Résumé en activités du quotidien de vidéos issues de caméras portées pour l'aide au diagnostique de démences

Svebor Karaman, Jenny Benois-Pineau - LaBRI

Rémi Megret, Vladislavs Dovgalecs – IMS

Yann Gaëstel, Jean-Francois Dartigues - INSERM U.897

Université de Bordeaux

SUPPORTED BY ANR

Résumé en activités du quotidien

1. Contexte de l'étude
2. Vidéos portées
3. Analyse automatique des activités
    1. Segmentation temporelle
    2. Espace de description
    3. Reconnaissance des activités (MMC)
4. Expériences
5. Conclusions et perspectives

1. Contexte de l'étude

- IMMED: Indexation de données MultiMédia Embarquées pour le Diagnostic et le suivi des traitements des démences: LABRI, IMS, IRIT, ISPED/U879 Inserm ANR-09-BLAN-0165

  http://immed.labri.fr → Démos: Vidéo

- Société vieillissante:
  - → Impacte grandissant des troubles liés à l'âge
  - → Démences, maladie d'Alzheimer…
- Diagnostique précoce:
  - → Apporter à temps des solutions aux patients et aidants
  - → Retarder la perte d'autonomie et le placement en institut médicalisé

1. Contexte de l'étude

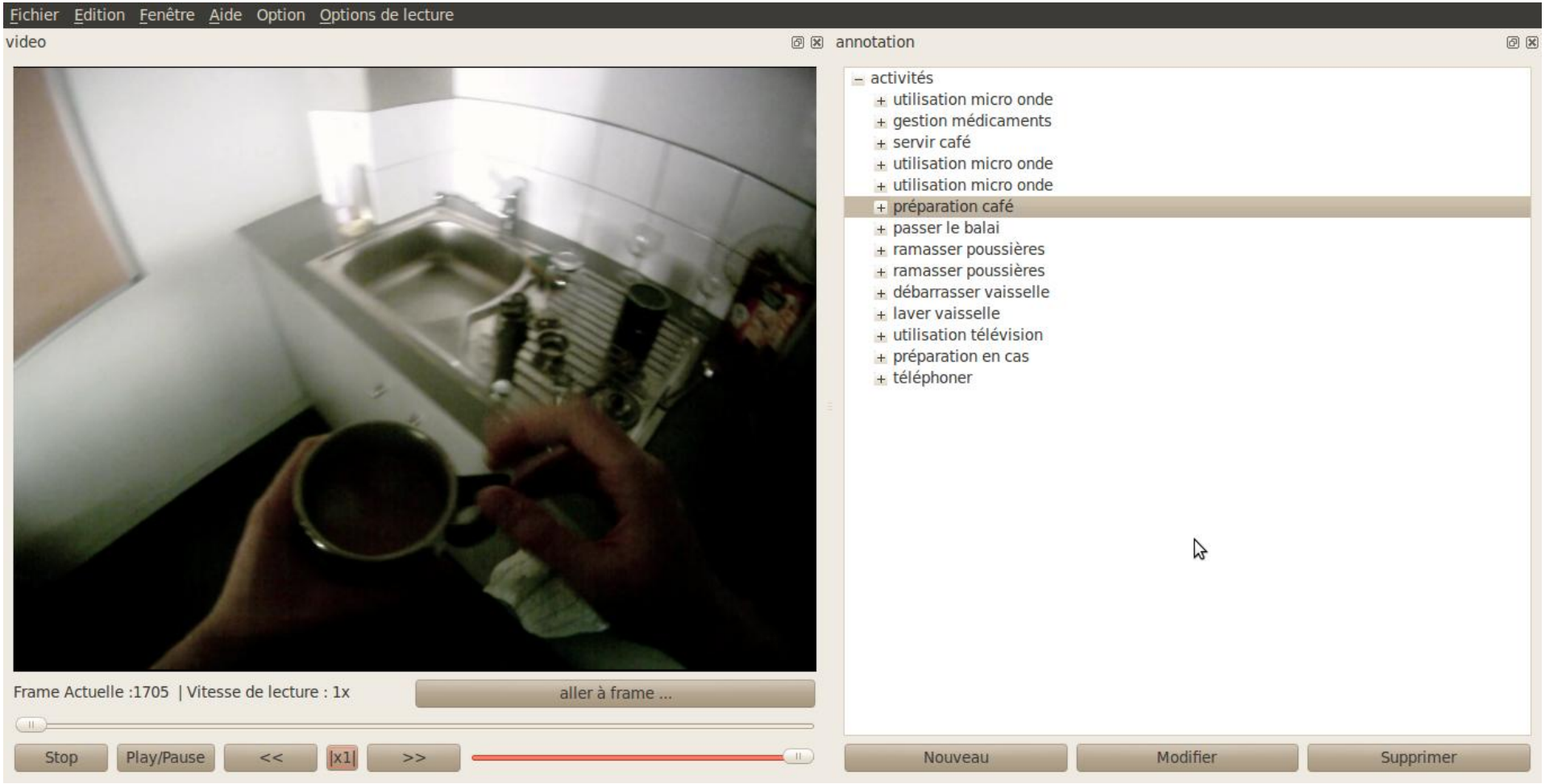
- Activités Instrumentales de la Vie Quotidienne (AIVQ)
  - Déclin dans AIVQ corrélé avec des démences futures

  PAQUID [Peres'2008]
- Analyse des AIVQ:
  - Questionnaire pour les patients et les aidants

  → Réponses subjectives
- Projet IMMED:
  - Observation des AIVQ par une caméra portée par le patient à domicile
- Observation objective de la situation et de l'évolution de la maladie
- Ajuster la thérapie pour chaque patient

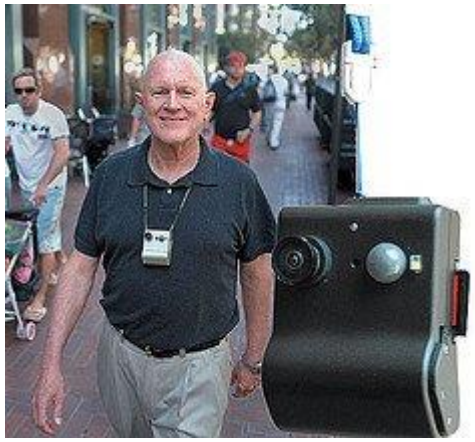
1. Contexte de l'étude

- Objectif: fournir un outil de visualisation ergonomique des activités du quotidien indexées automatiquement aux médecins qui résume les activités de la période d'observation



2. Vidéos portées

- Travaux apparentés:
- SenseCam
  - Images utilisées comme aide-mémoire



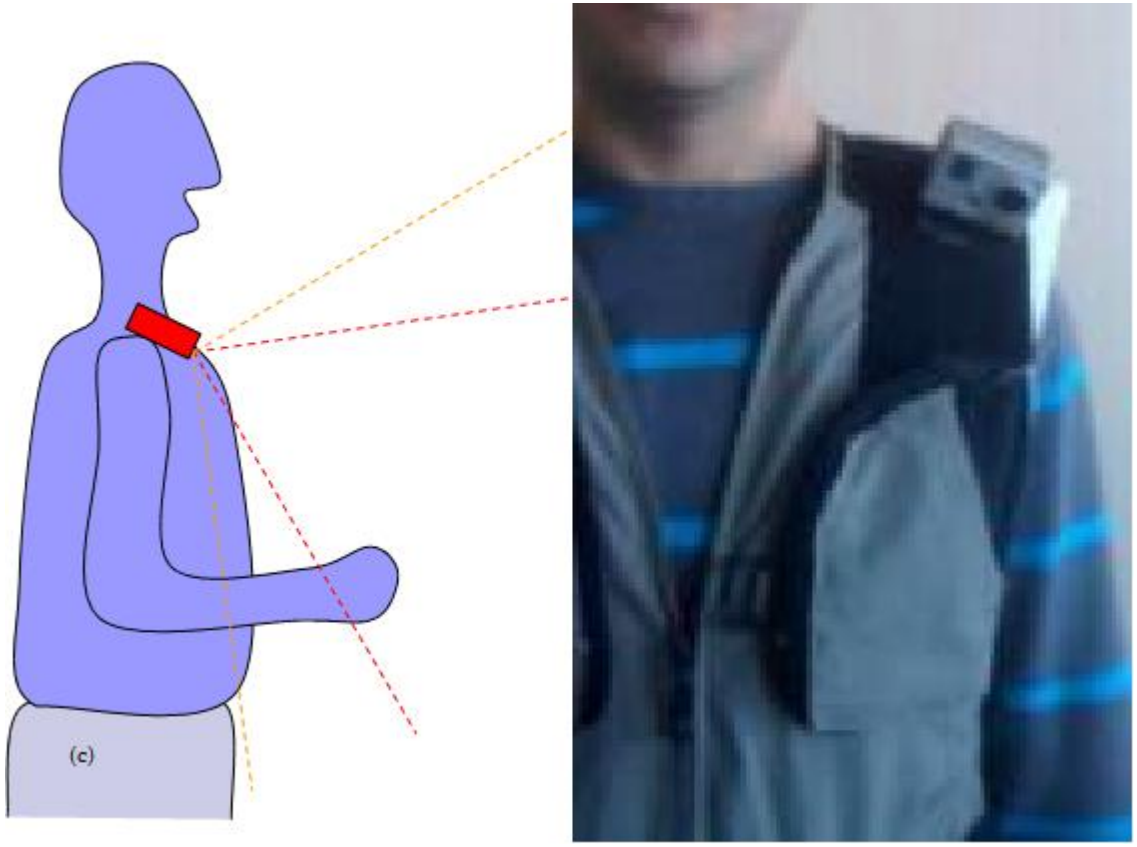
[Hodges et al.] “SenseCam: a Retrospective Memory Aid » UBICOMP’2006

- WearCam
  - Caméra positionnée sur la tête de jeunes enfants afin d'identifier de possibles déficiences comme l'autisme

[Picardi et al.] “WearCam: A Head Wireless Camera for Monitoring Gaze Attention and for the Diagnosis of Developmental Disorders in Young Children” International Symposium on Robot & Human Interactive Communication, 2007

2. Vidéos portées

- Dispositif d'acquisition vidéo



• Caméra grand angle positionné sur l'épaule

• Dispositif non intrusif et simple à utiliser

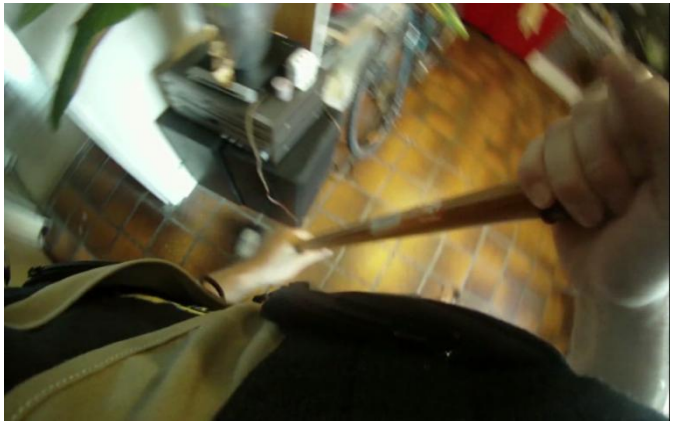
• Enregistrement des AIVQ: de 30 minutes jusqu'à 2,5 heures

2. Vidéos portées

4 exemples d'activités enregistrées avec cette caméra: video

Faire le lit, Laver la vaisselle, Passer le balai, Passer l'aspirateur






3.1 Segmentation temporelle

- Prétraitement: étape préliminaire à la structuration en activités
- Objectifs:
  - Réduire l'écart entre la quantité de données (images) et le nombre de détections visées (activités)
  - Associer une observation à chaque point de vue
- Principe:
  - Utiliser le mouvement global (ego motion) pour segmenter la vidéo en termes de points de vue
  - Une image-clé par segment: centre temporel
- Indexation brute pour une navigation au sein de la vidéo qui est un long plan séquence
- Résumé vidéo automatique de chaque nouvelle vidéo

3.1 Segmentation temporelle

- Modèle affine complet du mouvement global (a1, a2, a3, a4, a5, a6)

$$\begin{pmatrix} dx_i \\ dy_i \end{pmatrix} = \begin{pmatrix} a_1 \\ a_4 \end{pmatrix} + \begin{pmatrix} a_2 & a_3 \\ a_5 & a_6 \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix}$$

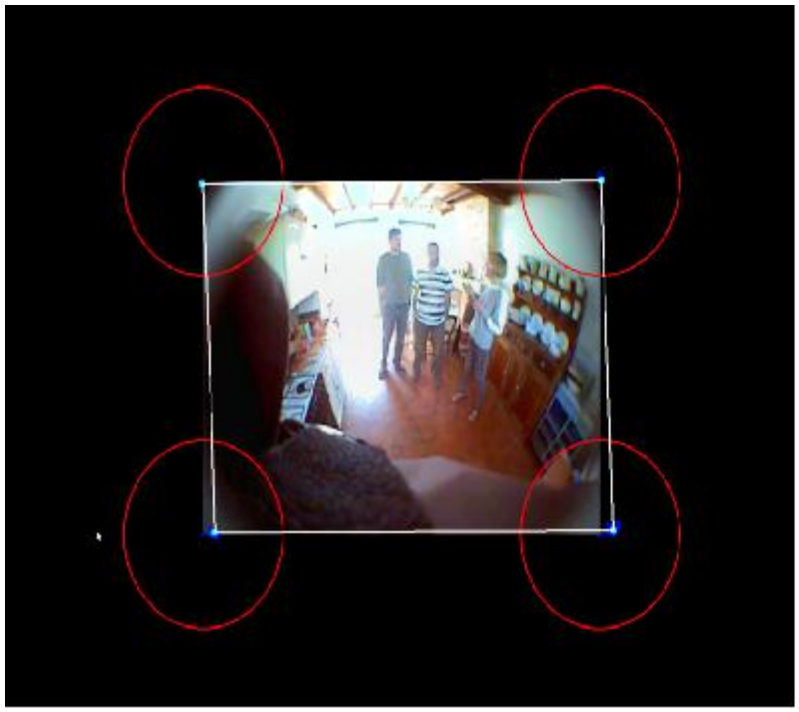
[Krämer et al.] Camera Motion Detection in the Rough Indexing Paradigm, TREC'2005.

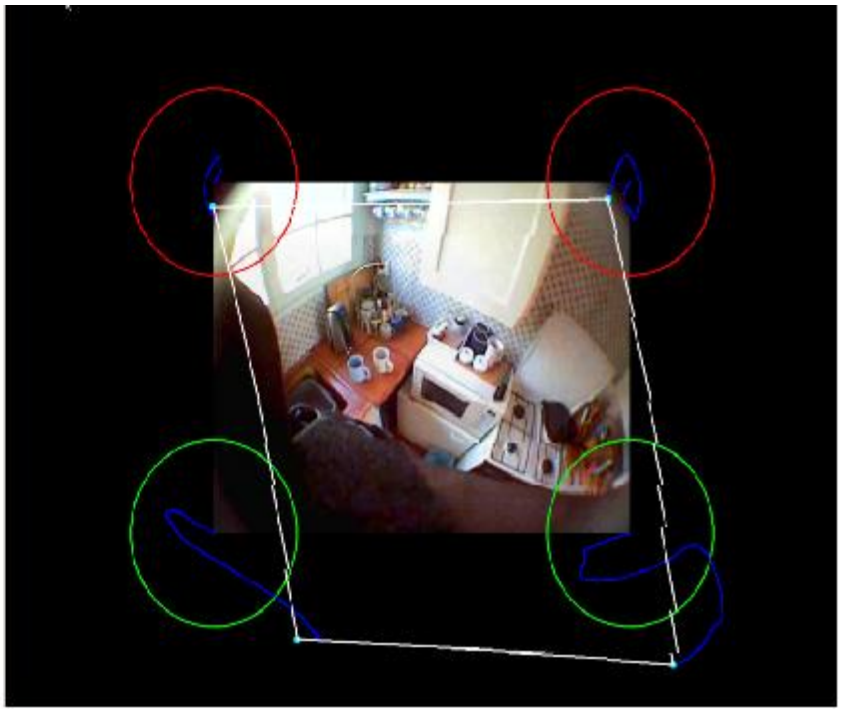
- Principe:
  - Trajectoires des coins à partir du modèle de mouvement global
  - Fin du segement: quand au moins 3 trajectoires de coins ont au moins une fois atteint une poisition hors limites

3.1 Segmentation temporelle

Seuil *t* définit comme un pourcentage p de la largeur de l'image *w.* p=0.2 … 0.25

$$t = p \times w$$

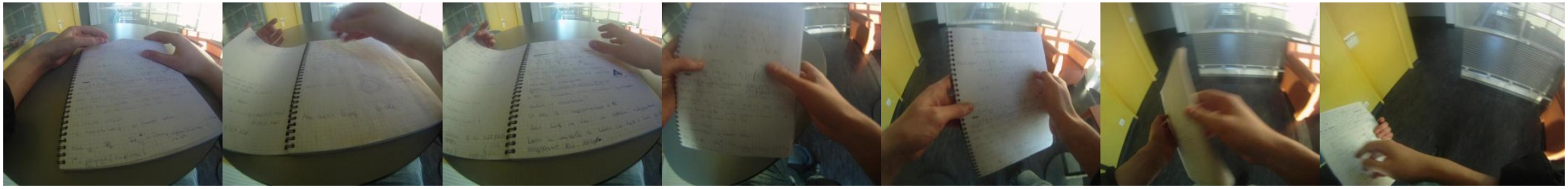




3.1 Segmentation temporelle
Résumé vidéo primaire

- 332 images-clés, 17772 images initialement

Résumé vidéo primaire (6 fps)



3.2 Résumé en activités . Espace de description

- Couleur: MPEG-7 Color Layout Descriptor (CLD)

  6 coefficients pour la luminance, 3 pour chaque chrominance

  - Pour un segment: CLD de l'image clé, x(CLD) $\in \Re^{12}$
- Localisation: vecteur de description adaptable à l'environnement personnel
- N_{home} localisations. x(Loc) $\in \Re^{N_{home}}$
- Localisation estimée pour chaque image
- Pour un segment: vecteur moyenné sur les images du segment

V. Dovgalecs, R. Mégret, H. Wannous, Y. Berthoumieu. "Semi-Supervised Learning for Location Recognition from Wearable Video". CBMI'2010, France.

3.2 Espace de description

- H_{tpe} histogramme en log de l'énergie des paramètres de translation
  - Caractérise la force du mouvement global
  - Distinguer les activités avec un fort ou un faible mouvement
- $N_e = 5$, $s_h = 0.2$. Vecteurs de description $x(H_{tpe},a_1)$ et $x(H_{tpe},a_4) \in \Re^5$

$$H_{tpe}[i] += 1 \quad if \quad \log(a^2) < i \times s_h \qquad for \quad i = 1$$

$$H_{tpe}[i] += 1 \quad if \quad (i-1) \times s_h \le \log(a^2) < i \times s_h \qquad for \quad i = 2..N_e - 1$$

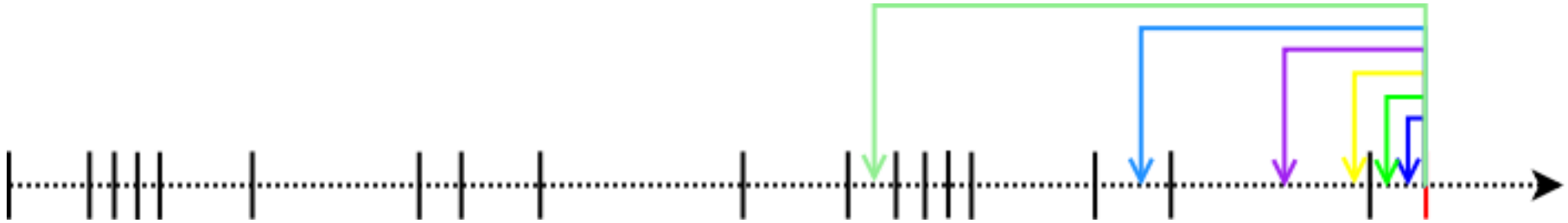
$$H_{tpe}[i] += 1 \quad if \quad \log(a^2) \ge i \times s_h \qquad for \quad i = N_e$$

- Histogrammes sont moyennés sur toutes les images du segment

	$x(H_{tpe}, a_1)$	$x(H_{tpe},a_4)$
Segment à faible mouvement	0,87 0,03 0,02 0 0,08	0,93 0,01 0,01 0 0,05
Segment à fort mouvement	0,05 0 0,01 0,11 0,83	0 0 0 0,06 0,94

3.2 Espace de description

- H_c: histogramme de coupures. La $i^{ème}$ classe de l'histogramme contient le nombre de coupures de la segmentation temporelle dans les 2^i dernières images



$H_c[1]=0$, $H_c[2]=0$, $H_c[3]=1$, $H_c[4]=1$, $H_c[5]=2$, $H_c[6]=7$

- Histogramme moyenné sur toutes les images du segment
- Caractérise l'historique du mouvement, la force du mouvement même en dehors du segment

2^6=64 images → 2s, 2^8=256 images → 8.5s

$x(H_c) \in \Re^6$ or $\Re^8$

3.2 Espace de description

- Fusion des vecteurs de description: fusion précoce
  - CLD $\rightarrow$ x(CLD) $\in \Re^{12}$
  - Mouvement
    - $x(H_{tpe}) \in \Re^{10}$
    - $x(H_c) \in \Re^6$ or $\Re^8$
  - Localisation: N_{home} entre 5 et 10.
    - $x(Loc) \in \Re^{Nhome}$
- Taille finale du vecteur de description: entre 33 et 40, si tous les descripteurs sont utilisés
- Par exemple:
  - $x \in \Re^{33} = (x(CLD), x(H_{tpe},a_1), x(H_{tpe},a_4), x(H_c), x(Loc))$

3.3 Reconnaissance des activités

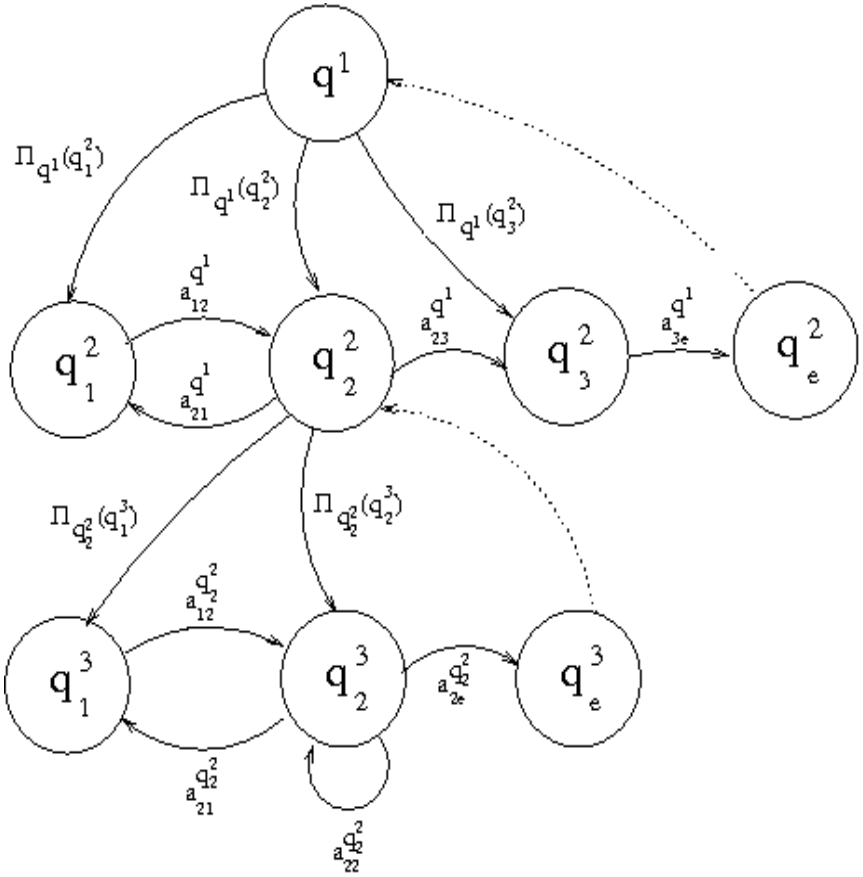
Modèles de Markov Cachés (MMC): efficaces pour la classification avec causalité temporelle

Une activité est complexe, difficilement modélisable par un seul état

MMC Hiérarchique? [Fine98], [Bui04]

- Niveaux multiples
- Coût calcul/apprentissage

- $Q^D=\{q_i^d\}$ ensemble des états
- $\Pi_{q_i^d}(q_j^{d+1})$ = probabilité initiale d'un fils q_j^{d+1} de l'état q_i^d
- $A_{ij}^{q^d}$ = probabilités de transitions entre les fils de q^d



3.3 Reconnaissance des activités

Un MMC hiérarchique à deux niveaux:

- Niveau supérieur:

transitions entre les activités

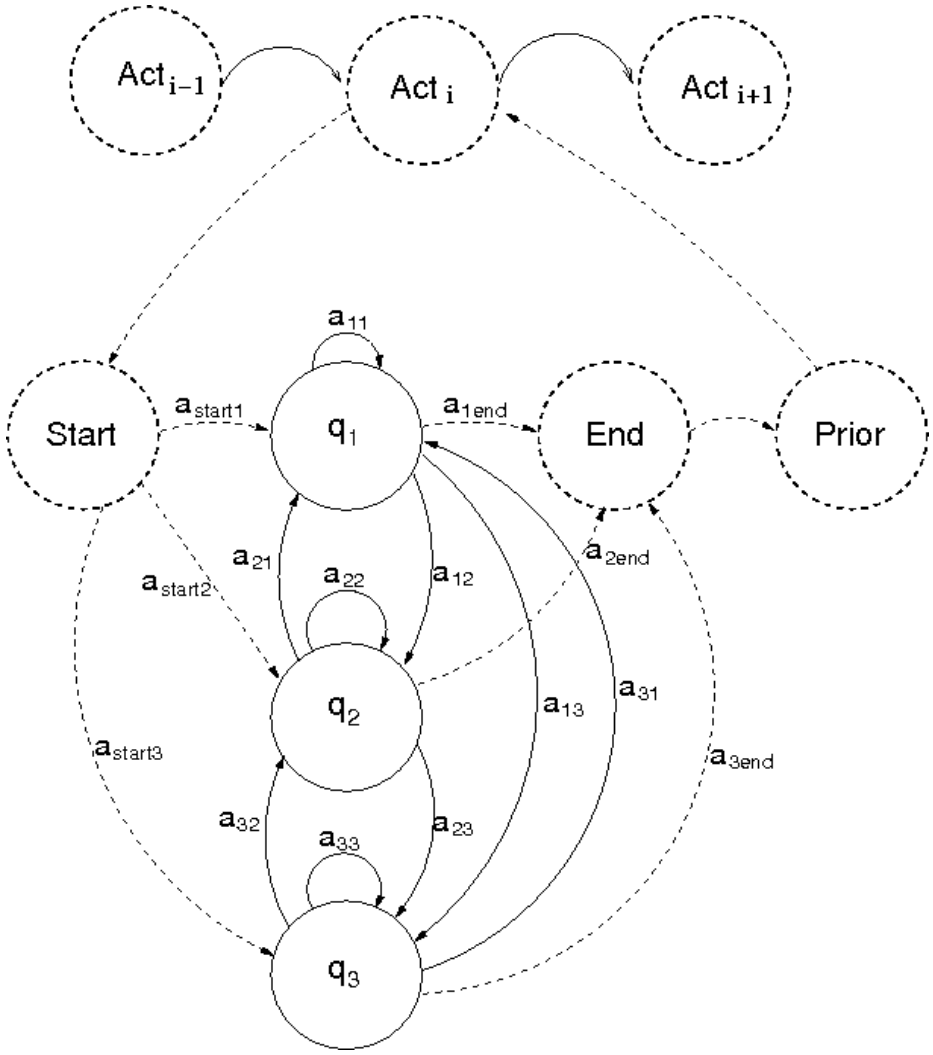
  - Exemples d'activités:

    Laver la vaisselle, Faire le thé,
    Passer l'aspirateur, Faire le café

- Niveau inférieur:

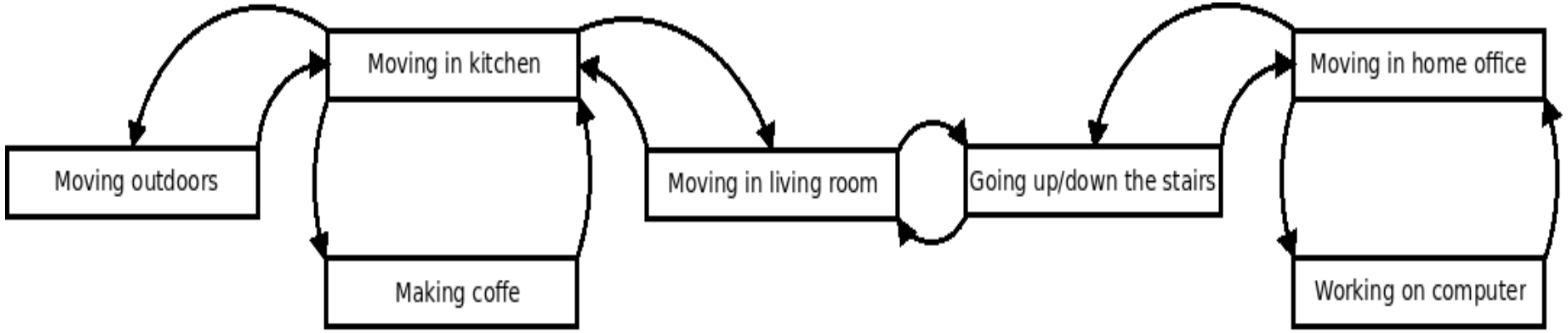
description de l'activité

  - Activité: MMC avec 3/5/7 états
  - Modèle d'observation: MMG
  - Probabilité à priori des activités



3.3 Reconnaissance des activités

- MMC du niveau supérieur
  - Connectivité du MMC peut être définit par les contraintes de l'environnement personnel



  - Transitions entre activités peuvent être pénalisées selon une connaissance a priori des transitions les plus fréquentes
  - Pas de réapprentissage des probabilités de transitions à ce niveau

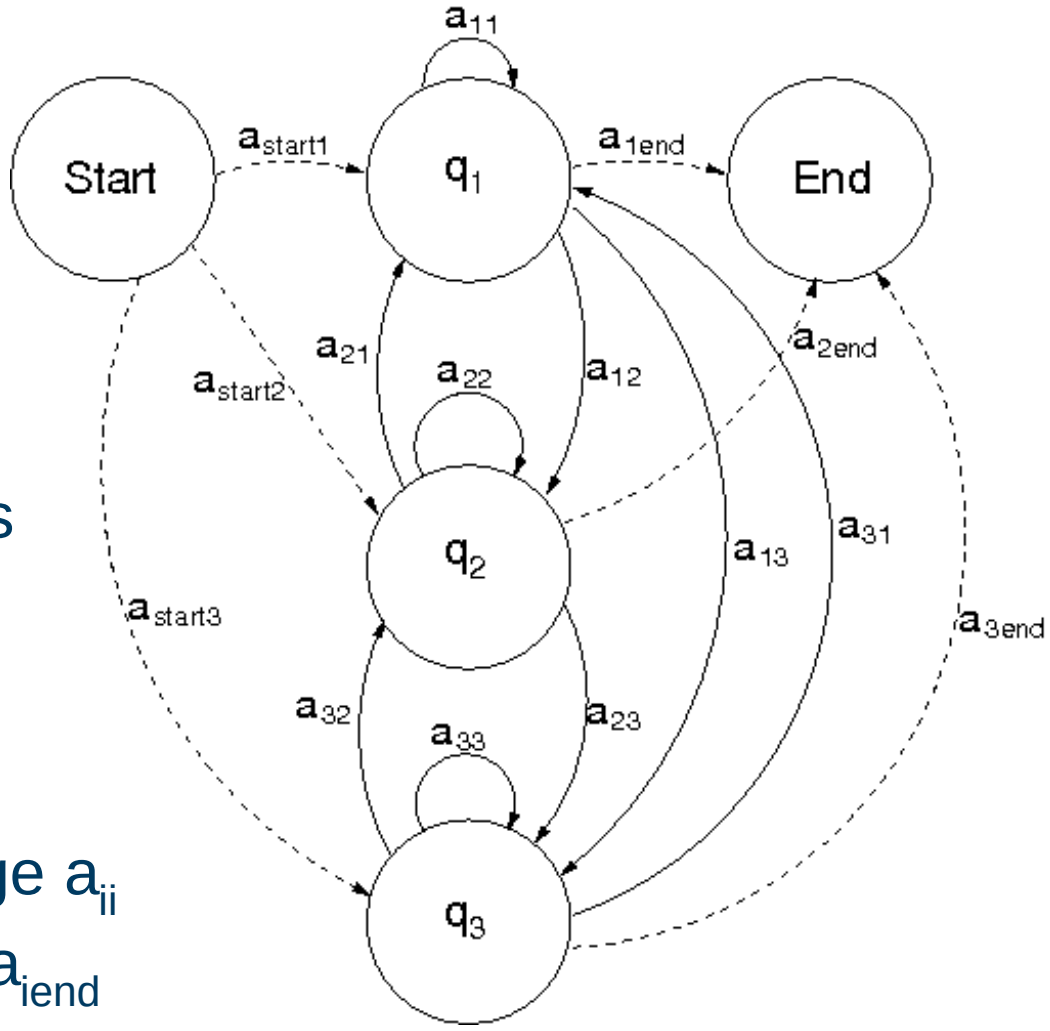
3.3 Reconnaissance des activités

MMC du niveau inférieur

- Start/End

→ Etat non émetteur

- Observation x seulement pour les états émetteurs q_i
- Probabilités de transitions et paramètres des MMG sont appris par l'algorithme de Baum-Welsh
- Nombre d'états fixés à priori
- Initialisation MMC:
  - Forte probabilité de bouclage a_{ii}
  - Faible probabilité de sortie a_{iend}



4. Résultats

- Pas de base de données disponible. Une vidéo. Total: 47489 images.
- Apprentissage sur 10% des images de chaque activité: 3974 images. Reconnaissance sur 310 segments
- Tests: nombre d'états du MMC et l'espace de descriptions étaient variables. Probabilités a priori égales.
- Meilleurs résultats:

Configuration	Nb States	F-Score	Recall	Precision
H_c + Localisation	5	**0.64**	0.66	**0.67**
H_c + CLD + Localisation	3	0.62	**0.7**	0.66

4. Résultats

- 7 activités:

  Déplacement bureau, Déplacement cuisine, Monter/Descendre l'escalier, Déplacement extérieur, Déplacement salon, Faire le café, Travailler sur l'ordinateur

- Confusion entre Déplacement bureau et Monter/Descendre l'escalier (1 et 3)

  → proximité

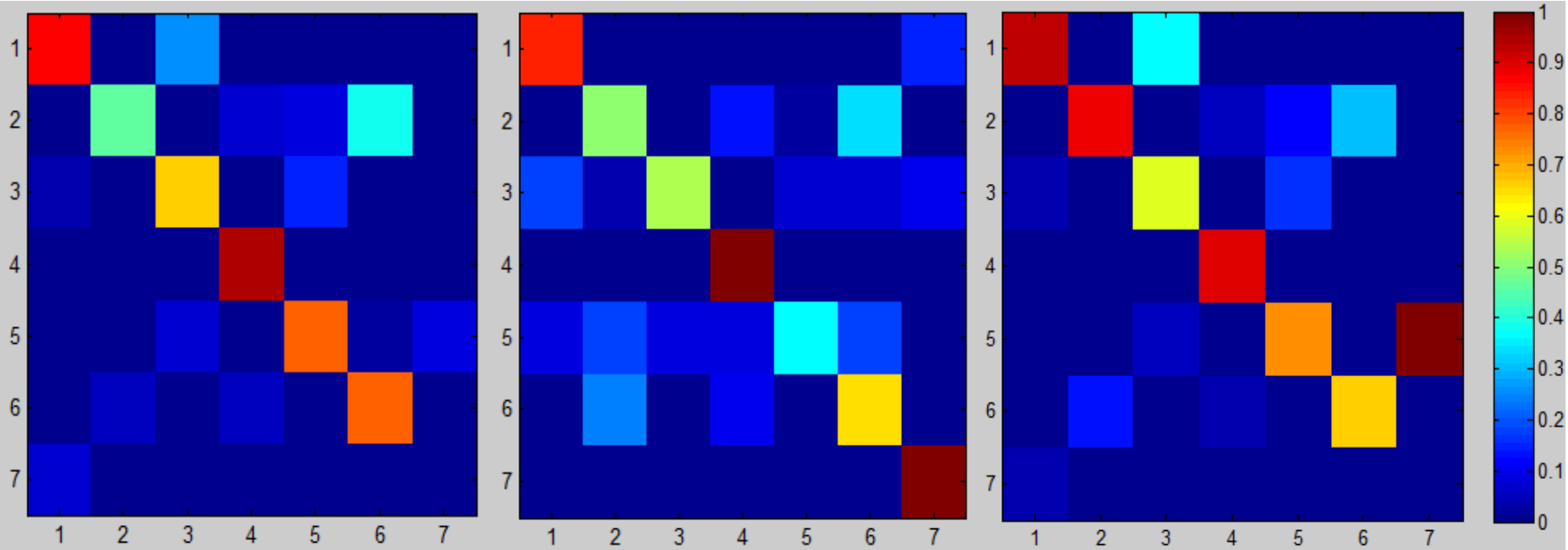
- Confusion entre Déplacement cuisine et Faire le café (2 et 6)

  → même localisation/environnement

4. Résultats

7 activités: Déplacement bureau, Déplacement cuisine, Monter/Descendre l'escalier, Déplacement extérieur, Déplacement salon, Faire le café, Travailler sur l'ordinateur

Matrices de confusion:



F-Score Recall Precision

5. Conclusions et perspectives

- Des méthodes de segmentation temporelle basée sur le mouvement et d'indexation d'activités humaines ont été présentées. Résultats encourageants
- Difficulté d'obtenir des vidéos (pas de base de données disponible) et coût d'annotation – psychologues INSERM
- Tests sur plus grand corpus: 6h de vidéos disponibles (travail en cours)
- Intégration de l'audio : résultats de classification du flux par IRIT (travail en cours)
- Descripteurs locaux et moyens niveaux
    - Analyse du mouvement local
    - Détection et reconnaissance d'objets
- Fusion tardive

Merci pour votre attention.

Questions?