

Extraction non supervisée de motifs temporels, multidimensionnels et hétérogènes

Application à la télésurveillance médicale à domicile

Florence Duchêne et Catherine Garbay



Laboratoire des Techniques de l'Imagerie, de la Modélisation et de la Cognition
TIMC-IMAG, UMR CNRS 5525, Grenoble, France

Introduction

Systeme d'apprentissage et de décision
pour l'évaluation de l'état de santé d'une personne

Dégradation de
l'état de santé

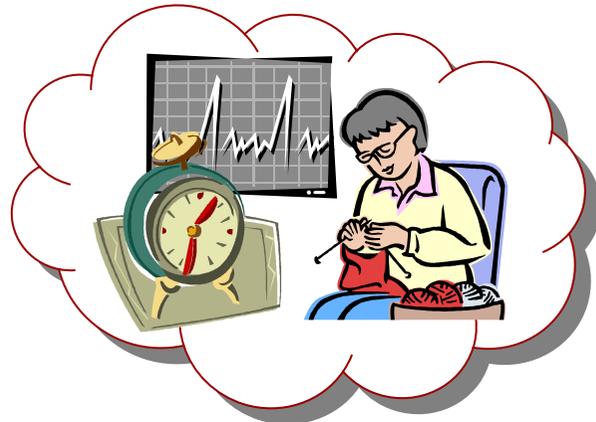
→ Troubles du comportement à domicile

→ *Évolution de la situation à long terme :*

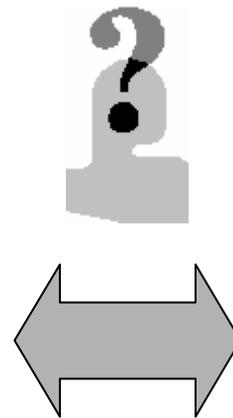
Étude des habitudes de la vie quotidienne



Apprentissage d'un
profil comportemental

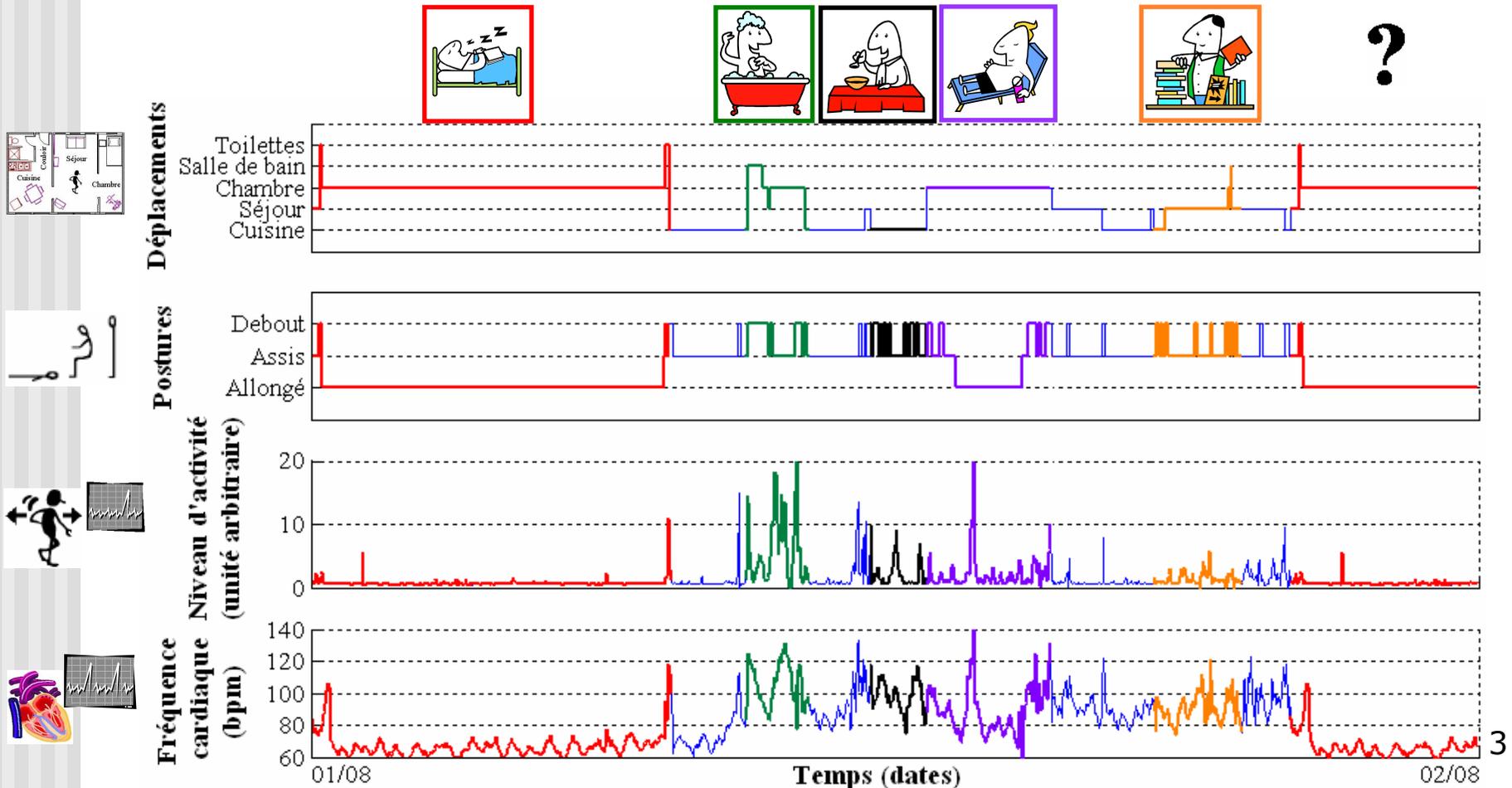
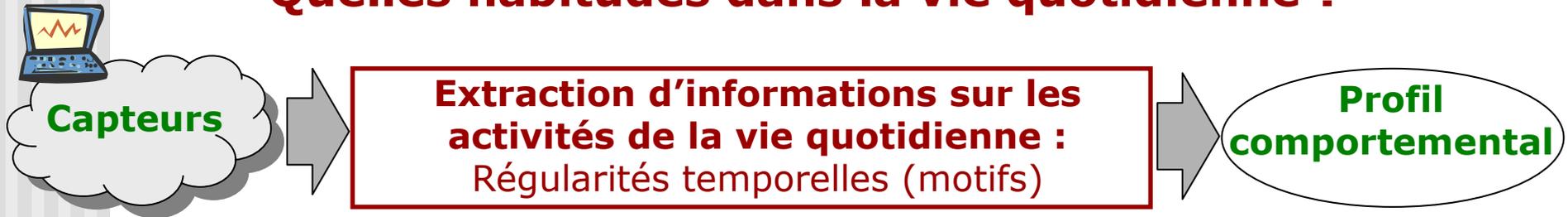


Décision par
comparaison au profil



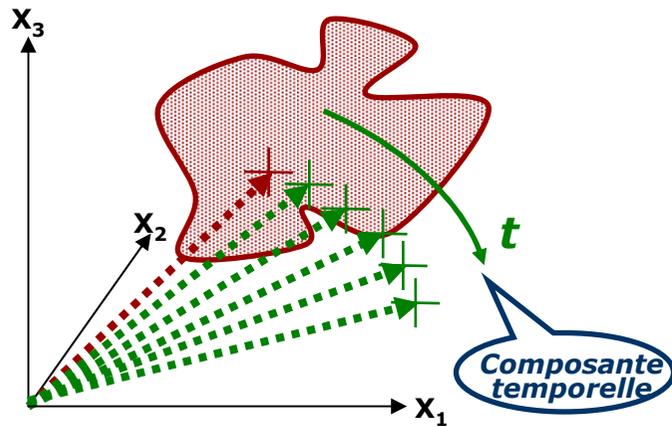
Contexte de l'apprentissage

Quelles habitudes dans la vie quotidienne ?



Caractéristiques des données

Observation de séquences de données temporelles



Espace multidimensionnel de données hétérogènes

Données issues de Capteurs



- Qualitatives
- Quantitatives



Observation de données complexes

- Dépendances mutuelles des paramètres observés
- Importants facteurs d'influence



Observation de séquences mixtes

- Sous-séquences représentatives de motifs
- Et d'autres de « non motifs »

Caractéristiques des motifs

Entre les représentants d'un motif multidimensionnel...



- Grande variabilité dans les valeurs
 - ▶ *Comportements humains*
 - ▶ *Observation à "haut niveau"*

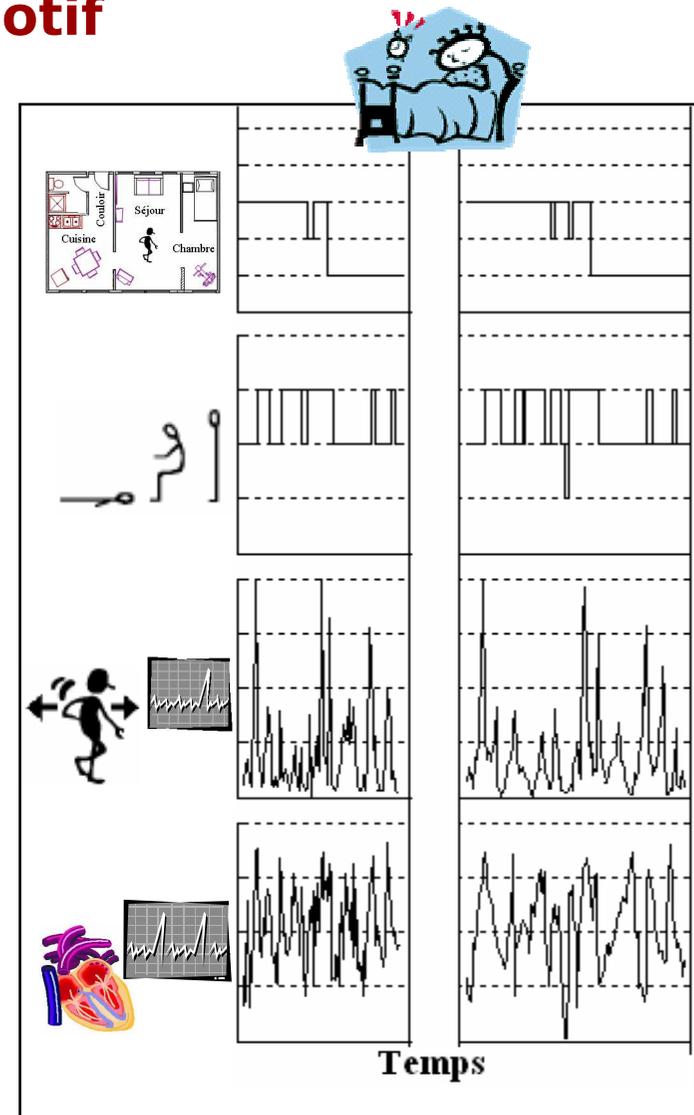


- « *Outliers* » (*Interruptions*)

- Translation dans le temps



- Déformation temporelle



Caractéristiques de l'apprentissage

Contraintes spécifiques au contexte



- Manque de données expérimentales
- Manque de connaissances *a priori*
 - ▶ Évolutions conjointes des paramètres
 - ▶ Situations critiques possibles
- Nécessaire personnalisation de l'analyse dans le contexte de chaque personne



➔ Apprentissage non supervisé

- **Pas d'exemples** des motifs présents dans les données
- Basé sur une **notion de similarité** entre sous-séquences

➔ Problèmes posés

- Identifier les sous-séquences récurrentes significatives
- Les catégoriser en motifs représentatifs des comportements récurrents au haut niveau de décision
- Définir une mesure de distance appropriée à la comparaison de séquences multidimensionnelles et hétérogènes

Méthodologie

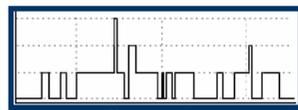
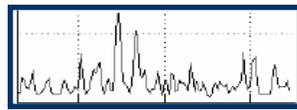
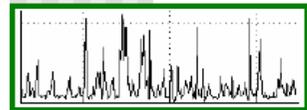
Systeme « classique » de reconnaissance

Représentation



- « Haut Niveau »

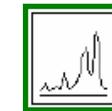
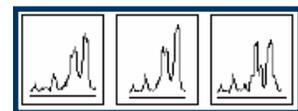
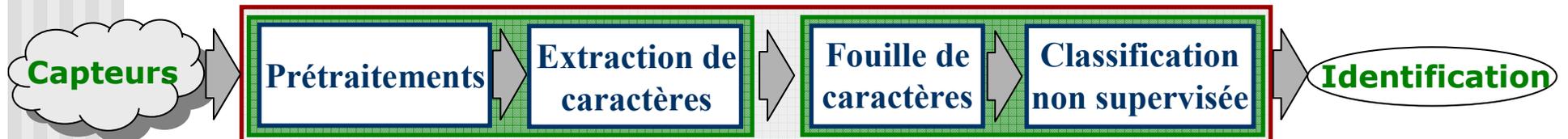
Abstraction



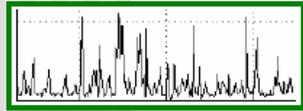
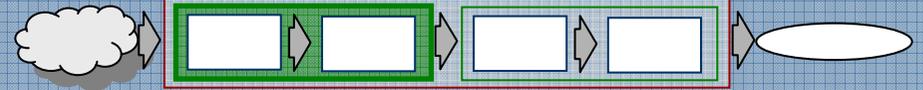
- Contexte non supervisé
- Grandes ensembles de données temporelles

Fouille de données temporelles

- Séquences « mixtes »



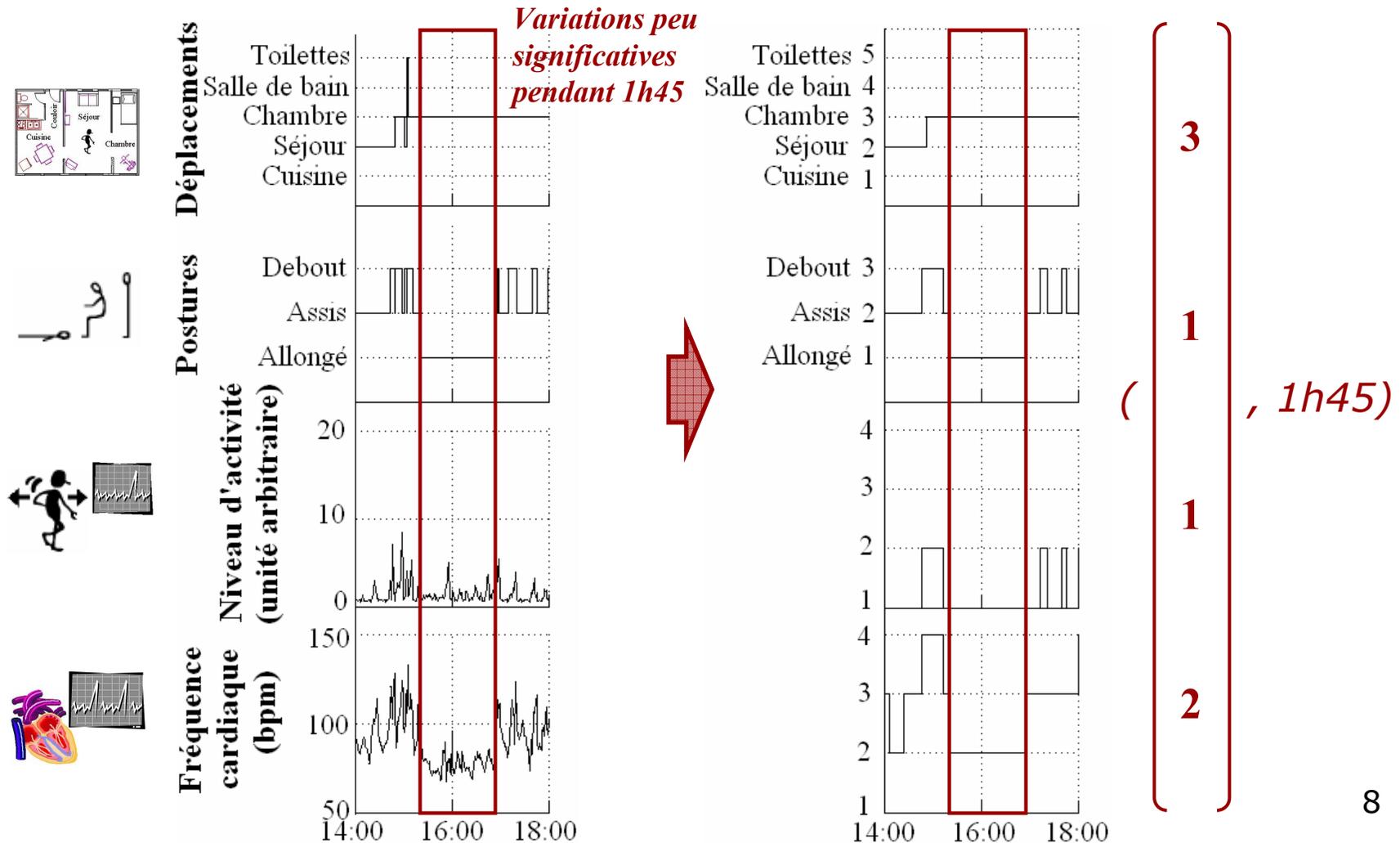
Abstraction



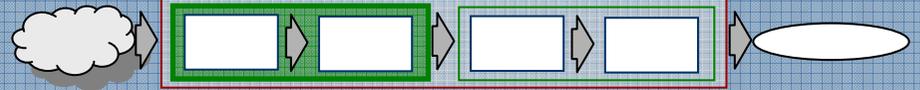
Données brutes

Notion d'interprétation de la représentation des données

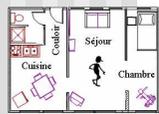
- Mise en évidence des situations « stationnaires »
- Description d'un signal par une succession de vecteurs discrets : *(symbole, durée)*



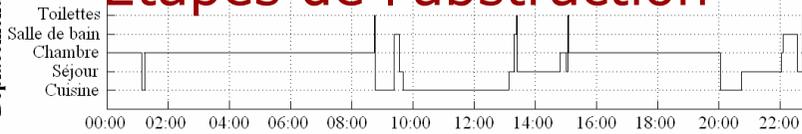
Abstraction



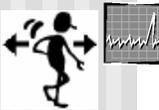
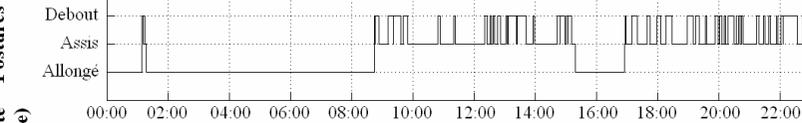
Étapes de l'abstraction



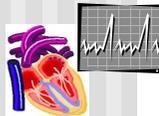
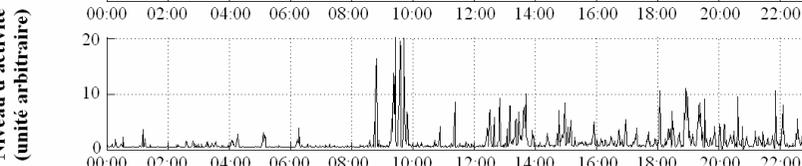
Déplacements



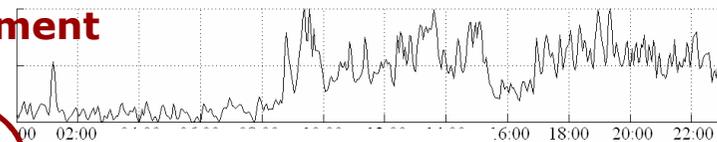
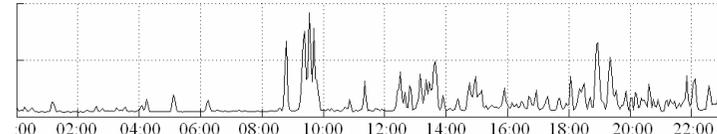
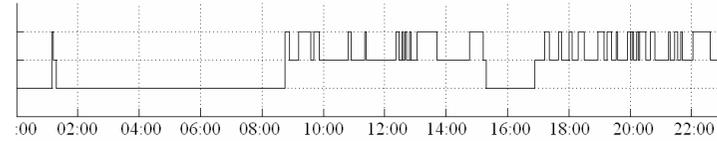
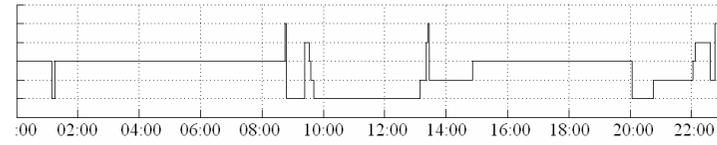
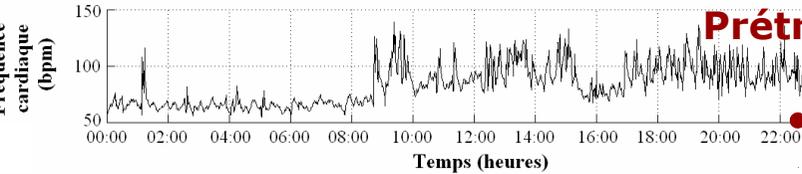
Postures



Niveau d'activité
(unité arbitraire)



Fréquence
cardiaque
(bpm)



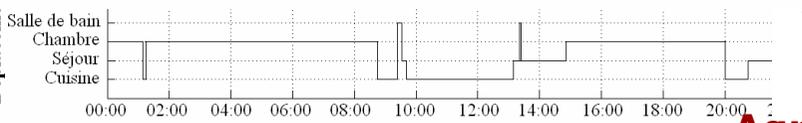
Prétraitement
(1)

(2) Discrétisation

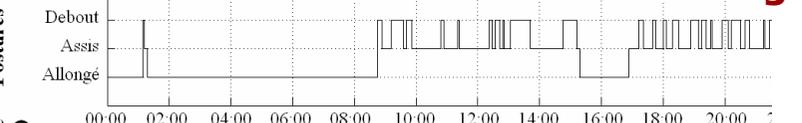
(3)

Agrégation

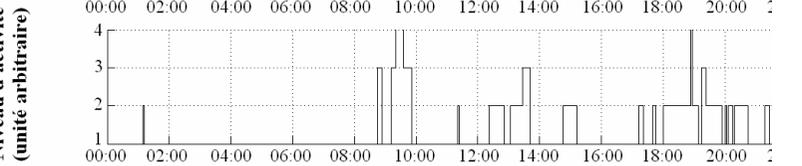
Déplacement



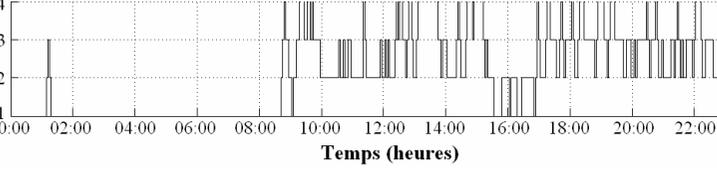
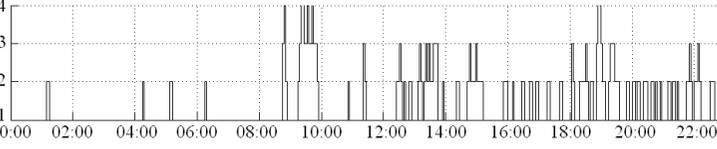
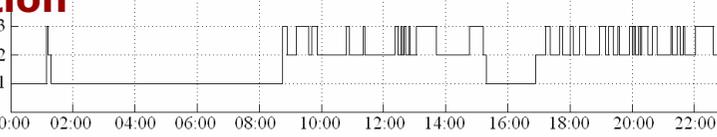
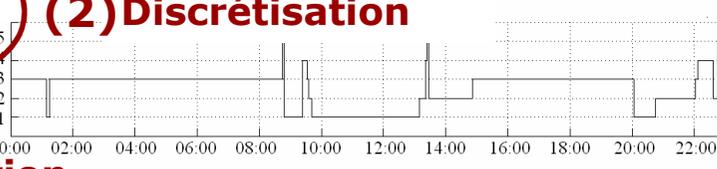
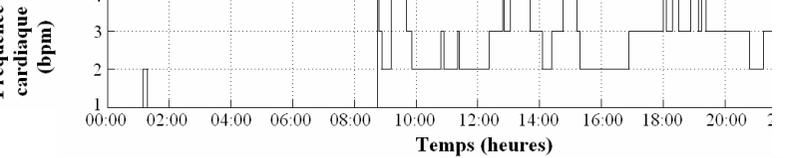
Postures

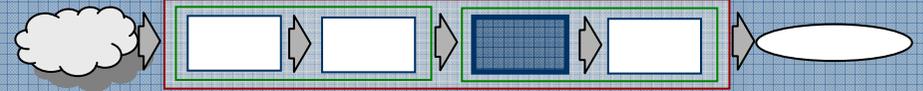


Niveau d'activité
(unité arbitraire)



Fréquence
cardiaque
(bpm)





Données abstraites

Extraction de sous-séquences récurrentes

- Réduire l'espace de recherche des instances de motifs

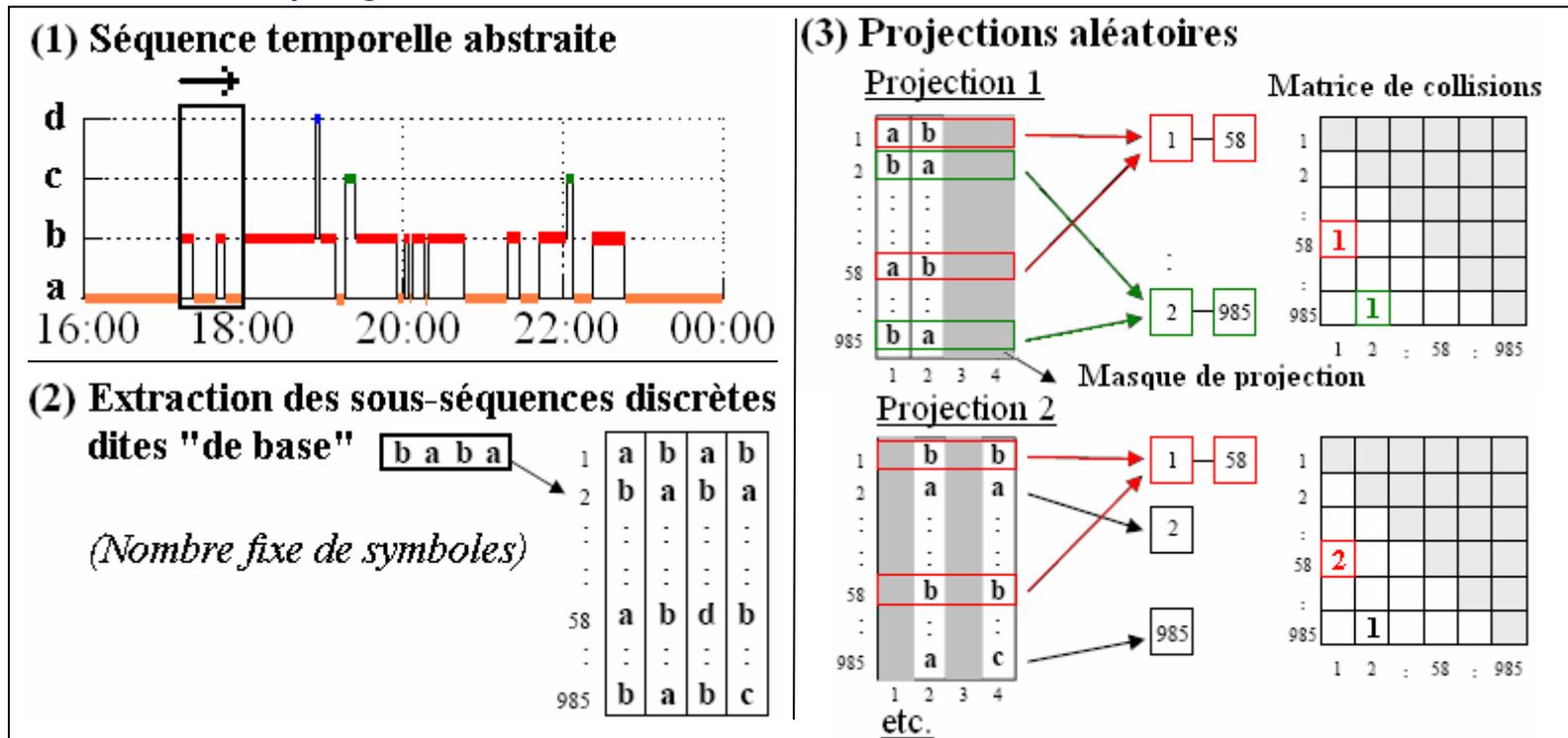
Critères de sélection

1. Fréquence

2. Signification

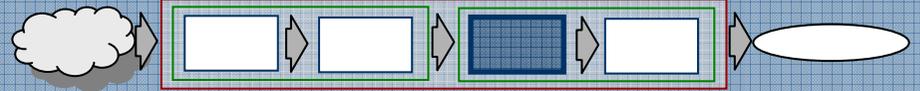
3. Non Redondance

- Méthode des projections aléatoires⁵



- Forte présomption de récurrence \equiv Valeurs de collisions $>$ seuil
- Extension aux séquences multidimensionnelles

⁵ Chiu, B. Keogh, E., & Lonardi, S. (2003), "Probabilistic Discovery of Time Series Motifs," In the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, August 24 - 27, 2003. Washington, DC, USA.

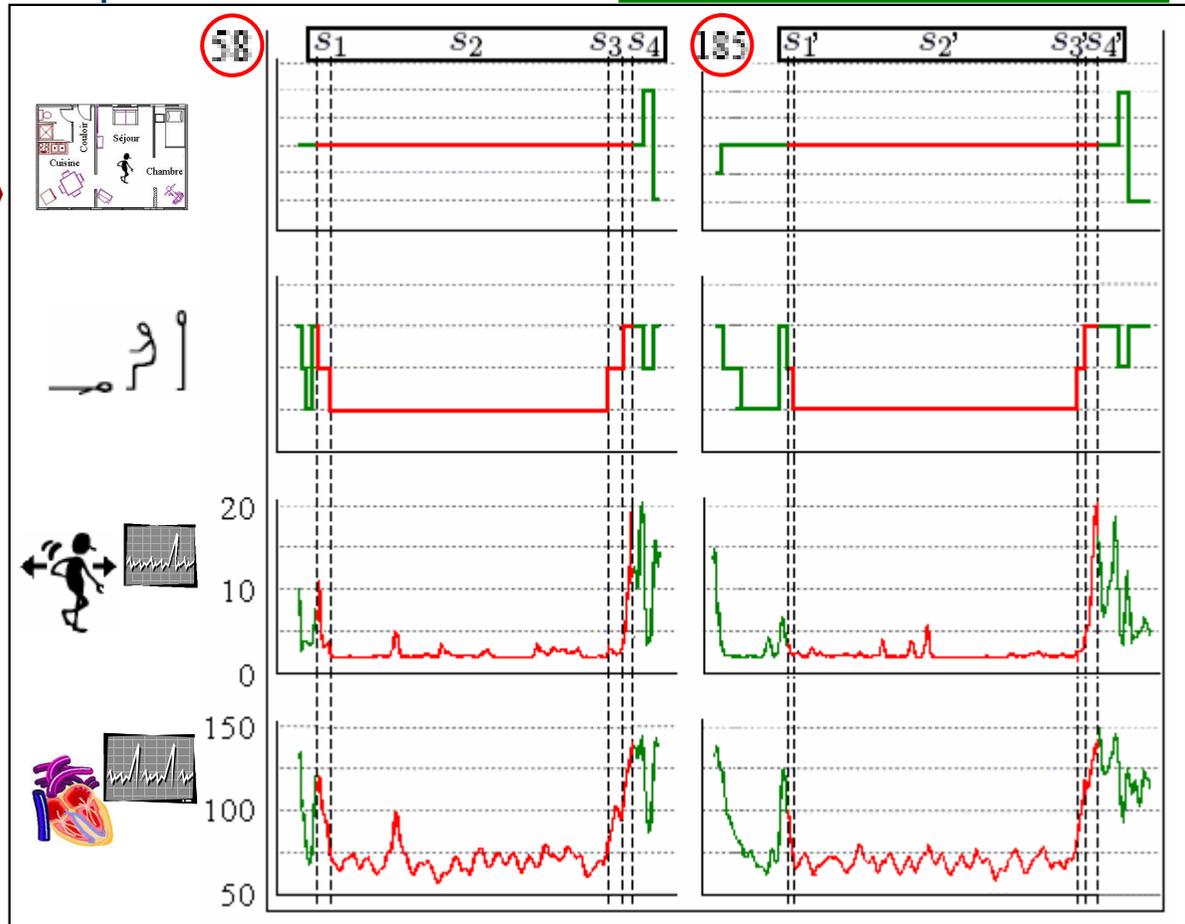
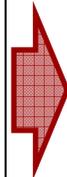
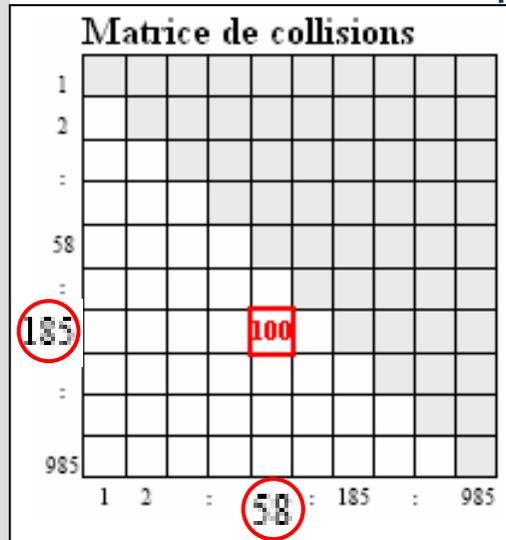


1. Fréquence

2. Signification

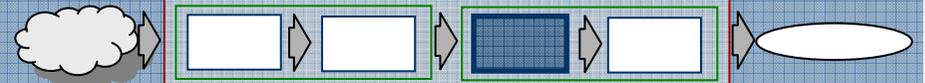
3. Non Redondance

- Vérification de la présomption de récurrence : $\text{Distance réelle} < \text{seuil}$



- Extension des sous-séquences « de base »
à partir de la matrice de collisions selon *deux critères* :
 - ▶ $\text{Valeurs de collisions} > \text{seuil}$ dans le voisinage des extensions
 - ▶ $\text{Distance réelle} < \text{seuil}$ entre les sous-séquences étendues

Fouille de caractères



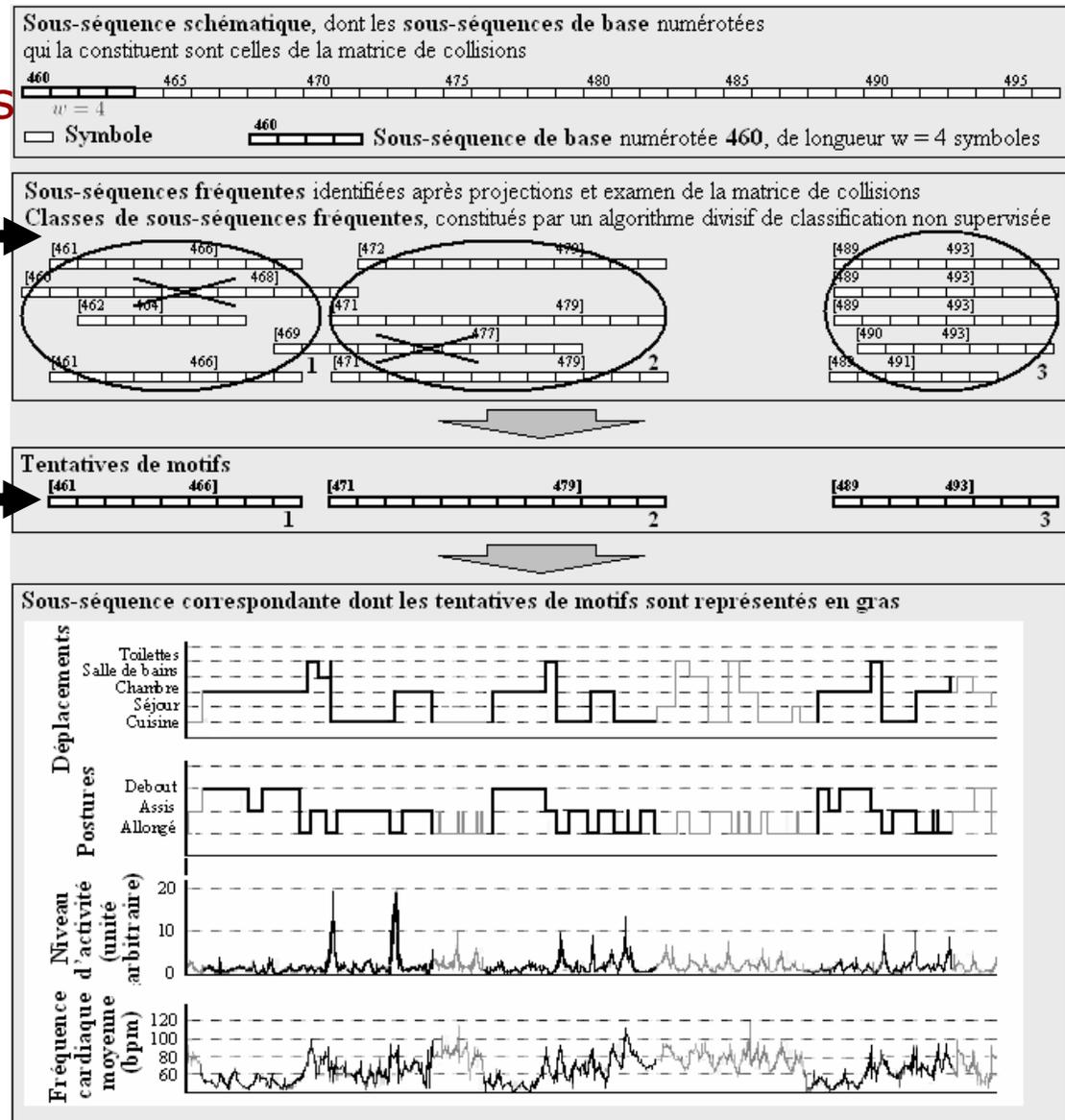
1. Fréquence

2. Signification

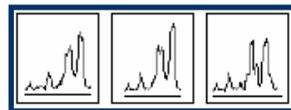
3. Non Redondance

Classification divisive des sous-séquences récurrentes

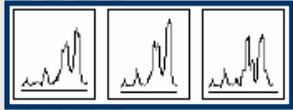
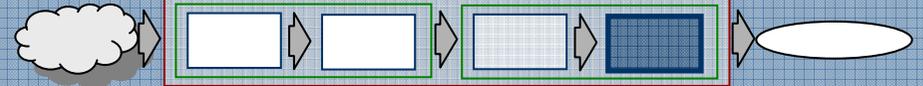
- Groupes de sous-séquences représentatives d'une même instance de motif : « *même activité au même moment* »
- Groupes représentés par des séquences toutes disjointes



Tentatives de motifs



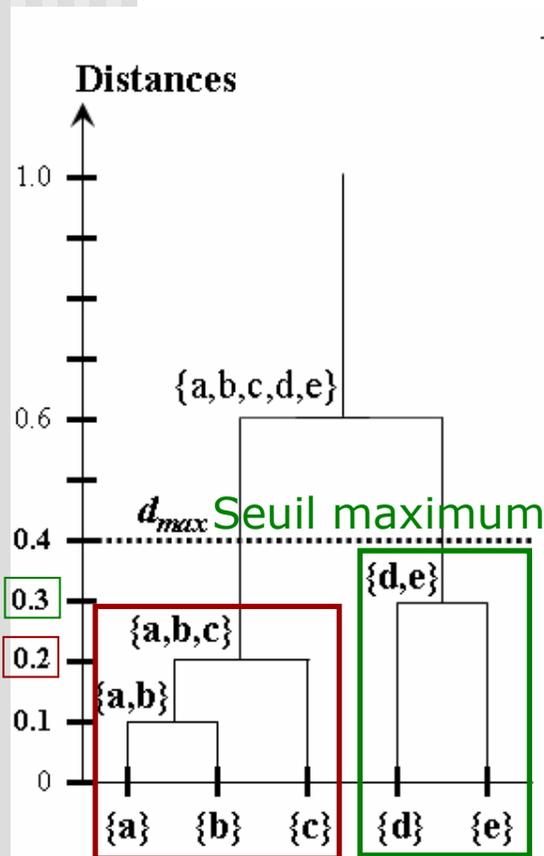
Classification



Tentatives de motifs

Classification ascendante hiérarchique

- Classification des tentatives de motifs en groupes de sous-séquences représentatives d'une même activité



$$D_0$$

	a	b	c	d	e
a	0				
b	0.1	0			
c	0.1	0.2	0		
d	0.3	0.3	0.4	0	
e	0.4	0.4	0.6	0.3	0

$$D_1$$

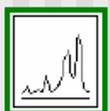
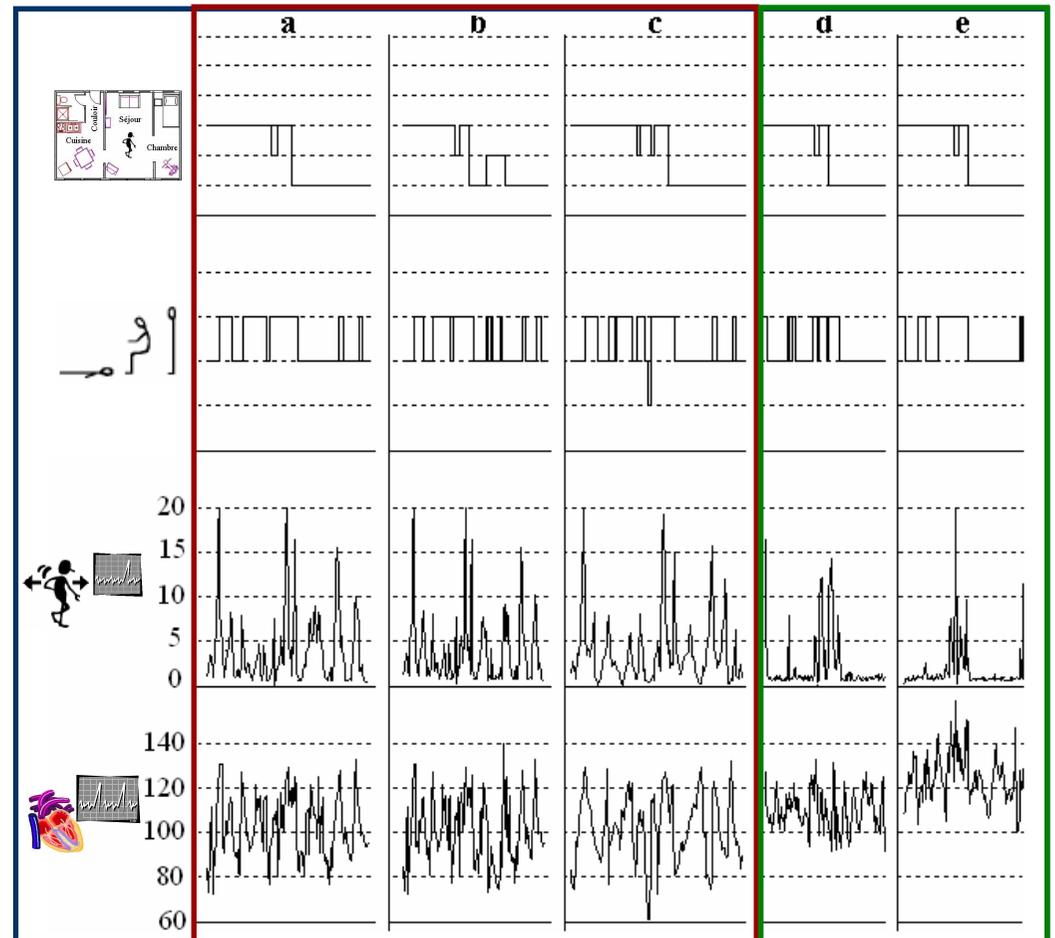
	{a,b}	c	d	e
{a,b}	0			
c	0.2	0		
d	0.3	0.4	0	
e	0.4	0.6	0.3	0

$$D_2$$

	{a,b,c}	d	e
{a,b,c}	0		
d	0.4	0	
e	0.6	0.3	0

$$D_3$$

	{a,b,c}	{d,e}
{a,b,c}	0	
{d,e}	0.6	0

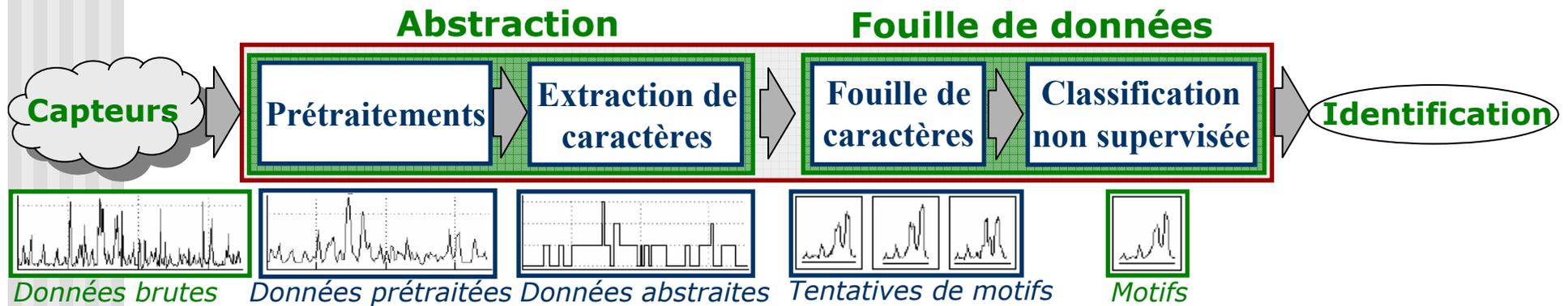
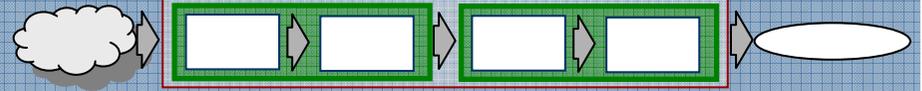


Motifs



Le représentant de chaque classe définit un **motif**

Mesures de Similarité



**Distance
approchée**

**Distance
réelle**

- Partitionner la séquence initiale en sous-séquences homogènes au regard de la décision
- Agrégation en un seul symbole décrivant la continuité d'une même situation

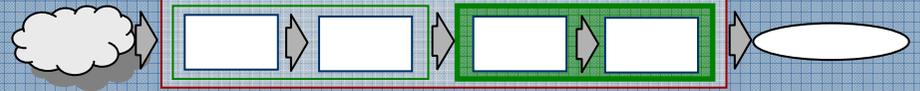
- Comparaison de séquences multidimensionnelles et hétérogènes
- Similarité entre séquences bruitées :
 - Variabilité dans les valeurs
 - Interruptions
 - Déformations temporelles



Distance discrète minimum

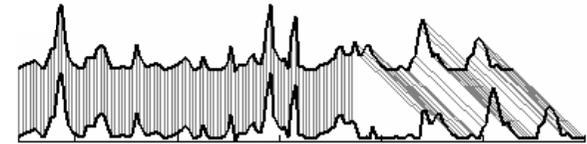
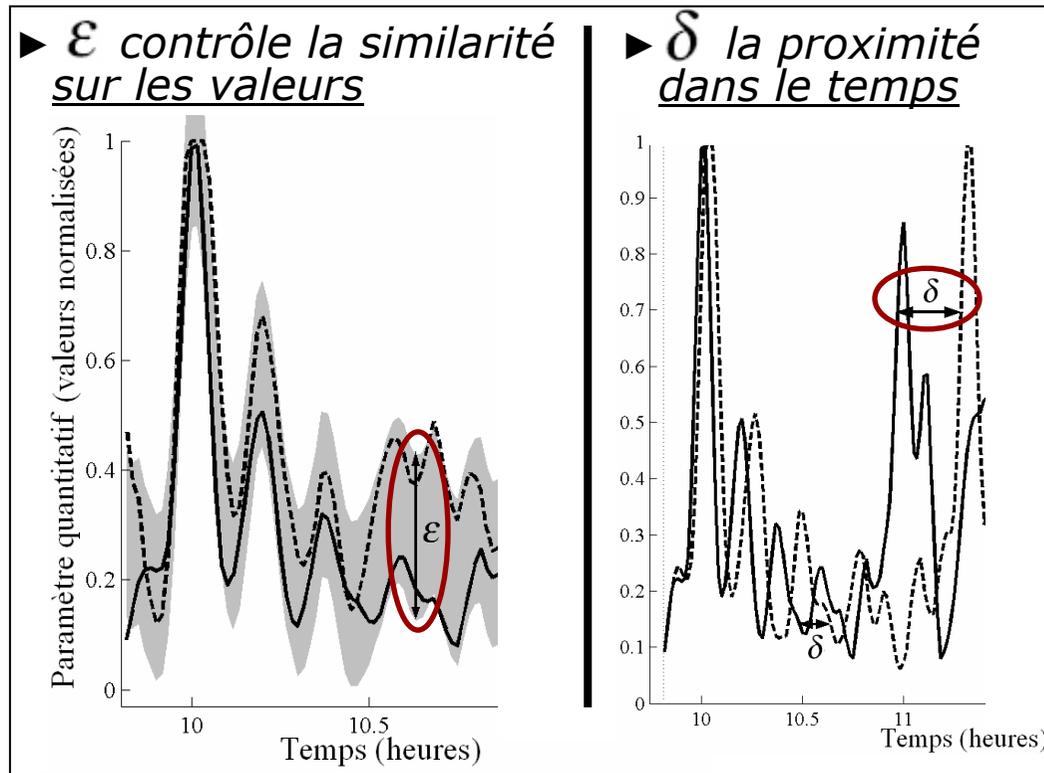
Distance non métrique

Distance réelle



Distance basée sur la plus longue sous-séquence commune – LCSS⁴

- Compter le nombre de « points » similaires : $LCSS_{\delta,\epsilon}(A,B)$ selon deux seuils de similarité pré-définis : ϵ, δ



- Distance dans $[0,1]$:

$$D_{\delta,\epsilon}(A,B) = 1 - \frac{LCSS_{\delta,\epsilon}(A,B)}{\min(n,m)}$$

Extension de la notion de similarité entre deux « points »

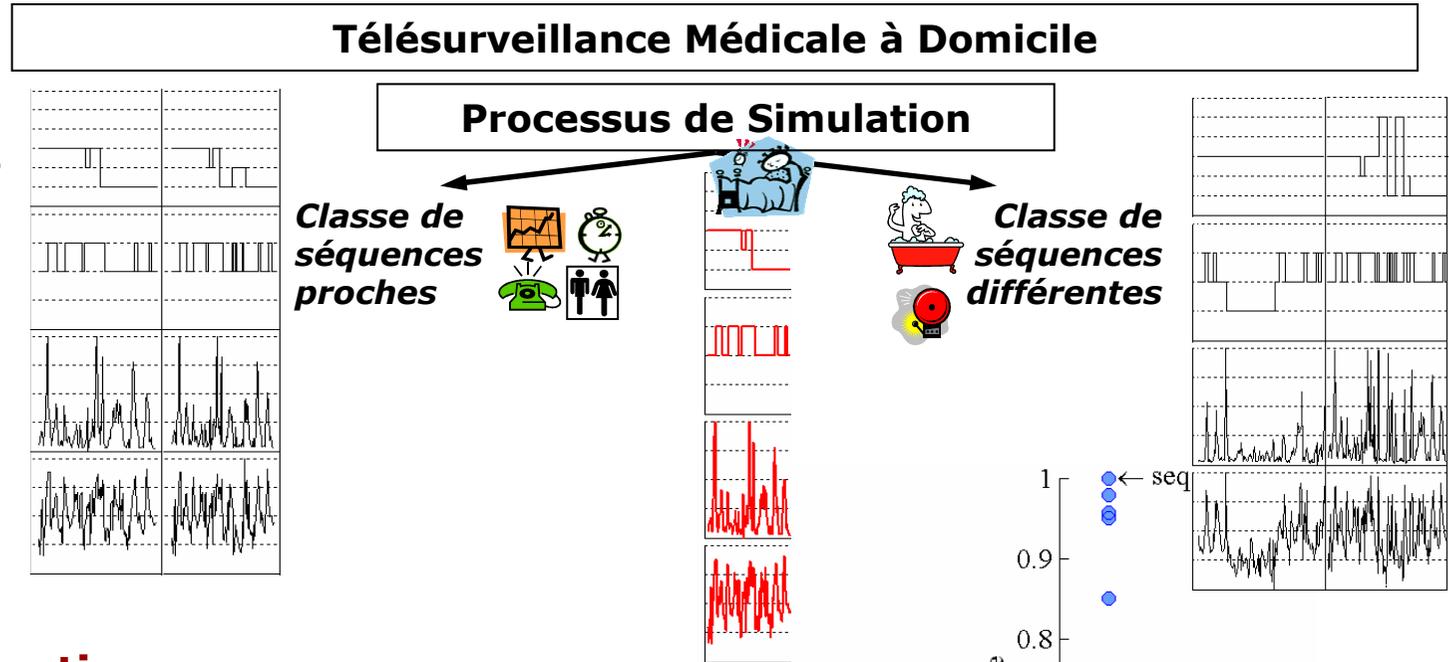
- Multidimensionnalité ► Similarité entre chaque composante
- Hétérogénéité ► Égalité des valeurs des paramètres qualitatifs

⁴ M. Vlachos, G. Kollios, and G. Gunopulos, "Discovering Similar Multidimensional Trajectories," in *Proc. of the 18th ICDE*, San Jose, CA, 2002, pp. 673–684.

Qualité de la mesure de distance réelle

1. Contexte

2. Données expérimentales



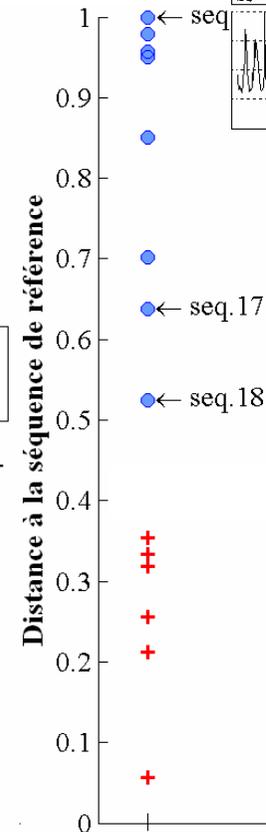
3. Méthode d'évaluation

Classification basée sur la mesure de distance réelle

- **Faibles distances** entre des séquences correspondant à la réalisation d'une même activité, dans de bonnes conditions
- **Distances plus élevées sinon** : réalisation d'activités différentes ou d'une même activité mais dans de mauvaises conditions

4. Mesures de performances

- **Bon classement des séquences**
- **Bonne discrimination des classes**



Qualité des résultats

1. Contexte

2. Données expérimentales

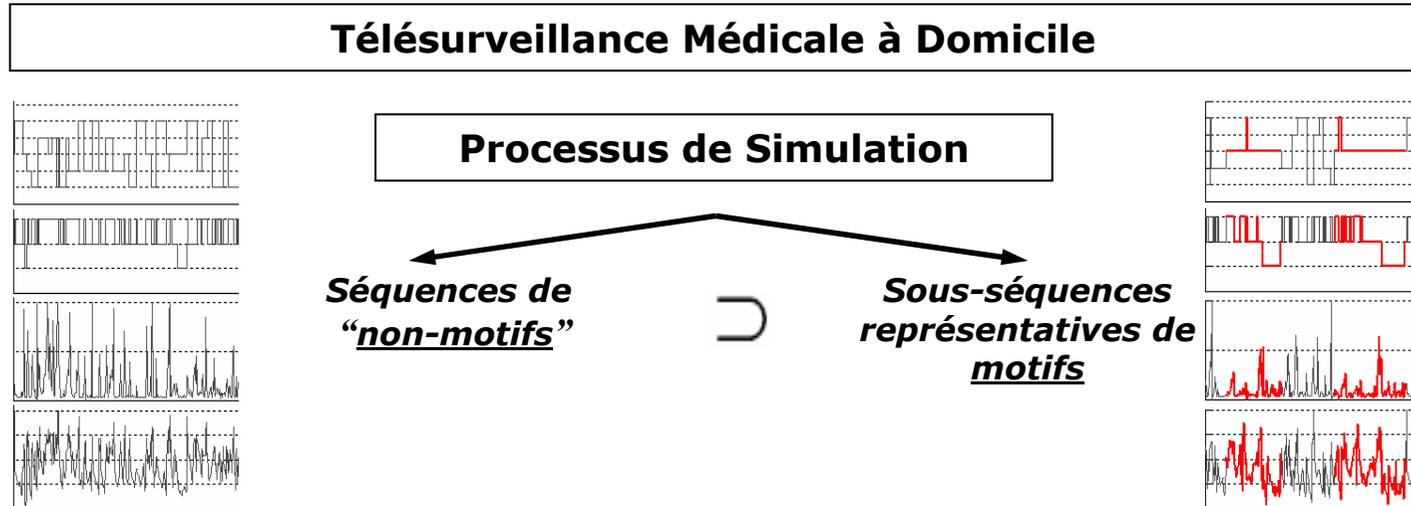
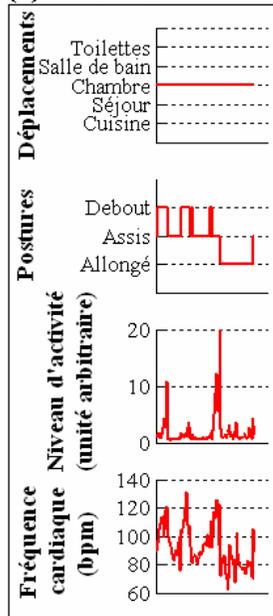
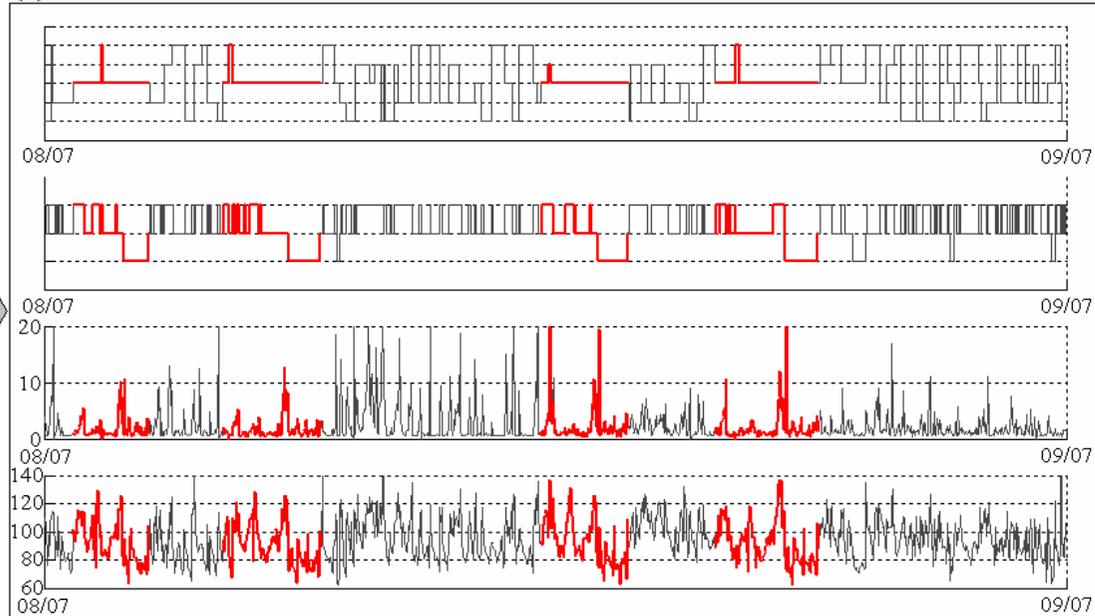


Illustration de l'extraction de motifs

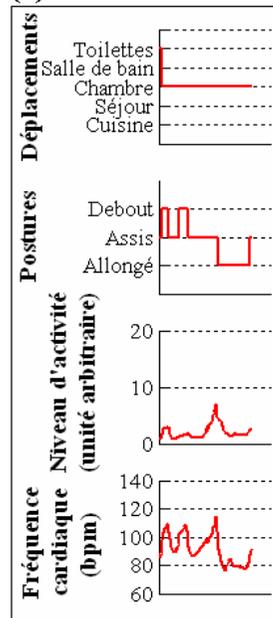
(1) MOTIF INITIAL



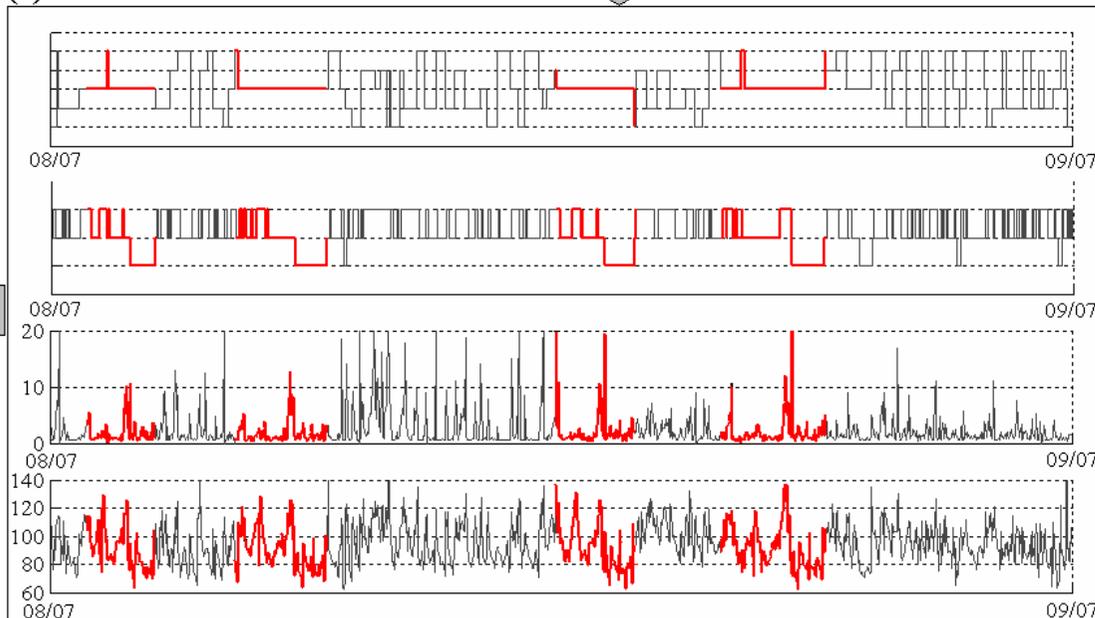
(2) INTRODUCTION DES INSTANCES BRUTEES DU MOTIF



(4) MOTIF EXTRAIT



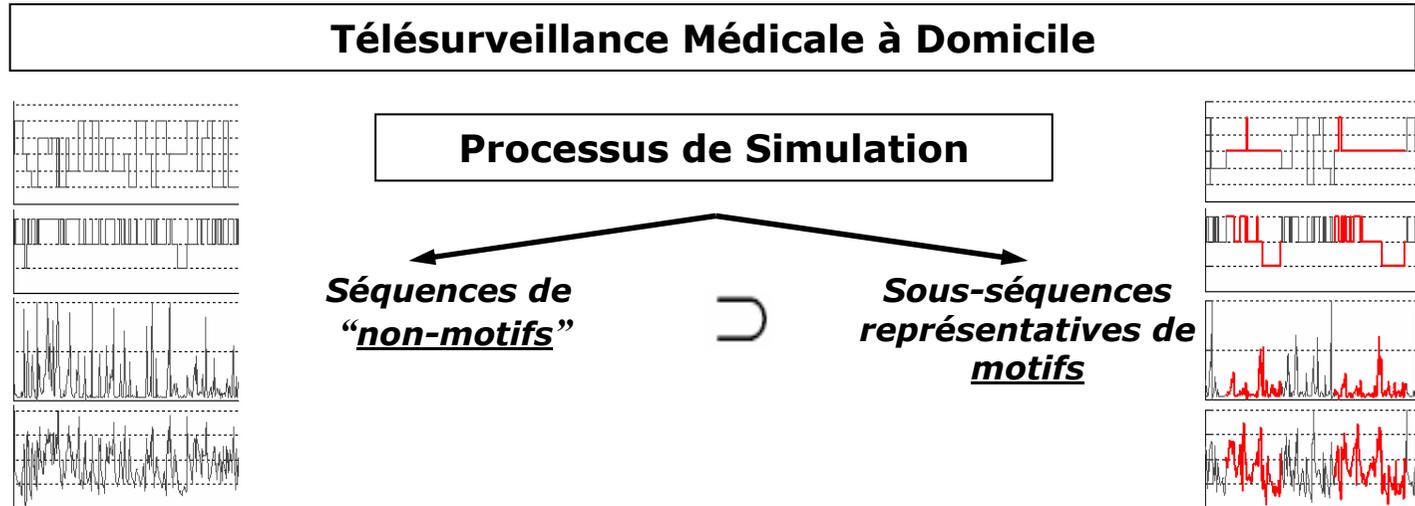
(3) EXTRACTION DES MOTIFS



Qualité des résultats

1. Contexte

2. Données expérimentales



3. Méthode d'évaluation

- Validation de l'extraction possible de sous-séquences récurrentes
- Réglage des paramètres clés du système – seuils de décision

- Mesure de similarité

Paramètre	Définition
δ_{LCSS}	Écart temporel maximum entre deux points similaires
ϵ_{LCSS}	Écart maximum sur les valeurs de deux points similaires

- Identification des motifs

c_{min}	Seuil minimum de collisions
d_{max}	Seuil maximum de distance

➔ **Maximiser les performances de la classification**

Qualité des résultats

4. Mesures de performances

Identification des tentatives de motifs

Classification en motifs

Sensibilité

“Identifier comme tentative de motifs des sous-séquences qui correspondent effectivement à des sous-séquences récurrentes dans la séquence initiale.”

“Toutes les instances d’un même motif doivent être regroupées dans une seule classe, sans fractionnement.”

Spécificité

“Ne pas identifier comme tentative de motifs des sous-séquences issues d’intervalles de non-motifs.”

“Tous les éléments d’une classe doivent être représentatifs d’un seul motif, sans fractionnement.”

➔ Performances en contexte faiblement bruité

Indices	Identification		Classification	
	S_e	S_p	S_e	S_p
Moyenne	0.71	0.92	0.66	0.79
Écart-type	0.18	0.07	0.34	0.26
Indices parfaits	–	–	35%	60%
Classification parfaite			20%	

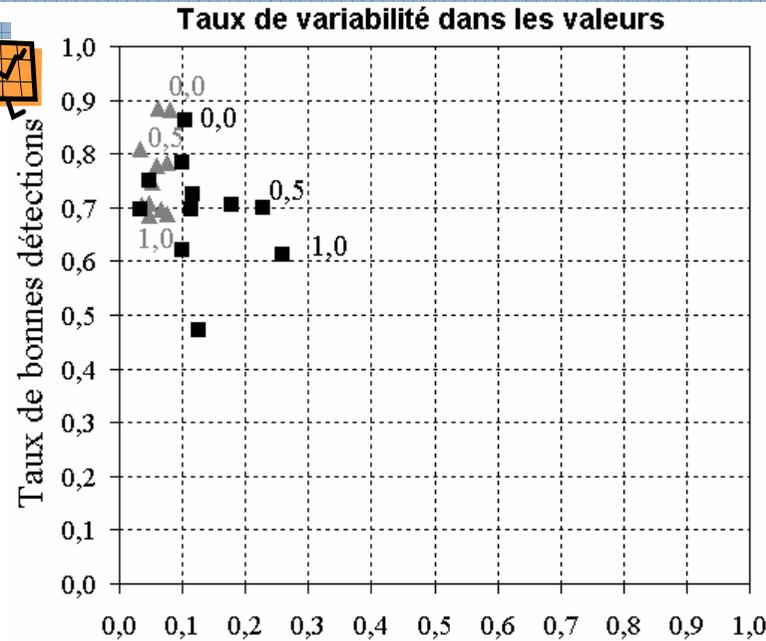
- Des résultats encourageants
 - ▶ **Identification et classification parfaite possible des motifs**
- Grande variabilité des résultats
 - ▶ **Diversité des motifs sélectionnés aléatoirement ?**

Test de Sensibilité

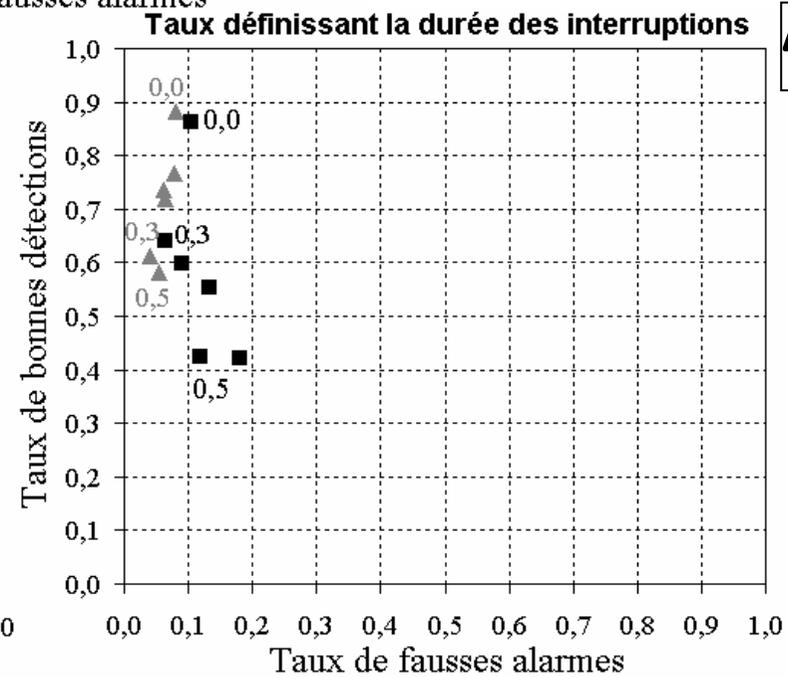
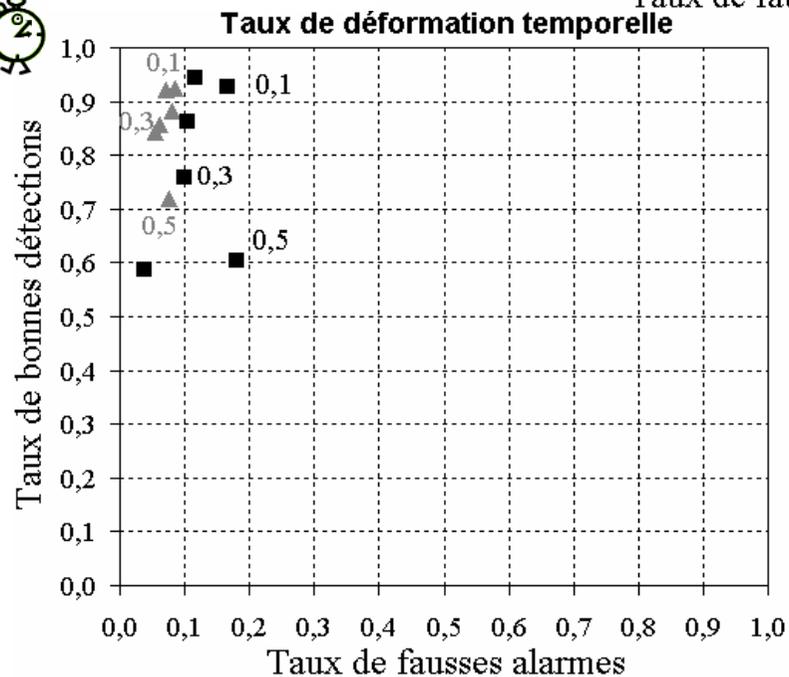
Modifications
"normales"



- *Courbes COR*

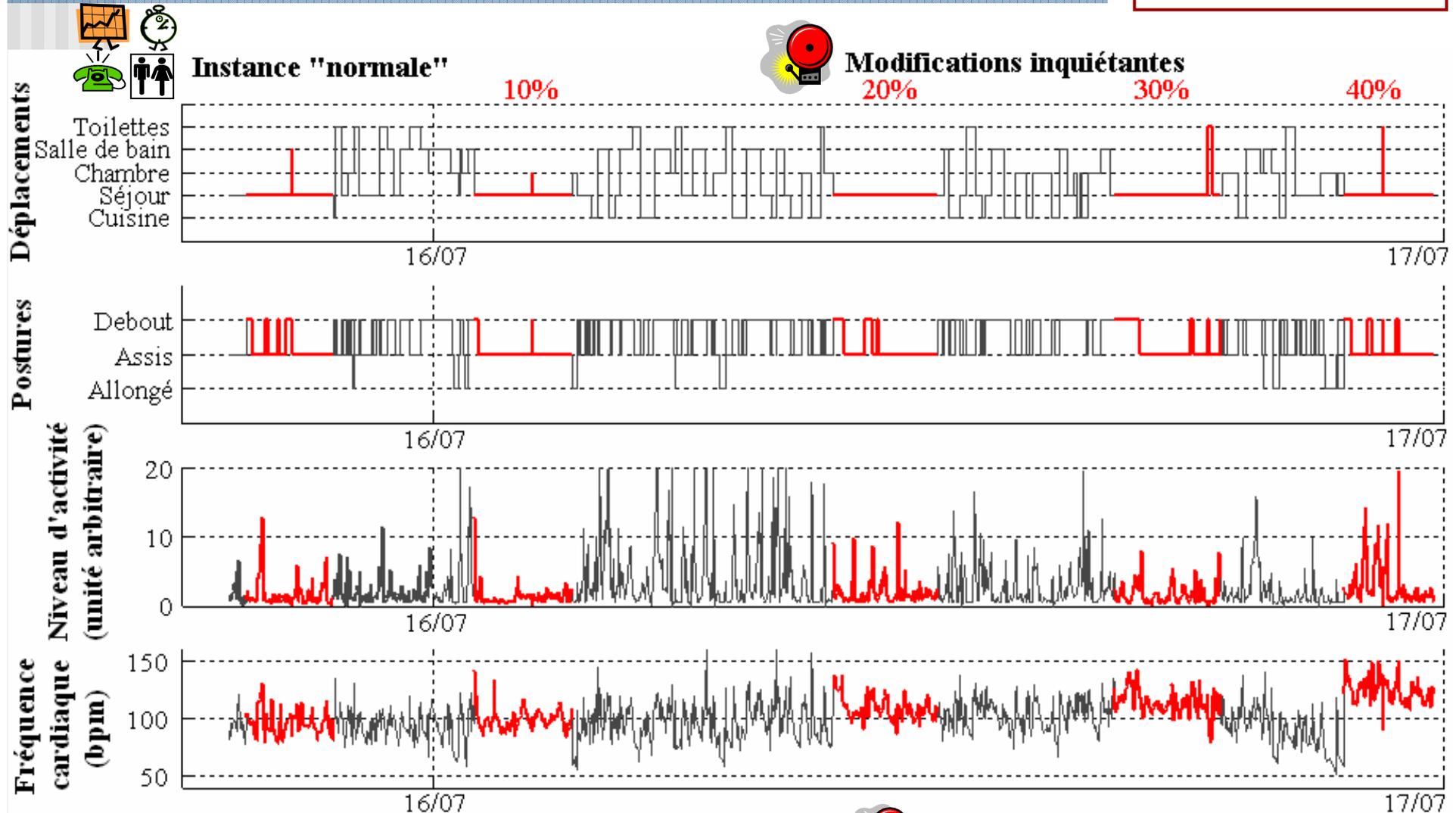


▲ Identification
 ■ Classification



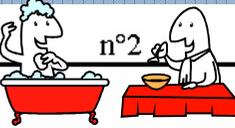
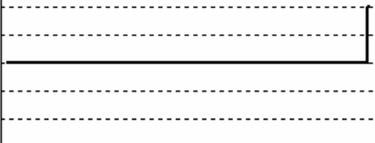
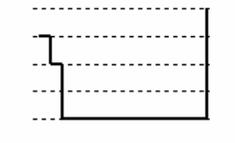
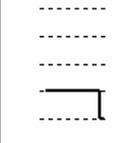
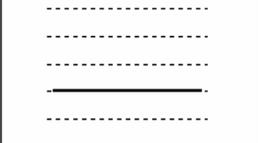
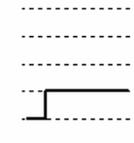
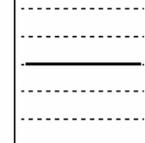
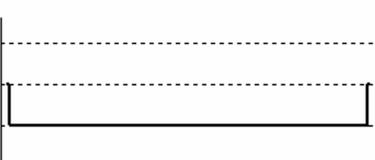
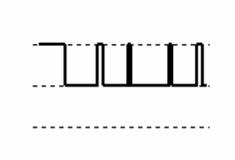
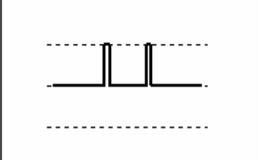
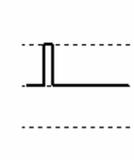
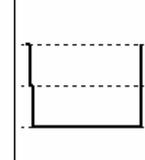
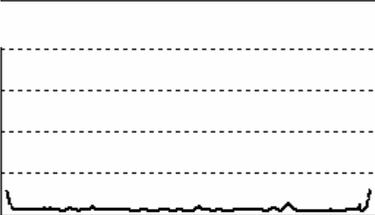
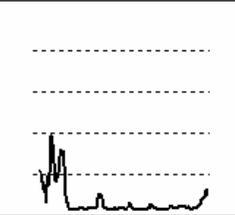
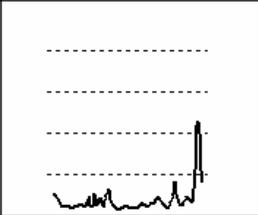
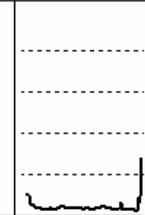
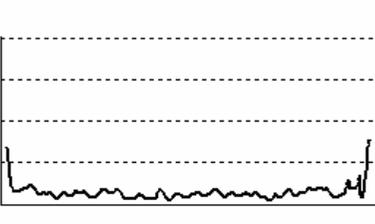
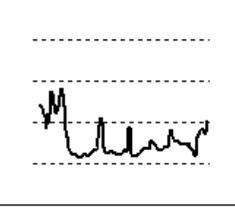
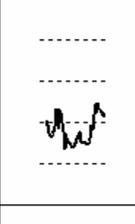
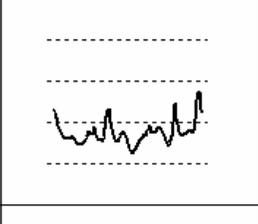
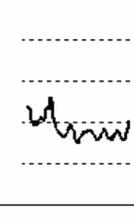
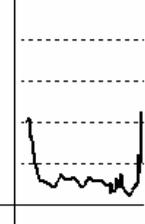
Test de Spécificité

Modifications inquiétantes



Modifications	Normales	Inquiétantes			
		10%	20%	30%	40%
Taux de modification	0%	10%	20%	30%	40%
Taux de reconnaissance	77.5%	25%	10%	10%	5%

Extraction de motifs de séquences simulées

Classes	 n°1	 n°2	 n°3	 n°4	 n°5	 n°6
Déplacements Toilettes Salle de bain Chambre Séjour Cuisine						
Postures Debout Assis Allongé						
Niveau d'activité (unité arbitraire)						
Fréquence cardiaque (bpm)						
Effectif	3	5	5	3	10	5
Durée	7h30	3h30	1h15	3h	2h	2h30
Instant d'occurrence	vers minuit	vers 9h	vers 13h ou 20h	vers 15h	entre 16h et 19h	vers 14h ou 21h
Fréquence moyenne reconnue	tous les deux jours	presque tous les matins	un jour sur deux	un jour sur deux	une à deux fois par jour	un jour sur deux
Interprétation de l'activité	<i>Nuit de sommeil</i>	<i>Toilette du matin Petite activité dans la cuisine et repas</i>	<i>Petite période calme au salon</i>	<i>Longue période calme au salon suivie d'une activité plus intense</i>	<i>Repas rapide ou collation suivie d'une activité calme</i>	<i>Période courte de repos</i>

Discussion et conclusion

Caractéristiques fondamentales de la méthode

- Analyse multidimensionnelle et hétérogène
- Aspect non supervisé de l'extraction de motifs
- Peu de données d'apprentissage nécessaires
- Grande variabilité possible entre les instances d'un motif
- Décision « haut niveau » à partir de données « bas niveau »

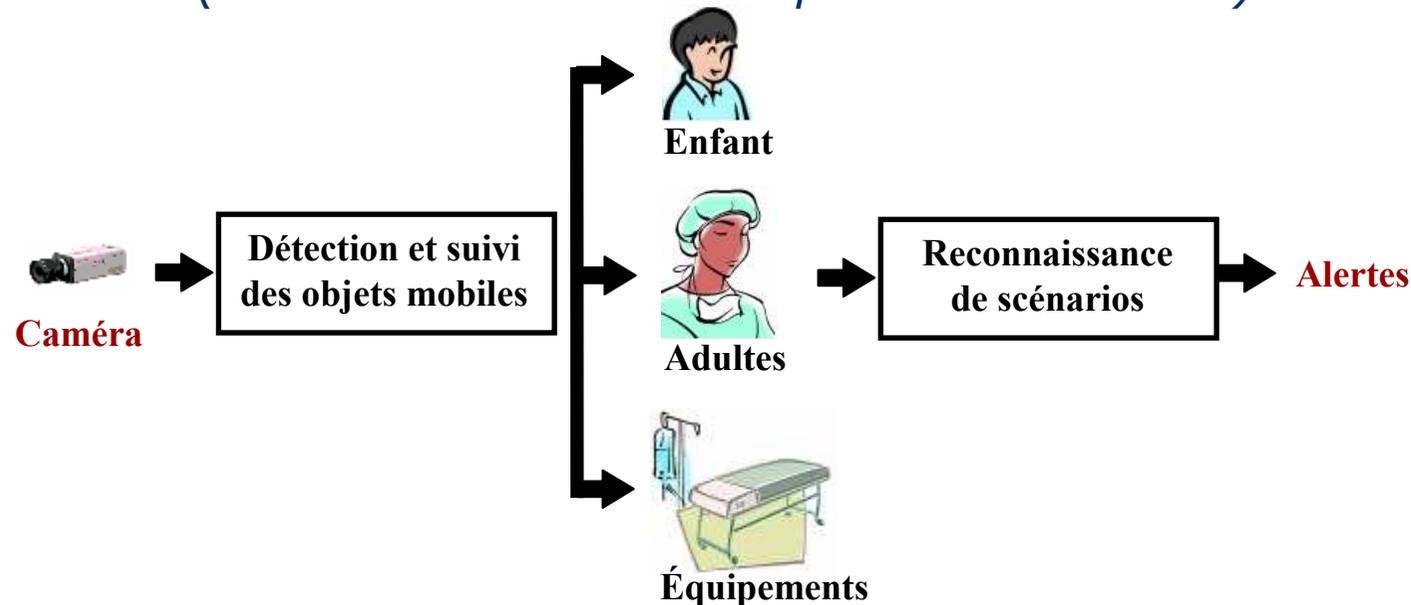
Expérimentation et Validation

- Complexité de la mise en pratique
 - ▶ *De nombreux paramètres, parfois interdépendants et impliqués à différents niveaux de l'analyse*
 - ▶ *Spécificité individuelle des seuils de décision ?*
- Potentialités de la méthode dans une configuration par défaut
 - ▶ *Identification et classification parfaite possible des motifs*
 - ▶ *Comportements récurrents identifiés à partir des données de simulation*
- Validation nécessaire de l'approche sur des données réelles
 - ▶ *Application à d'autres paramètres et niveaux de détail*

Validation de l'approche de résolution
dans un contexte différent de télésurveillance

➔ **Surveillance des enfants-bulles à l'hôpital l'Archet de Nice**

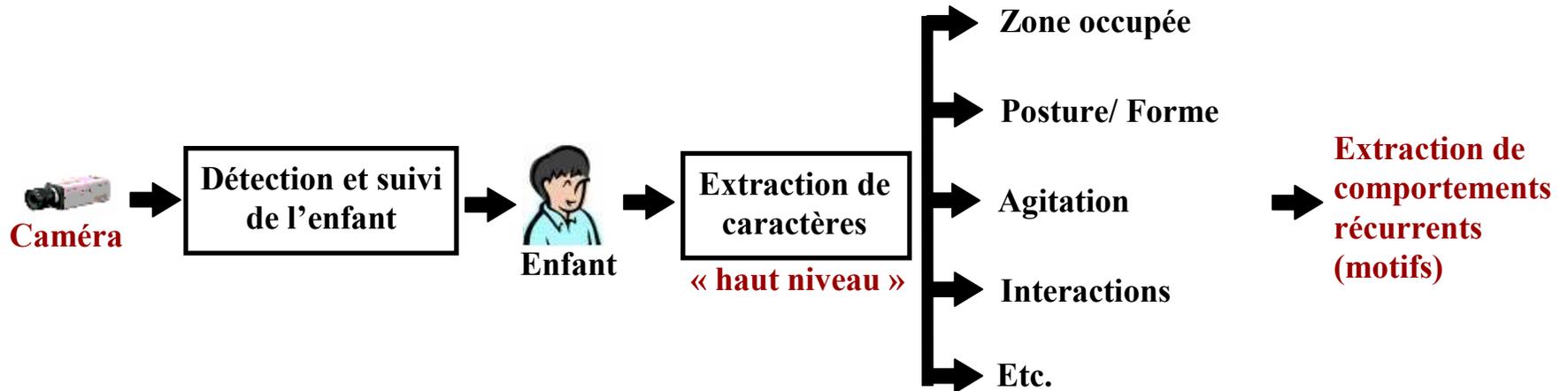
- Observation d'un enfant dans une chambre a partir de caméras
- Objectif de détection de situations inhabituelles
- Utilisation de la plateforme VSIP pour la compréhension des vidéos (*Video Surveillance Interpretation Platform*)



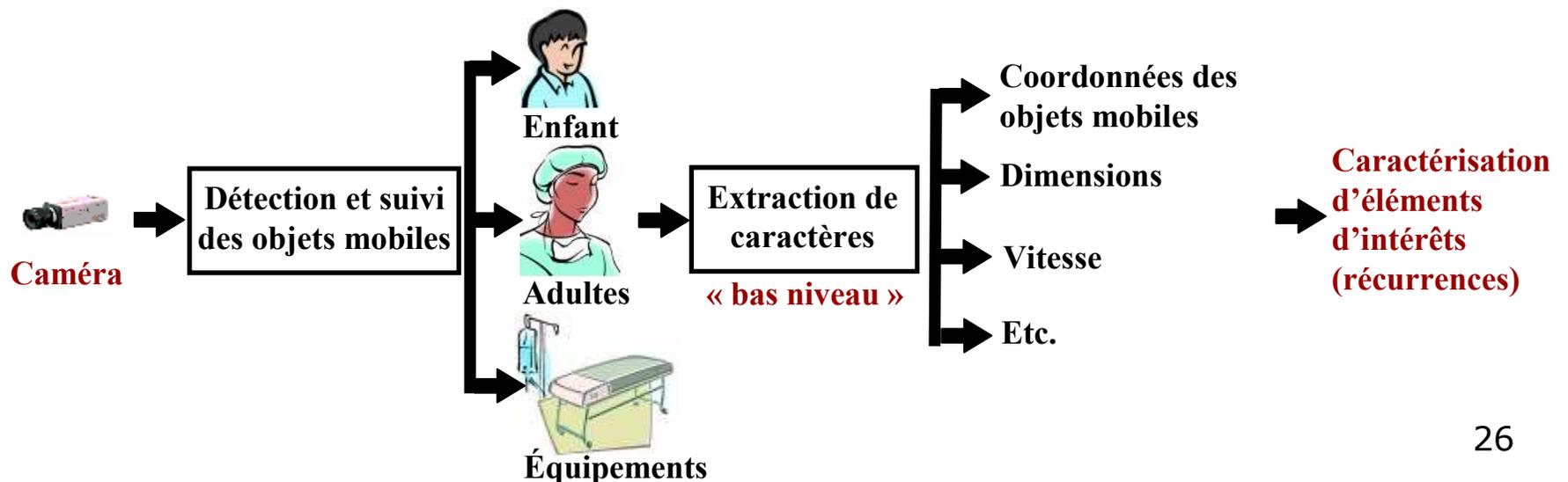
- ## ➔ **Complexité**
- Un seul capteur vidéo
 - Plusieurs personnes dans la chambre
 - Gestion des interactions

Objectifs

- Apprentissage du profil comportemental de l'enfant



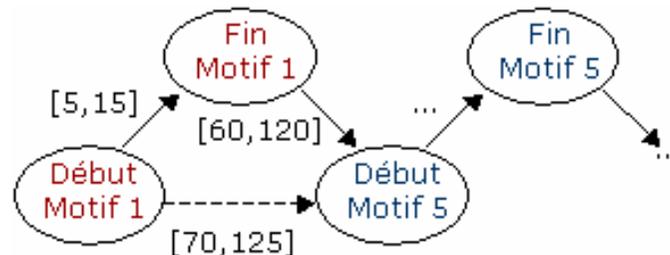
- Acquisition automatique de connaissances *a priori* utiles à la décision
 - zones d'intérêts, états, événements définissant les scénarios



Autres Perspectives

A plus long terme

- Caractérisation du profil comportemental en terme d'une succession de « *motifs* » et de « *non-motifs* »
 - ▶ *Graphes temporels*



- Détection des situations critiques
 - ▶ *Reconnaissance des motifs*
 - ▶ *Comparaison de graphes temporels*





Merci de votre attention.