



HAL
open science

Le projet CIRDO d'assistance aux personnes âgées isolées à domicile

Michel Vacher, Saïda Bouakaz, Marc-Eric Bobillier-Chaumon, François Portet, Frédéric Aman, Solange Rossato, Erwan Guillou

► To cite this version:

Michel Vacher, Saïda Bouakaz, Marc-Eric Bobillier-Chaumon, François Portet, Frédéric Aman, et al..
Le projet CIRDO d'assistance aux personnes âgées isolées à domicile. Congrès Handicap 2014 - 8ème
Édition - Les technologies d'assistance : de la compensation à l'autonomie, Jun 2014, Paris, France.
pp.229-235. hal-00990201

HAL Id: hal-00990201

<https://hal.science/hal-00990201>

Submitted on 13 May 2014

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Le projet CIRDO d'assistance aux personnes âgées isolées à domicile

Michel Vacher
and François Portet
and Frédéric Aman
and Solange Rossato
LIG, UMR5217 CNRS
Grenoble, France

Email : Michel.Vacher@imag.fr

Saïda Bouakaz
Erwan Guillou
LIRIS, UMR5205 CNRS
Université Claude Bernard
Lyon 1, France

Email : saida.bouakaz@univ-lyon1.fr

Marc-Eric Bobillier Chaumon
GRePS
Université Lyon 2, France

Email : marc-eric.bobillier-chaumon@univ-lyon2.fr

Résumé—CIRDO RI est un projet qui vise à mettre au point un système d'assistance pour faciliter le maintien à domicile des personnes âgées et/ou dépendantes. Il a pour objectif de sécuriser les seniors et de favoriser leur autonomie pour soulager les familles et les aidants. L'idée principale du projet CIRDO est d'utiliser à la fois l'analyse vidéo et la reconnaissance de la parole et ceci dans le même processus. Afin de répondre aux contraintes fixées dans le projet CIRDO comme le respect de l'intimité, la préservation de l'information privée et la réactivité dans la détection de situations de détresse, les solutions proposées sont conçues pour opérer en tâches de fond, sans l'intervention d'un opérateur. Outre l'aspect technique, ce projet porte également sur l'évaluation psycho-sociale et psycho-ergonomique de ces technologies par des utilisateurs potentiels et leur entourage (famille, aidants...) au niveau de leur utilisabilité et de leur acceptabilité notamment. Des expérimentations conduites dans un milieu réaliste ont permis l'enregistrement de données audio et vidéo correspondant au cas d'usage.

I. INTRODUCTION

Le maintien à domicile des Personnes Âgées (PA) dépendantes devient une priorité dans nos sociétés pour des raisons tant économiques que sociétales (liées au bien-être, à la prise en charge, à l'assistance de nos aînés). Ce concept, « vieillir chez soi », a vu le jour en raison de l'augmentation de l'espérance de vie dans tous les pays développés au cours de la dernière décennie.

Un moyen consiste à retarder autant que faire se peut le passage vers une maison médicalisée. C'est le but de l'*Ambient Assisted Living* (AAL) ou assistance à la Vie Autonome de développer des programmes qui visent à mettre en place des services innovants en vue d'améliorer la qualité de vie des personnes âgées, de maintenir leur indépendance dans un cadre de vie normal sachant que les personnes âgées sont les plus touchées par des maladies physiques ou cognitives.

C'est dans ce contexte que se positionne notre projet de recherche en gérontechnologie CIRDO (projet ANR-CNSA). Ce projet a pour objectif de conduire à la conception d'une technologie ambiante (disséminée dans le foyer) permettant, par des systèmes audio et vidéo, de détecter automatiquement les chutes et d'alerter les secours le cas échéant. Une approche de « conception centrée utilisateur » [1] et une démarche de recherche pluridisciplinaire associent les sciences informatiques, de psychologie ergonomique et de psychologie sociale, des

industriels de terrain, des formateurs et un réseau d'acteurs de la télésanté.

L'impact attendu de ce projet est l'amélioration de la sensation de sécurité et d'autonomie. Par contraste avec les systèmes existants, la personne n'a pas besoin de porter un capteur en permanence. En outre, si elle tombe en étant dans le champ de la caméra, le système peut réagir immédiatement y compris en cas de perte de conscience. Si la personne est hors du champ de la caméra, la personne qui reste consciente conserve la possibilité d'appeler à l'aide grâce à des microphones. Ainsi, le système proposé permet de surmonter une partie des limitations des systèmes existants.

Parmi les verrous scientifique que ce projet se propose de lever ; l'analyse et l'identification de la chute en milieu domestique reste un défi important en raison de la variabilité des profils et des fragilités des sujets, et de la configuration des contextes de vie [2]. Différents types de dispositifs ont été conçus et beaucoup de systèmes utilisent des gyroscopes [3] ou des accéléromètres parfois complétés par des microcontrôleurs [2]. La chute alors est caractérisée par la variation de la vitesse et l'inclinaison du porteur du dispositif.

En outre, l'acceptation réelle de capteurs vidéo / audio est très variable selon la situation de l'utilisateur, les intérêts perçus doivent l'emporter sur les inconvénients. En ce qui concerne ces aspect, il n'est pas suffisant de considérer seulement les besoins de l'utilisateur, il faudra aussi prendre en compte les desiderata de ses proches et des professionnels de santé. Les autres points concernent le traitement vidéo rendu difficile par les configurations d'éclairage non contrôlées, la difficulté de traitement dans le cas des voix atypiques et la présence de bruit, la prise de décision à partir de données incertaines (inférences à partir des résultats de l'analyse vidéo et audio), et l'interaction homme-machine (est-ce que le système réagit correctement ?).

Les étapes de l'étude sont les suivantes : (i) modélisation du processus de chute des personnes âgées au domicile et enquête sur les conditions d'acceptation du système CIRDO par les différents acteurs du foyer, (ii) conception des modèles d'analyse informatique nécessaires sur la base de cette étude, (iii) élaboration d'un système de surveillance complet et finalement, (iv) validation avec les utilisateurs ciblés.

Cet article présente l'approche du projet CIRDO et les principaux résultats des études menées jusqu'à présent pour conduire au développement d'un système adapté aux besoins spécifiques des utilisateurs et transférable dans leur propre environnement de vie psychosocial. Après un court état de l'art donné en section II, la section III décrit l'étude de terrain et les résultats qui ont conduit à la spécification du système CIRDO. Les aspects techniques de ce système sont détaillés dans la section IV. Enfin pour construire des modèles d'inférence, une expérience a été menée dans des contextes réalistes, impliquant des utilisateurs cibles et des "acteurs" équipés d'un simulateur de vieillissement corporel. Cette expérience est décrite dans la section V. Enfin une discussion sur les résultats obtenus et les travaux futurs permet de clore cet article.

II. ÉTAT DE L'ART

A. La situation de la personne âgée : entre dégradation et réhabilitation

Le processus de vieillissement se caractérise par un déclin des principales capacités biophysiques, et se traduit par un repli de la personne sur le domicile, un affaiblissement significatif de ses rôles sociaux et familiaux [4], et surtout une perte de but et d'identité [5]. La survenue des incapacités ne résiderait pas seulement dans l'individu, mais se situerait aussi dans l'interaction avec son environnement, et c'est l'incompatibilité de cet environnement avec les spécificités de la personne fragilisée qui créerait une situation d'empêchement [6]. Dans cette perspective, deux modèles se distinguent [7] : un modèle « intégratif » où la réduction des handicaps passe par une action sur l'individu (rééducation, appareillages...) et un modèle « participatif » où l'environnement supplée les déficiences. Le projet CIRDO s'inscrit clairement dans cette seconde approche en proposant d'adapter l'environnement de vie à la personne âgée dépendante. Cette technologie ambiante vise à transformer une situation d'empêchement en une situation adaptée.

B. La chute d'une personne et sa détection

Définie par l'OMS comme « le fait de se retrouver involontairement sur le sol ou dans une position de niveau inférieur par rapport à sa position de départ », la chute se traduit comme un phénomène difficilement appréhendable. Elle représente néanmoins 80% des accidents de la vie courante et est responsable d'environ 9300 décès chaque année chez les plus de 65 ans. Près d'une personne âgée de 65 à 75 ans sur quatre a indiqué qu'elle avait chuté dans les douze mois écoulés [8]. Si toutes les personnes qui font une chute sont exposées à un risque de traumatisme, l'âge, le sexe et l'état de santé peuvent avoir une incidence sur le type et la gravité de ce traumatisme [9]. Chez la personne âgée, elle est souvent entraînée par l'intrication de plusieurs facteurs dits à risques, de types intrinsèques (liés à des fragilités, son expérience, sa représentation du risque...) et/ou extrinsèques (éléments de l'environnement) [10] ainsi qu'à des éléments socio-culturels (expériences personnelles ou partagées de la chute, histoire de vie...) [11].

Étant donné les conséquences dramatiques que peuvent entraîner une chute, de nombreux projets ont développé des solutions pour les détecter rapidement. Un certain nombre utilisent des microphones en combinaison avec d'autres modalités

comme le mouvement grâce à l'utilisation d'un accéléromètre [12], [13]. Doukas et al. [13] ont proposé un système adapté au patient pour détecter une chute et le stress vocal dans la parole par l'analyse des données sonore et de mouvement, mais dans ce cas, la personne est contrainte de porter des capteurs en permanence. Un système de dialogue a été proposé pour remplacer les systèmes traditionnels d'appel d'urgence tout en s'adaptant au style de vie des personnes âgées [14]. Cependant, le vocabulaire du prototype était limité (« yes » ou « no ») et le système n'a pas été testé avec des utilisateurs âgés. Une autre approche classique est l'utilisation de robots compagnons comme dans le projet CompanionAble [15]. Ce robot compagnon est capable de se déplacer à proximité de l'utilisateur. Le système est capable de reconnaître des commandes vocales, la reconnaissance d'appels de détresse ne semble avoir été prévue. Ce système est limité par la capacité du robot de se déplacer près de la personne. Des études récentes ont montré que le microphone, tout en étant intrusif est malgré tout accepté par la population âgée [16].

L'originalité du projet CIRDO réside dans la combinaison des analyses audio et vidéo ainsi qu'à une conception qui s'est appuyée sur une étude de terrain. Le traitement de la vidéo est très précis pour détecter des mouvements de la personne, mais son action est limitée à l'angle de vision de la caméra et peut être gênée par la lumière ambiante, les microphones quant à eux peuvent détecter des sons émis hors du champ de la caméra et permettent la reconnaissance vocale, mais ils sont perturbés par les bruits ambiants.

III. ÉTUDES DE TERRAIN

L'objectif est d'analyser en profondeur les différents cas de chute afin d'identifier les poses clés (comportement, gestes et poses) et les phrases caractéristiques de situations à risque et qui pourront ensuite être utilisées pour construire les modèles du système CIRDO. Un certain nombre d'interrogations ont guidé notre analyse : quelles sont les Activités de la Vie Quotidienne (AVQ) réalisée à domicile par les personnes âgées ? Quels sont les caractéristiques des chutes et les facteurs à leur origine ? Quels descripteurs le système CIRDO peut-il utiliser pour détecter automatiquement les chutes ? Quelles sont les principaux obstacles à l'utilisation du système, à la fois d'un point technique mais aussi éthique (respect de la vie privée) ?

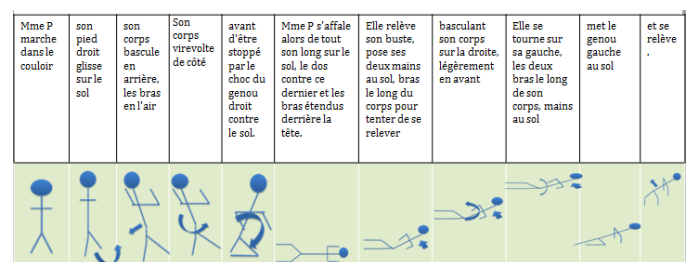


FIGURE 1. Différentes phases d'une chute

A partir des méthodes d'entretien et d'observation, nous avons ainsi pu reconstruire les « scripts » de ces différentes chutes en décrivant d'une part les conditions de la chute (caractéristiques de la personne, activité réalisée, lieu, moment, causes et circonstances de l'incident...) et d'autre part, les

Alors que c'était le milieu de la nuit, vers 3 heures du matin, Mme P., 70 ans, se lève pour aller boire un verre d'eau dans la cuisine. Elle avance dans la pénombre de la nuit, d'une bonne allure pour vite retourner se coucher. Alors qu'elle ne voit rien, son pied droit glisse sur le sol, elle perd l'équilibre et tout son corps part en arrière. A cause de la vitesse à laquelle elle marchait, son corps virevolte de côté. Son genou droit heurte alors violemment le sol. Et c'est tout le corps de Mme P qui heurte le sol ; elle se retrouve allongée dos au sol, les bras étendus derrière la tête. Elle relève alors son buste et se tourne légèrement vers la droite, en prenant appui sur son bras pour tenter de se relever. En vain, elle se tourne sur la gauche, prend appui sur son bras, plie sa jambe gauche afin de prendre appui sur son genou. Elle pousse sur ses bras et son genou afin de se relever.

FIGURE 2. Scénario de chute utilisant la méthode des *Personas*

modalités de la chute : quels sont les différents membres du corps mobilisés (membres supérieurs/inférieurs), la direction et l'amplitude de chaque mouvement (bras vers le haut/ bas, corps à droite/gauche...), la vélocité (vitesse et direction du corps) ses réactions au sol (cherche à se relever, rampe...) ainsi que les temps d'(in)action (ie. durée d'immobilisation au sol). La Figure 1 montre le script d'une chute typique. Des phrases ou mots d'alerte ont également été identifiés à différents moments de la chute : « AAhhh, zuuut, qu'est ce qui m'arrive, Oh merde, merde... ». (voir Figure 1)

Par la méthode des *Personas*, nous avons aussi décrit finement par des scénarios, les contextes de réalisation et de déroulement des chutes, qui visaient deux objectifs : (i) d'une part avoir une représentation assez claire et précise des différents chutes identifiées (28) de façon à pouvoir les catégoriser et à les regrouper dans trois grandes classes discriminantes : en l'occurrence tomber (s'affaler au sol, depuis une posture statique), glisser (déséquilibre du corps dans la continuité d'un déplacement) et trébucher (perte de verticalité induite par un choc contre un obstacle), (ii) d'autre part, faire rejouer ces scénarios par 20 sujets dans un contexte d'expérimentation de type living lab, afin d'affiner les paramétrages de détection et de valider ses algorithmes de repérage. Il était ainsi plus aisé de reconstituer les scènes d'incident en adaptant le cadre de vie selon les différents scénarios des chutes à simuler (tapis écorné, tomber d'un canapé). (voir Figure 2)

IV. LE SYSTÈME CIRDO

A. Architecture globale

Les différents composants du système CIRDO et leur connexion sont décrits sur la Figure 3. Le flux de données issus de la caméra est analysé en continu par la caméra vidéo en vue d'en déduire les postures et mouvements de la personne et de détecter les chutes. Quant au module audio, il analyse en permanence le flux de données issu du microphone afin d'en extraire en permanence les segments de parole ou de son produits dans la maison. Ces deux types d'événements sont envoyés ensuite au module de fusion et décision en charge de détecter les situations à risque qui décide si il y a lieu d'émettre un appel d'urgence grâce au système "e-lío"¹.

B. Module audio CirdoX : reconnaissance automatique de la parole âgée

Le module audio opère en tâche de fond sans que la personne n'ait à intervenir pour le configurer. Le traitement

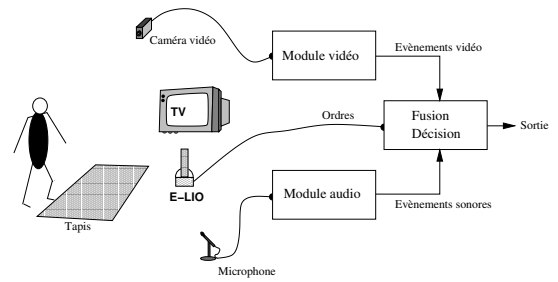


FIGURE 3. Le système CIRDO

se décompose en quatre étapes : (1) détection de l'événement sonore dans le bruit de fond, (2) discrimination entre parole et son de la vie courante, (3) reconnaissance de la parole prononcée et (4) filtrage des paroles relatives à une détresse pour envoi au système de fusion. Les phrases qui ne sont pas relatives à une situation de détresse sont immédiatement effacées pour respecter l'intimité de la personne. Les deux premiers étages font appel à une détection d'énergie sur la transformée en ondelettes du signal et à la classification du signal entre classe de parole et classe de son [17]. Le contexte applicatif rend la réalisation du troisième étage difficile car le vieillissement physique et cognitif des personnes fait que la voix âgée est particulièrement difficile à traiter par les systèmes de reconnaissance automatique de la parole (RAP) qui doivent être adaptés en conséquence [18] [19]. Il convient aussi de prendre en compte le fait qu'une personne appelant à l'aide lorsqu'elle est en situation de détresse aura une voix affectée par l'émotion qu'elle éprouve dans cette circonstance, ce qui va là aussi compliquer la tâche du système de reconnaissance [20]. Par ailleurs, l'analyse sonore dans un milieu non contrôlé comme le domicile impose de relever d'autres défis [21].

La réalisation d'un système de reconnaissance automatique de la voix âgée nécessiterait l'enregistrement d'un corpus comprenant une centaine d'heures de paroles prononcées par plus de cent locuteurs. Étant donné les difficultés rencontrées auprès de ces personnes, état de fatigue et difficultés à se déplacer vers un studio d'enregistrement, nous avons décidé d'enregistrer un corpus appelé AD80 de taille plus modeste, mais suffisante pour permettre d'adapter un système existant à la voix âgée, l'étude correspondante est décrite avec plus de précisions dans [22]. Le Taux d'Erreur de Mot (TEM) qui était au départ de 43,5% a pu être ainsi ramené à 14,5%.

| | | |
|----|-----------|---|
| 1 | Détresse | Qu'est-ce qui m'arrive ! |
| 2 | Détresse | Oh là ! |
| 3 | Détresse | Oh là ! Je saigne ! Je me suis blessé ! |
| 4 | Détresse | Aïe ! J'ai mal ! |
| 5 | Détresse | Je peux pas me relever ! |
| 6 | Détresse | Aidez-moi ! |
| 7 | Détresse | Au secours ! |
| 8 | Détresse | Je me sens mal ! |
| 9 | Détresse | Je suis tombé ! |
| 10 | Détresse | Du secours s'il vous plaît ! |
| 11 | Domotique | Appelle quelqu'un e-lío ! |
| 12 | Domotique | e-lío appelle les secours ! |
| 13 | Domotique | e-lío appelle ma fille ! |
| 14 | Anodin | Bonjour madame ! |
| 15 | Anodin | Il fait soleil. |
| 16 | Anodin | J'ai ouvert la porte. |

TABLE I. EXEMPLES DE PHRASES DU CORPUS AD80

1. <http://www.technosens.fr/>

Le contenu des phrases prononcées a été déterminé pour partie à partir de l'étude de terrain et pour partie en prolongement d'études précédentes [23], il comprend donc à la fois des phrases de détresse, des ordres domotiques et des phrases anodines. Quelques exemples sont donnés dans le Tableau I. Le corpus de parole lue AD80 a été enregistré par 95 locuteurs dont 43 personnes âgées (62 à 94 ans) et a une durée totale de 4h 42min. Le système de Reconnaissance Automatique de la Parole (RAP) utilisé est Sphinx3 [24], il utilise des modèles acoustiques appris sur le corpus BREF120 et ensuite adaptés à la voix âgée. Le modèle de langage utilisé est une combinaison du modèle obtenu à partir des phrases de détresse de AD80 (poids de 90%) et du modèle générique *Gigaword* (poids 10%). Cette combinaison permet de continuer à reconnaître les phrases prononcées lorsque le locuteur s'écarte des phrases type et d'éviter de les confondre avec des phrases de détresse [23].

Le filtrage des phrases de détresse se fait grâce au calcul de la distance de Levenstein entre l'hypothèse de reconnaissance et chacune des phrases de détresse (se reporter à la Table I). Ce système a été testé avec les enregistrements des 43 locuteurs âgés du corpus AD80, soit sur 2796 phrases de détresse ou d'ordre domotique et 3006 phrases anodines c'est à dire 2h 12 min de signal. La présence de phrases anodines est nécessaire pour vérifier que le système ne se déclenche pas en dehors des cas de détresse. Le TEM moyen était 14,5% pour les phrases de détresse et 87,5% pour les phrases anodines car le modèle de langage n'était pas adapté à ces dernières. Les résultats obtenus au point d'égale erreur (seuil=0,7) sont affichés sur la Table II, les taux de rappel, précision et F-mesure étant 88,8, 86,9 et 87,2%.

C. Module vidéo : détection automatique des chutes

L'acquisition vidéo dans un espace privé intérieur peut bénéficier de quelques conditions d'exploitation favorables : des caméras stables non soumises à des facteurs provoquant des vibrations ni à des phénomènes météorologiques directs. Toutefois, les scènes d'un environnement de vie quotidienne sont caractérisées par d'autres particularités qui augmentent les difficultés de traitements et d'interprétation des situations [25]. Parmi les difficultés les plus importantes, nous citerons : (i) le changement brusque d'éclairage et de luminosité (une lampe qui s'allume, un rideau qui bouge), (ii) la proximité des couleurs et des motifs ainsi que le manque de contraste entre les éléments du décor (tapisserie, rideaux...) et l'habillement du sujet d'intérêt induisant des phénomènes de camouflage, (iii) la diversité des sources d'éclairage –souvent artificielles– qui peuvent générer des ombres connectées aux objets et ainsi affecter la détection et l'identification du sujet d'intérêt, (iv) la diversité des objets (meubles, bibelots, tapis, etc.) présents dans l'environnement de vie (salon, cuisine, chambre...) et dont certains sont susceptibles d'être déplacés.

Compte tenu du contexte applicatif, on se situe dans le cas où les caméras sont fixes. Il est alors possible de procéder

| Seuil = 0.7 | Détresse | Anodin |
|-----------------------|-----------|-----------|
| $d \leq \text{seuil}$ | VP = 2472 | FP = 374 |
| $d > \text{seuil}$ | FN = 324 | VN = 2632 |

TABLE II. Test des appels de détresse.

au calibrage géométrique du système de vision ainsi qu'à la limitation du champ de vision. L'acquisition vidéo se fait en permanence, le module vidéo opère en tâche de fond, l'interprétation est réalisée de façon automatique et en temps réel sans l'intervention d'un opérateur. Dans ce contexte nous nous intéressons à la détection d'événements dangereux, dont le plus courant est la chute, à partir de séquences vidéo. Notre système se base sur l'extraction de la silhouette versus la séparation de l'arrière-plan. La silhouette extraite est utilisée pour obtenir des caractéristiques de mouvements utiles à la détection d'événements « à risque ». Le suivi du mouvement de cette silhouette nous permet d'identifier des cas de chute. Le traitement se décompose en quatre étapes : (1) apprentissage de l'arrière-plan, (2) extraction de la silhouette, (3) analyse de la posture et suivi du mouvement, (4) reconnaissance d'une posture de détresse/chute pour envoi au système de fusion. A l'instar du traitement audio, pour respecter l'intimité de la personne, les scènes qui ne sont pas relatives à une situation de détresse sont effacées.

Le modèle utilisé pour l'apprentissage de l'arrière-plan est une variante de mélange de gaussiennes ou MOG (*Mixture Of Gaussian*) adaptée à une représentation multi-niveaux : pixel, région, image qui est calculée en utilisant des informations de luminance et chrominance. Il s'agit d'un modèle bien adapté à la modélisation du fond d'image [26] et constitue un bon compromis entre précision des résultats et temps de calcul. Le modèle développé est une amélioration de celui proposé par Stauffer et Grimson [27]. Le modèle appris est mis à jour au cours du temps par un algorithme classique d'*Expectation-Maximisation* (EM).

L'extraction de la silhouette se fait se fait d'abord au niveau du pixel par l'estimation d'une distance de Mahalanobis calculées sur tous les canaux. Un traitement au niveau région permet d'éliminer les éléments du premier plan, susceptibles de correspondre aux ombres ou aux changements rapides de lumière. Enfin l'analyse au niveau image permet de détecter les changements globaux dans la scène tout en assurant la conservation de la silhouette jusqu'à la fin du processus d'adaptation. La silhouette est ensuite segmentée en six parties : la tête, le tronc, les quatre membres².

En s'appuyant sur une modélisation de type *blobs*, nous réalisons le suivi des différentes parties du corps identifiées dans l'image. La détection de chute est réalisée par une identification de la posture de la silhouette de la personne filmée [28]. L'algorithme de suivi/identification s'appuie sur un processus itératif markovien qui calcule une énergie. Enfin, la fonction d'énergie obtenue est ensuite minimisée par un algorithme de coupe de graphe [29]. Cette étape est renforcée par le suivi de la tête et la surveillance de sa trajectoire [30].

V. ENREGISTREMENT D'UN CORPUS EN CONDITIONS RÉALISTES

Des expériences ont été menées dans une salle configurée pour ressembler à un petit salon standard (chaises, tapis, table basse, TV...), elles visaient à la collecte d'un échantillon significatif de données pour apprendre et tester les algorithmes

2. les membres inférieurs peuvent être fusionnés pour gérer différents types d'habillement

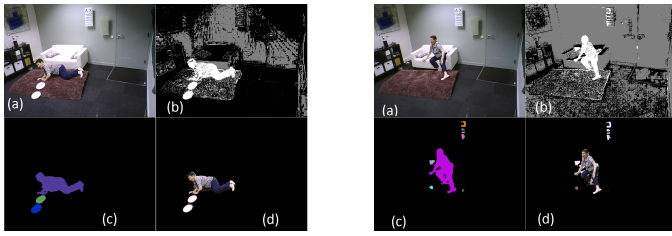


FIGURE 4. Résultat de l'analyse vidéo : détection d'une chute, situation de détresse (difficulté de se lever) a) acquisition initiale, b) discrimination fond/premier plan, c) extraction de la silhouette d) identification de la situation

en demandant à un utilisateur de rejouer les scénarios définis lors de l'étude de terrain.

A. Protocole expérimental

Le protocole était le suivant. Chaque participant était introduit au contexte de la recherche et a été invité à signer un formulaire de consentement. Il a ensuite été invité à jouer cinq types de chutes choisis parmi les 28 situations à risque identifiées (glisser, trébucher, tomber dans une position fixe et être sur le canapé en position de hanche bloquée). La Figure 4 montre un participant simulant une chute. Ces situations ont été choisies parce qu'elles étaient représentatives des chutes à la maison et parce qu'elles pouvaient être jouées en toute sécurité par les participants. Deux autres scénarios, appelés « Vrais-Négatif » (VN), ont été ajoutés pour tester la détection automatique des chutes. Le premier VN consistait à ramasser rapidement un magazine posé sur le sol (proche d'une situation de chute), tandis que la deuxième était d'essayer d'attraper une télécommande sur une table de café lorsque la personne est assise sur le canapé (situation proche de celle dans laquelle la personne a une hanche bloquée).

Avant de jouer une scène, l'expérimentateur précisait le contexte (lieu, moment de la journée...), l'activité (ce que la personne était en train de faire et ce qu'elle allait faire avant que l'incident ne se produise) et le geste et les mots à produire avant, pendant et après la chute. Ensuite, la personne devait répéter plusieurs fois la scène avant de commencer les enregistrements.

B. Configuration matérielle

La pièce expérimentale et le matériel utilisé sont représentés sur la Figure 5. En ce qui concerne l'analyse de l'audio, nous avons utilisé deux microphones sans fil Sennheiser : un EW-300-G2 ME2 fixé au plafond de la pièce, et un SKM-300-G2 posé sur un meuble à proximité du participant. Un enregistrement a été réalisé sur la durée de chaque expérience par un ordinateur PC équipé du logiciel StreamHIS et d'une carte National Instruments PCI-6220 à 8 canaux. Un haut-parleur dans la salle de contrôle permettait aux expérimentateurs de suivre la progression de la scène. Le logiciel CirdoX a été installé sur un ordinateur situé dans la salle d'expérimentation et connecté à un microphone unidirectionnel à condensateur Sennheiser. Il était d'autre part relié au système e-lío placé à côté. En outre, l'acquisition audio était également faite par le microphone d'un système Kinect.

Pour l'analyse vidéo, nous avons utilisé deux caméras de type webcam (Sony PSeye : acquisition en 640x480 60Hz et

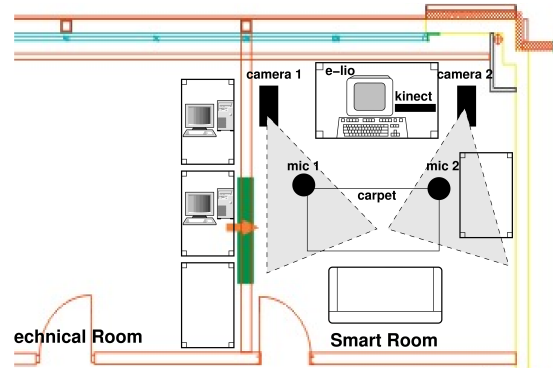


FIGURE 5. Le local d'expérimentation et ses équipements

connectées USB 2.0) fixées sur le mur face au canapé et une caméra à profondeur (Microsoft Kinect). La synchronisation et l'enregistrement des flux vidéos ont été réalisés sur un ordinateur PC (processeur intel i3 3.2GHz, 4Go Ram, carte graphique NVidia GeForce GTS 450 512Mo).

C. Corpus

Les participants ciblés étaient des personnes âgées qui étaient encore en mesure de jouer le scénario défini à la section III. Le recrutement d'une telle population s'avérant difficile et étant donné que le but était surtout de recueillir des données réalistes, mais pas nécessairement réelles, nous avons aussi recruté des personnes de moins de 60 en les invitant à porter un équipement (voir Figure 4) permettant de réduire leur mobilité, leur vision et leur audition afin de simuler les conditions dans lesquelles se trouve une personne âgée. Au total, ce sont 13 personnes jeunes (de 16 à 52 ans, 32 ans en moyenne, 7H, 6F) et 4 personnes âgées (de 61 à 83 ans, 68,5 en moyenne, 2H, 2F) qui ont participé.

Le détail du corpus est donné dans le tableau III, C1 à C4 sont les scénarios de chute, B celui de la hanche bloquée et F1, F2 les vrais négatifs. La durée totale des enregistrements est égale à 1h 55min, avec un total de 162 scénarios joués.

| No | Âge | Scénarios | | | | | | | Total | Durée |
|-----|-----|-----------|----|----|----|---|----|----|-------|----------|
| | | C1 | C2 | C3 | C4 | B | F1 | F2 | | |
| J01 | 30 | 1 | 2 | 1 | 1 | 1 | 1 | 1 | 8 | 8min 40s |
| J02 | 24 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 | 4min 35s |
| J03 | 29 | 1 | 1 | 1 | 3 | 1 | 1 | 2 | 10 | 6min 30s |
| J04 | 44 | 1 | 1 | 1 | 2 | 1 | 1 | 1 | 8 | 5min 54s |
| J05 | 16 | 1 | 1 | 1 | 3 | 2 | 1 | 2 | 11 | 8min 50s |
| J06 | 16 | 1 | 1 | 1 | 1 | 1 | 1 | 1 | 7 | 5min 07s |
| J07 | 52 | 1 | 1 | 1 | 1 | 1 | 1 | 2 | 8 | 5min 17s |
| J08 | 28 | 1 | 1 | 2 | 1 | 1 | 1 | 2 | 9 | 7min 04s |
| J09 | 52 | 3 | 1 | 1 | 1 | 1 | 1 | 2 | 10 | 6min 48s |
| J10 | 23 | 1 | 1 | 2 | 3 | 1 | 1 | 1 | 10 | 5min 50s |
| J11 | 40 | 2 | 1 | 2 | 3 | 1 | 1 | 1 | 11 | 7min 31s |
| J12 | 40 | 1 | 1 | 1 | 2 | 2 | 1 | 2 | 10 | 8min 01s |
| J13 | 25 | 2 | 2 | 1 | 1 | 1 | 1 | 1 | 9 | 5min 54s |
| A01 | 83 | 1 | 1 | 3 | 2 | 3 | 2 | 2 | 14 | 9min 07s |
| A02 | 64 | 1 | 2 | 1 | 2 | 1 | 1 | 2 | 10 | 6min 31s |
| A03 | 61 | 1 | 1 | 1 | 3 | 1 | 1 | 1 | 9 | 6min 00s |
| A04 | 66 | 1 | 1 | 2 | 1 | 2 | 2 | 2 | 11 | 7min 16s |

TABLE III. COMPOSITION DU CORPUS AUDIO/VIDÉO

VI. CONCLUSION

L'ambition du projet CIRDO est de sécuriser les seniors et de favoriser leur autonomie pour soulager les familles et

les aidants au quotidien en tenant compte des besoins et usages, et en adaptant, de manière évolutive et individuelle, les technologies les plus innovantes à leur usager. Par l'étude d'usage menée par les partenaires du domaine des sciences humaines, CIRDO fait le lien entre les besoins, les usages, les personnes et les technologies, et propose des solutions pour y répondre en fédérant divers partenaires. Le projet CIRDO se trouve à la croisée des sciences humaines et des sciences de l'ingénieur et s'adresse à une population en perte d'autonomie dans le but d'adapter les actions et réactions des équipements en fonctions de l'autonomie de son utilisateur. Sur le plan technique, CIRDO propose une solution originale qui s'appuie à la fois sur l'analyse vidéo et la reconnaissance de la parole et ceci dans le même processus. Pour respecter l'intimité de la personne et préserver les données privées, les traitements sont exécutés de façon automatique sans une intervention humaine. Le système développé viendra compléter la plateforme de services "e-lio", pour en faire un produit de télélien social, augmenté et automatisé en intégrant des services innovants.

REMERCIEMENTS

Cette étude a été financée par l'Agence Nationale de la Recherche et le Centre National pour la Santé et l'Autonomie (contrat ANR-2010-TECS-012). Les auteurs tiennent à remercier toutes les personnes âgées ou non âgées qui ont accepté de participer aux expérimentations. Ils remercient aussi Brigitte Meillon au montage des expérimentations.

RÉFÉRENCES

- [1] D. A. Norman and S. W. Draper, *User Centered System Design; New Perspectives on Human-Computer Interaction*. Hillsdale, NJ, USA : L. Erlbaum Associates Inc., 1986.
- [2] N. Noury, P. Rumeau, A. Bourke, G. ÓLaighin, and J. Lundy, "A proposal for the classification and evaluation of fall detectors," *[IRBM]*, vol. 29, no. 6, pp. 340 – 349, 2008.
- [3] M. Nyan, F. E. Tay, and E. Murugasu, "A wearable system for pre-impact fall detection," *Journal of Biomechanics*, vol. 41, no. 16, pp. 3475 – 3481, 2008.
- [4] I. Rosow, *Socialization to old age*. Berkeley and Los Angeles, California : University of California Press, 1974.
- [5] E. Cumming and W. E. Henry, *Growing Old : The Process of Disengagement*. New York : Basic Books, 1961.
- [6] A. Newel and P. Gregor, "User Sensitive inclusive Design," in *Actes du Colloque Interaction Homme Machine et Assistance*, Metz, France, 2001, pp. 18–20.
- [7] E. Ebersold, "Le champ du handicap, ses enjeux et ses mutations : du désavantage à la participation sociale," *Handicap*, vol. 94–95, pp. 149–164, 2002.
- [8] INPES, "Mieux prévenir les chutes chez les personnes âgées," Institut National de Prévention et d'Éducation pour la Santé, 381," La santé de L'homme, 2006.
- [9] C. J. Todd, C. Ballinger, and S. Whitehead, "A global report on falls prevention : reviews of socio-demographic factors related to falls and environmental interventions to prevent falls amongst older people living in the community," WHO, Tech. Rep., 2007.
- [10] C. Ballinger and S. Payne, "The construction of the risk of falling, among and by older people," *Ageing and Society*, vol. 22, no. 3, pp. 305–321, 2002.
- [11] M.-E. Bobillier-Chaumon, C. Michel, F. Tarpin-Bernard, and B. Croisile, "Can ict improve the quality of life of elderly adults living in residential home care units? from actual impacts to hidden artefacts," *Behaviour & Information Technology*, pp. 1–17, (to appear).
- [12] M. Popescu, Y. Li, M. Skubic, and M. Rantz, "An acoustic fall detector system that uses sound height information to reduce the false alarm rate," in *Proc. 30th Annual Int. Conference of the IEEE-EMBS 2008*, 20–25 Aug. 2008, pp. 4628–4631.
- [13] D. Charalampos and I. Maglogiannis, "Enabling human status awareness in assistive environments based on advanced sound and motion data classification," in *Proceedings of the 1st international conference on Pervasive Technologies Related to Assistive Environments*, 2008, pp. 1 :1–1 :8.
- [14] M. Hamill, V. Young, J. Boger, and A. Mihailidis, "Development of an automated speech recognition interface for personal emergency response systems," *Journal of NeuroEngineering and Rehabilitation*, vol. 6, no. 1, p. 26, 2009.
- [15] P. Milhorat, D. Istrate, J. Boudy, and G. Chollet, "Companionable Hands-free speech-sound," in *EUSIPCO (European Signal Processing Conference)*, aug 2012, pp. 1678–1682.
- [16] F. Portet, M. Vacher, C. Golanski, C. Roux, and B. Meillon, "Design and evaluation of a smart home voice interface for the elderly : acceptability and objection aspects," *Personal and Ubiquitous Computing*, vol. 17, pp. 127–144, 2013.
- [17] D. Istrate, E. Castelli, M. Vacher, L. Besacier, and J.-F. Serignat, "Information Extraction from Sound for Medical Telemonitoring," *IEEE Transactions on Information Technology in Biomedicine*, vol. 10, no. 2, pp. 264–274, 2006.
- [18] A. Baba, S. Yoshizawa, M. Yamada, A. Lee, and K. Shikano, "Acoustic models of the elderly for large-vocabulary continuous speech recognition," *Electronics and Communications in Japan, Part 2*, vol. 87, pp. 49–57, 2004.
- [19] R. C. Vipperla, S. Renals, and J. Frankel, "Ageing voices : The effect of changes in voice parameters on ASR performance," *EURASIP Journal on Audio, Speech, and Music Processing*, pp. 1–10, 2010.
- [20] F. Aman, V. Auberge, and M. Vacher, "How affects can perturb the automatic speech recognition of domotic interactions," in *Workshop on Affective Social Speech Signals*, Grenoble, France, 2013, pp. 1–5.
- [21] M. Vacher, F. Portet, A. Fleury, and N. Noury, "Development of Audio Sensing Technology for Ambient Assisted Living : Applications and Challenges," *International Journal of E-Health and Medical Communications*, vol. 2, no. 1, pp. 35–54, mar 2011.
- [22] F. Aman, M. Vacher, S. Rossato, and F. Portet, "Speech Recognition of Aged Voices in the AAL Context : Detection of Distress Sentences," in *The 7th International Conference on Speech Technology and Human-Computer Dialogue, SpeD 2013*, oct 2013, pp. 177–184.
- [23] M. Vacher, B. Lecouteux, and F. Portet, "Recognition of Voice Commands by Multisource ASR and Noise Cancellation in a Smart Home Environment," in *EUSIPCO (European Signal Processing Conference)*, aug 2012, pp. 1663–1667.
- [24] K. Seymore, C. Stanley, S. Doh, M. Eskenazi, E. Gouvea, B. Raj, M. Ravishankar, R. Rosenfeld, M. Siegler, R. Stern, and E. Thayer, "The 1997 CMU Sphinx-3 English broadcast news transcription system," *DARPA Broadcast News Transcript. and Understand. Workshop*, 1998.
- [25] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani, "Computer vision system for in-house video surveillance," *Vision, Image and Signal Processing, IEE Proceedings -*, vol. 152, no. 2, pp. 242–249, 2005.
- [26] R. Deeb, E. Desserée, and S. Bouakaz, "Real-time two-level foreground detection and person-silhouette extraction enhanced by body-parts tracking," in *Proc. SPIE*, vol. 83010R, 2012, pp. 1–8.
- [27] C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Computer Vision and Pattern Recognition, 1999. IEEE Computer Society Conference on.*, vol. 2, 1999.
- [28] R. Deeb, D. Elodie, and S. Bouakaz, "Extraction de silhouette et suivi des parties du corps dans un environnement non contrôlé," in *CORESA 2012*, 24–25 May. 2012.
- [29] C. Rother, V. Kolmogorov, and A. Blake, "'grabcut' : Interactive foreground extraction using iterated graph cuts," in *ACM SIGGRAPH 2004 Papers*, ser. SIGGRAPH '04. ACM, 2004, pp. 309–314.
- [30] R. Deeb, F. Lédée, E. Desserée, and S. Bouakaz, "Méthode robuste pour la détection de chute dans un environnement non contrôlé," in *RFIA 2012 - RFA 2012 - Atelier VISAGES*, 2012.