into a single line ended by a square. The connections between accessory contextual objects and the activity are represented by lines coming from the contextual objects which then merge into a single line ended by a triangle.
  Figure 5.39 displays two examples of sensor network type: Prepare cold meal and Prepare hot meal. It is upon the ability to formalise the representation of ontology networks that we can now proceed and manage uncertainty within sensed contexts.

5.8.3 Evidential inference of activities

Based on the simplified ontology example as previously introduced in section 5.4.3, we draw a scenario which will be used throughout this section to help illustrate the Dempster-Shafer concepts and evidential operations.
Case study: There are many activities that can be performed in the kitchen, such as "Prepare meal" ("Prepare Cold Meal" or "Prepare Hot Meal"). Based on the simplified ontology of activities in the kitchen as shown in Figure 5.6, we can derive the evidential networks for "Prepare Meal", "Prepare Cold Meal", and "Prepare Hot Meal", as shown in Figures 5.38(a), 5.39(a) and 5.39(b) respectively. Inference through the evidential networks can then find out what activity is most likely to have been performed in the kitchen.

Observations occur at the sensor nodes as shown in Figure 5.39. For example, in "Prepare Cold Meal" activity the sensors sFrid, sCupb, sWatr and sVideo were fired and were activated (see figure 5.39(a)), the other sensors were not activated (i.e. sMicr and sStov). The activity "Prepare Cold Meal" in Figure 5.39, the activity "Prepare Hot Meal" in Figure 5.39(b) and the activity "Prepare meal" in Figure 5.38 are the hypotheses to be deduced.

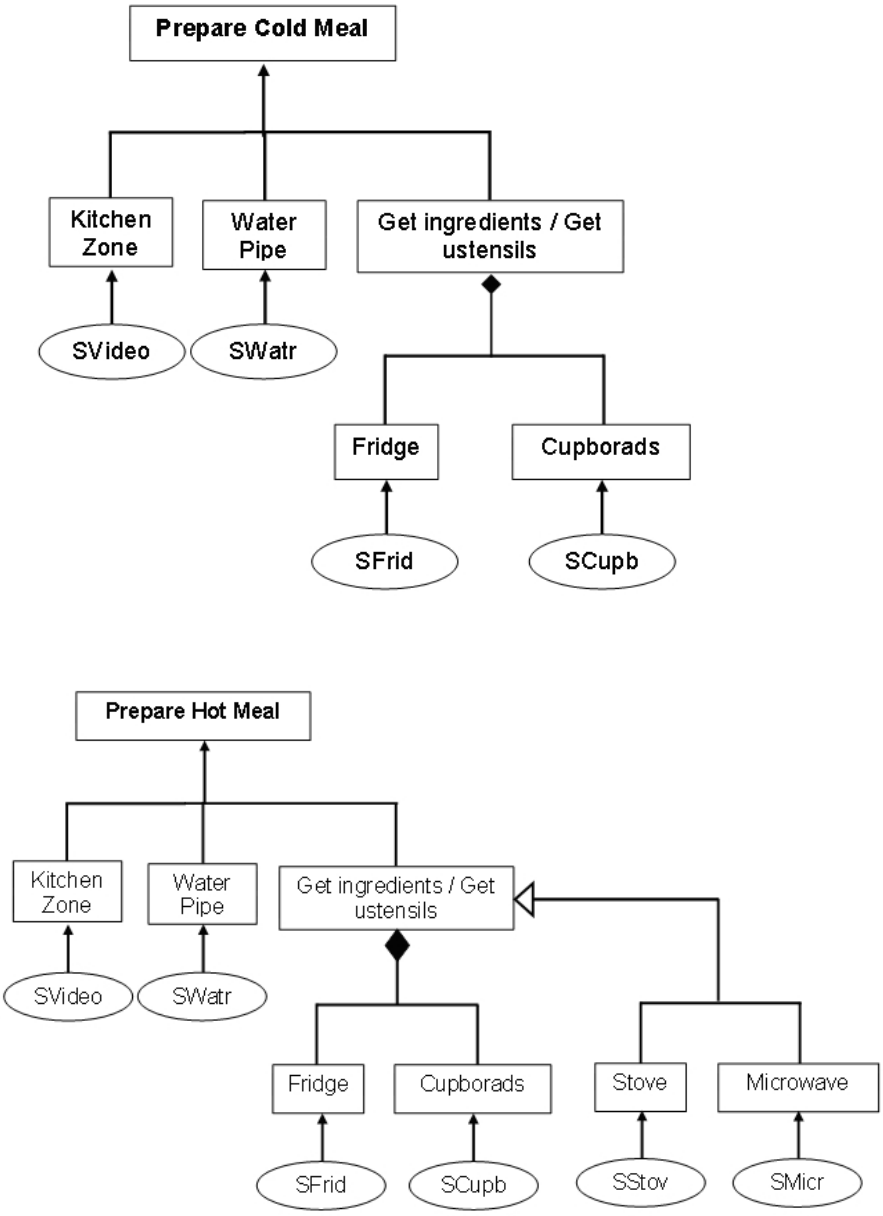


Figure 5.39: Examples of evidential networks of sensor type; (a) Prepare Cold Meal, (b) Prepare Hot Meal. Sensor abbreviations: SFrid: fridge sensor, SCupb: cupboard sensor, SStov: stove sensor, SMicr: microwave sensor, SWatr: water sensor, SVideo: video sensor. (The graphical notations are summarized in figure 5.4)

5.8.4 Evidential network representation

Inferring activities starts from representing the evidential networks in evidential forms. Each node is represented by the frame of discernment. For the case

Name	Type	Frame of discernment
SFrid	Sensor	$\{SFrid, \neg SFrid\}$
Frid	Context (i.e. equipment)	$\{Frid, \neg Frid\}$
Prepare Cold Meal	Activity	$\{PrepareColdMeal, \neg PrepareColdMeal\}$

Table 5.7: Examples of frames of discernment

study, table 5.7 shows an example of the frame of discernment for each type of node. Sensor nodes can have two values: active and inactive, hence the frame of discernment for a sensor is comprised of two elements. Each arc in an evidential network represents the relationship between one node to another, which can be represented by a multivalued mapping or an evidential mapping.

In the evidential networks of the case study, all relationships between a sensor and its associated contextual object node, and an activity and its sub-activity are compatible. Given this compatibility they are represented by multivalued mappings. Table 5.8 shows examples of multivalued mappings.

5.8.5 Activity Inference on Evidential Network

Activity inference starts with the evidential networks of sensors, followed by reasoning on activities networks.
In a sensor network, evidence appears on a sensor node associated with a contextual object, which can be summed up onto a composite contextual object node by an equally weighted sum operation that is then translated to the relevant activity node, or propagated to a connected activity node by an evidential mapping.

On an activity node, several belief distributions can be combined by Dempster's combination rule.
In the example showing "Prepare Cold Meal", firstly, evidence on sensor nodes are represented by mass functions as follows:

$$m_{SFrid}(\{SFrid\}) = 1; \tag{5.9}$$
$$m_{SCupb}(\{SCupb\}) = 1; \tag{5.10}$$
$$m_{SWatr}(\{SWatr\}) = 1; \tag{5.11}$$
$$m_{SVideo}(\{SVideo\}) = 1; \tag{5.12}$$
$$m_{SStov}(\{\neg SStov\}) = 1; \tag{5.13}$$
$$m_{SMicro}(\{\neg SMicro\}) = 1; \tag{5.14}$$

In the example showing "Prepare Hot Meal", firstly, evidence on sensor nodes

Relationship	**Multivalued mappings**
$SFrid \rightarrow Frid$	$\{SFrid\} \rightarrow \{Frid\}$; $\{\neg SFrid\} \rightarrow \{\neg Frid\}$;
$SCupb \rightarrow Cupb$	$\{SCupb\} \rightarrow \{Cupb\}$; $\{\neg SCupb\} \rightarrow \{\neg Cupb\}$;
$SWatr \rightarrow Watr$	$\{SWatr\} \rightarrow \{Watr\}$; $\{\neg SWatr\} \rightarrow \{\neg Watr\}$;
$(SFrid, SCupb, SWatr) \rightarrow (Frid, Cupb, Watr)$	$\{(SFrid, SCupb, SWatr)\} \rightarrow \{(Frid, Cupb, Watr)\}$; $\{\neg(SFrid, SCupb, SWatr)\} \rightarrow \{\neg(Frid, Cupb, Watr)\}$;
$(Frid, Cupb, Watr) \rightarrow PrepareColdMeal$	$\{(Frid, Cupb, Watr)\} \rightarrow \{PrepareColdMeal\}$; $\{\neg(Frid, Cupb, Watr)\} \rightarrow \{\neg PrepareColdMeal\}$;
$PrepareColdMeal \rightarrow PrepareMeal$	$\{PrepareColdMeal\} \rightarrow \{PrepareMeal\}$;

Table 5.8: Examples of multivalued mappings; SFrid: represents a fridge sensor (a contact sensor associated to the contextual object "Fridge"), SCupb: represents a cupboard sensor (a contact sensor associated to the contextual object "Cupboard"), and SWatr: represents a water sensor (a water sensor associated to the contextual object "Water Pipe"), Frid: represents the contextual object "Fridge", Cupb: represents the contextual object "Cupboard", and Watr: represents the contextual object "Water Pipe"

are represented by mass functions as follows:

$$m_{SFrid}(\{SFrid\}) = 1; \tag{5.15}$$
$$m_{SCupb}(\{SCupb\}) = 1; \tag{5.16}$$
$$m_{SWatr}(\{SWatr\}) = 1; \tag{5.17}$$
$$m_{SVideo}(\{SVideo\}) = 1; \tag{5.18}$$
$$m_{SStov}(\{SStov\}) = 1; \tag{5.19}$$
$$m_{SMicro}(\{SMicro\}) = 1; \tag{5.20}$$

The inference procedure consists of four steps of evidential operations.

- **Step 1:** Discounting sensor evidence.
  In this study we assume that the video sensor is reliable at 100%.
  Statistics (see section 6.4.2) using ground truth of 20 video sequences of one human actor show that the used environmental sensors (in the kitchen and in the livingroom) are working correctly at different rates: 95% for contact sensors including fridge and cupboard sensors, 90% for electrical sensors including microwave and stove sensors, 85% for water flow sensors including water pipes sensors, 70% for pressure sensors and 70% for presence sensors. So a discount rate of 5% is assigned to fridge and cupboard sensors, 10% is assigned to electrical sensors including microwave and stove sensors, 15% is assigned to water flow sensors including water pipe sensors, and 20% is assigned to pressure sensors and presence sensors. The discounted mass functions of fridge, cupboard, stove, microwave, water and video sensors are calculated as following:

$$m^r_{SFrid}(\{SFrid\}) = 0.95;\ m^r_{SFrid}(\{SFrid, \neg SFrid\}) = 0.05; \tag{5.21}$$
$$m^r_{SCupb}(\{SCupb\}) = 0.95;\ m^r_{SCupb}(\{SCupb, \neg SCupb\}) = 0.05; \tag{5.22}$$
$$m^r_{SStov}(\{SStov\}) = 0.90;\ m^r_{SStov}(\{SStov, \neg SStov\}) = 0.10; \tag{5.23}$$
$$m^r_{SMicro}(\{SMicro\}) = 0.90;\ m^r_{SMicro}(\{SMicro, \neg SMicro\}) = 0.10; \tag{5.24}$$
$$m^r_{SWatr}(\{SWatr\}) = 0.85;\ m^r_{SWatr}(\{SWatr, \neg SWatr\}) = 0.15; \tag{5.25}$$
$$m^r_{SVideo}(\{SVideo\}) = 1.00;\ m^r_{SVideo}(\{SVideo, \neg SVideo\}) = 0.00; \tag{5.26}$$

- **Step 2:** Translating mass functions from sensors to associated contextual objects. A sensor being active indicates the associated contextual object has been interacted with. A sensor and the associated contextual object maintain a compatible relationship which can be represented by a multivalued mapping as the examples shown in Table 5.9. The mass function on a

Relationship	**Evidence mappings**
$SFrid \rightarrow Frid$	$\{SFrid\} \rightarrow \{(\{Frid\}, 0.95), (\{SFrid, \neg Frid\}, 0.05)\}$
$SCupb \rightarrow Cupb$	$\{SCupb\} \rightarrow \{(\{Cupb\}, 0.95), (\{SCupb, \neg Cupb\}, 0.05)\};$
$SWatr \rightarrow Watr$	$\{SWatr\} \rightarrow \{(\{Watr\}, 0.85), (\{SWatr, \neg Watr\}, 0.15)\};$
$SStov \rightarrow Stov$	$\{SStov\} \rightarrow \{(\{Stov\}, 0.90), (\{SStov, \neg Stov\}, 0.10)\};$
$SMicr \rightarrow Micr$	$\{SMicr\} \rightarrow \{(\{Micr\}, 0.90), (\{SMicr, \neg Micr\}, 0.10)\};$

Table 5.9: Examples of evidence mappings

sensor node can then be translated to the associated contextual object node by using the multivalued mapping.

$$m_{Frid}(\{Frid\}) = m^r_{SFrid}(\{SFrid\}) = 0.95; \quad (5.27)$$
$$m_{Frid}(\{Frid, \neg Frid\}) = m^r_{SFrid}(\{SFrid, \neg SFrid\}) = 0.05; \quad (5.28)$$
$$m_{Cupb}(\{Cupb\}) = m^r_{SCupb}(\{SCupb\}) = 0.95; \quad (5.29)$$
$$m_{Cupb}(\{Cupb, \neg Cupb\}) = m^r_{SCupb}(\{SCupb, \neg SCupb\}) = 0.05; \quad (5.30)$$
$$m_{Watr}(\{Watr\}) = m^r_{SWatr}(\{SWatr\}) = 0.85; \quad (5.31)$$
$$m_{Watr}(\{Watr, \neg Watr\}) = m^r_{SWatr}(\{SWatr, \neg SWatr\}) = 0.15; \quad (5.32)$$
$$m_{Stov}(\{Stov\}) = m^r_{SStov}(\{SStov\}) = 0.90; \quad (5.33)$$
$$m_{Stov}(\{Stov, \neg Stov\}) = m^r_{SStov}(\{SStov, \neg SStov\}) = 0.10; \quad (5.34)$$
$$m_{Micr}(\{Micr\}) = m^r_{SMicr}(\{SMicr\}) = 0.90; \quad (5.35)$$
$$m_{Micr}(\{Micr, \neg Micr\}) = m^r_{SMicr}(\{SMicr, \neg SMicr\}) = 0.10; \quad (5.36)$$

- **Step 3:** Summing up on a composite contextual object node. On the evidential network "Prepare Cold Meal", "Fridge, Cupb, Watr" is the composite node formed by "Fridge", "Cupb" and "Watr". The three mass functions translated from "Fridge", "Cupb" and "Watr" onto "Fridge, Cupb, Watr" as calculated in the previous step are then summed up by the equally weighted sum operator.

$$
\begin{aligned}
& m_{Frid,Cupb,Watr}(\{Frid,Cupb,Watr\}) \\
& = \frac{1}{3}(m_{Frid}(\{Frid\}) + m_{Cupb}(\{Cupb\}) \\
& + m_{Watr}(\{Watr\})) = \frac{1}{3}(0.95 + 0.95 + 0.85) = 0.916 \qquad (5.37)
\end{aligned}
$$

$$
\begin{aligned}
& m_{Frid,Cupb,Watr}(\{(Frid,Cupb,Watr), \neg(Frid,Cupb,Watr)\}) \\
& = \frac{1}{3}(m_{Frid}(\{Frid, \neg Frid\}) \\
& + m_{Cupb}(\{Cupb, \neg Cupb\}) \\
& + m_{Watr}(\{Watr, \neg Watr\})) \\
& = \frac{1}{3}(0.05 + 0.05 + 0.15) = 0.083 \qquad (5.38)
\end{aligned}
$$

The contexts "Microwave" and "Stove" are the two alternatives of context "Microwave/Stove". The mass function on "Microwave/Stove" can be calculated by the maximization operator as follows:

$$
\begin{aligned}
& m_{Micr,Stov}(\{Micr,Stov\}) = max(m_{Micr}(\{Micr\}), m_{Stov}(\{Stov\})) \\
& = max(0.90, 0.90) = 0.90 \qquad (5.39) \\
& m_{Micr,Stov}(\{Micr,Stov\}, \neg\{Micr,Stov\}) \\
& = max(m_{Micr}(\{Micr, \neg Micr\}), m_{Stov}(\{Stov, \neg Stov\})) \\
& = max(0.10, 0.10) = 0.10 \qquad (5.40) \\
& \qquad (5.41)
\end{aligned}
$$

On the evidential network "Prepare Hot Meal", "Fridge, Cupb, Watr, Micr, Stov" is the composite node formed by "Fridge", "Cupb", "Watr", "Micr" and "Stov". The five mass functions translated from "Fridge", "Cupb", "Watr", "Micr" and "Stov" onto "Fridge, Cupb, Watr, Micr, Stov" as calculated in the previous step are then summed up by the equally weighted sum operator.

$$
\begin{aligned}
& m_{Frid,Cupb,Watr,Micr,Stov}(\{Frid,Cupb,Watr,Micr,Stov\}) \\
& = \tfrac{1}{4}(m_{Frid}(\{Frid\}) + m_{Cupb}(\{Cupb\}) \\
& + m_{Watr}(\{Watr\}) + m_{Micr,Stov}(\{Micr,Stov\})) \\
& = \tfrac{1}{4}(0.95 + 0.95 + 0.85 + 0.90) = 0.912
\end{aligned} \qquad (5.42)
$$

$$
\begin{aligned}
& m_{Frid,Cupb,Watr,Micr,Stov}(\{(Frid,Cupb,Watr,Micr,Stov), \neg(Frid,Cupb,Watr,Micr,Stov)\}) \\
& = \tfrac{1}{4}(m_{Frid}(\{Frid, \neg Frid\}) \\
& + m_{Cupb}(\{Cupb, \neg Cupb\}) \\
& + m_{Watr}(\{Watr, \neg Watr\})) \\
& + m_{Micr,Stov}(\{Micr,Stov\}, \neg\{Micr,Stov\}) \\
& = \tfrac{1}{4}(0.05 + 0.05 + 0.15 + 0.10) = 0.087
\end{aligned} \qquad (5.43)
$$

- **Step 4:** Translating from a composite contextual object node or propagating from an accessory contextual object node, to an activity node.

On network "Prepare Cold Meal", the mass function on "Fridge, Cupboard, Water" is translated to "Prepare Cold Meal".

$$\begin{aligned} & m_{PrepareColdMeal}(\{PrepareColdMeal\}) \\ & = m_{Frid,Cupb,Watr}(\{Frid, Cupb, Watr\}) \\ & = 0.916 \end{aligned} \tag{5.44}$$

$$\begin{aligned} & m_{PrepareColdMeal}(\{PrepareColdMeal, \neg PrepareColdMeal\}) \\ & = m_{Frid,Cupb,Watr}(\{(Frid, Cupb, Watr), \neg(Frid, Cupb, Watr)\}) = 0.083 \end{aligned} \tag{5.45}$$

On network "Prepare Hot Meal", the mass function on "Fridge, Cupboard, Water, Microwave, Stove" is translated to "Prepare Hot Meal".

$$\begin{aligned} & m_{PrepareHotMeal}(\{PrepareHotMeal\}) \\ & = m_{Frid,Cupb,Watr,Micr,Stov}(\{Frid, Cupb, Watr, Micr, Stov\}) \\ & = 0.912 \end{aligned} \tag{5.46}$$

$$\begin{aligned} & m_{PrepareHotMeal}(\{PrepareHotMeal, \neg PrepareHotMeal\}) \\ & = m_{Frid,Cupb,Watr,Micr,Stov}(\{(Frid, Cupb, Watr, Micr, Stov), \neg(Frid, Cupb, Watr, Micr, Stov)\}) \\ & = 0.087 \end{aligned} \tag{5.47}$$

Calculating Bel and Pls: From mass function on "Prepare Cold Meal" and "Prepare Hot Meal", we calculate the beliefs for "Prepare Cold Meal" and "Prepare Hot Meal" as follows:

$$Bel(PrepareColdMeal) = m(PrepareColdMeal) = 0.916 \tag{5.48}$$

$$\begin{aligned} Pls(PrepareColdMeal) &= m(PrepareColdMeal) \\ &+ m(PrepareColdMeal, \neg PrepareColdMeal) \\ &= 0.916 + 0.083 = 0.999 \end{aligned} \tag{5.49}$$

Then the uncertainty μ of "Prepare Cold Meal" is calculated using the following equation:

$$\begin{aligned} \mu\{PrepareColdMeal\} &= Pls(PrepareColdMeal) - Bel(PrepareColdMeal) \\ &= 0.999 - 0.916 = 0.083 \end{aligned} \tag{5.50}$$

$$Bel(PrepareHotMeal) = m(PrepareHotMeal) = 0.912 \tag{5.51}$$

$$\begin{aligned} Pls(PrepareHotMeal) &= m(PrepareHotMeal) \\ &+ m(PrepareHotMeal, \neg PrepareHotMeal) \\ &= 0.912 + 0.087 = 0.999 \end{aligned} \tag{5.52}$$

Then the uncertainty μ of "Prepare Hot Meal" is calculated using the following equation:

$$\begin{aligned}\mu\{PrepareHotMeal\} &= Pls(PrepareHotMeal) - Bel(PrepareHotMeal) \\ &= 0.999 - 0.912 = 0.087\end{aligned} \tag{5.53}$$

Bel on "Preparing Cold Meal" is 0.916 with a value of 0.004 greater than that on "Preparing Hot Meal", and ($PlsBel$) is smaller on "Preparing Cold Meal" than "Preparing Hot Meal" (0.083 vs. 0.087). These results indicate that with a high confidence we can identify that the activity "Preparing Cold Meal" has been performed in the kitchen.
In an evidential network of activity of type 1, the belief of an activity is the maximum of beliefs over its sub-activities.
On the evidential network of "Preparing Meal", "Preparing Cold Meal" and "Preparing Hot Meal" are the two alternative sub-activities of "Preparing Meal" activity. With the beliefs on "Preparing Cold Meal" and "Preparing Hot Meal" calculated above, the belief about that the person is "Preparing Meal" is calculated by the maximization operator (see equation 5.9) as follows:

$$\begin{aligned}Bel(PreparingMeal) &= max(Bel(PreparingColdMeal), Bel(PreparingHotMeal)) \\ &= max(0.916, 0.912) = 0.916\end{aligned} \tag{5.54}$$

$$\begin{aligned}Pls(PreparingMeal) &= max(Pls(PreparingColdMeal), Pls(PreparingHotMeal)) \\ &= max(0.999, 0.999) = 0.999\end{aligned} \tag{5.55}$$

5.9 Conclusion

In this chapter we have introduced a framework within multisensor data can be processed and fused for activity recognition. This fusion consists in combining video events with environmental events. The proposed framework allows to recognize a set of human activities at home with a low rate of false alarms.
We have done a strong effort in event modeling. The result is 100 models which is our knowledge base of events.
We have also proposed an ontology for daily activities which can be used in other applications in the same domain.
To handle with uncertainty in sensor measurements, we have proposed the use of Dempster-Shafer theory of evidence. We have proposed evidential networks to represent the hierarchy of inferring activities based on sensor data. Four evidential operations have been formalized for activity inference on evidential networks which can accommodate the fusion of different types and sources of data.

In the next chapter, we evaluate the proposed approach using a set of video and environmental sensors data.

Chapter 6

Evaluation and Results of the Proposed Approach

In order to evaluate the whole proposed activity monitoring framework, several experiments have been performed. The main objectives of these experiments are to validate the different phases of the activity monitoring framework, to highlight interesting characteristic of the approach, and to evaluate the potential of the framework for real world applications.

The performed evaluations consist of:

- First, an evaluation of the vision-based framework for real world applications. In this experiment, 15 videos were tested in an experimental laboratory (called Gerhome) for elderly care at home. For more details, refer to section 6.4.1.
- Second, an evaluation of the proposed sensor-based model for real world applications. In this experiment, we have used 20 video sequences of one human actor to calculate the a posteriori probability for each environmental sensor. For more details, refer to section 6.4.2.
- Third, an evaluation of the multisensor-based fusion framework in a real world application is performed. It consists in analyzing sensors data and video sequences in the same experimental site (i.e. Gerhome). This experiment has multiple objectives, such as evaluating the influence of the utilization of multiple sensors to recognize activities at home. The experiment is detailed in section 6.4.3.
- Finally, medical evaluation using real data for 9 observed elderly volunteers is performed. It consists in comparing the behavioral profile for 9 elderly persons using results of 6 daily activities. The experiment is detailed in section 6.5.

This chapter is organized as follows. First, section 6.1 describes the experimental site, including the number and placement of installed sensors. Second, the

metrics utilized in the evaluation of our framework are described in section 6.2. Third, the different performed experiments are described in section 6.3. Fourth, the performance evaluation and the obtained results are described in section 6.4. Fifth, medical evaluation is presented in section 6.5 and finally, section 6.6 presents a conclusion about the experiments.

6.1 Experimental Site

Developing and testing the impact of the activity monitoring solutions requires a realistic environment in which training and evaluation can be performed. To attain this goal we have set up an experimental laboratory (called Gerhome, see figure 6.1). This laboratory is located in the CSTB (Scientific Center of Technical Building) at Sophia Antipolis in France.

Figure 6.1: External views of the Gerhome laboratory.

6.1.1 Gerhome Laboratory

Gerhome laboratory is equipped with the different sensors previously cited in section 5.2.1. This laboratory has been built to evaluate the performance of the multisensor fusion approach and to explore the activities that can be recognized by such approach. This laboratory looks like a typical apartment of an elderly person: $41m^2$ with an entrance, a livingroom, a bedroom, a bathroom, and a kitchen. The kitchen includes an electric stove, a microwave, a fridge, cupboards, and drawers.

The Gerhome laboratory plays an important role in research and system development in the domain of activity monitoring and of assisted living. Firstly, it is used to collect data from the different installed sensors. Secondly, it is used as a demonstration platform in order to visualize the system results. Finally, it is used to assess and test the usability of the system with elderly. Figure 6.2 and figure 6.3 show respectively pictures and a 3D visualization of the Gerhome laboratory.

(a) The kitchen (b) The livingroom (c) The bathroom

Figure 6.2: Internal views of the Gerhome laboratory.

6.1.2 Video Cameras and Environmental Sensors

Gerhome is equipped with different sensors (see figure 6.4) to evaluate ADL scenarios predefined by investigating gerontologists from Nice hospital. Commercially available sensing devices were used for data gathering including video cameras, and environmental sensors embedded in the home infrastructure. We call environmental sensors each sensor that measures environmental information such as pressure, temperature, light (e.g. pressure sensors, electrical sensors, light sensors).

To detect and track a person in Gerhome laboratory and to recognize his/her activities and postures, four video cameras are installed in this laboratory. One video camera is installed in the kitchen, two video cameras are installed in the livingroom and the last one is installed in the bedroom. Figure 6.5 shows the different views of the installed video cameras.

Twelve contact sensors are mounted on many devices in the apartment for detecting the opening and closing of cupboard doors, fridge door, drawers and closet doors. Two electrical sensors for detecting electrical appliance use (i.e. microwave, stove, phone and TV). Three presence sensors to detect the presence of people near sinks, cooking stoves and washbowls. Four hot and cold water consumption sensors in the kitchen and bathroom. Four pressure sensors located beneath 2 chairs, an armchair and a bed to detect when a person is sitting, sleeping or not. Figure 6.6 shows some pictures of these installed sensors. The selected sensors can easily and quickly be installed in home environments and are removable without damage to the cabinets or furniture. These environmental sensors are based on RF (radio frequency) transceiver with low battery consumption (i.e. lithium battery with a lifespan of 1 year). The number of installed sensors varies depending on the room and the areas of interest (see Table 6.1). For example, in the bedroom there are only one video camera, 2 contact sensors installed on closet doors and one pressure detector that is installed under the bed. We have not installed a video camera in the entrance and in the bathroom. In the entrance because the second installed video camera in the livingroom has a field view to the entrance. In the bathroom to save the privacy of the observed person.

(a) View of the kitchen from the livingroom

(b) A top view

Figure 6.3: 3D visualization of the Gerhome laboratory.

6.2 Evaluation Metrics

Different metrics have been used according to the nature of the experiment.

For the recognition of states and events, the utilized metrics are:

- **True Positive (TP)**: An event E_i is correctly detected according to the

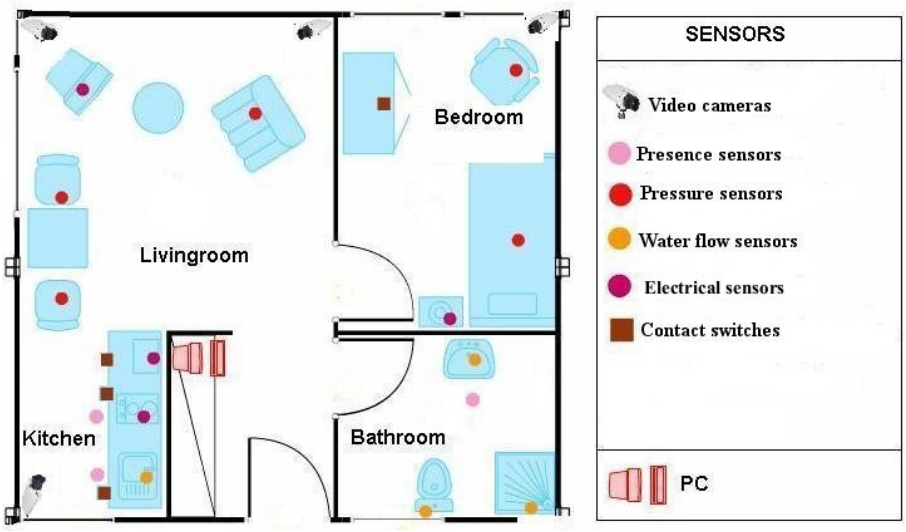
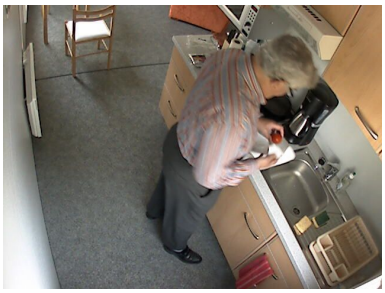
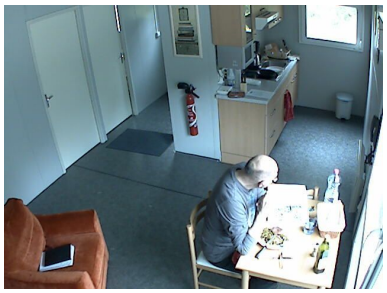


Figure 6.4: Position of the sensors in the Gerhome laboratory.

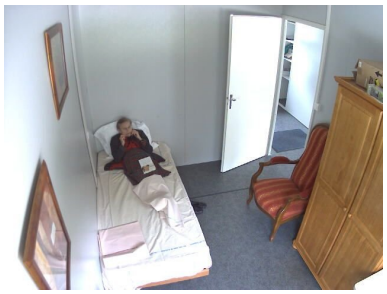


(a) Video camera 1 in the kitchen



(b) Video camera 2 in the living-room

(c) Video camera 3 in the living-room



(d) Video camera 4 in the bedroom

Figure 6.5: Views from the installed video cameras in the Gerhome laboratory.

ground truth.

- **False Positive (FP)**: An event E_i is wrongly detected according to the

Figure 6.6: Views of some environmental sensors installed in the Gerhome laboratory. Sensors are circled. (a) Contact sensor on cupboard door in the kitchen; (b) Electrical sensor on electrical outlet in the kitchen; (c) Presence sensor in front of the washbowl in the bathroom; (d) Water sensor on water pipe in the kitchen; (e) Pressure sensor under the armchair in the livingroom.

Sensors	Entrance	Livingroom	Kitchen	Bathroom	Bedroom	Total
Video Camera	0	2	1	0	1	4
Contact Sensor	0	0	9	0	2	11
Pressure Sensor	0	3	0	0	1	4
Water Flow Sensor	0	0	2	2	0	4
Electrical Sensor	0	1	1	0	0	2
Presence Sensor	0	0	2	1	0	3

Table 6.1: List of installed sensors per room

ground truth.

- **False Negative (FN)**: A false negative occurs when an event E_i occurs and a system does not report it.

- **Precision (P)**: The precision metric can be seen as a measure of exactness or fidelity. The precision corresponds to the number of events correctly detected divided by the total number of detected events. This metric is formally defined as:

$$P = \frac{TP}{TP + FP} \tag{6.1}$$

- **Sensitivity (S)**: A sensitivity corresponds to the number of events correctly detected divided by the total number of occurred events. A sensitivity of 100% means the recognition of the all occurring events. This metric is formally defined as:

$$S = \frac{TP}{TP + FN} \tag{6.2}$$

For the comparison of behaviors of 2 or more persons (see section 6.4.3.2), we have used the following metrics:

- **Number of instance (i.e. n1, n2, ...)**: It corresponds to the number of instance of a certain activity occurred during the experiment.

- **Mean duration (i.e. m1, m2, ...)**: It corresponds to the mean duration of a certain activity occurred during the experiment.

- **Normalized Difference of mean durations of Activity (NDA)**: It corresponds to the difference of two mean durations divided by their sum. This metric is formally defined as:

$$NDA = \frac{|m1 - m2|}{m1 + m2} \tag{6.3}$$

- **Normalized Difference of Instance number (NDI)**: It corresponds to the difference of two number of instance divided by their sum. This metric is formally defined as:

$$NDI = \frac{|n1 - n2|}{n1 + n2} \tag{6.4}$$

In addition to the metrics already cited, we have defined a 3D visualization tool in order to visualize the recognized events.

A 3D Visualization Tool

In collaboration with Bernard Boulay from Pulsar team, we have developed a prototype of a 3D visualization tool which is useful for a demonstration and debugging purposes. For example, we can verify the coherence of a proposed event by visualizing it. A 3D visualization tool displays a 3D scene environment, mobile objects (usually persons) and recognized events.

We have proposed a 3D engine based on OpenGL to display the 3D scene environment. Each contextual object observable in the scene is manually modeled with 3D colored and textured parallelepipeds (e.g. floor, walls, table, cupboard). A specific property is associated to the objects which can have interaction with people evolving in the scene (e.g. microwave, fridge).

These objects are then highlighted as soon as a detected event involves these objects. A 3D human model can be displayed with the recognized posture at the detected 3D position. Finally, the different recognized events are displayed as overlay (see figure 6.10(b) for example) in the 3D virtual scene: the location of the detected person, the involved sensors (video camera or other sensors) and the current detected activity. The tool takes as input the video and environmental processing results. An illustration of a 3D visualization tool for the Gerhome laboratory is shown in figure 6.7.

Figure 6.7: An illustration of a 3D visualization.

6.3 Performed Experiments

To evaluate the proposed activity monitoring framework we have tested a set of human activities in the Gerhome laboratory.
In this section, we describe the predefined scenarios and the data collection.

6.3.1 Predefined Scenarios and Data Collection

In this thesis we mainly focus on activities taking place in the kitchen and in the livingroom (e.g. person location in each zone in the laboratory, open kitchen equipment, prepare a meal, take a meal). We study a range of activities that are useful in a home health monitoring system.
Two validation experiments are performed (using our datasets acquired in the Gerhome laboratory): The first one with one human actor, and the second one with fourteen elderly people.

6.3.2 With One Human Actor

In the first experiment, one human actor (i.e. woman of 33 years) has tested some household activities in the Gerhome laboratory such as: using fridge, using microwave, preparing a meal. This actor has also tested some abnormal situations of elderly living alone in his/her own home such as fainting and falling down. She has also tested some activities occurring at the same time, for example

slumping on armchair during preparing a meal. This may indicate that the person feels ill or is sleepy (see figure 6.8).

The given scenario is described as follow: "A person takes a ready meal made from the fridge, and puts the meal in the microwave oven to warm it. After that, the person leaves the kitchen and goes to the livingroom to sit in the armchair. After some minutes the person slumped in the armchair with closed eyes. The microwave oven is still running and this can cause fire".

Figure 6.8: A person is slumping in the armchair when he/she warms up a meal in the microwave oven.

6.3.3 With Fourteen Elderly Volunteers

In the second experiment, fourteen volunteers (i.e. 6 women and 8 men aged from 60 years to 85 years) were recruited by advertisements for a study of ways to make sensing technologies easier to use in the home. A major goal of this experiment is to analyze behavioral data that are as natural as possible.

While evolving in the Gerhome laboratory, the fourteen volunteers have been observed, each one during 4 hours, and 56 video sequences have been acquired by 4 video cameras (about ten frames per second), each video sequence contains about 144 000 frames. The collected data includes the 56 video streams, and also sensors data provided by the 24 environmental sensors. The access of these data is limited (a password is necessary to have access to these data) and not yet available [1].

The volunteers were encouraged to behave freely and to maintain as normal as possible their behaviors and were asked to perform a set of household activities (for more detail about the proposed scenario see Appendix B), such

[1] http://www-sop.inria.fr/members/Francois.Bremond/topicsText/gerhomeProject.html

as preparing meal, taking meal, washing dishes, cleaning the kitchen, watching TV and taking a nap while staying at Gerhome laboratory. Each volunteer was alone in the laboratory during the observation period and was observed during 4 hours (i.e. between 10h and 14h) by using the 4 installed video cameras (see section 6.1.2).

All the volunteers were interviewed separately after the study about the experience of living in the Gerhome laboratory. Volunteers were asked questions about the proposed scenario, the acceptance of the sensor technologies and about the Gerhome laboratory. The post study interviews with the volunteers who have participated in the experiment indicate that the sensors do not impact of their everyday behavior.

In this second experiment, using the video camera 2 installed in the livingroom, we have collected 56 hours of video data of the 14 volunteers. Our data collection has several limitations. We mentioned here two limitations:

- The instrumented home was not the volunteers real home, volunteer's behavior was not completely natural (e.g. many volunteers have opened several kitchen equipments before executing the predefined scenario).
- Our dataset is missing some activities (e.g. activities taking place in the bedroom and in the bathroom). The bathroom was not observable by the video camera, so many activities of potential interest related to personal hygiene, and grooming are not collected.

Due to the tedious and therefore costly nature of annotation, our results use a subset of 36 hours (i.e. 9x4) of the collected data. We have annotated only 20 hours (i.e. 5x4) from these 36 hours.

6.4 Performance Evaluation

In this section we describe the different evaluations. First, we describe the evaluation of the vision-based framework with the obtained results. Second, we describe the evaluation of the sensor-based modeling framework with the obtained results. Finally, we describe the evaluation of the multisensor-based framework with the obtained results and also the obtained results with presence of uncertainty in sensor measurements.

6.4.1 Evaluation of the Vision-Based Framework

To evaluate the vision-based framework, we have acquired 15 video sequences with one human actor (woman of 33 years) (see section 6.3.2), and 14 video sequences with fourteen elderly volunteers (see section 6.3.3).
The duration of each video, with the human actor, is about 20 minutes and each video contains about 9600 frames (about eight frames per second).

States and events	GT	TP	FN	FP	P	S
In the kitchen	45	40	5	3	**93%**	88%
In the livingroom	35	32	3	5	86%	**91%**
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%**

Table 6.2: Results for recognition of a set of states and events by using video camera; Recognition of person location in the kitchen and in the livingroom. Recognition of the different human postures.

The duration of each video, with the elderly volunteers, is about 4 hours and each video contains about 144 000 frames (about ten frames per second).
Using only video cameras and video sequences with one human actor, we have tested some normal activities of a person such as: different human postures, different person location in the different zones in the laboratory, different person location versus the different equipments in the laboratory. We have also tested two abnormal activities: "fainting" and "falling down".

Results and Discussion

Table 6.2 summarizes the ground truth (GT), the true positive (TP), the false negative (FN), the false positive (FP), the precision (P) and the sensitivity (S) of the recognition of a set of primitive states and events (i.e. person location in the laboratory and the different human postures).

The primitive states "in the kitchen" and "in the livingroom" are well recognized by video cameras. The few errors in the recognition occur at the border between livingroom and kitchen. These errors are due to noise and shadow problems. The results of the recognition of the different postures show a sensitivity of 63-79% and a precision of 62-85% (see Table 6.2). When the system fails in the recognition of postures, it mixes postures such as bending and sitting (i.e. the bending posture instead the sitting one) due to segmentation errors (shadow, light change) and object occlusions.

We have also tested a set of human postures for 5 elderly volunteers. Figure 6.9 shows the recognition of "bending in the kitchen" activity for one elderly

volunteer (man of 64 years) among the 5 other, and the corresponding 3D visualization of this recognition.

Results of recognition of "slumping in the armchair" (for the human actor) is

Figure 6.9: (a) Recognition of "bending in the kitchen" activity and (b) the 3D visualization of this recognition.

shown in figure 6.10, the person is recognized with the posture "slumping" and "located in the livingroom". We have visualized the recognized events with the 3D visualization tool described in section 6.2.

In the 15 acquired videos with one actor, we have filmed one "falling down" event and two "fainting" events which have been correctly recognized. These abnormal activities have only tested with the human actor, the geriatrics have not accepted to test these abnormal activities with the elderly volunteers for fear they would hurt.

Figure 6.11 and figure 6.12 show respectively the recognition of "fainting" and "falling down" abnormal activities, and the 3D visualization of these recognition.

(a) Original image showing a person slumping in the armchair

(b) 3D visualization of the recognition of "slumping in the armchair" activity

Figure 6.10: Visualization of the recognition of "slumping in the armchair" activity in the Gerhome laboratory.

(a) The recognition of: (1) "standing", (2) "bending", and (3) "sitting on the floor with outstretched legs" postures

(b) The 3D visualization of the recognition of: (1) "standing", (2) "bending", and (3) "sitting on the floor with outstretched legs" postures

Figure 6.11: Recognition and the 3D visualization of the recognition of "fainting" situation.

(a) The recognition of: (1) "standing", (2) "sitting on the floor with flexed legs", and (3) "lying on the floor with outstretched legs" postures

(1) (2) (3)

(b) The recognition of: (1) "standing", (2) "sitting on the floor with flexed legs", and (3) "lying on the floor with outstretched legs" postures

Figure 6.12: Recognition and the 3D visualization of the recognition of "falling down" situation.

States and events	GT	TP	FN	FP	P	S
In the kitchen	45	25	20	3	89%	55%
Sitting	80	64	16	5	92%	80%

Table 6.3: Results for recognition of a set of states and events by using environmental sensors

6.4.2 Evaluation of the Sensor-Based Framework

To evaluate the proposed sensor model, we have used a ground truth of 20 video sequences of one human actor which contains data of environmental sensors. We calculate the a posteriori probability $P(\Theta = \theta|y)$ for each environmental sensor. θ represents the true value of the variable of interest Θ and $y = (y_1^T, y_2^T, ..., y_N^T)^T$ denotes the vector of N sensor measures.
The obtained results show that the used sensors (in the kitchen and in the livingroom) are working correctly at different rates: 95% for contact sensors including fridge and cupboard sensors, 90% for electrical sensors including microwave and stove sensors, 85% for water flow sensors including water pipes sensors, 70% for pressure sensors and 70% for presence sensors.

To evaluate the sensor-based framework, we have tested with one human actor (woman of 33 years) a set of human activities in Gerhome laboratory. We used the same data as in section 6.4.1, those with one human actor.
Table 6.3 summarizes the ground truth (GT), the true positive (TP), the false negative (FN), the false positive (FP), the precision (P) and the sensitivity (S) of the recognition of a set of primitive states and events (i.e. person location in the kitchen and sitting posture) by using the environmental sensors (i.e. presence sensors installed near sink and near stove in the kitchen, pressure sensors installed under chair and armchair).

Results and Discussion

The obtained results show that the primitive state "in the kitchen" is not well recognized by environmental sensors. This is due to the fact that the presence sensor detects the presence of a person in the kitchen only when this person is near stove or is near a sink. This is also due to the fact that the presence sensor is activated when it detects variations of the illumination (e.g. natural illumination like the sun).
The few errors in the recognition of the sitting posture is due to sensor failures and also to the fact that the pressure sensor is active when a person puts his bags on a chair.

Comparison between the obtained results using the environmental sensors (see table 6.3) with the obtained results using the video camera (see table 6.2)

Multimodal events	GT	TP	FN	FP	P	S
Using fridge	13	11	2	3	78%	84%
Using stove	40	35	5	2	**94%**	87%
Sitting on a Chair	12	9	3	4	69%	75%
Sitting in an Armchair	2	1	1	1	**50%**	**50%**

Table 6.4: Results for recognition of a set of multimodal events of one volunteer among the 5 volunteers with ground truth

shows:
- the primitive state "in the kitchen" has been better recognized by the video sensor than by environmental sensor (sensitivity 88% vs. 55).
- the human posture "sitting" has been better recognized by the environmental sensor than by video sensor (precision 92% vs. 72).

6.4.3 Evaluation of the Multisensor-Based Fusion Framework

To evaluate the multisensor-based fusion framework, we have used the same video sequences as in section 6.4.1.
Using both video cameras and environmental sensors, we have tested a set of the daily activities of a person such as: using kitchen equipment, using TV, preparing meal, and taking meal. We have also compared two behavioral profiles of two elderly volunteers (see section 6.4.3.2). We have done this comparison in order to bring out the possible differences in the behaviors of the two volunteers.

In the next section, we present firstly the obtained results by using multisensor data without taking into account the uncertainties of sensors. After that we present the obtained results by using multisensor data with uncertainty in sensor measurements.

6.4.3.1 Results of Recognition

Results of recognition of "using microwave" is shown in figure 6.13, the person is recognized with the posture "standing with one arm up", "located in the kitchen" and "opening the microwave". We have visualized the recognized events with the 3D visualization tool described in section 6.2.

In the experiment with the fourteen volunteers, among all analyzed data, results for one volunteer (person P2) observed during 4 hours are shown in table 6.4. This table summarizes the ground truth (GT), the true positive (TP), the false negative (FN), the false positive (FP), the precision (P) and the sensitivity (S) of the recognition of a set of multimodal events.

Table 6.5 and table 6.6 show the obtained results for 4 volunteers (persons P1, P3, P4 and P9) also observed during 4 hours. These tables summarize the

(a) Original image showing a person using a microwave

(b) 3D visualization of the recognition of "use microwave" activity

Figure 6.13: Visualization of the recognized events in the Gerhome laboratory.

ground truth (GT), the true positive (TP), the false negative (FN), the false positive (FP), the precision (P) and the sensitivity (S) of the recognition of a set of multimodal events.

	P1						P3					
Multimodal events	GT	TP	FN	FP	P	S	GT	TP	FN	FP	P	S
Using Fridge	24	18	6	3	85%	75%	10	7	8	2	87%	70%
Using Stove	18	15	3	4	78%	83%	9	5	4	2	71%	55%
Sitting on a Chair	29	25	4	7	78%	86%	6	3	3	2	60%	**50%**
Sitting on an Armchair	8	6	2	5	**54%**	75%	15	12	2	6	66%	80%

Table 6.5: Results of recognition of a set of daily activities for 2 observed elderly persons

	P4						P9					
Multimodal events	GT	TP	FN	FP	P(%)	S(%)	GT	TP	FN	FP	P(%)	S(%)
Use Fridge	12	9	3	3	75%	75%	6	5	1	1	83%	83%
Use Stove	2	1	1	1	**100%**	50%	108	106	2	10	84%	**98%**
Sitting on a Chair	13	9	4	5	**64%**	69%	6	4	2	3	57%	66%
Sitting on an Armchair	17	15	2	3	83%	88%	14	12	2	5	70%	85%

Table 6.6: Results of recognition of a set of daily activities for 2 observed elderly persons

The recognition of the multimodal events showed:

- a sensitivity of 75-86% and a precision of 54-85% (Table 6.5) for the volunteer P1 (man, 71 years).
- a sensitivity of 50-87% and a precision of 50-91% (Table 6.4) for the volunteer P2 (man, 64 years).
- a sensitivity of 50-80% and a precision of 60-87% (Table 6.5) for the volunteer P3 (man, 66 years).
- a sensitivity of 50-88% and a precision of 64-100% (Table 6.6) for the volunteer P4 (man, 68 years).
- a sensitivity of 66-98% and a precision of 57-84% (Table 6.6) for the volunteer P9 (woman, 85 years).

The multimodal events are well recognized, the errors in the recognition are due to the sensor measurement errors (e.g. contact sensor is still active when a person close the fridge, or when a person does not correctly closed the drawer (see figure 6.14) or another kitchen equipment) and to the fact that the person which drops his bag on the chair or on the armchair, may activate the chair (or armchair) sensor (sensor installed under the chair or under armchair) and gives a false result.

Figure 6.15 shows the recognition of "preparing a meal" activities for the volunteer P2, and the corresponding 3D visualization of this recognition.

Figure 6.16 shows the recognition of "taking a meal" activity for the volunteer P9, and the corresponding 3D visualization of this recognition.

The obtained results demonstrate that the proposed method allows to detect and recognize a set of activities of a person by using the data provided by the combination of the selected sensors.

6.4.3.2 Results on Behavior for 2 Elderly Volunteers

Results comparing volunteer P2 (man of 64 years) and volunteer P9 (woman of 85 years), observed during 4 hours are shown in table 6.7 and table 6.8. These tables summarize the mean duration, the total duration and the number of instances of each monitored activity. Time unit is hh:mm:ss.

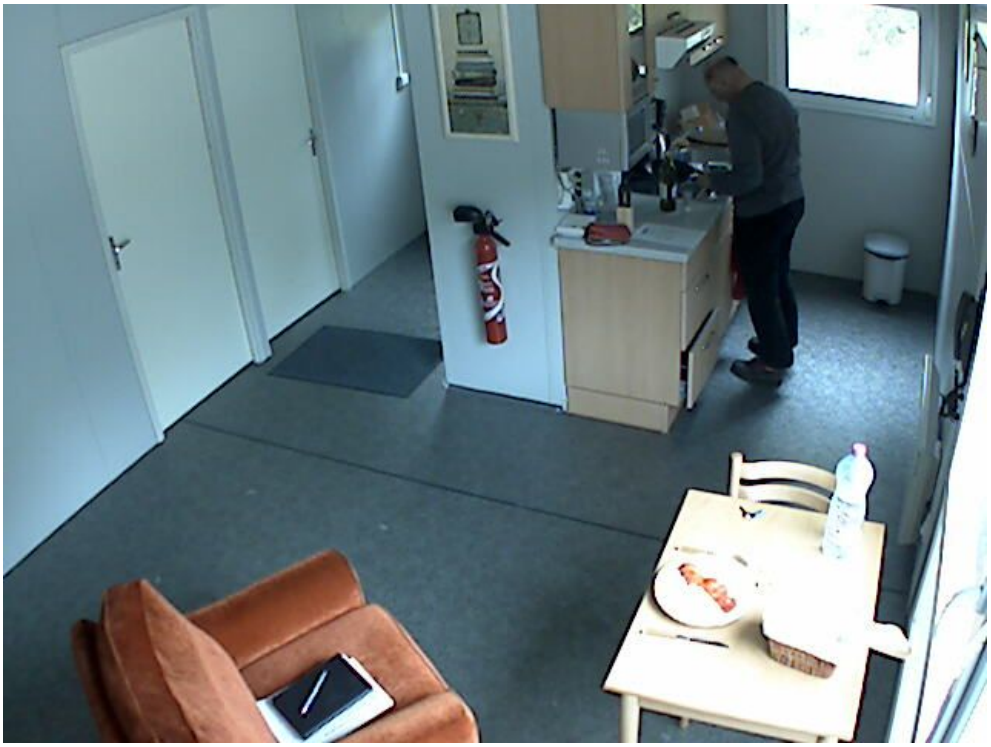
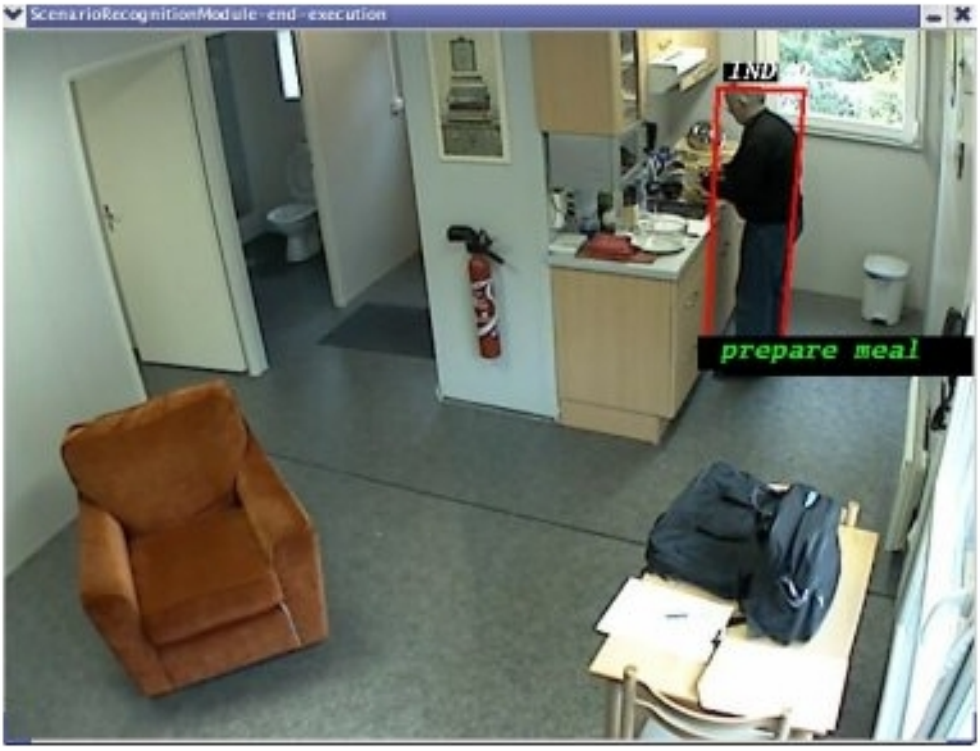


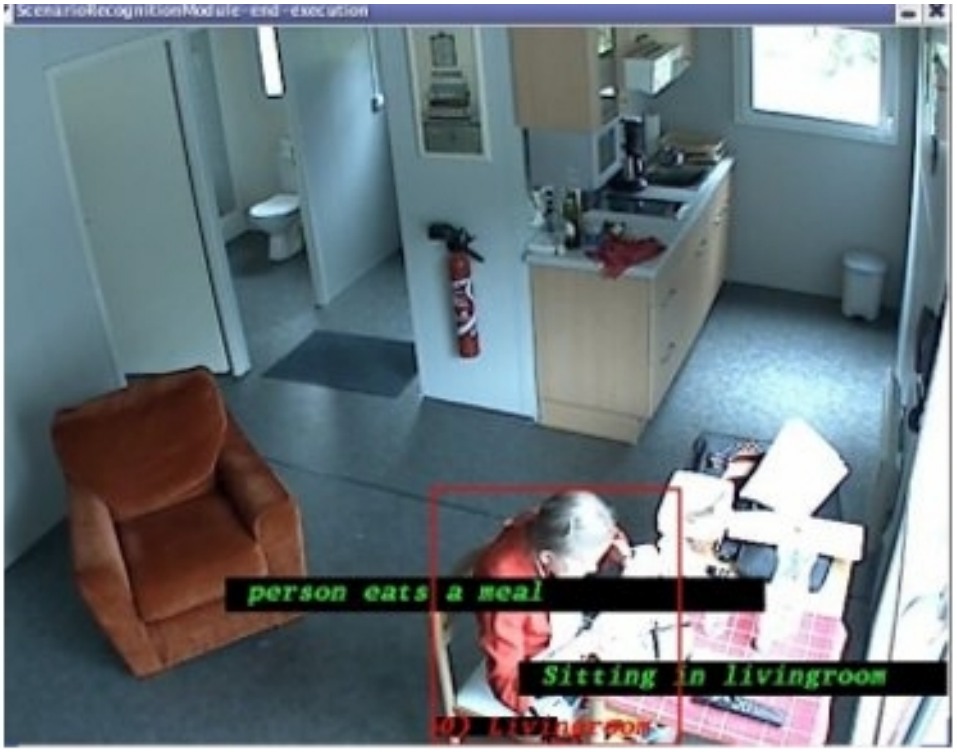
Figure 6.14: The drawer is still open when the person does not use it



(a) The recognition of: "prepare meal" activity

(b) The 3D visualization of the recognition of "prepare meal" activity

Figure 6.15: The recognition and the 3D visualization of the recognition of "preparing a meal" activity.



(a) The recognition of: "sitting in the livingroom" and of "person eats a meal" activities

(b) The 3D visualization of the recognition of "sitting in the livingroom" activity

Figure 6.16: The recognition and the 3D visualization of the recognition of "taking a meal" activity.

Volunteers	Volunteer P2 (64 years)			Volunteer P9 (85 years)			NDA and NDI	
Activity	Mean Duration (hh:mm:ss)	Total Duration (hh:mm:ss)	Number Instance	Mean Duration (hh:mm:ss)	Total Duration (hh:mm:ss)	Number Instance	NDA (%)	NDI (%)
Using Fridge	00:00:10	00:01:45	11	00:00:14	00:01:09	5	17	37
Using Stove	00:00:08	00:04:52	35	00:00:17	00:29:13	106	36	50
Using Kitchen Hot-Water	00:00:20	00:12:36	38	00:00:10	00:03:50	24	33	22
Using Kitchen Cold-Water	00:00:29	00:09:36	20	00:00:03	00:01:15	23	81	7
Using the Upper Cupboard	00:00:57	00:21:01	22	00:04:43	00:42:24	9	65	42
Using the Lower Cupboard	00:00:21	00:03:09	9	00:01:51	00:07:23	4	68	38
Using the Middle Cupboard	00:00:10	00:01:51	11	00:00:03	00:00:15	5	54	37
Sitting on a Chair	00:06:27	01:36:43	15	00:52:37	03:30:29	4	78	58
Sitting on armchair	00:12:04	00:24:09	2	00:00:29	00:05:46	12	92	71
Using the Toilet	00:00:56	00:00:56	1	00:00:56	00:00:56	1	0	0

Table 6.7: Monitored activities, their frequencies (n1 and n2), mean (m1 and m2) and total duration of 2 volunteers staying in the GERHOME laboratory for 4 hours; NDA =Normalized Difference of mean durations of Activities= $|mean1 - mean2|/(mean1 + mean2)$; NDI =Normalized Difference of Instances number= $|n1 - n2|/(n1 + n2)$. Time unit is hh:mm:ss

Volunteers	Volunteer P2 (64 years)			Volunteer P9 (85 years)			NDA and NDI	
Activity	Mean Duration (hh:mm:ss)	Total Duration (hh:mm:ss)	Number Instance	Mean Duration (hh:mm:ss)	Total Duration (hh:mm:ss)	Number Instance	NDA (%)	NDI (%)
Using TV	00:33:59	02:49:53	5	00:13:53	01:09:24	5	42	0
Using the Bathroom Cupboard	00:00:04	00:00:07	2	00:00:04	00:00:07	2	0	0
Using the Bathroom Hot-Water	00:00:41	00:01:22	2	00:00:12	00:00:36	3	55	20
Using the Bathroom Cold-Water	00:00:12	00:00:24	2	00:00:07	00:00:07	1	26	33
Entering in the Kitchen	00:02:24	00:12:00	5	00:02:40	00:08:00	3	5	25
Entering in the Livingroom	00:01:15	00:25:00	20	00:02:42	00:35:00	13	37	21
Entering in the Entrance	00:02:00	00:08:00	4	00:01:51	00:11:07	6	4	20
Entering in the Bedroom	00:01:15	00:05:00	4	00:03:17	00:23:00	7	45	27
Entering in the Bathroom	00:01:00	00:02:00	2	00:01:00	00:05:00	5	0	43
Standing	00:00:09	00:30:07	200	00:00:16	00:12:00	45	28	63
Bending	00:00:03	00:01:20	30	00:00:12	00:03:00	15	60	33

Table 6.8: Monitored activities, their frequencies (n1 and n2), mean (m1 and m2) and total duration of 2 volunteers staying in the GERHOME laboratory for 4 hours; NDA =Normalized Difference of mean durations of Activities= $|mean1 - mean2|/(mean1 + mean2)$; NDI =Normalized Difference of Instances number= $|n1 - n2|/(n1 + n2)$. Time unit is hh:mm:ss

.

Among the 21 activities for which the 2 older volunteers were compared (see Table 6.7 and Table 6.8) 10 activities show differences. Five activities among the 10 activities are considered meaningful and discriminative.

- Volunteer P2 of 64 years changed zones more often than the volunteer P9 of 85 years (for "entering livingroom" 20 vs. 13), and did this at a quicker pace (00:01:15 vs. 00:02:42), showing a greater ability to walk.

- Volunteer P2 was more often seen "sitting on chair" (15 vs. 4, NDI=58%), but volunteer P9 was "sitting on chair" for a longer duration (00:52:37 vs. 00:06:27, NDA=78%), showing also a greater ability for the volunteer P2 to move in the apartment.

- Similarly volunteer P2 was "bending" twice as much as volunteer P9 (30 vs. 15, NDI=33%), and in a quicker way (00:00:03 vs. 00:00:12, NDA=60%), showing greater dynamism for the younger volunteer.

- Volunteer P2 was using more the "upper cupboard" than the volunteer P9 (22 vs. 9, NDI=42%), and in a quicker way (00:00:57 vs. 00:04:43, NDA=65%).

- Volunteer P2 was more able to using the stove (less trials for "using stove" 35 vs. 106, NDI=50%).

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

6.4.3.3 Results of the Recognition using DS Uncertainty

Using the Dempster-Shafer theory like described in chapter 5 in section 5.8, we have calculate the uncertainty in sensor measurements of 4 activities for volunteer P1 and volunteer P3.

The obtained results are shown in table 6.9.

6.4.3.4 Discussion

Comparison between the results obtained without using uncertainty (see table 6.6) and the results obtained with using uncertainty (see table 6.9) shows some improvements in the recognition of activities. For example, the new results (using uncertainty in sensor measurements) show a good recognition, compared to the results obtained without using uncertainty in sensor measurements, of the "sitting on a chair" (for volunteer P1: P= 93% vs. 78% and S= 93% vs. 86%, for volunteer P3: P= 83% vs. 60% and S= 83% vs. 50%) and of "sitting on an armchair" (for volunteer P1: P= 77% vs. 54% and S= 87% vs. 75%, for volunteer P3: P= 93% vs. 66% and S= 93% vs. 80%) activities.

	P1						P3					
Multimodal events	GT	TP	FN	FP	P	S	GT	TP	FN	FP	P	S
Using Fridge	24	21	3	1	95%	87%	10	8	2	1	88%	80%
Using Stove	18	16	2	1	94%	88%	9	7	1	1	87%	77%
Sitting on a Chair	29	27	2	2	93%	93%	6	5	1	1	83%	83%
Sitting in an Armchair	8	7	1	2	77%	87%	15	14	2	1	93%	93%

Table 6.9: Results of recognition (using uncertainty) of a set of daily activities for 2 observed elderly persons

6.5 Medical Evaluation

In this section, we compare the behavioral profile for 9 observed elderly volunteers using results of 6 daily activities.

6.5.1 Events Durations for 9 Elderly Persons

Table 6.10 summarizes the duration of 6 daily activities for the 9 observed elderly persons.

Event (Ei)	**Duration of each event for the 9 observed Persons**								
	$d_{Ei,P1}$	$d_{Ei,P2}$	$d_{Ei,P3}$	$d_{Ei,P4}$	$d_{Ei,P5}$	$d_{Ei,P6}$	$d_{Ei,P7}$	$d_{Ei,P8}$	$d_{Ei,P9}$
Using Fridge	00:03:27	00:01:45	00:00:40	00:05:47	00:01:25	00:01:04	00:02:19	00:01:14	00:01:09
Using Stove	00:03:41	00:04:52	00:02:27	00:00:08	00:05:49	00:03:31	00:04:14	00:02:34	00:29:13
Sitting on a Chair	01:58:00	01:36:43	01:08:48	00:58:51	00:50:39	00:02:40	00:26:25	00:18:07	03:30:29
Sitting on an Armchair	00:12:57	00:24:09	00:52:58	01:50:19	00:57:14	01:48:35	00:16:07	00:12:27	00:05:46
Using TV	01:31:56	02:49:53	02:25:55	02:47:44	02:12:08	00:05:12	02:18:44	01:47:34	01:09:24
Using Upper Cupboard	00:03:40	00:06:14	00:00:43	00:01:04	00:23:15	00:03:02	00:17:31	00:02:22	00:00:15

Table 6.10: Durations of six daily activities for the 9 observed elderly persons. Time unit is hh:mm:ss

Figure 6.17 shows the duration of each activity for the 9 elderly persons.

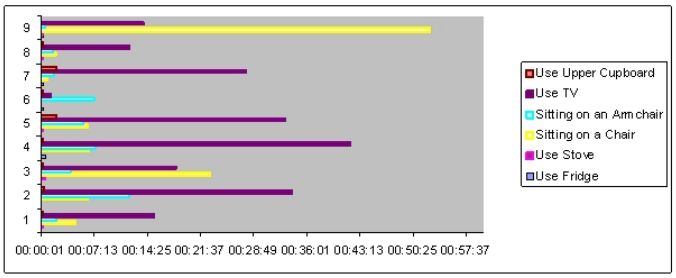


Figure 6.17: Duration of each activity for the 9 observed elderly persons

6.5.2 Leave-One-Out Cross Validation

In this section, for each observed elderly person among the 9 observed elderly persons, we have done the leave-one-out cross validation on the activity duration for the 9 observed old persons. When using the leave-one-out method, the learning algorithm is trained multiple times, using all but one of the training set of data. The form of the algorithm is as follows:

To do that, we calculate firstly the mean duration of each activity for R persons

Algorithm 4 $Leave-One-Out-Cross-ValidationAlgorithm$

For $K = 1$ to R (where R is the number of training set of data)
- Temporarily **remove** the j^{th} data from the training set
- **Train** the learning algorithm on the remaining R data
- **Test** the removed data

Calculate the mean error over all R data

among the all observed persons by using the following equation:

$$MD_{Ei,Pj} = \frac{\sum^{Pk \in P, Pk \neq Pj} d_{Ei,Pk}}{R}, \quad \forall Pj \in P \tag{6.5}$$

Where:

- $MD_{Ei,Pj}$ represents the mean duration for a given event Ei for each person without a person Pj;
- $d_{Ei,Pk}$ represents the duration for each event Ei for each person Pk;
- $P = \{P1, P2, P3, P4, P4, P5, P6, P7, P8, P9\}$;
- R represents the number of the training set of data (i.e. R=8 in this case).

Table 6.11 summarizes the mean durations of 6 activities.

Activity	$MD_{Ei,P1}$	$MD_{Ei,P2}$	$MD_{Ei,P3}$	$MD_{Ei,P4}$	$MD_{Ei,P5}$	$MD_{Ei,P6}$	$MD_{Ei,P7}$	$MD_{Ei,P8}$	$MD_{Ei,P9}$
Using Fridge (E1)	00:01:55	00:02:08	00:02:16	**00:01:38**	00:02:11	00:02:13	00:02:04	00:02:12	00:02:13
Using Stove (E2)	00:06:36	00:06:27	00:06:45	00:07:03	00:06:20	00:06:37	00:06:32	00:06:44	**00:03:24**
Sitting on a Chair (E3)	01:06:35	01:09:15	01:12:44	01:13:59	01:15:00	01:21:00	01:18:02	01:19:04	00:55:02
Sitting on an Armchair (E4)	00:48:27	00:47:03	00:43:27	00:36:17	00:42:55	00:36:30	00:48:03	00:48:31	00:49:21
Using TV (E5)	01:57:04	01:47:20	01:50:19	01:47:36	01:52:03	**02:07:55**	01:51:13	01:55:07	01:59:53
Using Upper Cupboard (E6)	00:06:48	00:06:29	00:07:10	00:07:08	**00:04:21**	00:06:53	00:05:04	00:06:58	00:07:14

Table 6.11: The mean duration of each activity. $MD_{Ei,P1}$ represents the mean duration using the data for 8 persons (i.e. by removing data from the person P1), and so on

Secondly, we calculate the standard deviation $\sigma_{Ei,Pj}$ (see table 6.12) for each event Ei for each person without a person Pj by using the following equation:

$$\begin{aligned}\sigma_{Ei,Pj} &= \sqrt{\frac{1}{R}\sum_{k=1}^{R}(d_{Ei,Pk} - MD_{Ei,Pj})^2} \\ &= \sqrt{\frac{1}{R}\sum_{k=1}^{R} d_{Ei,Pk}^2 - MD_{Ei,Pj}^2}\end{aligned} \tag{6.6}$$

Where:

- $\sigma_{Ei,Pj}$ represents the standart deviation for each event Ei for each person without a person Pj;
- $d_{Ei,Pk}$ represents the duration of an event Ei for a person Pk;
- $MD_{Ei,Pj}$ represents the mean duration for each event Ei for each person without a person Pj;
- R represents the number of the training set of data (i.e. R=8 in this case).

Removed person	$\sigma_{E1,Pj}$	$\sigma_{E2,Pj}$	$\sigma_{E3,Pj}$	$\sigma_{E4,Pj}$	$\sigma_{E5,Pj}$	$\sigma_{E6,Pj}$
P1	00:01:32	00:08:42	01:01:13	00:39:12	00:52:09	00:08:09
P2	00:01:36	00:08:44	01:02:55	00:40:14	00:48:32	00:08:13
P3	00:01:31	00:08:38	01:03:34	00:40:50	00:51:29	00:07:56
P4	**00:00:50**	00:08:27	01:03:22	00:32:41	00:48:52	00:07:58
P5	00:01:36	00:08:45	01:03:03	00:40:40	00:52:24	**00:05:17**
P6	00:01:34	00:08:42	00:57:58	00:33:09	**00:33:26**	00:08:07
P7	00:01:37	00:08:43	01:01:12	00:39:33	00:52:01	00:07:06
P8	00:01:35	00:08:39	01:00:14	00:39:09	00:52:46	00:08:05
P9	00:01:34	00:01:37	**00:36:50**	00:38:17	00:50:04	00:07:53

Table 6.12: Standard deviations $\sigma_{Ei,Pj}$ of each event Ei for each person Pj

Thirdly, we calculate for each person and for each event the interval $I_{Ei,Pj} = [MD_{Ei,Pj} - \sigma_{Ei,Pj}; MD_{Ei,Pj} + \sigma_{Ei,Pj}]$

We have obtained 54 intervals as described in tables 6.13, 6.14.

Activity	**Intervals; Time unit is hh:mm:ss**				
	$I_{Ei,P1}$	$I_{Ei,P2}$	$I_{Ei,P3}$	$I_{Ei,P4}$	$I_{Ei,P5}$
Fridge	[00:00:23;00:03:27]	[00:00:32;00:03:44]	[00:00:45;00:03:47]	[00:00:48;00:02:28]	[00:00:35;00:03:47]
Stove	[00:00:00;00:15:18]	[00:00:00;00:15:11]	[00:00:00;00:15:23]	[00:00:00;00:15:30]	[00:00:00;00:15:05]
Chair	[00:05:22;02:07:48]	[00:06:20;02:12:10]	[00:09:10;02:16:18]	[00:10:37;02:17:21]	[00:11:57;02:18:03]
Armchair	[00:09:15;01:27:39]	[00:06:49;01:27:17]	[00:02:37;01:24:17]	[00:03:36;01:08:58]	[00:02:15;01:23:35]
TV	[01:01:55;02:46:13]	[00:58:48;02:35:52]	[00:58:50;02:41:48]	[00:58:44;02:36:28]	[00:59:39;02:44:27]
UpperCupboard	[00:00:00;00:14:57]	[00:00:00;00:14:42]	[00:00:00;00:15:06]	[00:00:00;00:15:06]	[00:00:00;00:09:38]

Table 6.13: Intervals $I_{Ei,Pj} = [MD_{Ei,Pj} - \sigma_{Ei,Pj}; MD_{Ei,Pj} + \sigma_{Ei,Pj}]$

Activity	**Intervals; Time unit is hh:mm:ss**			
	$I_{Ei,P6}$	$I_{Ei,P7}$	$I_{Ei,P8}$	$I_{Ei,P9}$
Fridge	[00:00:39;00:03:47]	[00:00:27;00:03:41]	[00:00:37;00:03:47]	[00:00:39;00:03:47]
Stove	[00:00:00;00:15:19]	[00:00:00;00:15:15]	[00:00:00;00:15:23]	[00:01:47;00:05:01]
Chair	[00:23:02;02:18:58]	[00:16:50;02:19:14]	[00:18:50;02:19:18]	[00:18:12;01:31:52]
Armchair	[00:03:21;01:09:39]	[00:08:30;01:27:36]	[00:09:22;01:27:40]	[00:11:04;01:27:38]
TV	[01:34:29;02:41:21]	[00:59:12;02:43:14]	[01:02:21;02:47:53]	[01:09:49;02:49:57]
UpperCupboard	[00:00:00;00:15:00]	[00:00:00;00:12:10]	[00:00:00;00:15:03]	[00:00:00;00:15:07]

Table 6.14: Intervals $I_{Ei,Pj} = [MD_{Ei,Pj} - \sigma_{Ei,Pj}; MD_{Ei,Pj} + \sigma_{Ei,Pj}]$

And finally, we validate each activity by comparing the duration of each activity for each person with the corresponding interval.
For example, for the person $P1$, to validate an event $E1$ which represents the event "Using Fridge", we compare his duration value to the interval $I_{E1,P1} = [MD_{E1,P1} - \sigma_{E1,P1}; MD_{E1,P1} + \sigma_{E1,P1}]$.

- Table 6.15 shows the validation of the event "Using Fridge" of the 9 observed persons, by comparing the value of the duration $d_{E1,Pj}$, with the interval $I_{E1,Pj}$.
  If this value belongs to interval $I_{E1,Pj}$, then the person has a normal behavior compared to the average. If not, then the person has a deviated (i.e. different) behavior compared to the average.
  This table shows that the person P4 has used a fridge for long time compared to the others, and person P3 has used a fridge for small time compared to the others. We can deduce that the person P4 is more slowly than person P3.

"Using Fridge" durations compared to interval $I_{E1,Pj}$	**Person Profile on "Using Fridge"**
$d_{E1,P1} = 00:03:27 \in I_{E1,P1}$	A person P1 has a normal profile
$d_{E1,P2} = 00:01:45 \in I_{E1,P2}$	A person P2 has a normal profile
$d_{E1,P3} = 00:00:40 \notin I_{E1,P3}$	A person P3 has a different profile
$d_{E1,P4} = 00:05:47 \notin I_{E1,P4}$	A person P4 has a different profile
$d_{E1,P5} = 00:01:25 \in I_{E1,P5}$	A person P5 has a normal profile
$d_{E1,P6} = 00:01:04 \in I_{E1,P6}$	A person P6 has a normal profile
$d_{E1,P7} = 00:02:19 \in I_{E1,P7}$	A person P7 has a normal profile
$d_{E1,P8} = 00:01:14 \in I_{E1,P8}$	A person P8 has a normal profile
$d_{E1,P9} = 00:01:09 \in I_{E1,P9}$	A person P9 has a normal profile

Table 6.15: Comparison between the duration of the event "Using Fridge" and the interval $I_{E1,Pj}$

- Table 6.16 shows the validation of the event "Using Stove" of the 9 observed persons, by comparing the value of the duration $d_{E2,Pj}$, with the interval $I_{E2,Pj}$.
  If this value belongs to interval $I_{E2,Pj}$, then the person has a normal behavior compared to the average. If not, then the person has a deviated (i.e. different) behavior compared to the average.

- Table 6.17 shows the validation of the event "Sitting on a Chair" of the 9 observed persons, by comparing the value of the duration $d_{E3,Pj}$, with the interval $I_{E3,Pj}$.
  If this value belongs to interval $I_{E3,Pj}$, then the person has a normal behavior compared to the average. If not, then the person has a deviated (i.e.

"Using Stove" durations compared to interval $I_{E2,Pj}$	**Person Profile on "Using Stove"**
$d_{E2,1} = 00:03:41 \in I_{E2,P1}$	A person P1 has a normal profile
$d_{E2,2} = 00:04:52 \in I_{E2,P2}$	A person P2 has a normal profile
$d_{E2,3} = 00:02:27 \in I_{E2,P3}$	A person P3 has a normal profile
$d_{E2,4} = 00:00:08 \in I_{E2,P4}$	A person P4 has a normal profile
$d_{E2,5} = 00:05:49 \in I_{E2,P5}$	A person P5 has a normal profile
$d_{E2,6} = 00:03:31 \in I_{E2,P6}$	A person P6 has a normal profile
$d_{E2,7} = 00:04:14 \in I_{E2,P7}$	A person P7 has a normal profile
$d_{E2,8} = 00:02:34 \in I_{E2,P8}$	A person P8 has a normal profile
$d_{E2,9} = 00:29:13 \notin I_{E2,P9}$	A person P9 has a different profile

Table 6.16: Comparison between the duration of the event "Using Stove" and the interval $I_{E2,Pj}$

different) behavior compared to the average.

This table shows that a person P9 was "sitting on a chair" for a longer

"Sitting on a Chair" durations compared to interval $I_{E3,Pj}$	**Person Profile on "Sitting on a Chair"**
$d_{E3,1} = 01:58:00 \in I_{E3,P1}$	A person P1 has a normal profile
$d_{E3,2} = 01:36:43 \in I_{E3,P2}$	A person P2 has a normal profile
$d_{E3,3} = 01:08:48 \in I_{E3,P3}$	A person P3 has a normal profile
$d_{E3,4} = 00:58:51 \in I_{E3,P4}$	A person P4 has a normal profile
$d_{E3,5} = 00:50:39 \in I_{E3,P5}$	A person P5 has a normal profile
$d_{E3,6} = 00:02:40 \notin I_{E3,P6}$	A person P6 has a different profile
$d_{E3,7} = 00:26:25 \in I_{E3,P7}$	A person P7 has a normal profile
$d_{E3,8} = 00:18:07 \in I_{E3,P8}$	A person P8 has a normal profile
$d_{E3,9} = 03:30:29 \notin I_{E3,P9}$	A person P9 has a different profile

Table 6.17: Comparison between the duration of the event "Sitting on a Chair" and the interval $I_{E3,Pj}$

duration than the others, and person P6 was "sitting on a chair" for a short time. Using these results, we can deduce that a person P6 is more able to the person P9 to move in the apartment.

- Table 6.18 shows the validation of the event "Sitting on an Armchair" of the 9 observed persons, by comparing the value of the duration $d_{E4,Pj}$, with the interval $I_{E4,Pj}$.
  If this value belongs to interval $I_{E4,Pj}$, then the person has a normal behavior compared to the average. If not, then the person has a deviated (i.e. different) behavior compared to the average.
  This table shows that persons P4 and P6 were "sitting on a chair" for a longer duration than the others, and person P9 was "sitting on a chair" for a short time. Using these results, we can deduce that a person P6 prefers

"Sitting on an Armchair" durations compared to interval $I_{E4,Pj}$	**Person Profile on "Sitting on an Armchair"**
$d_{E4,1} = 00:12:57 \in I_{E4,P1}$	A person P1 has a normal profile
$d_{E4,2} = 00:24:09 \in I_{E4,P2}$	A person P2 has a normal profile
$d_{E4,3} = 00:52:58 \in I_{E4,P3}$	A person P3 has a normal profile
$d_{E4,4} = 01:50:19 \notin I_{E4,P4}$	A person P4 has a different profile
$d_{E4,5} = 00:57:14 \in I_{E4,P5}$	A person P5 has a normal profile
$d_{E4,6} = 01:48:35 \notin I_{E4,P6}$	A person P6 has a different profile
$d_{E4,7} = 00:16:07 \in I_{E4,P7}$	A person P7 has a normal profile
$d_{E4,8} = 00:12:27 \in I_{E4,P8}$	A person P8 has a normal profile
$d_{E4,9} = 00:05:46 \notin I_{E4,P9}$	A person P9 has a different profile

Table 6.18: Comparison between the duration of the event "Sitting on an Armchair" and the interval $I_{E4,Pj}$

to sit in an armchair instead the chair.

- Table 6.19 shows the validation of the event "Using TV" of the 9 observed persons, by comparing the value of the duration $d_{E5,Pj}$, with the interval $I_{E5,Pj}$.
  If this value belongs to interval $I_{E5,Pj}$, then the person has a normal behavior compared to the average. If not, then the person has a deviated (i.e. different) behavior compared to the average.
  This table shows that person P6 has used a TV for a short time (e.g.

"Using TV" durations compared to interval $I_{E5,Pj}$	**Person Profile on "Using TV"**
$d_{E5,1} = 01:31:56 \in I_{E5,P1}$	A person P1 has a normal profile
$d_{E5,2} = 02:49:53 \notin I_{E5,P2}$	A person P2 has a different profile
$d_{E5,3} = 02:25:55 \in I_{E5,P3}$	A person P3 has a normal profile
$d_{E5,4} = 02:47:44 \notin I_{E5,P4}$	A person P4 has a different profile
$d_{E5,5} = 02:12:08 \in I_{E5,P5}$	A person P5 has a normal profile
$d_{E5,6} = 00:05:12 \notin I_{E5,P6}$	A person P6 has a different profile
$d_{E5,7} = 02:18:44 \in I_{E5,P7}$	A person P7 has a normal profile
$d_{E5,8} = 01:47:34 \in I_{E5,P8}$	A person P8 has a normal profile
$d_{E5,9} = 01:09:24 \notin I_{E5,P9}$	A person P9 has a different profile

Table 6.19: Comparison between the duration of the event "Using TV" and the interval $I_{E5,Pj}$

00:05:12 for P6 vs. 02:25:55 for P3) than the others, and person P2 has used a TV for longer duration (e.g. 02:49:53 for P2 vs. 01:31:56 for P1).

- Table 6.20 shows the validation of the event "Using Upper Cupboard" of the 9 observed persons, by comparing the value of the duration $d_{E6,Pj}$, with the interval $I_{E6,Pj}$.

If this value belongs to interval $I_{E6,Pj}$, then the person has a normal behavior compared to the average. If not, then the person has a deviated (i.e. different) behavior compared to the average.

This table shows that person P9 has used uppercupboard for a very short

"Using Upper Cupboard" durations compared to interval $I_{E6,Pj}$	**Person Profile on "Using Upper Cupboard"**
$d_{E6,1} = 00:03:40 \in I_{E6,P1}$	A person P1 has a normal profile
$d_{E6,2} = 00:06:14 \in I_{E6,P2}$	A person P2 has a normal profile
$d_{E6,3} = 00:00:43 \in I_{E6,P3}$	A person P3 has a normal profile
$d_{E6,4} = 00:01:04 \in I_{E6,P4}$	A person P4 has a normal profile
$d_{E6,5} = 00:23:15 \notin I_{E6,P5}$	A person P5 has a different profile
$d_{E6,6} = 00:03:02 \in I_{E6,P6}$	A person P6 has a normal profile
$d_{E6,7} = 00:17:31 \notin I_{E6,P7}$	A person P7 has a different profile
$d_{E6,8} = 00:02:22 \in I_{E6,P8}$	A person P8 has a normal profile
$d_{E6,9} = 00:00:15 \in I_{E6,P9}$	A person P9 has a normal profile

Table 6.20: Comparison between the duration of the event "Using Upper Cupboard" and the interval $I_{E6,Pj}$

time (e.g. 00:00:15 for P9 vs. 00:06:14 for P1) than the others, and person P5 and P7 has used uppercupboard for longer duration (e.g. 00:23:15 for P5 and 00:17:31 for P7).

6.5.3 Discussion

The main deductions of all the obtained results show that:

- The person P9 (woman of 85 years) has a fairly different profile from the others. This person shows some inabilities in using kitchen equipment (e.g. on using stove) and also shows some difficulties to move in the laboratory (e.g. sitting on a chair for a long duration), which may be the first sign of the frailty of this person.
- The person P7 (woman of 66 years) and the person P5 (woman of 69 years) show different profile in using uppercupboard. After viewing the videos, we found that these persons have forgotten to close the uppercupboard. It could be due to the fact that these persons start a new activity and they forgot to finish the original activity (challenge of false starts introduced in section 3.1.2 in chapter 3).
- The person P6 (woman of 70 years) shows different profile in using TV, in sitting on a chair and in sitting in an armchair. After viewing the videos, we found that this person had difficulties in turning on the TV. This may be due to the fact that this person does not have TV in her own home or has difficulties using the remote control. About sitting on a chair and sitting in an armchair it is due to the sensor failures.

- The person P5 (woman of 69 years) shows different profile in using uppercupboard. After viewing the videos, we found that this person has forgotten to close the uppercupboard. It could be due to the fact that this person starts a new activity and she forgot to finish the original activity (challenge of false starts introduced in section 3.1.2 in chapter 3).
- The person P4 (man of 66 years) shows different profile in using TV, using fridge and sitting in an armchair. After viewing the videos, we found that this person was behaving weirdly but we do not know why. We may deduce that this person was not motivated to do the experiments.

6.6 Conclusion

Several tests containing a large number of complex activities, including a test lasting over two weeks have been realized. The obtained results show that the proposed approach allows to recognize reliably with a low false alarm rate a set of interesting activities at home from multisensor data (i.e. video and environmental).

In the next chapter, we conclude our work and we propose some future work to improve the activity recognition framework.

Chapter 7

Conclusion and Future Work

In this thesis we have proposed a new approach for monitoring human activities at home. This approach includes an algorithm for real-time (video rate) recognition of primitive and composite activities that have occurred in the scene observed by video cameras and sensors attached to house furnishings. The proposed approach is based on combining video events with environmental events to recognize human activities. The proposed approach consists in detecting people, tracking people as they move, and recognizing activities of interest based on multisensor analysis and human activity recognition.
The proposed approach takes as input the data provided by the different sensors and exploits three major sources of knowledge: the 3D model of the scene (i.e. an apartment), the 3D model of mobile objects (e.g. person), and the models of activities.

An overview of the contributions of this work is described in the next section. Then a discussion is made to show the limitations of the proposed approach. Finally, future works are proposed in section 7.3 to improve the proposed approach.

7.1 Overview of the Contributions

In this section we describe an overview of our contributions in this work:

- A **sensor model** has been introduced as described in chapter 4. This sensor model is able to give a coherent and efficient representation of the information provided by various physical sensors. The introduction of uncertainty modeling in this sensor model is inspired by the real-world environment. Consideration of the uncertainty is crucial in order to maintain a robust sensor management performance.
  The proposed sensor model contains six attributes which are the major characteristics of the sensors in the world. This sensor model is independent from the type of physical sensors installed in the observed scene. Un-

certainty manifests itself in the sensor probabilities of detection and false alarm.

- A **multisensor activity recognition framework** is proposed in chapter 5 to recognize interesting human activities at home. A proposed approach for multisensor activity recognition is based on fusing video events with environmental events on the decision level. This approach is well adapted to the fusion of heterogeneous data provided by different types of sensor. We have used Dempster-Shafer theory to model the uncertainties on the sensor measurements. A set of mass functions are associated with each combination of sensor measurement.

- A **knowledge base** of elderly activities is proposed in section 5.4.1.2 in chapter 5. This knowledge base is based on modeling a set of interesting activities at home. In this work we have modeled 100 events which include 58 video events, 26 environmental events and 16 multimodal events.

- An **experimental study in a real world** environment is described in chapter 6. The results of the proposed approach, the recognized postures and activities, have been described in section 6.3 in chapter 6. The approach has been successfully tested for a set of ADLs of 9 elderly volunteers observed in the Gerhome laboratory. The proposed posture-based event models are tested with a human actor and with the volunteers in the Gerhome laboratory. We have also tested the two abnormal activities: fainting and falling down with a human actor. We have obtained good results with few false alarms. We have proposed a new dataset which contain 224 hours of video stream for 14 elderly persons which have performed a set of household activities. This dataset contains also 14 log files of the non-video sensors.

7.2 Discussion

The proposed activity recognition approach shows the ability to help experts to represent easily interesting events and the capacity of recognizing events models related to daily activities at home. The proposed approach for activity recognition gives good results. However, the approach has some limitations and can be extended in a number of new studies and of new research directions.

The first limitation is that we are limited in terms of detection due to segmentation errors (e.g. shadow, light change, strong illumination changes as turning on the light) and to object occlusion. To solve this problem, currently we use a set of background images to take into account the various changes. More work on vision algorithms in particular in image segmentation is required to solve this kind of problem.

The second limitation concerns the used of environmental sensors which

give information about context only at an abstract level. For example, a contact sensor is installed on the door of the fridge. There are many food items contained in the fridge such as milk, juice, and butter. When the fridge sensor is triggered, the state of the fridge is changed which indicates that the person interacts with the fridge (opening the fridge and getting food out of the fridge). However, it is not possible currently to infer which food item is removed from the fridge by simply considering the current state of the fridge door. Knowing which food item is removed from the fridge can help us to recognize finer activities (e.g. recognize a person taking a milk or taking a juice) than activities we currently recognize.

The mapping from the sensed fridge to the item removed from the fridge is dynamic and uncertain. Some improvements can be done to solve this limitation. In particular a set of radio-frequency-identification (RFID) tags [Tapia et al., 2004] can be installed on objects of interest to detect object interactions (e.g. detect what food item is removed from the fridge). Nevertheless, the constraint imposed to ware a glove to sense tags makes it potentially less desirable to elderly or disabled people in terms of their perceived desire to use such a solution.

Another solution to recognize finer activities at home is image segmentation based on texture, colors and shape to distinguish between objects of interest (e.g. using texture, colors and shape for example to distinguish between tomatoes and cucumbers).

The proposed approach for activity monitoring can be applied in other environment equipped with the same requirements: stationary video camera, sensor data with timestamps, tracking only one Individual.

7.3 Future Work

The purpose of this section is to analyze the future work, as extensions to the approach and as possible solutions to its limitations. In this section we present firstly the proposed short-term perspectives, after that we present the proposed long-term perspectives.

7.3.1 Short-Term Perspectives

In short term, the activity monitoring approach can be extended in several ways:

7.3.1.1 Improving Object Detection

The detection of errors in the segmentation task can be an interesting extension of the approach. Reliability measures could be associated to the detected moving regions in order to account for the quality of segmentation in terms of the influence of illumination changes, level of contrast between the moving objects and the

background of the scene, and the possibility of the presence of shadows, object occlusion, among other aspects.

7.3.1.2 Learning Event Models and Learning Temporal Information of Events

Both normal and abnormal behaviors can be modeled by experts of application domains using the presented event description language and a dedicated ontology. Nevertheless, defining event models is time consuming and an error prone process. Thus, it will be interesting to learn automatically normal behaviors of every day data, because normal behaviors are frequent and can be extracted from everyday activities.

In everyday environments, any particular event may take variable time to finish. In a household kitchen for instance, the event of taking something out of the refrigerator may take longer or shorter time depending on how many items are being taken out and also depending on the individual who did it (e.g. age and health of the person may influence the duration of that event). This duration over which an event takes place can be an important discriminating factor to distinguish amongst various activity classes. Furthermore, the event duration can be an important indicator about whether the event was performed correctly or not. At present, we only calculate the duration of events by using leave-one-out method (see section 6.5.1 in chapter 6) and we are not learning individual duration variation of each event depending on the person and the number of items being taken out. Learning these durations needs an observation of the old person during at least 2 weeks. A potential future direction of our work might be to investigate the extent to which considering such temporal information of events is useful for activity analysis.

7.3.1.3 Incorporate Another Uncertainty

The proposed uncertainty in sensor measurements is useful in multisensor systems but it does not take into account the identity of the person using the sensor. There is uncertainty which occurs when several persons trigger the same set of sensors. In this type of uncertainty the system does not know which person has triggered which sensor (i.e. which data to associate to which person). Managing this type of uncertainty requires to identify the person by using for example an RFID tag or by using face detection techniques. If the sensors were able to distinguish the identity of the person activating them, it would be possible to create systems that recognize activities in multiple person environments.

7.3.2 Long Term Perspectives

In long term, the activity monitoring approach can be extended in several ways:

7.3.2.1 Activity Monitoring in Other Environment

In the current work, the proposed activity recognition approach was evaluated in the experimental laboratory with fourteen elderly people. The next step of this work requires to test this approach in nursing homes and in hospital environment involving more people with different wellness and different health status. The main advantage of these tests would be to develop a knowledge database so that rules to monitor the functional health status of elderly people could be driven. Possibilities of studies include:

- In nursing homes: Tests to validate the use of the proposed monitoring activities for any change in the health status. This could include healthy and frail elderly. At least 50 persons would be required for a period of at least 6 months. This study would help to compare the obtained results in the nursing homes to those obtained with the fourteen volunteers in the Gerhome laboratory.

- In hospital environment: Tests to validate the proposed monitoring activities for different persons with different diseases. This could include patients with chronic diseases (e.g. Alzheimer). At least 50 to 100 persons would be required for a period of at least 6 months to one year. This study would help to compare the obtained results with persons with different diseases to those obtained from the healthy persons in the nursing homes.
  Currently, in Pulsar team a new PhD thesis has started which consists in monitoring Alzheimer patient activities in Nice hospital. In this application, (in plus of the environmental sensors) they also use an actimetry sensor.

7.3.2.2 Improve Activity Recognition Algorithms

In the future, it will be interesting to improve the activity recognition algorithms to explore the following questions:

- Can activity recognition algorithms be improved to recognize not only the activity but also the style of the activity? For instance, can we develop algorithm that can detect not only "preparing dinner" but "preparing dinner slowly"? Also, can we develop algorithm that can detect the way a person takes his/her meal?

- Can multitasking activities be detected? For example when a person performs several activities at the same time .

- How can algorithms that work for one individual at home can be extended to multiple persons?

7.3.2.3 Embedding the sensors into common architectural components

Strategies for embedding the environmental sensors in objects such as cupboards, drawers and light switches could further simplify the installation in new environments. Ultimately these sensors might be built into the architectural components, and furniture at time of manufacture.

Appendix A

Publications of the Author

- **International Journal:**

1. A computer system to monitor older adults at home: Preliminary results. ZOUBA, N. and BREMOND, F. and THONNAT,M. and ANFOSSO, A. and PASCUAL, E. and MALLEA, P. and MAILLAND, V. and GUERIN, O. Gerontechnology Journal. July 2009, 8(3), pp 129-139.

- **World Congress:**

1. Assessing Computer Systems for the Real Time Monitoring of Elderly People Living at Home. ZOUBA, N. and BREMOND, F. and THONNAT,M. and ANFOSSO, A. and PASCUAL, E. and MALLEA, P. and MAILLAND, V. and GUERIN, O. 19th IAGG World Congress of Gerontology and Geriatrics (IAGG 2009). July 2009.

- **International Conferences:**

1. Multisensor Fusion for Monitoring Elderly Activities at Home. ZOUBA, N. and BREMOND, F. and THONNAT,M. IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS 2009). September 2009.

2. Monitoring Activities of Daily Living (ADLs) of Elderly Based on 3D Key Human Postures. ZOUBA, N. and BREMOND, F. and THONNAT,M. 4th International Cognitive Vision Workshop (ICVW 2008), pp 37-50. May 2008.

3. Multi-sensors Analysis for Everyday Activity Monitoring. ZOUBA, N. and BREMOND, F. and THONNAT,M. and VU, V.T. 4rth International Conference, Sciences of Electronic, Technologies of Information and Telecommunications (SETIT 2007). March 2007.

Appendix B

Predefined Scenario

Here is a detail of the predefined scenario for the fourteen volunteers (see figures B.1, B.2, B.3). This scenario is defined in collaboration with Geriatrics and Gerontologists from Nice hospital in France.

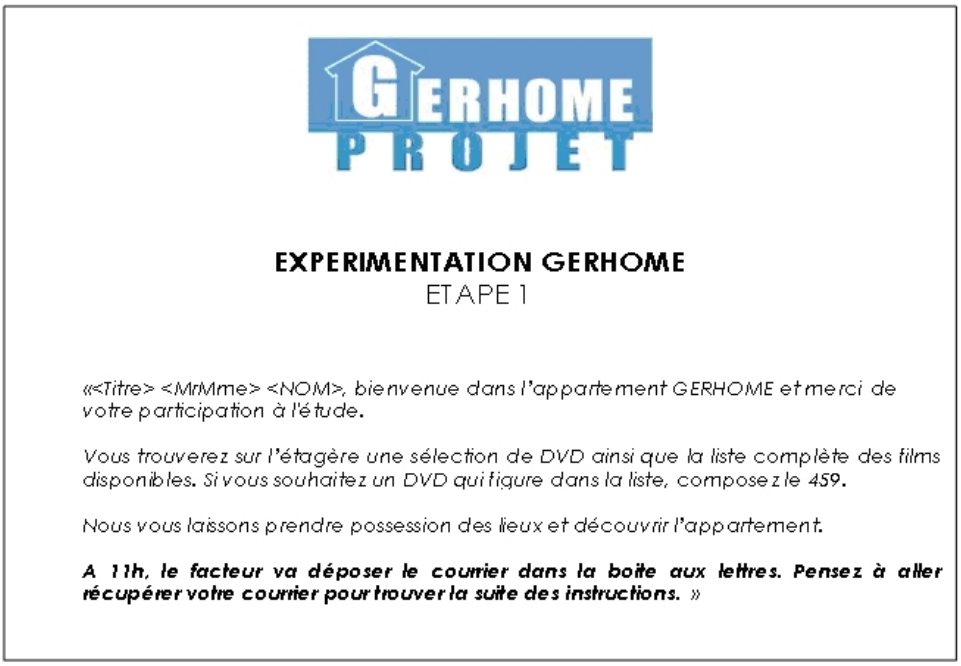


Figure B.1: A predefined scenario (step 1) for the fourteen volunteers.

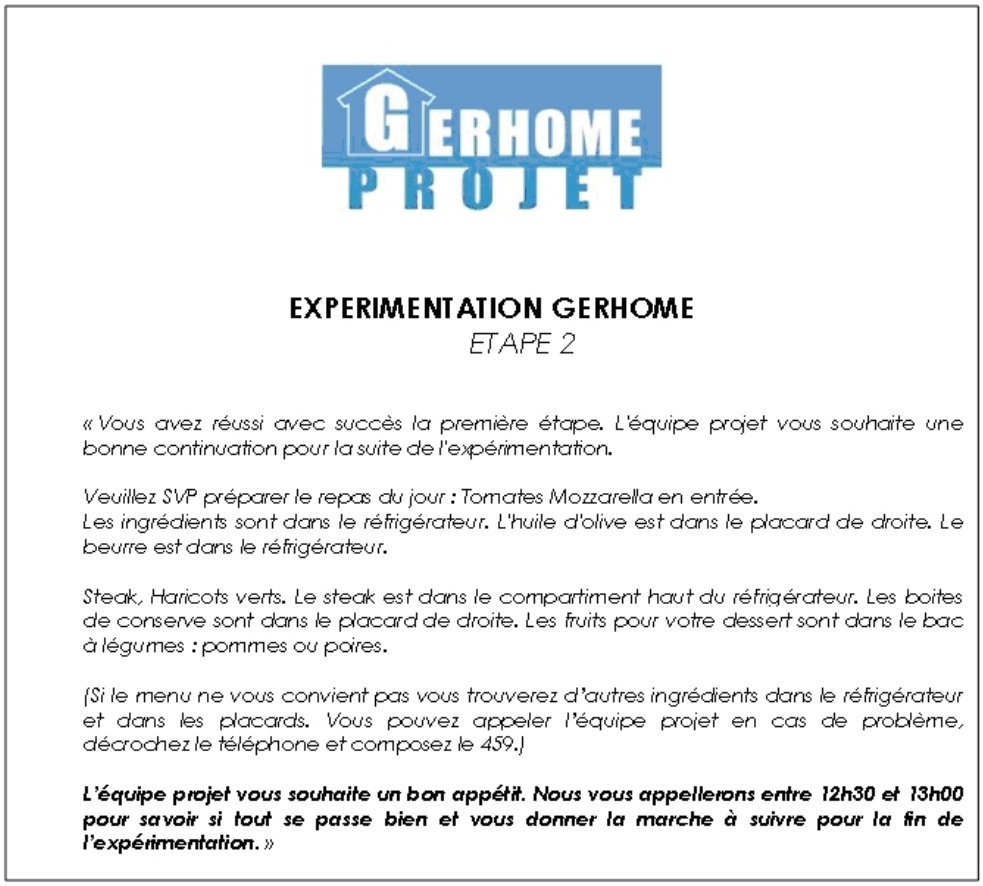


Figure B.2: A predefined scenario (step 2) for the fourteen volunteers.

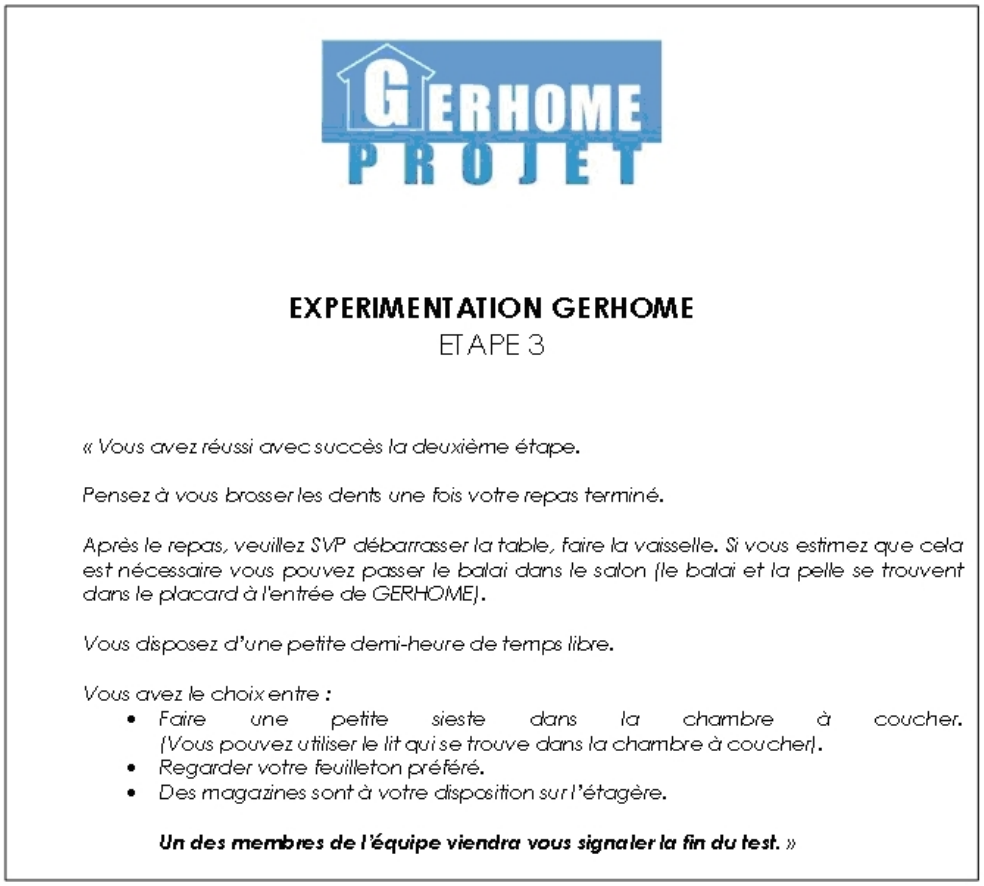


Figure B.3: A predefined scenario (step 3) for the fourteen volunteers.

Bibliography

[Abowd et al., 2002] Abowd, G., Bobick, A., Essa, I., Mynatt, E., and Rogers, W. (2002). The aware home: Developing technologies for successful aging. In *Proceedings of AAAI Workshop and Automation as a Care Giver: The Role of Intelligent Technology in Elder Care.*

[Aggarwal and Cai, 1999] Aggarwal, J. K. and Cai, Q. (1999). Human motion analysis: A review. *IEEE Transaction on Pattern Analysis and Machine Intelligence.*, vol.21, no.11:pp.123–154.

[Allen, 1983] Allen, J. F. (1983). Maintaining knowledge about temporal intervals. In *Communications of the ACM.*

[Avanzi et al., 2005] Avanzi, A., Bremond, F., Tornieri, C., and Thonnat, M. (2005). Design and assessement of an intelligent activity monitoring platform. *EURASIP Journal on Applied Signal Processing, Special Issue on "Advances in Intelligent Vision Systems: Methods and Applications".*

[Bajcsy et al., 1996] Bajcsy, R., Kamberova, G., Mandelbaum, R., and Mintz, M. (1996). Robust fusion of position data. In *Proceeding of the workshop on Foundations of Information/Decision Fusion.*

[Bamiv and Casasent, 1981] Bamiv, Y. and Casasent, D. (1981). Multisensor image registration: Experimental verification. In *Process. Images and Datafrom Optical Sensors.*

[Bao and Intille, 2004] Bao, L. and Intille, S. (2004). Activity recognition in the home setting using simple and ubiquitous sensors. *Proceedings of Pervasive*, vol.LNCS3001:Springer-Verlag:pp.1–17.

[Bardram and Christensen, 2004] Bardram, J. and Christensen, H. (2004). Open issues in activity-based and task-level computing. In *Proceedings of first international workshop on computer support for human tasks and activities, CfPC.*

[Bellot et al., 2002] Bellot, D., Boyer, A., and Charpillet, F. (2002). A new definition of qualified gain in a data fusion process: application to telemedicine. In *Proceedings Fifth International Conference on Information Fusion.*

[Benaim et al., 2005] Benaim, C., Froger, J., Compan, B., and Pélissier, J. (2005). Evaluation de l'autonomie de la personne âgée. *Annales de réadaptation et de médecine physique*, vol.48:pp.336–340.

[Berlo, 1998] Berlo, A. V. (1998). A smart model house as research and demonstration tool for telematics development. In *Proceedings of the 3rd TIDE Congress: Technology for Inclusive Design and Equality Improving the Quality of Life for the European Citizen.*

[Boulay et al., 2006] Boulay, B., Bremond, F., and Thonnat, M. (2006). Applying 3d human model in a posture recognition system. *Pattern Recognition Letter.*, vol.27, no.15:pp.1785–1796.

[Bouma and Graafmans, 1993] Bouma, H. and Graafmans, J. (1993). Gerontechnology: A framework on technology and aging. *Gerontechnology, ISO Press*, pages pp.1–6.

[Bracio et al., 1997] Bracio, B., Horn, W., and Moler, D. (1997). Sensor fusion in biomedical systems. In *Proceedings of the 19th Annual International Conference of the IEEE Engineering in Medicine and Biology Society.*

[Brey, 2005] Brey, P. (2005). Freedom and privacy in ambient intelligence. *Ethics and Information Technology.*, vol.7, no.3:pp.157–166.

[Brown et al., 1992] Brown, C., Durrant-Whyte, H., Leonard, J., Rao, B., and Steer, B. (1992). Distributed data fusion using kalman filtering: A robotics application. *In Data Fusion in Robotics and Machine Intelligence, Academic Press, Inc.*, vol.7:pp.267–309.

[Burgio et al., 2001] Burgio, L., Scilley, K., Hardin, J. M., and Hsu, C. (2001). Temporal patterns of disruptive vocalization in elderly nursing home residents. *Journal of Geriatric Psychiatry*, vol.16(1):pp.378–386.

[C. Castel and Tessier, 1996] C. Castel, L. C. and Tessier, C. (1996). 1st order c-cubes for the interpretation of petri nets: an application to dynamic scene understanding. In *Proceedings of the 8th International Conference on Tools with Artificial Intelligence (TAI).*

[C. Dousson and Ghallab, 1993] C. Dousson, P. G. and Ghallab, M. (1993). Situation recognition: Representation and algorithms. In *Proceedings of the 13th International Joint Conference on Artificial Intelligence (IJCAI).*

[C. J. Needham and Cohn, 2005] C. J. Needham, P. E. Santos, D. R. M. V. D. D. C. H. and Cohn, A. (2005). Protocols from perceptual observations. *Artificial Intelligence, Elsevier Science Publishers Ltd.*, vol.167, no.3:pp.103–136.

[Carter et al., 2006] Carter, N., Young, D., and Ferryman, J. (2006). A combined bayesian markovian approach for behaviour recognition. In *18th International Conference on Pattern Recognition (ICPR).*

[Chan et al., 1999] Chan, M., Bocquet, H., Campo, E., Val, T., and Pous, J. (1999). Alarm communication network to help carers of the elderly for safety purposes: a survey of a project. *International Journal of Rehabilitation Research.*, vol.22:pp.131–136.

[Chleq and Thonnat, 1996] Chleq, N. and Thonnat, M. (1996). Real-time image sequence interpretation for video-surveillance applications. In *Proceedings of IEEE International Conference on Image Processing (ICIP).*

[Chomat and Crowley, 1999] Chomat, O. and Crowley, J. (1999.). Probabilistic recognition of activity using local appearance. In *International Conference on Computer Vision and Pattern Recognition (CVPR).*, Vancouver, Canada.

[Chowdhury and Chellappa, 2003] Chowdhury, A. R. and Chellappa, R. (2003). Advanced a factorization approach for activity recognition. In *Computer Vision and Pattern Recognition Workshop on Event Mining(CVPR).*

[Clarkson et al., 1998] Clarkson, B., Sawhney, N., and Pentland, A. (1998). Auditory context awareness via wearable computing. In *Proceedings of the Perceptual User Interfaces Workshop (PUI).*

[CNEG, 2000] CNEG (2000). Collège national des enseignants de gériatrie. chapitre 8 - autonomie et dépendance. *Corpus de Gériatrie*, vol.1:pp.185.

[Cohen and Medioni, 1999] Cohen, I. and Medioni, G. (1999). Detecting and tracking moving objects for video surveillance. In *Computer Vision and Pattern Recognition (CVPR).*

[Colin and Coutton, 2000] Colin, C. and Coutton, V. (2000). Le nombre de personnes âgées dépendantes d'après l'enquête handicaps, incapacités, dépendance. *DREES. Études et Résultats*, vol.94.

[Cook and Das, 2007] Cook, D. and Das, S. (2007). How smart are our environments? an updated look at the state of the art. *Pervasive Mobile Computer*, vol.3, no.2:pp.53–73.

[Coue et al., 2002] Coue, C., Fraichard, T., Bessiere, P., and Mazer, E. (2002). Multi-sensor data fusion using bayesian programming: An automotive application. In *IEEE/RSJ International Conference on Intelligent Robots and System.*

[Dasarathy, 1996] Dasarathy, B. V. (1996). Information/decision fusion - principles and paradigms. In *Proceeding of the workshop on Foundations of Information /Decision Fusion: Applications to Engineering Problems.*

[Dasarathy, 1997] Dasarathy, B. V. (1997). Sensor fusion potential exploitation - innovative architectures and illustrative applications. In *Proceedings of the IEEE.*

[Demiris et al., 2001] Demiris, G., Speedie, S., and Finkelstein, S. (2001). Change of patients perception of telehomecare. *Telemedicine Journal and e-Health.*, vol.7, no.3:pp.241–248.

[Dempster, 1968] Dempster, A. (1968). A generalization of bayesian inference. *Journal of the Royal Statistical Society.*, vol.30:pp.205–247.

[Elger and Furugren, 1998] Elger, G. and Furugren, B. (1998). Smartbo-an ict and computer-based demonstration home for disabled people. In *Proceedings of the 3rd TIDE Congress: Technology for Inclusive Design and Equality Improving the Quality of Life for the European Citizen.*

[Elmenreich et al., 2001] Elmenreich, W., Haidinger, W., and Kopetz, H. (2001). Interface design for smart transducers. In *IEEE Instrumentation and Measurement Technology Conference.*

[Elmenreich and Pitzek, 2001] Elmenreich, W. and Pitzek, S. (2001). Using sensor fusion in a time-triggered network. In *Proceedings of the 27th Annual Conference of the IEEE Industrial Electronics Society.*

[Fogarty et al., 2006] Fogarty, J., Au, C., and Hudson, S. (2006). Sensing from the basement: a feasibility study of unobtrusive and low-cost home activity recognition. In *UIST'06: Proceedings of the 19th annual ACM symposium on User interface software and technology.*

[Fowler and Schmalzel, 2004] Fowler, K. and Schmalzel, J. (2004). Sensors: The first stage in the measurement chain. *IEEE Instrumentation and Measurement Magazine*, pages pp.60–65.

[Gavrila, 1999] Gavrila, D. M. (1999). The visual analysis of human movement: a survey. *Computer Vision Image Understanding (CVIU)*, vol.73, no.1:pp.82–98.

[Ghallab, 1996] Ghallab, M. (1996). On chronicles: Representation, on-line recognition and learning. In *Proceedings of the 5th International Conference on Principles of Knowledge Representation and Reasoning (KR).*

[Glascock and Kutzik, 2006] Glascock, A. and Kutzik, D. (2006). The impact of behavioral monitoring technology on the provision of health care in the home. *Journal of Universal Computer Science*, vol.12, no.1:pp.59–79.

[Goodridge and Kay, 1996] Goodridge, S. G. and Kay, M. G. (1996). Multimedia sensor fusion for intelligent camera control. In *Proceeding of IEEE international conference on Multisensor Fusion and Integration for Intelligent System.*

[Grossmann, 1998] Grossmann, P. (1998). Multisensor data fusion. *The GEC journal of Technology.*, vol.15:pp.27–37.

[Guralnik and Haigh, 2002] Guralnik, V. and Haigh, K. (2002). Learning models of human behavior with sequential patterns. In *Proceedings of AAAI-02 workshop on Artificial Intelligence 'Automation as Caregiver'.*

[Hall and Llinas, 2001] Hall, L. D. and Llinas, J. (2001). Multisensor data fusion. In *Handbook of Multisensor Data Fusion. CRC Press.*

[Harmo et al., 2005] Harmo, P., Taipalus, T., Knuuttila, J., Wallet, J., and Halme, A. (2005). Needs and solutions- home automation and service robots for the elderly and disabled. In *International Conference on Intelligent Robot Systems.*

[Heikkila and Silven, 1999] Heikkila, J. and Silven, O. (1999). A real-time system for monitoring of cyclists and pedestrians. In *In Proceedings of the Second IEEE Workshop on Visual Surveillance.*, pages 74–81.

[Henricksen and Indulska, 2006] Henricksen, K. and Indulska, J. (2006). Developing context-aware pervasive computing applications: Models and approach. In *Proceedings of Pervasive and Mobile Computing Conference.*

[Hoey et al., 2007] Hoey, J., Bertoldi, A. V., and Mihailidis, A. (2007). Assisting persons with dementia during handwashing using a partially observable markov decision process. In *International Conference on Computer Vision Systems (ICVS).*

[Hsiao, 1988] Hsiao, M. (1988). Geometric registration method for sensor fusion. In *Senror Fusion: Spatial Reasoning and Scene Interpretation.*

[i pot, 2005] i pot (2005). Zojirushi corporation. www.mimamori.net.

[Ivanov and Bobick, 2000] Ivanov, Y. and Bobick, A. (2000). Recognition of visual activities and interactions by stochastic parsing. In *IEEE Transactions on Patterns Analysis and Machine Intelligence.*

[Jones, 2006] Jones, I. (2006). Aging, can we stop the clock? *Trust Report http://www.wellcome.ac.uk/.*

[Kabre, 1995] Kabre, H. (1995). Performance and competence models for audio-visual data fusion. In *SPIE international symposium on intelligent systems and advanced manufacturing.*

[Kalman, 1960] Kalman, R. (1960). A new approach to linear filtering and prediction problems. *ASME Journal on Basic Engineering.*, vol.82:pp.35–45.

[Katz, 1983] Katz, S. (1983). Assessing self-maintenance: Activities of daily living, mobility, and instrumental activities of daily living. *Journal of the American Geriatrics Society*, vol.31, no.12:pp.721–727.

[Katz et al., 1970] Katz, S., Downs, T. D., Cash, H. R., and Grotz, R. C. (1970). Progress in development of the index of adl. *The Gerontologist*, pages pp.20–30.

[Katz et al., 1963] Katz, S., Ford, A. B., Moskowitz, R. W., Jackson, B. A., and Jaffe, M. W. (1963). Studies of illness in the aged: The index of adl: A standardized measure of biological and psychosocial function. *Journal of the American Medical Association*, vol.185, no.12:pp.914–919.

[Kautz and Allen, 1986] Kautz, H. A. and Allen, J. F. (1986). Generalized plan recognition. In *Proceedings of the 5th National Conference on Artificial Intelligence (AAAI).*

[Kern et al., 2003] Kern, N., Schiele, B., and Schmidt, A. (2003). Multi-sensor activity context detection for wearable computing. In *Proceedings EUSAI, LNCS.*

[Kowalski and Sergot, 1986] Kowalski, R. and Sergot, M. (1986). A logic-based calculus of events. *New Generation Computing.*, vol.4:pp. 67–95.

[Krishnamachari and Iyengar, 2004] Krishnamachari, B. and Iyengar, S. (2004). Distributed bayesian algorithms for fault-tolerant event region detection in wireless sensor networks. In *IEEE Transactions on Computers.*

[Kumar and Mukerjee, 1987] Kumar, K. and Mukerjee, A. (1987). Temporal event conceptualization. In *Proceedings of the 10th International Joint Conference on Artificial Intelligence (IJCAI).*

[Kumar et al., 2005] Kumar, P., Ranganath, S., Weimin, H., and Sengupta, K. (2005). Framework for realtime behavior interpretation from traffic video. *IEEE Transactions on Intelligent Transportation Systems*, vol.6:pp.43–53.

[L. Davis and Shet, 2005] L. Davis, D. H. and Shet, V. D. (2005). Vidmap: Video monitoring of activity with prolog. In *Proceedings of Advanced Video and Signal-Based Surveillance (AVSS).*

[Lawton, 1990] Lawton, M. P. (1990). Aging and performance of home tasks. *Human Factors*, vol.32:pp.527–536.

[Lawton and Brody, 1969] Lawton, M. P. and Brody, E. M. (1969). Assessment of older people: Self-maintaining and instrumental activities of daily living. *The Gerontologist*, vol.9, no.3:pp.179–186.

[Lesire and Tessier, 2005] Lesire, C. and Tessier, C. (2005). Particle petri nets for aircraft procedure monitoring under uncertainty. In *Proceedings of the 26th International Conference on Application and Theory of Petri Nets and Other Models of Concurrency (ATPN).*

[Lester et al., 2005] Lester, J., Choudhury, T., Kern, N., Borriello, G., and Hannaford, B. (2005). A hybrid discriminative/generative approach for modelling human activities. In *Proceedings of the 19th International Joint Conference on Artificial Intelligence (IJCAI).*

[Loke, 2007] Loke, S. (2007). Context-aware pervasive systems. *Auerbach Publications.*

[Luo and Kay, 1992] Luo, R. and Kay, M. (1992). Data fusion and sensor integration: state-of-the-art 1990's. *Data Fusion in Robotics and Machine Intelligence*, pages pp.7–136.

[Luo and Kay, 1990] Luo, R. C. and Kay, M. G. (1990). A tutorial on multisensor integration and fusion. In *Proceedings of the 16th Annual Conference IEEE Industrial Electronics.*

[Luo et al., 2006] Luo, X., Dong, M., and Huang, Y. (2006). On distributed fault-tolerant detection in wireless sensor networks. In *IEEE Transactions on Computers.*

[M. Zúñiga, 2006] M. Zúñiga, F. Brémond, M. T. (2006). Fast and reliable object classification in video based on a 3d generic model. In *The 3rd International Conference on Visual Information Engineering (VIE 2006).*, pages 433–440.

[MacLeod and Summerfield, 1987] MacLeod, B. and Summerfield, A. (1987). Quantifying the contribution of vision to speech perception in noise. *British Journal of Audiology*, vol.21:pp.131–141.

[Manabe et al., 2000] Manabe, K., Matsui, T., Yamaya, M., Sato-Nakagawa, T., Okamura, N., Ari, H., and H.Sasaki (2000). Sleeping patterns and mortality among elderly patients in geriatric hospitals. *Gerontology*, vol.46:pp.318–322.

[Maybeck, 1982] Maybeck, P. (1982). Stochastic models, estimation, and control. In *Proceedings of SPIE on Sensor Fusion.*

[McIvor, 2000] McIvor, A. (2000). Background subtraction techniques. In *In Proceedings of the Conference on Image and Vision Computing (IVC 2000).*

[Mehboob et al., 1997] Mehboob, H., Jeffrey, M., and Izhak, B. (1997). A robust sensor fusion method for heart rate estimation. *Journal of clinical monitoring and computing.*, vol.13, no.6:pp.385–393.

[Mesrine, 2003] Mesrine, A. (2003). Les places dans les établissements pour personnes âgées en 2001-2002. *DREES. Études et Résultats*, vol.263.

[Moeslund et al., 2000] Moeslund, T., Hilton, A., and Krüger, V. (2000). A survey of advances in vision based human motion capture and analysis. *Behavior Research Methods, Instruments, and Computers*, vol.32, no.3.

[Moeslund et al., 2006] Moeslund, T., Hilton, A., and Krüger, V. (2006). A survey of advances in vision based human motion capture and analysis. *Computer Vision Image Understanding (CVIU)*, vol.104, no.2:pp.90–126.

[Moeslund and Granum, 2001] Moeslund, T. B. and Granum, E. (2001). A survey of computer vision based human motion capture. *Computer Vision and Image Understanding*, vol.81, no.3:pp.231–268.

[Moore et al., 1999] Moore, D., Essa, I., and Hayes, M. (1999). Exploiting human actions and object context for recognition tasks. In *Proceedings of IEEE International Conference on Computer Vision (ICCV99).*

[Munguia-Tapia et al., 2006] Munguia-Tapia, E., Choudhury, T., and Philipose, M. (2006). Building reliable activity models using hierarchical shrinkage and mined ontology. In *Proceedings of the 4th International Conference on Pervasive Computing.*

[Najafi et al., 2003] Najafi, B., Aminian, K., Paraschiv-Ionescu, A., Loew, F., Bula, C., and Robert, P. (2003). Ambulatory system for human motion analysis using a kinematic sensor: Monitoring of daily physical activity in the elderly. *IEEE Transactions on Biomedical Engineering.*, vol.50, no.6:pp.711–723.

[Nambu et al., 2005] Nambu, M., Nakajima, K., Noshira, M., and Tamura, T. (2005). An algorithm for the automatic detection of health conditions. *IEEE Enginering Medicine Biology Magazine*, vol.24, no.4:pp.38–42.

[Noury et al., 2001] Noury, N., Rialle, V., and G.Virone (2001). The telemedecine home care station : a model and some technical hints. In *Proceedings of Healtcomm2001.*

[Ogawa et al., 2002] Ogawa, M., Suzuki, R., Otake, S., Izutsu, T., Iwaya, T., and Togawa, T. (2002). Long term remote behavioral monitoring of elderly by using sensors installed in ordering houses. In *IEEE Engineering in Medicine and Biology Society (EMBS) special topic conference on microtechnologies in medicine and biology.*

[Oliver et al., 2002] Oliver, N., Horvitz, E., and Garg, A. (2002). Layered representations for human activity recognition. In *Fourth IEEE International Conference on Multimodal Interfaces.*

[Oliver et al., 1999] Oliver, N., Rosario, B., and Pentland, A. (1999). A bayesian computer vision system for modeling human interactions. In *International Conference on Vision Systems (ICVS).*

[Pentland, 2004] Pentland, A. (2004). Healthwear: Medical technology becomes wearable. *IEEE Computer.*, vol.37:pp.42–49.

[Philipose et al., 2004] Philipose, M., Fishkin, K., Perkowitz, M., Patterson, D., Kautz, H., and Hahnel, D. (2004). Inferring activities from interactions with objects. *IEEE Prevasive Computing Magazine.*, vol.3, no.4:pp.50–57.

[Phillips and Ancoli-Israel, 2001] Phillips, B. and Ancoli-Israel, S. (2001). Sleep disorders in the elderly.review. *Sleep Medicine.*, vol.2:pp.99–114.

[Pollack, 2005] Pollack, M. (2005). Intelligent technology for an aging population: the use of ai to assist elders with cognitive impairment. *Intelligence Artificiel Magazine.*, pages pp.1–27.

[Porteus and Brownsell, 2002] Porteus, J. and Brownsell, S. (2002). Exploring technologies for independent living for older people. *Anchor Trust/Housing Corporation, UK.*

[Provan, 1992] Provan, G. (1992). The validity of dempster-shafer belief functions. *International Journal of Approximate Reasoning.*, vol.6, no.3:pp.389–399.

[QuietCare, 2002] QuietCare (2002). Quiet care systems generation health and elder care company. www.quietcaresystems.com.

[Ranganathan et al., 2004] Ranganathan, A., Al-Muhtadi, J., and Campbell, R. (2004). Reasoning about uncertain contexts in pervasive computing environments. *IEEE Pervasive Computing.*, pages pp.62–70.

[Rota and Thonnat, 2000] Rota, N. and Thonnat, M. (2000). Activity recognition from video sequences using declarative models. In *Proceedings of 14th European Conference on Artificial Intelligence (ECAI).*

[Saphe, 2006] Saphe (2006). Saphe project. http://ubimon.doc.ic.ac.uk/saphe/.

[Sarela et al., 2003] Sarela, A., Korhonen, I., Lotjonen, L., Sola, M., and Myllymaki, M. (2003). Ist vivago - an intelligent social and remote wellness monitoring system for the elderly. In *Proceedings of the 4th Annual IEEE EMBS Special Topic Conference on Information Technology Applications in Biomedicine (ITAB2003).*

[Shafer, 1976] Shafer, G. (1976). A mathematical theory of evidence. *Princeton University Press.*

[Sidenbladh and Black, 2001] Sidenbladh, H. and Black, M. (2001). Learning image statistics for bayesian tracking. In *IEEE International Conference on Computer Vision (ICCV).*

[Tamura, 2005] Tamura, T. (2005). Biomedical engineering at the forefront in japan. *Engineering in Medicine and Biology Magazine, IEEE*, vol.24, no.4:pp.23–26.

[Tapia et al., 2004] Tapia, E. M., Intille, S., and Larson, K. (2004). Activity recognition in the home using simple and ubiquitous sensors. *IEEE Pervasive*, pages pp.158–175.

[T.E. Bullock and Boudreau, 1988] T.E. Bullock, S. Sangsuk-iam, R. P. and Boudreau, E. (1988). Sensor fusion applied to system performance under sensor failures. In *Proceedings of SPIE on Sensor Fusion.*

[Tran et al., 2004] Tran, D., Phung, D., Bui, H., and S.Venkatesh (2004). A probabilistic model with parsimonious representation for sensor fusion in recognizing activity in pervasive environment. In *Proceedings of the 18th International Conference on Pattern Recognition.*

[Tsai, 1986] Tsai, R. (1986). An efficient and accurate camera calibration technique for 3d machine vision. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition.*, pages 364–374.

[Vincent et al., 2002] Vincent, C., Drouin, G., and Routhier, F. (2002). Examination of new environmental control applications. *Assistive Technology Journal.*, vol.14, no.2:pp.98–111.

[Virone et al., 2003] Virone, G., Istrate, D., Vacher, M., Serignat, J., Noury, N., and Demongeot, J. (2003). First steps in data fusion between a multichannel audio acquisition and an information system for home healthcare. In *Proceedings of IEEE Engineering in Medicine and Biology Society.*

[Vu et al., 2003] Vu, V., Bremond, F., and Thonnat, M. (2003). Automatic video interpretation: A novel algorithm based for temporal scenario recognition. In *The Eighteenth International Joint Conference on Artificial Intelligence (IJCAI).*

[W. Elmenreich, 2003] W. Elmenreich, S. P. (2003). Smart transducers-principles, communications, and configuration. In *Proceedings 7th IEEE International Conference on Intelligent Engineering Systems (INES).*, page 510Ű515.

[Wang et al., 2007] Wang, S., Pentney, W., Popescu, A.-M., Choudhury, T., and Philipose, M. (2007). Common sense joint training of human activity recognizers. In *Proceedings of the Twentieth International Joint Conference on Artificial Intelligence (IJCAI).*

[Wilson and Atkeson, 2005] Wilson, D. and Atkeson, C. (2005). Simultaneous tracking and activity recognition (star) using many anonymous, binary sensors. In *Proceedings of the 3rd International Conference on Pervasive Computing.*

[Wilson et al., 2005] Wilson, D., Wyaat, D., and Philipose, M. (2005). Using context history for data collection in the home. *PERVASIVE LNCS.*, vol.3468.

[Wu et al., 2002] Wu, H., Siegel, M., and Ablay, S. (2002). Sensor fusion for context understanding. In *Proceedings of IEEE Instrumentation and Measurement Technology Conference.*

[Wyatt et al., 2005] Wyatt, D., Philipose, M., and Choudhury, T. (2005). Unsupervised activity recognition using automatically mined common sense. In *Proceedings of Twentieth National Conference on Artificial Intelligence (AAAI).*

[Yamato et al., 1992] Yamato, J., Ohya, J., and Ishii, K. (1992). Recognizing human action in time-sequential images using hidden markov model. In *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)*.

[Yang, 2006] Yang, G. (2006). Body sensor networks. *Springer-Verlag UK*.

[Yang et al., 2004] Yang, G., Lo, B., Wang, J. L., Rans, M., Thiemjarus, S., Ng, J., Garner, P., Brown, S., Majeed, B., and Neid, I. (2004). From sensor networks to behavior profiling: A homecare perspective of intelligent building. In *The IEE Seminar for Intelligent Buildings*.

[Zelnik-Manor and Irani, 2001] Zelnik-Manor, L. and Irani, M. (2001). Event-based video analysis. In *Computer Vision and Pattern Recognition (CVPR)*.

Résumé

Dans cette thèse, une approche combinant des données issues de capteurs hétérogènes pour la reconnaissance d'activités des personnes âgées à domicile est proposée. Cette approche consiste à combiner les données fournies par des capteurs vidéo avec des données fournies par des capteurs environnementaux pour suivre l'interaction des personnes avec l'environnement. La première contribution est un nouveau modèle de capteur capable de donner une représentation cohérente et efficace des informations fournies par différents types de capteurs physiques. Ce modèle inclue l'incertitude sur la mesure. La deuxième contribution est une approche, basée sur une fusion multicapteurs, pour la reconnaissance d'activités. Cette approche consiste à détecter la personne, suivre ses mouvements, reconnaître ses postures et ses activités d'intérêt, par une analyse multicapteurs et une reconnaissance d'activités humaines. Pour résoudre le problème de la présence de capteurs hétérogènes, nous avons choisi de réaliser la fusion à haut niveau (niveau événement) des différentes données issues des différents capteurs, en combinant les événements vidéo avec les événements environnementaux. La troisième contribution est l'extension d'un langage de description qui permet aux utilisateurs (ex. le corps médical) de décrire les activités d'intérêt dans des modèles formels. Les résultats de cette approche sont montrés pour la reconnaissance des AVQ pour de vraies personnes agées évoluant dans un appartement expérimental appelé GERHOME équipé de capteurs vidéo et de capteurs environnementaux. Les résultats obtenus de la reconnaissance des différentes AVQ sont encourageants.

Mots-clés: Activités de la Vie Quotidienne (AVQ), modèle de capteur, fonction de densité de probabilité (PDF), événements vidéo, événements environnementaux, événement multimodale, reconnaissance d'activités, théorie de Dempster Schäfer (DST).

Abstract

In this thesis, an approach combining heterogeneous sensor data for recognizing elderly activities at home is proposed. This approach consists in combining data provided by video cameras with data provided by environmental sensors to monitor the interaction of people with the environment. The first contribution is a new sensor model able to give a coherent and efficient representation of the information provided by various types of physical sensors. This sensor model includes an uncertainty in sensor measurement. The second contribution is a multisensor based activity recognition approach. This approach consists in detecting people, tracking people as they move, recognizing human postures and recognizing activities of interest based on multisensor analysis and human activity recognition. To address the problem of heterogeneous sensor system, we choose to perform fusion at the high-level (event level) by combining video events with environmental events. The third contribution is the extension of a description language which lets users (i.e. medical staff) to describe the activities of interest into formal models. The results of this approach are shown for the recognition of ADLs of real elderly people evolving in an experimental apartment called Gerhome equipped with video sensors and environmental sensors. The obtained results of the recognition of the different ADLs are encouraging.

Keywords: Activities of Daily Living (ADLs), sensor model, probability density function (PDF), video events, environmental events, multimodal events, multisensor activity recognition, Dempster Schäfer Theory (DST).