

Techniques visuelles pour la détection et le suivi d'objets 2D

Rafiq SEKKAL

Directrice de thèse : Marie Babel

Lagadic group

Inria Rennes Bretagne Atlantique & Irisa

<http://www.irisa.fr/lagadic>

Contexte: Navigation fauteuil électrique

- **Assistance à la mobilité :**
 - Perte d'autonomie des personnes âgées et/ou handicapées
 - Utilisation d'un fauteuil roulant électrique
 - Aide d'une tierce personne
- **Difficultés de navigation :**
 - Difficulté liée au contrôle (adapté à l'handicape)
 - Fatigue d'utilisation
- **Aide à la navigation :**
 - Assistance technique lors des tâches quotidiennes
 - Correction de trajectoire si nécessaire (sécurité)
 - Fauteuil électrique intelligent

Navigation en milieu intérieur

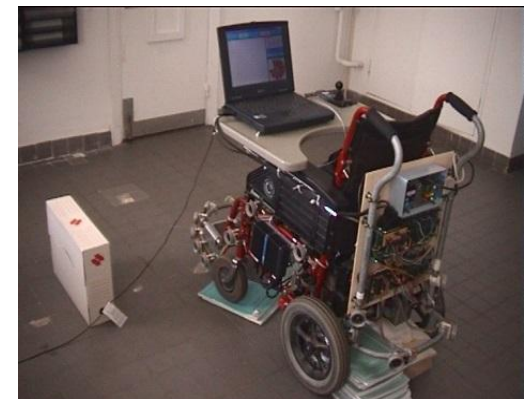
- Projets

capteurs	projets
Infrarouge	WAD [Mallet02]
Lidar	Muraka [Muraka09]
Ultrasonique	SYSIASS [Kokosy12]
Caméra	Muraka

- Concilier confort, prix et qualité
- Solution embarquée (ressources limitées)
- Solution économe en énergie



Muraka – 2 Lidars



WAD – Capteur de distance

Projet APASH

- **Projet APASH** : (Assistance au Pilotage pour l'Autonomie et la Sécurité des personnes Handicapées)
 - INSA leader projet
 - Industriels (Ergovie, AdvanSEE)
- **Objectifs** :
 - Module embarqué sur le fauteuil
 - Coût de réalisation raisonnable/coût du fauteuil
- **Deux tâches de navigation** :
 - Franchissement de portes
 - Evitement d'obstacles
- **Capteur utilisé** :
 - **Caméra monoculaire** placée à l'avant du fauteuil

caméra



Fauteuil 6-roues utilisé dans le projet APASH

Contexte: aide à la navigation

- Solutions de détection/représentation et suivi d'objets 2D
 - Solution basée vision
 - Complexité maîtrisée

Franchissement de portes

1. Détection de portes
2. Suivi de portes
3. Loi de commande
4. Navigation



Tâche de franchissement de porte

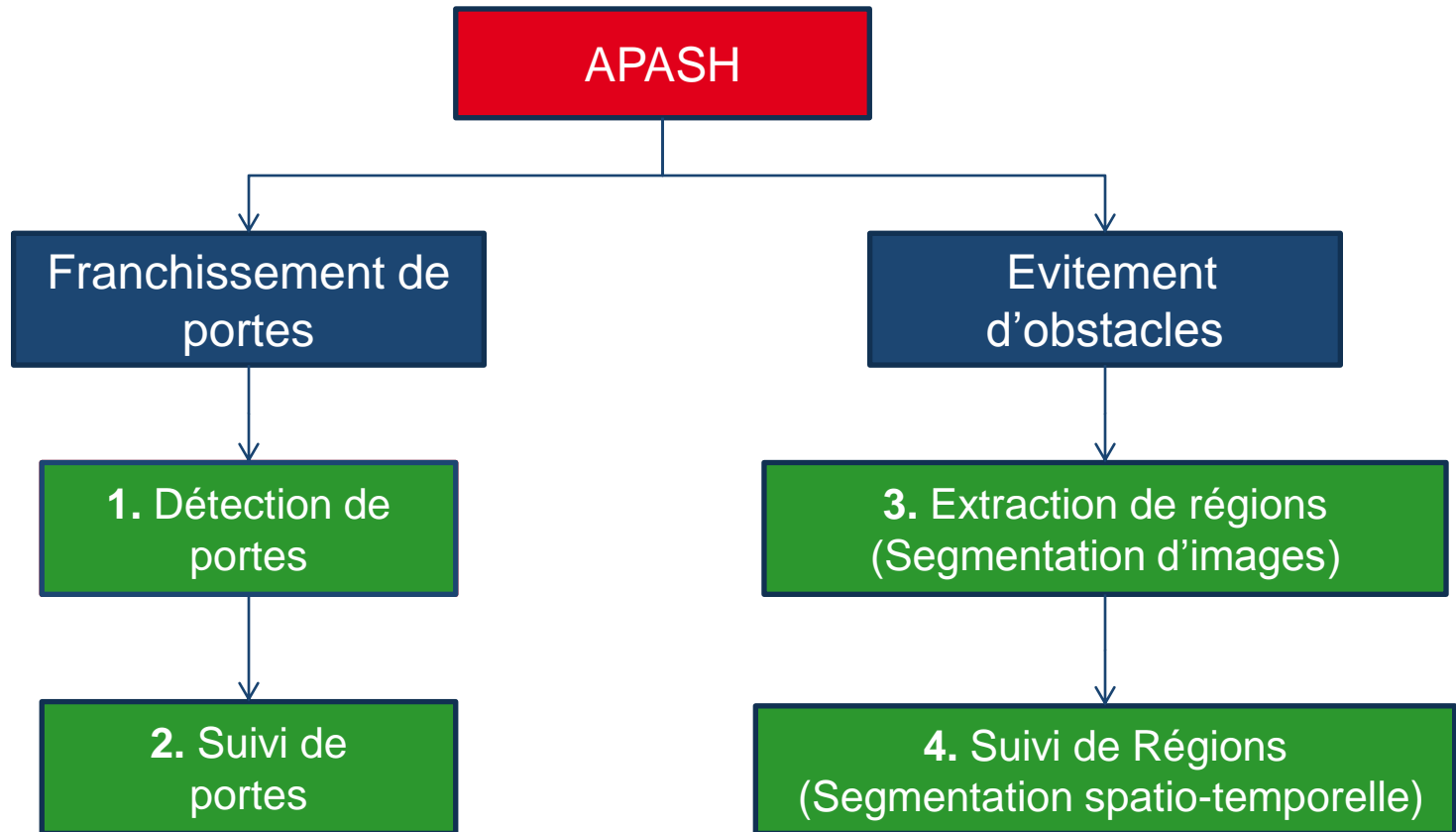
Evitement d'obstacles

1. Détection d'objets (représentation pseudo-sémantique)
2. Suivi d'objets
3. Estimation de la profondeur
4. Détection de collision
5. Correction de trajectoire



Tâche d'évitement d'obstacles

Plan



Contexte : détection de portes

- **Besoins :**
 - Localisation du fauteuil pour le franchissement de portes
- **Problématiques :**
 - Etat des portes (ouvertes ou fermées)
 - Position/orientation
- Détection de portes dans un couloir
 - A priori de la scène
 - Dimensions de la portes

Détection de portes

- Approches de détection de portes basées vision [Murillo08, Yang10, Chen08, Munoz04]

- **Descripteurs :**

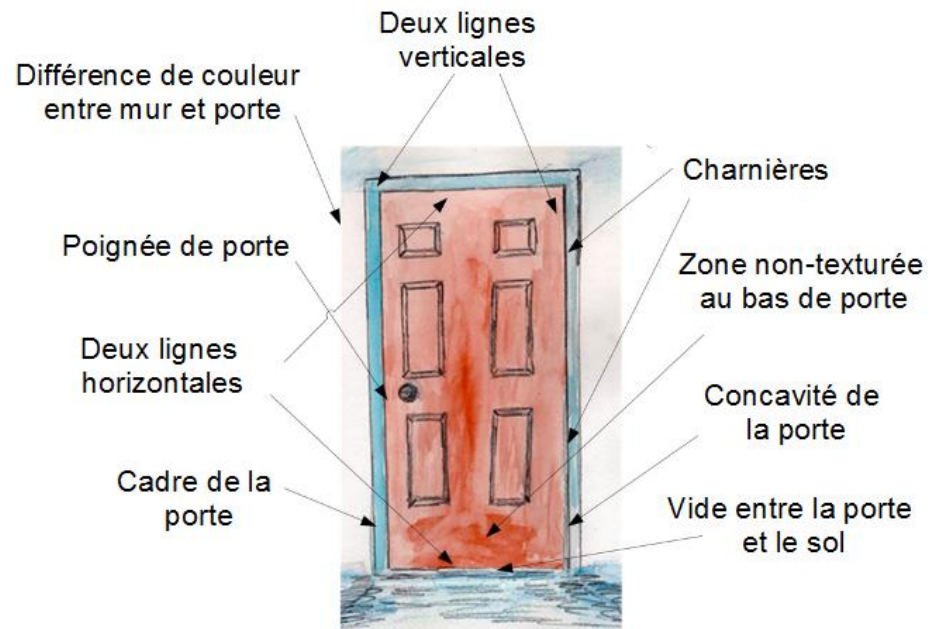
- Géométriques: lignes, coins...
- Photométriques: couleur

- **Approches :**

- Logique floue [Munoz04]
- Probabiliste [Murillo08]
- Détection de forme [Yang10]

- **Limitations :**

- Etat/couleur de la porte
- Contraintes sur la position de la camera
- Aucune information de position/orientation n'est exploitée

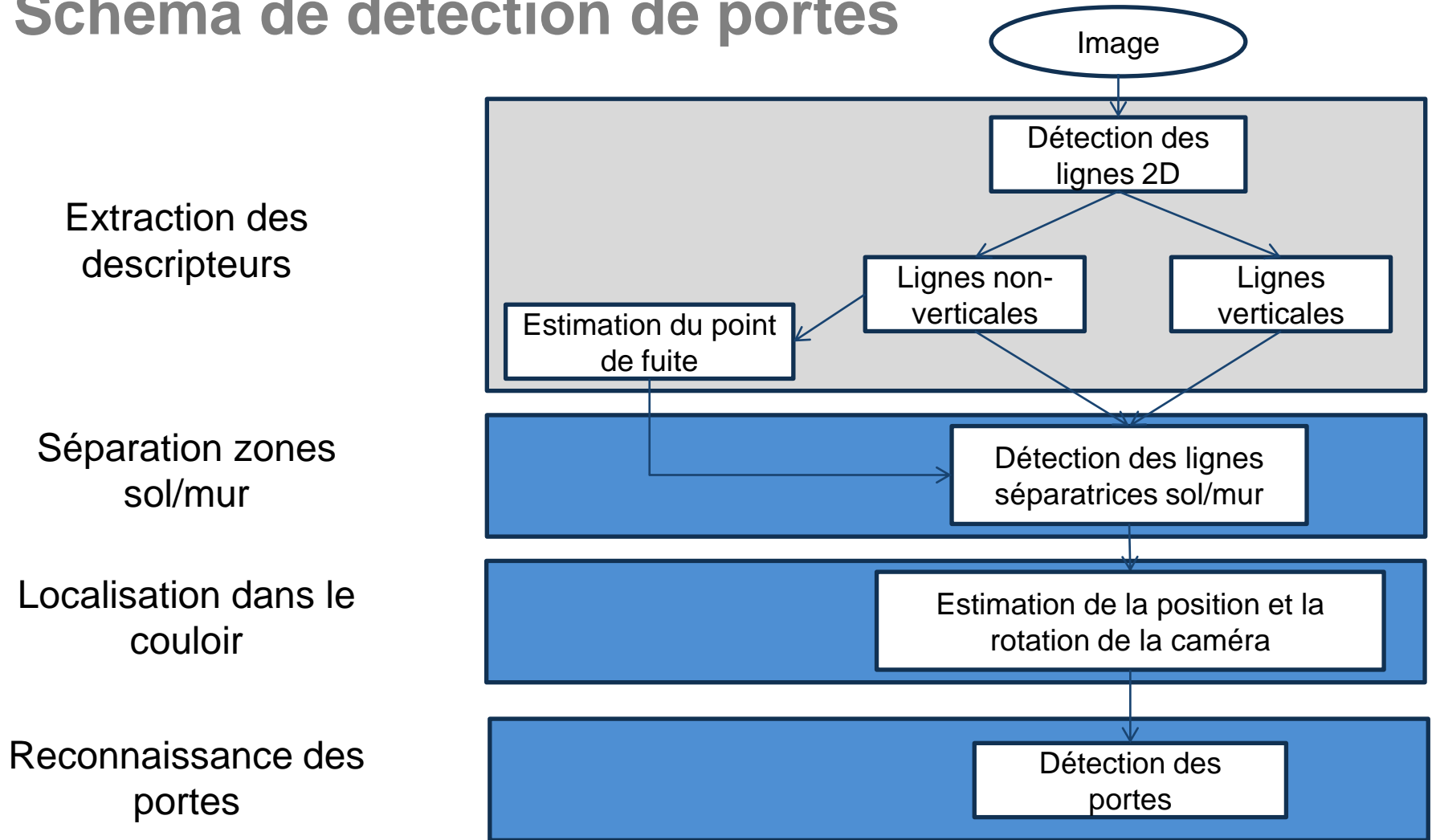


Descripteurs de porte

Approche proposée de détection de portes

- **Solution :**
 - Utiliser des descripteurs géométriques
 - Estimer la position et orientation de la caméra par rapport à la porte
 - Estimer la forme de la porte dans le plan image
- **Type d'environnement : couloir**
 - Murs parallèles
 - Portes partageant les mêmes caractéristiques
- **Descripteurs géométriques :**
 - Lignes droites verticales (montants)
 - Point de fuite (localisation)

Schéma de détection de portes

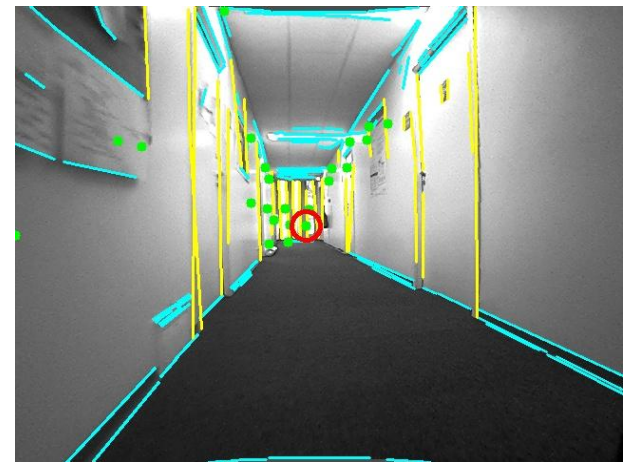


Extraction des descripteurs

- **Extraction des lignes:**
 - Utilisation de l'algorithme LSD [VonGioi12]
« Line Segment Detector »
 - Critère de direction de gradient
 - **Plus robuste et rapide** que la transformée de Hough [Oliver06]
- **Détection des lignes de montants candidates**
 - Classification: en fonction des angles
 - **Lignes verticales**
 - potentiellement montants
 - **Lignes non verticales**
 - pour l'estimation du point de fuite [Rother00]



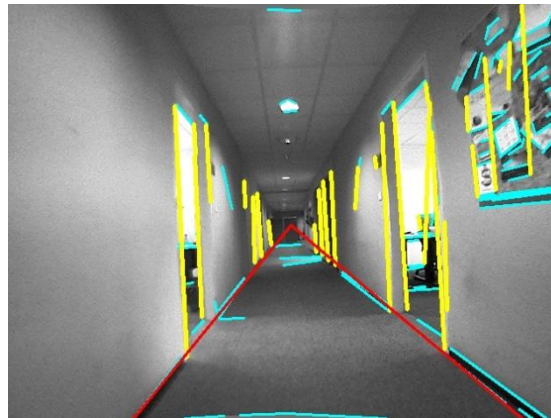
Détection de lignes LSD



Classification des lignes
Détection de point de fuite

Estimation des lignes sol/mur

- **Intérêt :**
 - Permettre la localisation dans le couloir
 - Définir les zones où peuvent se trouver les portes
- **Objectif :**
 - Estimer les lignes de fuite sol/mur et mur/plafond
- **Problème :** discontinuités (portes)
 - Fusionner les lignes non verticales en fonction de l'angle
 - Reconstituer la ligne de fuite sol/mur principale



Détection de lignes sol/mur

Localisation de la caméra dans le couloir

- Localisation de la caméra pour estimer la projection du modèle de la porte dans le plan image
- Rotation** (point de fuite)

$$\phi_Z = \text{atan}\left(\frac{x_{vp}}{p_x}\right)$$

$vp(x_{vp}, y_{vp})$: coordonnées du point de fuite

(u_0, v_0, p_x, p_y) : paramètres intrinsèques de la caméra

- Position** latérale relative dans le couloir (ligne de fuite)

$$Y_c = \frac{\theta_g + \theta_d}{\theta_g - \theta_d} \frac{L_c}{2}$$

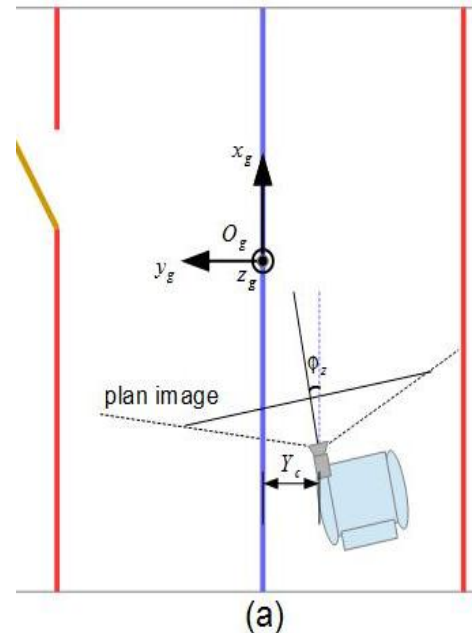
(θ_g, ρ_g) : paramètres de la ligne de fuite gauche

(θ_d, ρ_d) : paramètres de la ligne de fuite droite

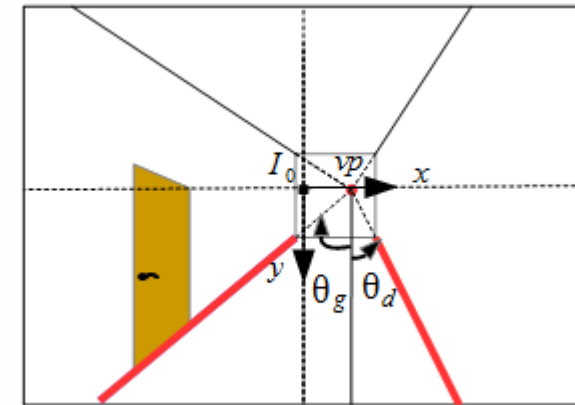
L_c : largeur estimée du couloir

- Suivi de couloir

- [Pasteau13] Corridor following wheelchair by visual servoing IROS13



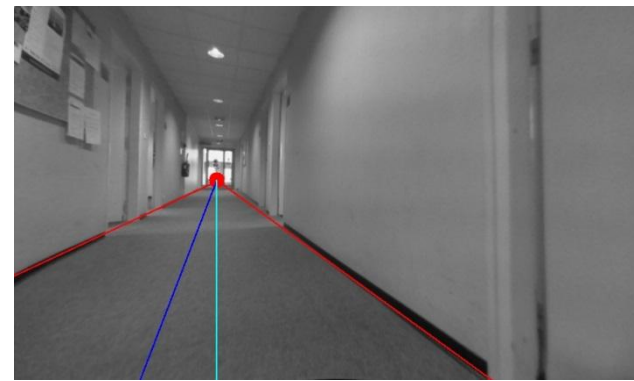
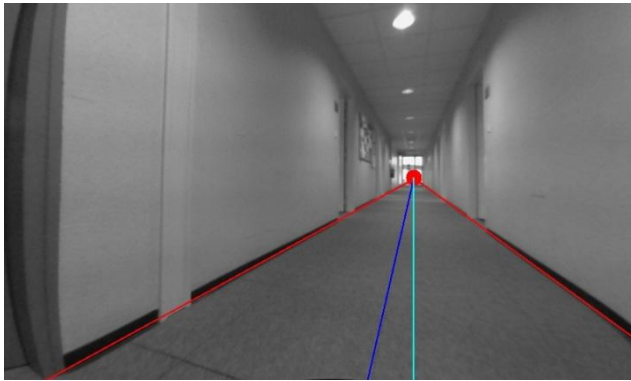
Localisation : vue de haut



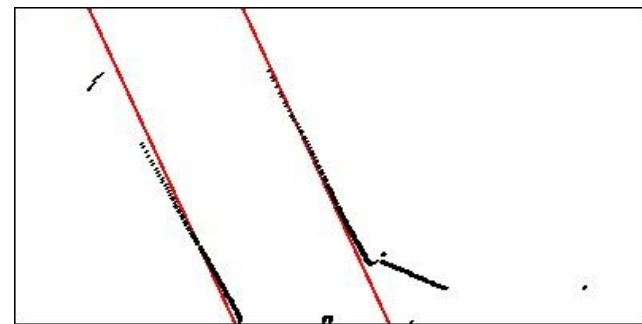
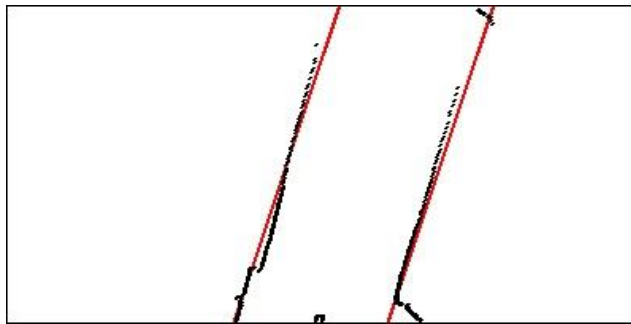
Localisation : scène de couloir

Localisation de la caméra dans le couloir

- Rouge : position estimée
- Noir : vérité terrain (laser)



Scène de couloir



Vue de Haut

Détection de portes

- **Technique :**

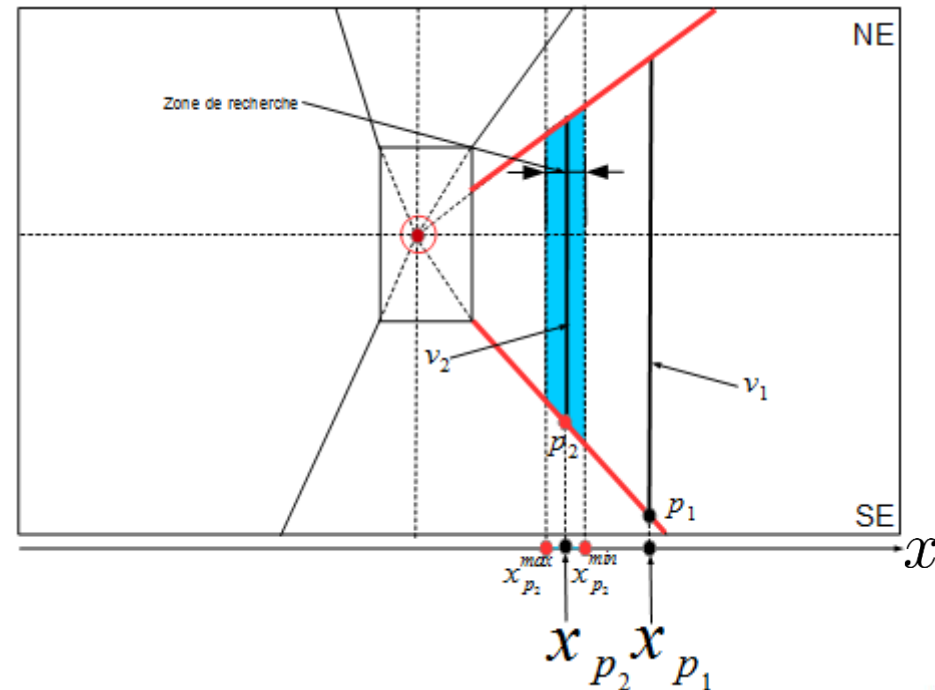
- Sélectionner 1^{ère} ligne verticale x_{p_2} (1^{er} montant de la porte)
- Estimer la position du 2^{ème} montant dans l'image
- S'intéresser aux coordonnées sur l'axe des x

- x_{p_1} dépend de :

- D : **Distance** caméra-mur
- ϕ_Z : **Rotation** de la caméra
- W : **Largeur** réelle de la porte

$$x_{p_2} = p_x \tan \left(\text{atan} \left(\frac{D}{X_{p_1} + W} \right) - \phi_Z \right)$$

$p_2(x_{p_2}, y_{p_2})$: coordonnées du 2eme montant p_2



Estimation de la largeur de la porte dans l'image

Résultats de détection

- **Expérimentations :**

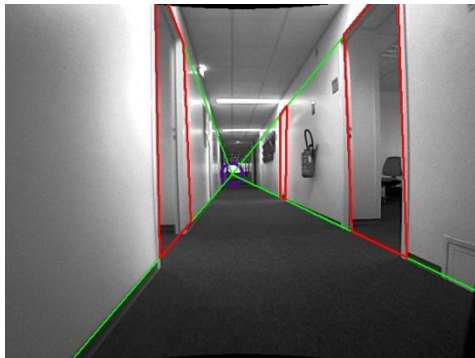
- Pas de benchmark ni de base de données dédiées
- 8 séquences tournées dans les locaux de IRISA
- Différentes caractéristiques : couleur de portes, murs, nombre de portes

- **Résultats :**

	Images	Portes	Bonnes détections	Fausse alarmes
total	3414	6356	5212	459

- 82% des portes sont détectées
- 12% de faux positifs
- Les portes manquées lors d'une image sont détectées après en moyenne de 8 images (< 1 seconde).

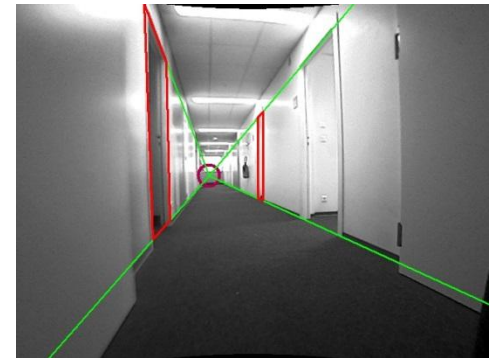
Résultats



Portes ouvertes



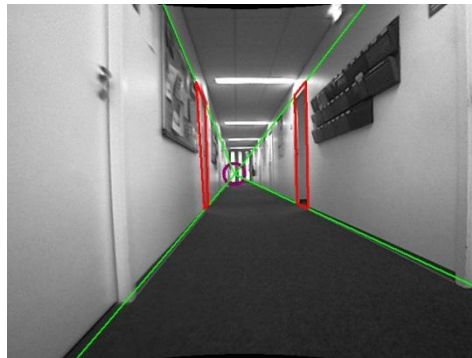
Portes fermées (même couleur)



Echec de détection



Caméra proche du mur

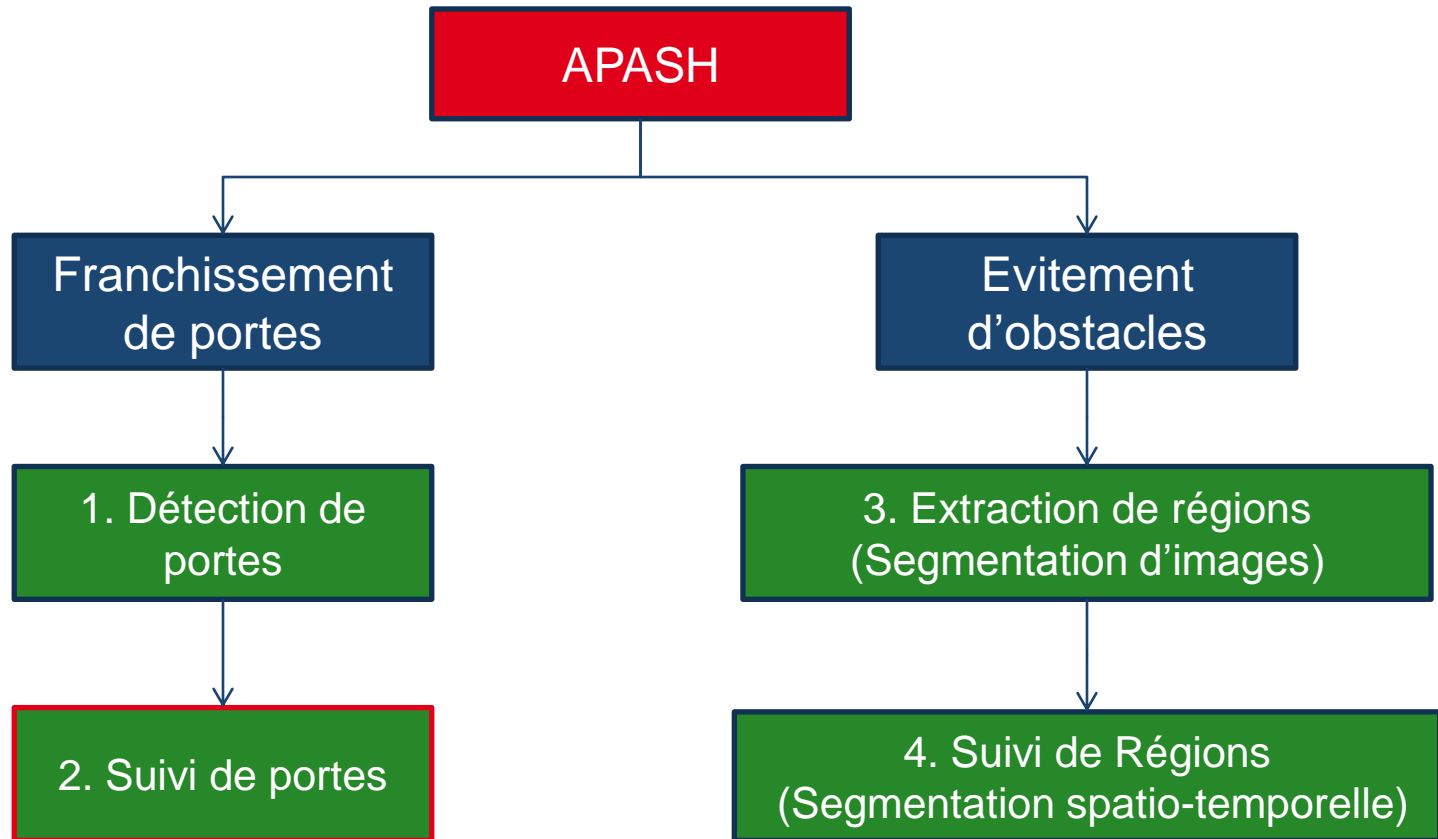


Portes éloignées



Fausse détection

Plan



Suivi de portes

- Initialiser un suivi de porte pour chaque nouvelle porte détectée

- Suivre les portes par leurs **montants**

- Suivi 2D de lignes verticales
- Utilisation de Moving Edges [Boutheimy89]
- Adaptation au suivi des montants

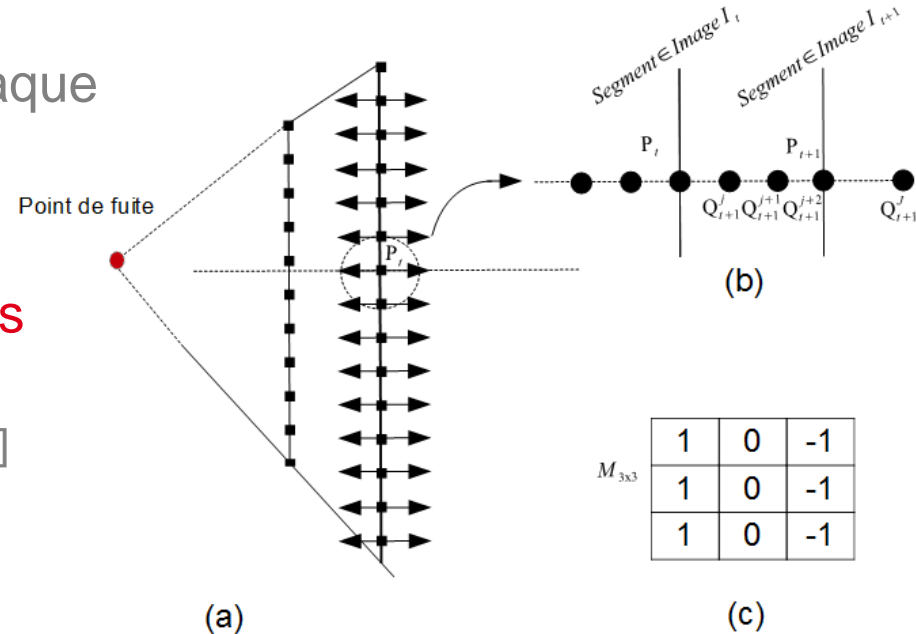
- **Principe :**

- Estimation du déplacement des points de contours
- Déplacement horizontal des lignes
- Convolution avec le masque M pour estimer le gradient local

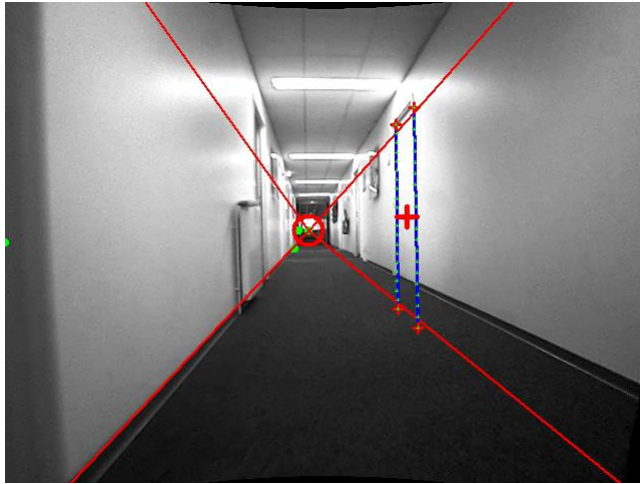
$$P_{t+1} = Q_{t+1}^{j*} = \operatorname{argmax}_{j \in [-J, J]} G_{t+1}^j$$

$$G_{t+1}^j = |I_{t+1}^{v(P_t)} * M + I_t^{v(Q_{t+1}^j)} * M|$$

- [Sekkal2013] Simple Monocular door detection and tracking. ICIP2013



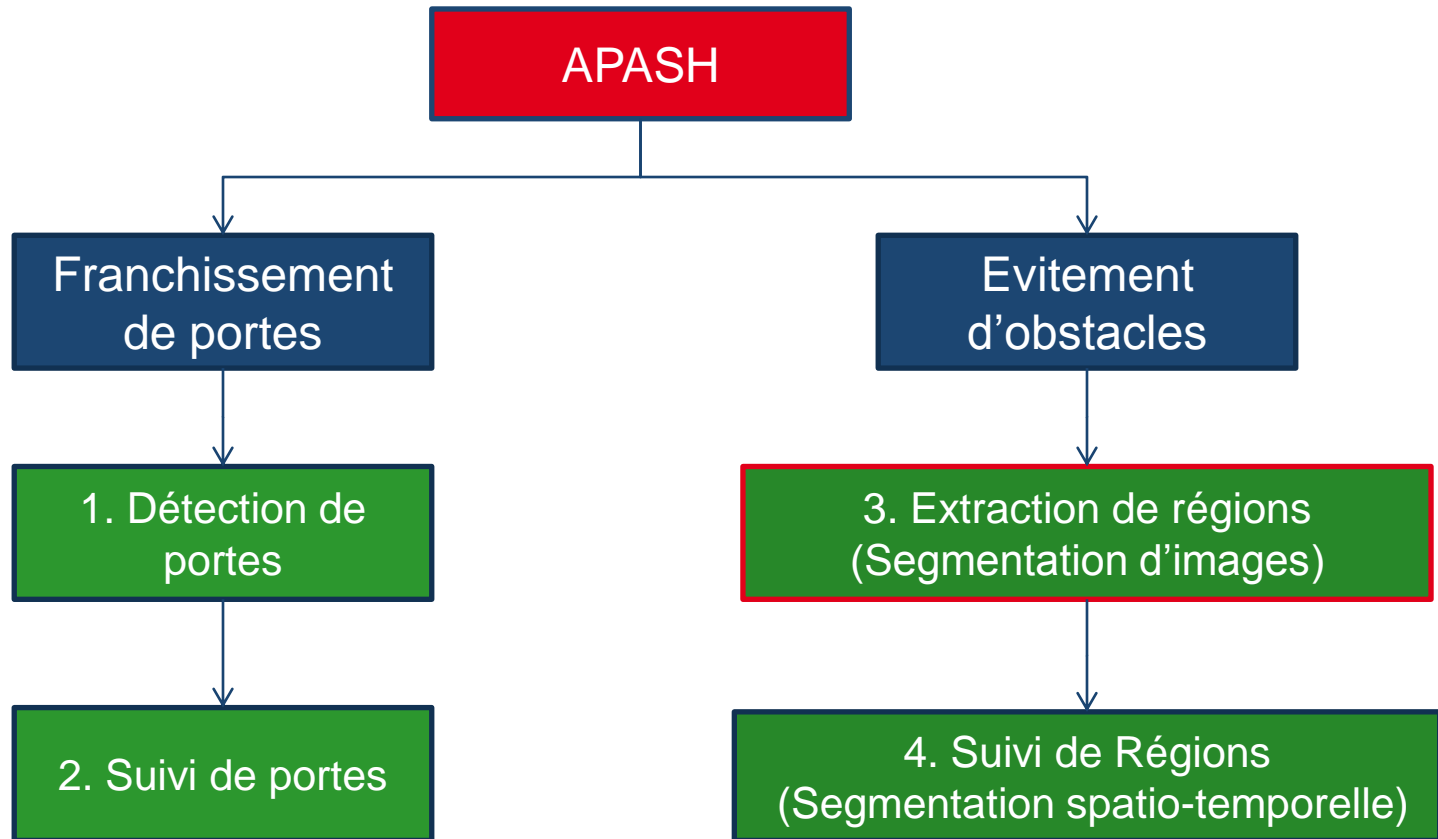
Suivi de portes



Conclusion : détection et suivi de porte

- **Détection des portes :**
 - Utilisation des primitives géométriques
 - Estimation de la localisation et l'orientation de la caméra
 - Détection des formes rectangulaires respectant la taille des portes
 - **Taux de détection de 82%**
- **Suivi des portes :**
 - Estimation du déplacement des montants
 - Suivi 2D des lignes
 - Suivi **temps réel**

Plan



Contexte

- **Objectifs :**
 - Extraire les régions pour initialiser leur suivi
 - Représentation **pseudo-sémantique** du contenu
 - Assurer une complexité raisonnable
 - Réduire la quantité d'information utilisée
 - Travailler avec la scalabilité **hiérarchie et multirésolution**
 - Offrir un résultat à plusieurs niveaux de la sémantique
- **JHMS** (Joint **H**ierarchical and **M**ultiresolution **S**egmentation)
 - [Sekkal12]: Fast pseudo-semantic segmentation for joint region-based hierarchical and multiresolution representation SPIE2012

Représentation pseudo-sémantique

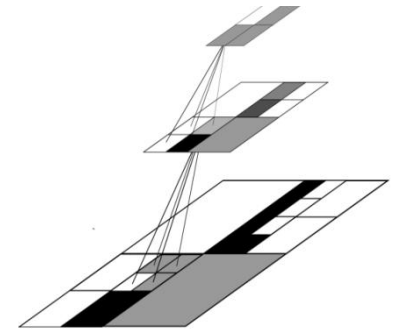
- **Segmentation d'image :**
 - Passer du niveau **pixel** au intermédiaire **riche en sémantique**
 - **Régions homogènes** en termes de :
 - Couleur, Texture, Gradient
- **Différentes approches :**
 - Approche par **recherche de discontinuités**
 - Détection des contours [Martin04, Maire08]
 - Contours actifs [Herbulot07]
 - Approche par **détection zones homogènes**
 - Croissance de régions [Shih05, Tang10]
 - Division fusion [Merigot03, Aneja09]



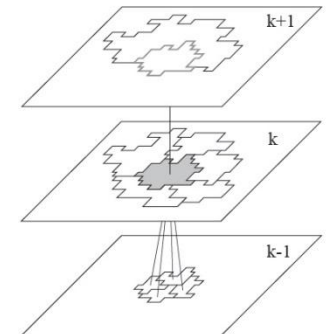
Représentation pseudo-sémantique

Segmentation d'images: JHMS

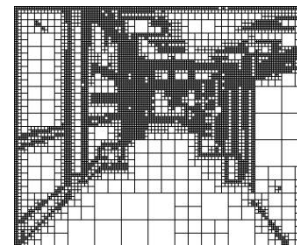
- Deux types de scalabilité :
 - **Spatiale**: multirésolution (taille de l'image)
 - Pyramides régulières [Brun02, Stojmenovic10, Vantaram09]
 - **Sémantique**: hiérarchie (niveaux de granularité)
 - Pyramides irrégulières [Cheng00, Arbelaez11]
- **Solution proposée**:
 - Travailler sur une **pyramide d'images**
 - Utiliser une représentation intermédiaire (**quadtree**)
 - Affiner la segmentation d'un niveau à un autre (**RAG multirésolution**)



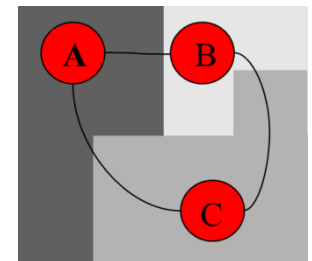
Pyramide régulière (multirésolution)



Pyramide irrégulière (hiérarchie)



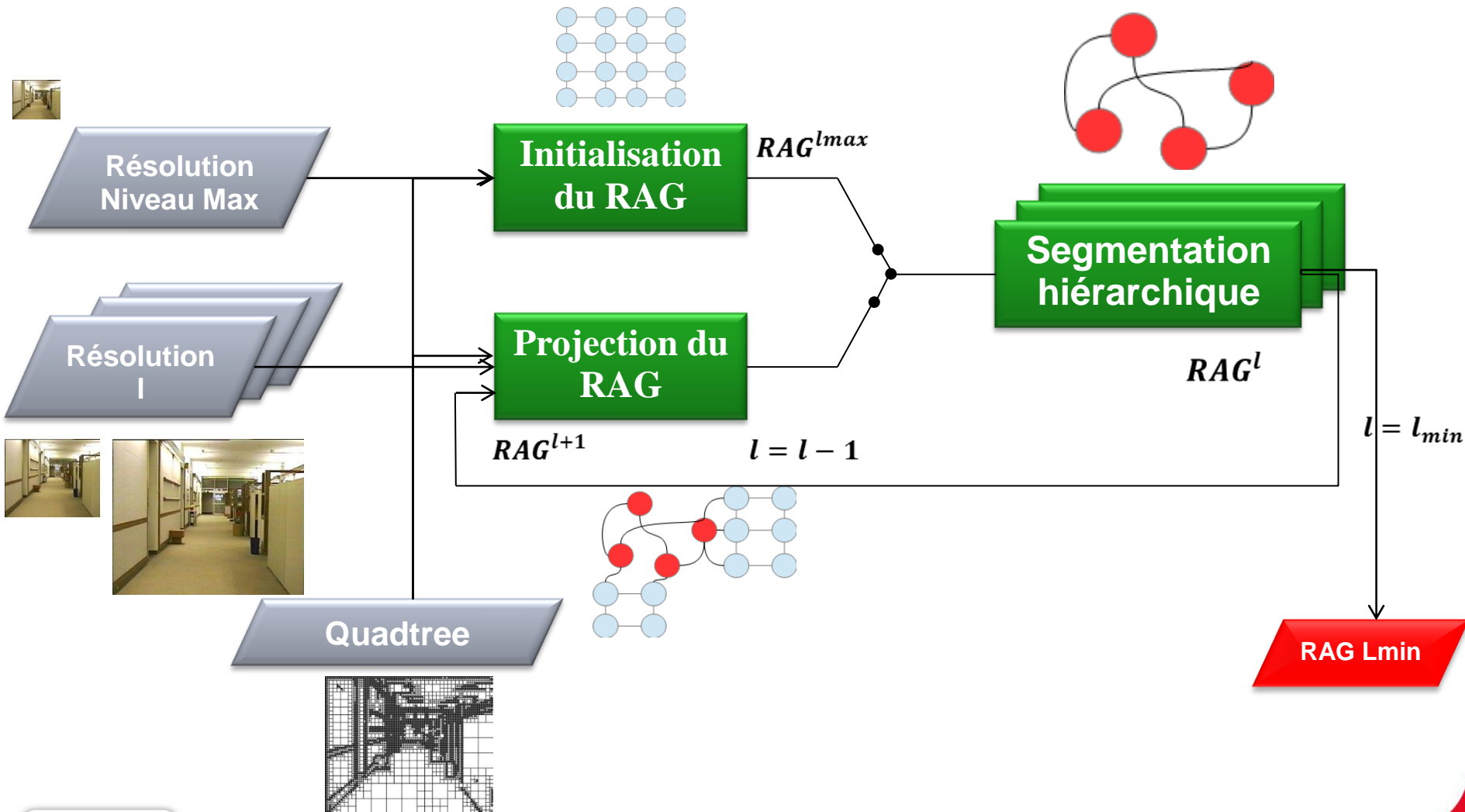
Quadtree



RAG

T_y seuil de décomposition Region Adjacency Graph

Schéma général de segmentation JHMS

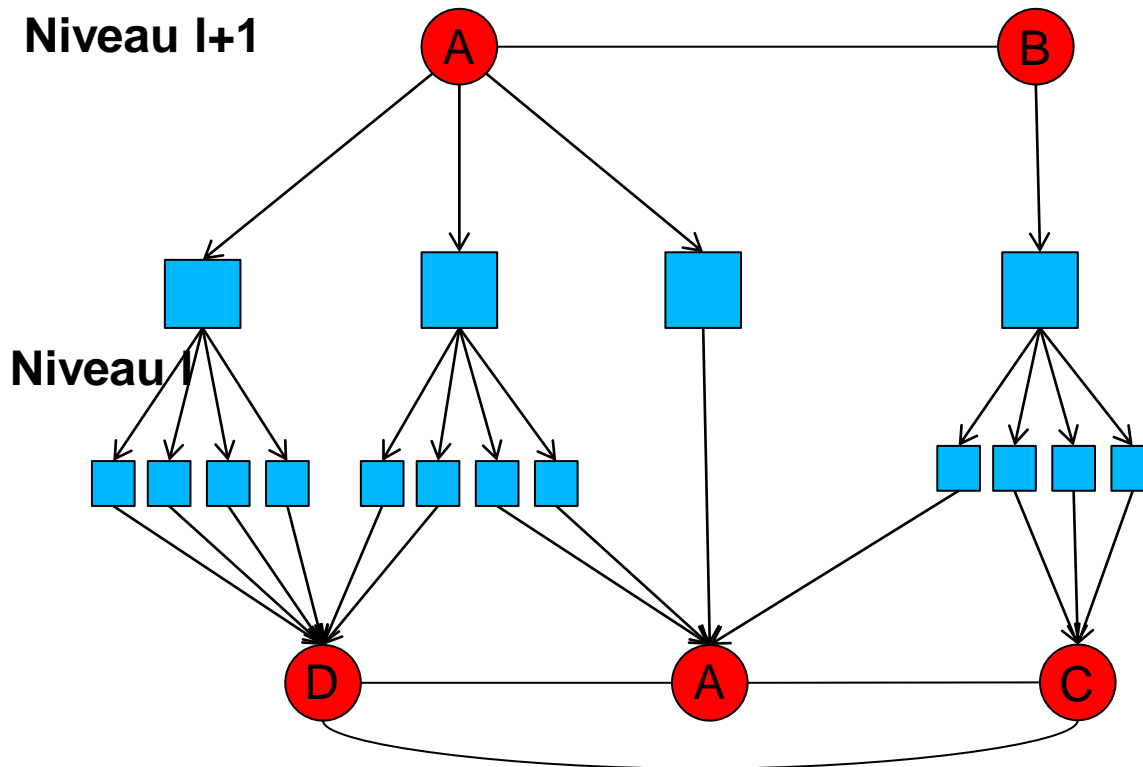


Segmentation hiérarchique

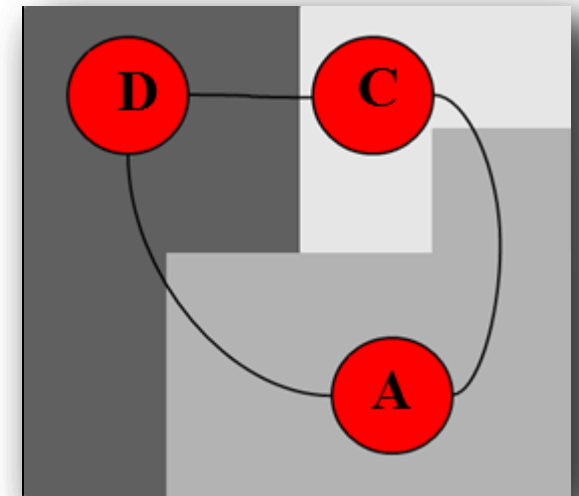
- Passer d'une représentation sur-segmentée à une représentation grossière (2 à 3 régions) à travers processus de fusion
- **Paramètres de fusion :**
 - T_{moy} : différence des moyennes de deux régions
 - T_{grad} : gradient sur les frontières de deux régions
 - Nb_{hier} : nombre de hiérarchies dans la pyramide
- Processus fusion [Strauss10]



Projection du RAG Multirésolution



Décomposition



Résultats de segmentation

- **Benchmark SEISM [Pont2013] :**

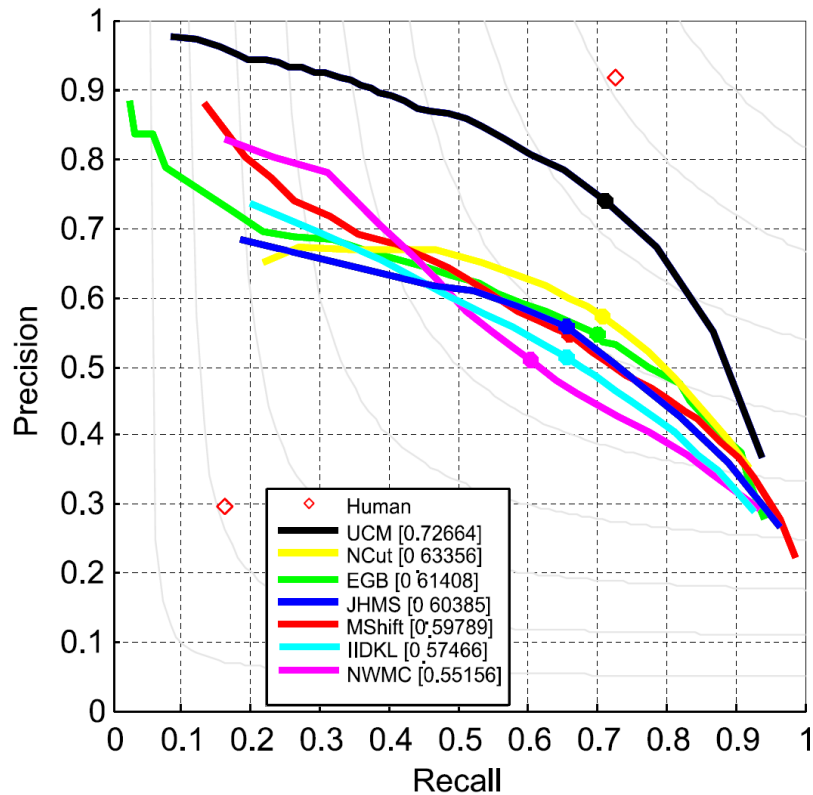
- Technique d'évaluation des algorithmes de segmentation d'images
- BSDS500 [Arbelaez11] (300 images d'apprentissage, 200 images de validation)
- Vérité terrain dessinée par des sujets humains (des objets sémantiques)
- Deux mesures complémentaires
 - F_b → qualité des contours (précision rappel des contours détectés)
 - F_{op} → qualité des régions (recouvrement des régions segmentées avec les régions vérité terrain)

- **Protocole d'apprentissage :**

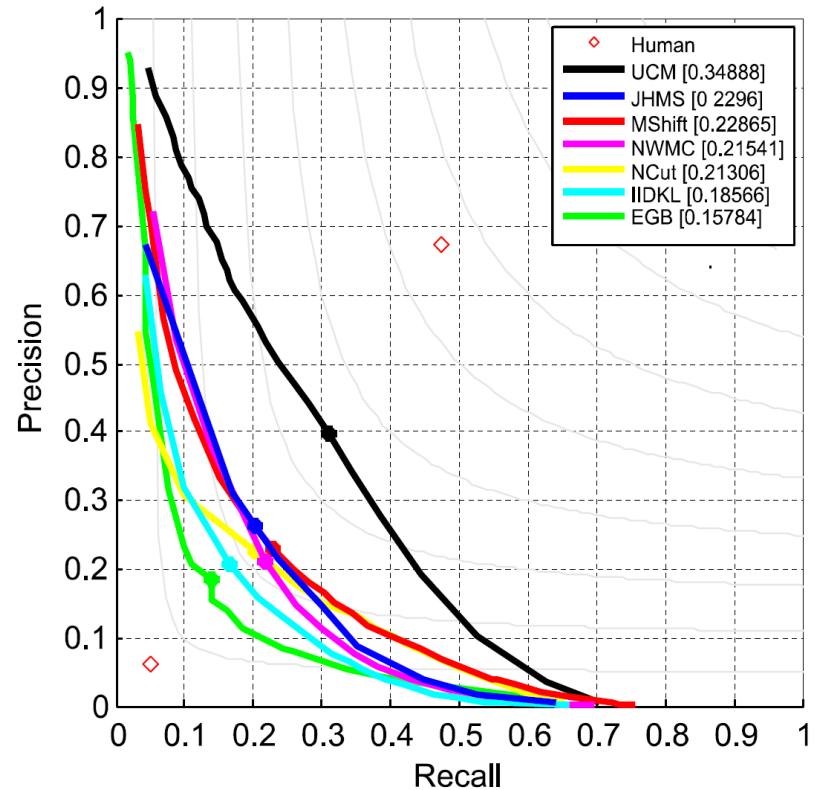
- Recherche de l'ensemble de paramètres optimaux
(T_y , T_{moy} , T_{grad} , nb_{Hier})
- T_y : Seuil de décomposition du quadtree

Résultats SEISM

- Les paramètres retenus sont :
 - 1^{er} niveau de hiérarchie $T_y = 20, T_{moy} = 6, T_{grad} = 2$
 - Varier le nombre de hiérarchies de $nb_{Hier} = \{1..5\}$



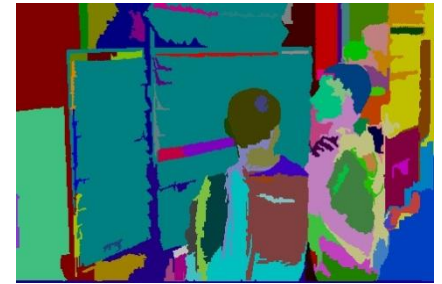
F_b : Contours



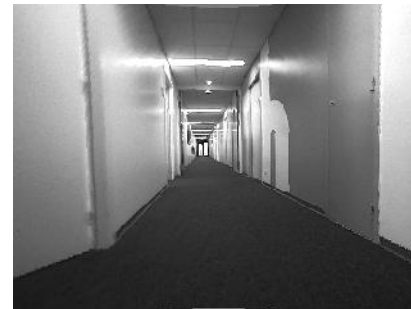
F_{op} : Régions

Résultats visuels

$$T_y = 20, T_{moy} = \{6, 12, 18, 24\}, T_{grad} = \{2, 4, 6, 8\}, nb_{Hier} = 4$$



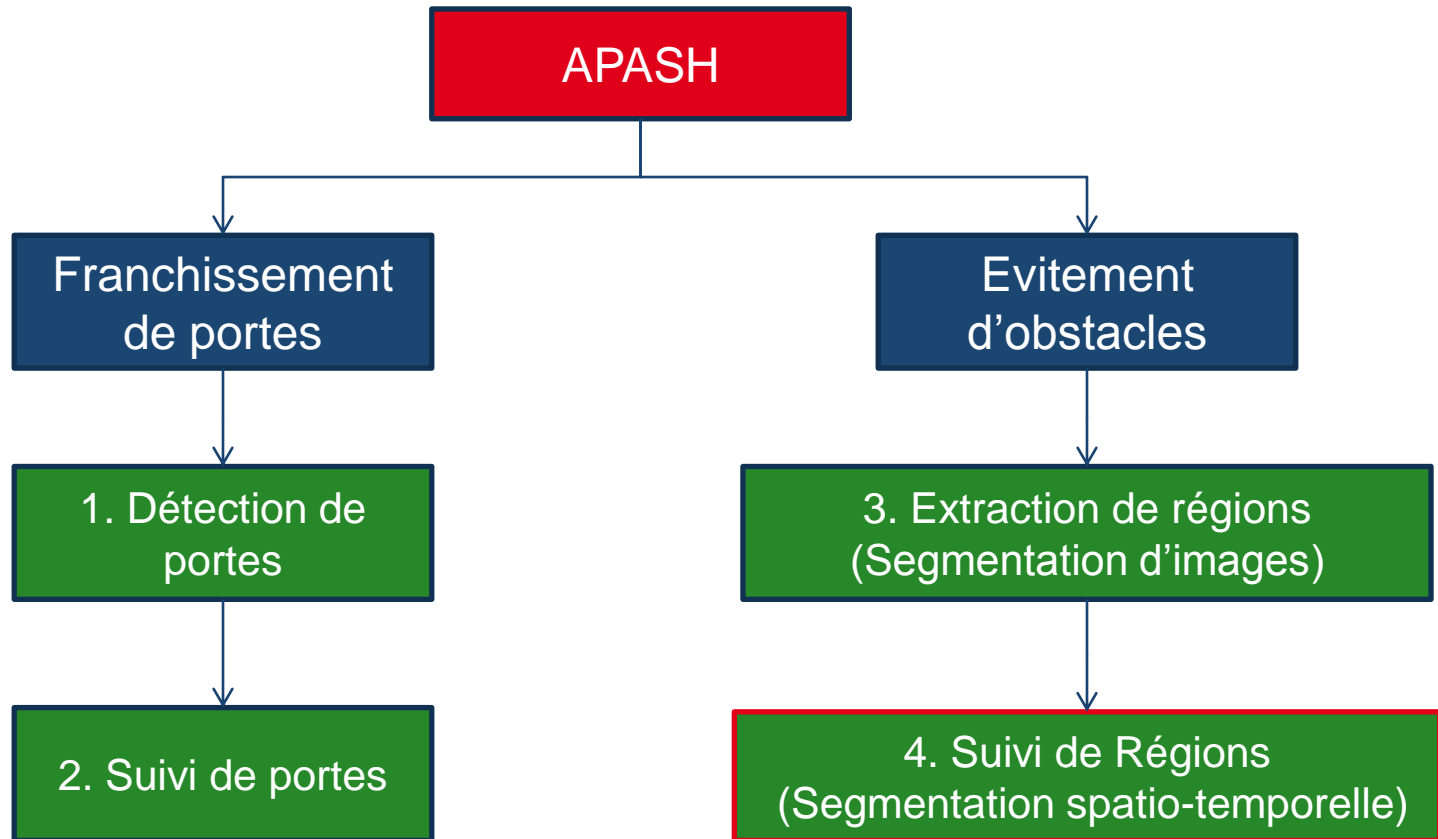
- Images à faible éclairage
- Réduction des seuils de fusion
- $T_y = 20, T_{moy} = \{2, 4, 6, 8, 10\}, T_{grad} = \{1, 2, 3, 4, 5\}, nb_{Hier} = 5$



Conclusion

- **Segmentation d'images JHMS**
 - Extraction des régions pour une initialisation du suivi
 - Représentation pseudo-sémantique
 - Scalabilité multirésolution et hiérarchique
 - Utilisation du **quadtree** et une **pyramide régulière**
 - Proposition du **RAG multirésolution** afin de raffiner la segmentation d'un niveau de résolution à un autre
- Résultats obtenus sur le benchmark SEISM
 - Une bonne détection des contours
 - Une bonne qualité des régions extraites

Plan



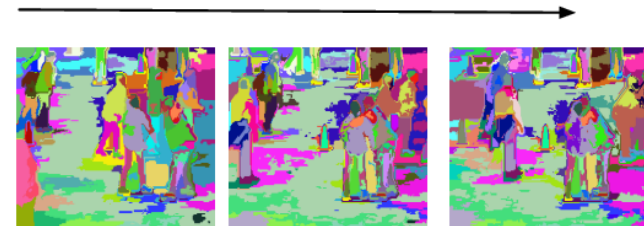
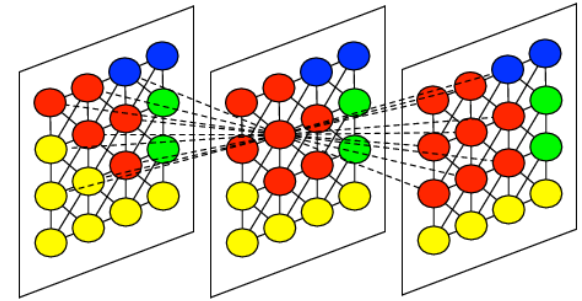
Segmentation spatio-temporelle

- **Objectif:**
 - Suivre des régions pseudo-sémantique
 - Assurer la cohérence spatio-temporelle
- **Deux types d'information:**
 - **Information spatiale:** couleur, gradient
 - **Information temporelle:** mouvement
- **Problèmes**
 - Complexité de la scène (mouvement)
 - Complexité maîtrisée (consommation mémoire)
 - Traitement au fil de l'eau (navigation)

Segmentation spatio-temporelle

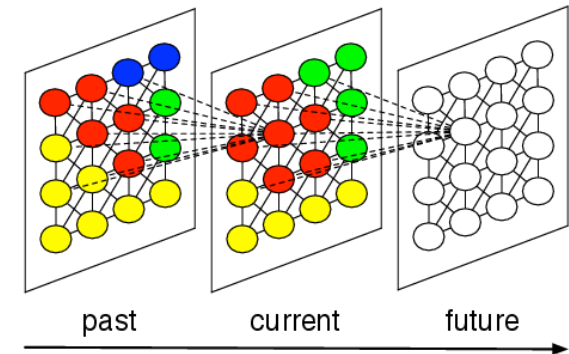
- **Techniques 3D :**

- Segmentation de l'ensemble des images comme un volume [Grundmann10, Palou13]
 - + trajectoire complète du pixel
→ plus de précision
 - - chargement de la séquence et les descripteurs dans la mémoire



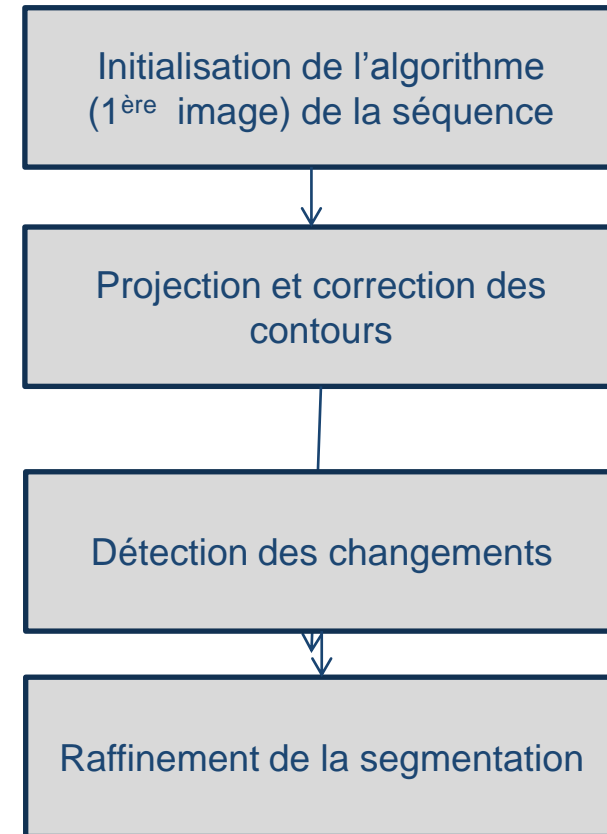
- **Techniques 2D+T :**

- Projection des régions sur deux frames successives [DelBimbo01, Galmar05, Xu12]
 - + processus séquentiel
 - - petite portée de l'information
→ moins de précision



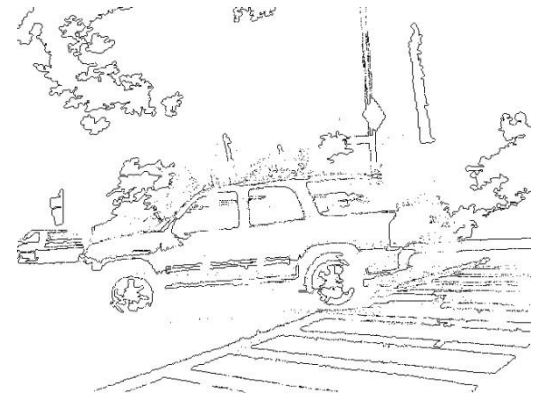
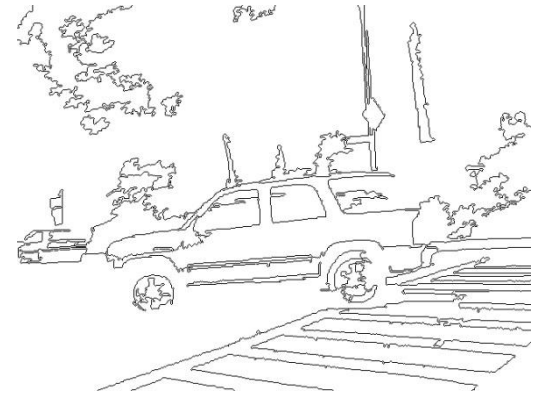
Segmentation par projection de contours

- **Approche proposée: JHMS+T**
 - Représenter les régions par leurs contours
 - Suivre les contours
 - **Propriété** : zones à fort gradient
→ des bons points à suivre
- **Principe :**
 - **Initialisation**: extraire les contours
 - **Estimation le mouvement** du contour
 - **Détection des changements** :
zones de mouvement et nouvelles régions
 - **Raffinement de la segmentation** :
Mettre à jour le RAG (ajout, suppression et modification)



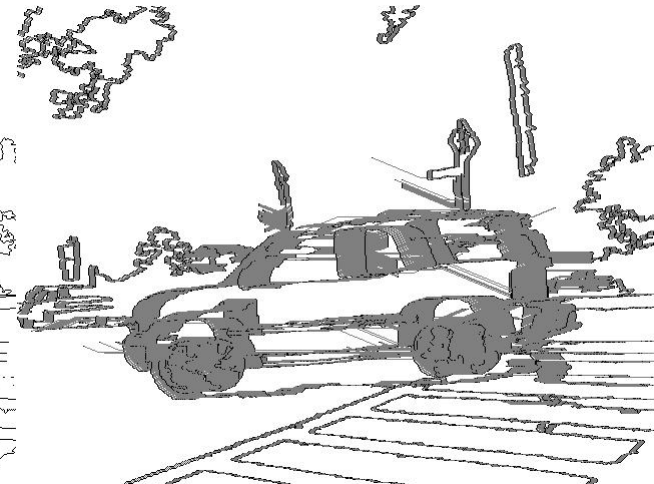
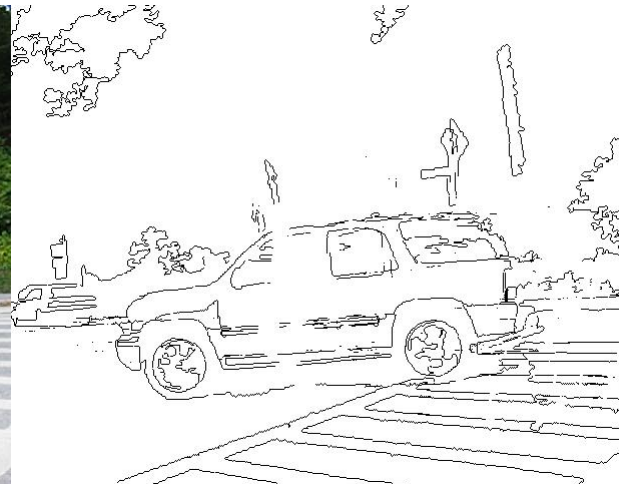
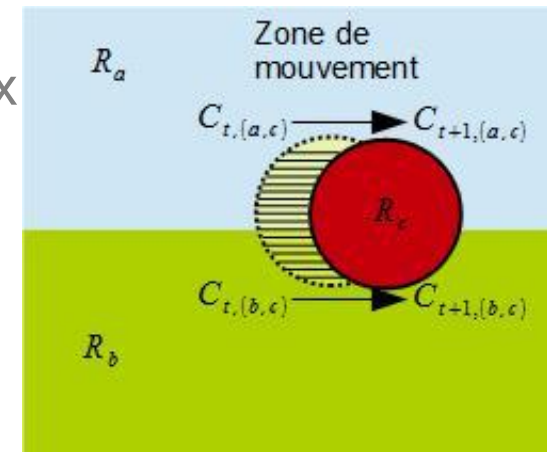
Projection des contours

- **Initialisation :**
 - Extraction des contours : **JHMS**
- **Projection :**
 - Fenêtre temporelle de deux frames successives
 - Projection en avant des points de contours
 - Utilisation du **KLT** Kanade-Lucas-Tomasi [Tomasi91]
 - Correction des contours à travers la modélisation de leur trajectoires (mélange de gaussiennes)



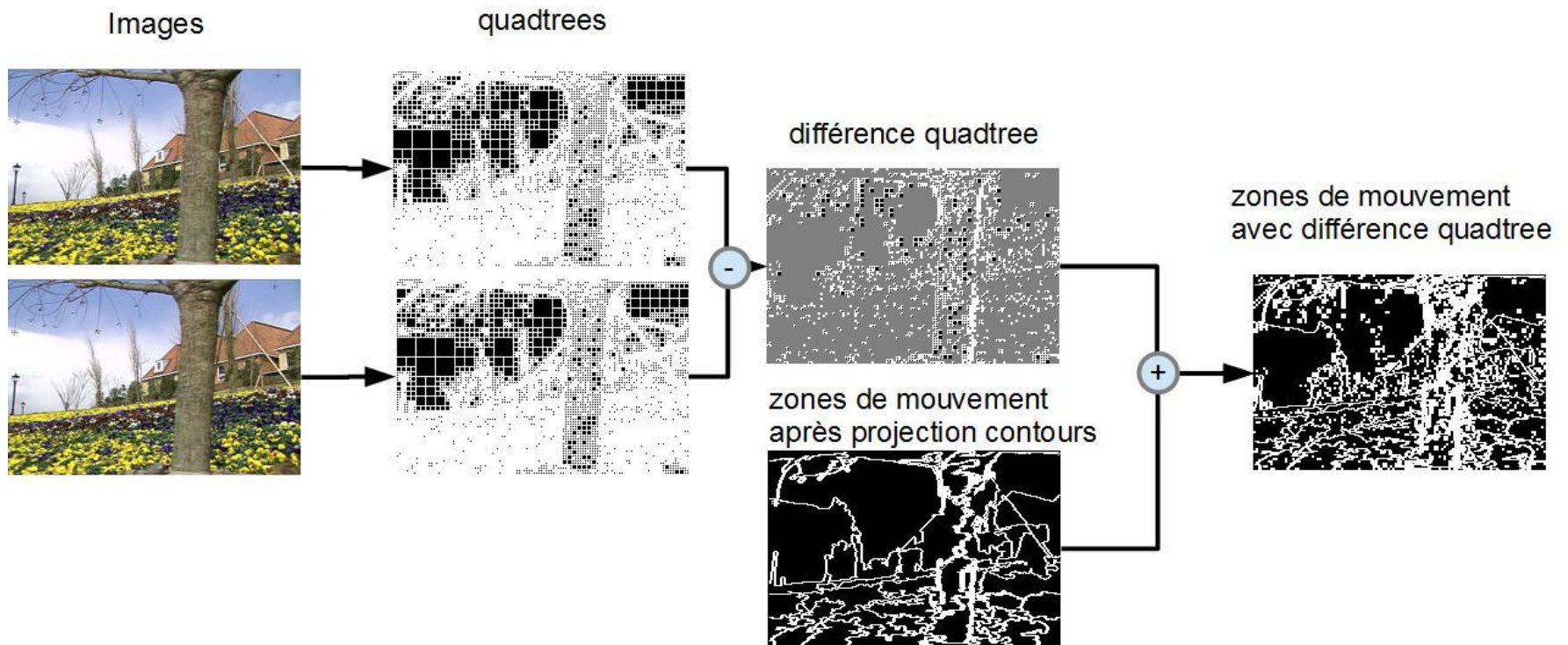
Zones de mouvement

- Zones correspondant aux changements dues aux mouvements des régions
- **Détection des zones de mouvements**
 - Pixels se trouvant **entre le contour initial** et le **contour projeté**



Détection des nouvelles régions

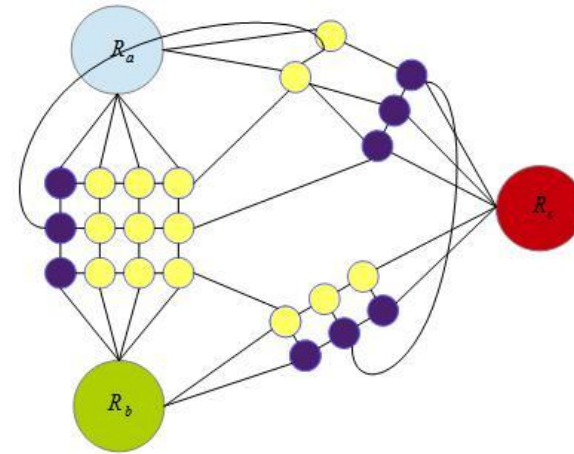
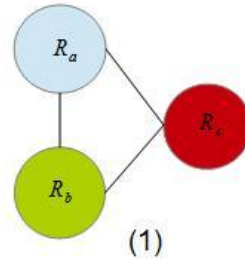
- Appliquer une différence de quadrees entre deux images successives
- Ajouter aux zones de mouvements



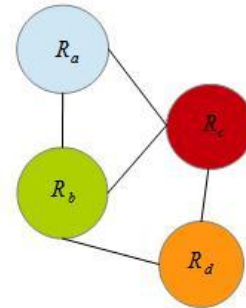
Raffinement de la segmentation

- **Etat du RAG:**

- Zones de mouvement
- Nouvelles régions
- Régions projetées



● Zones de mouvement
● Nouvelles régions



- **Mise à jour du RAG**

- Technique de **croissance de régions**
- Pour chaque **région projetée**, on fusionne tous les éléments autour qui respectent les critères de fusion
- Créer de **nouvelles régions** si nécessaire

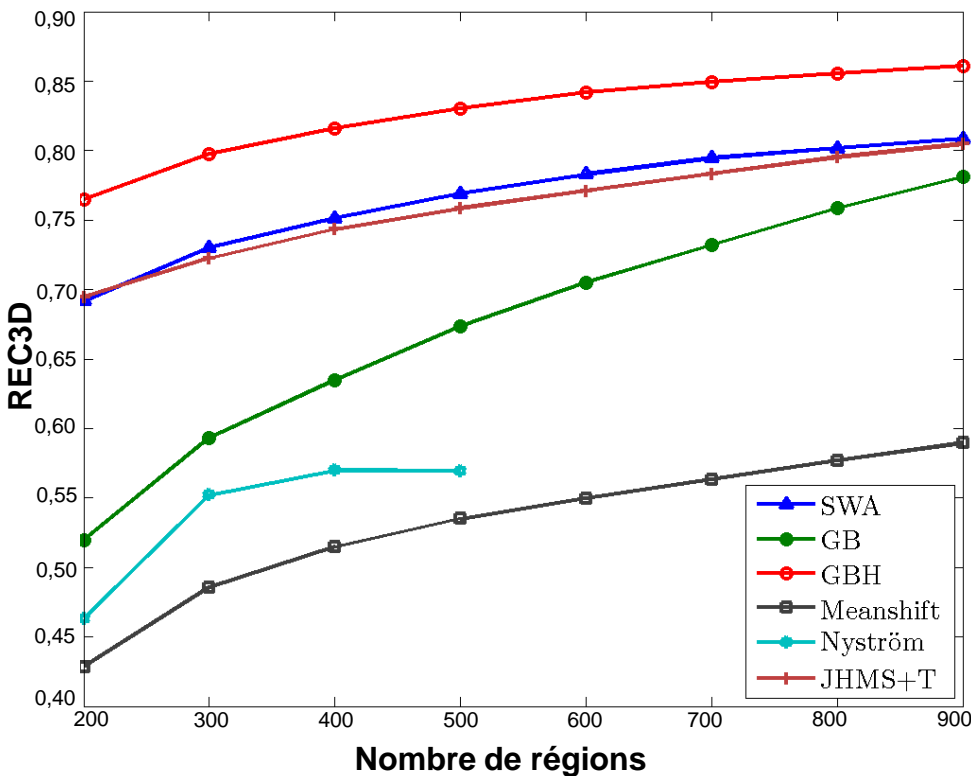
➔ Assurer la **cohérence des étiquettes** des régions existantes

Résultats de la segmentation

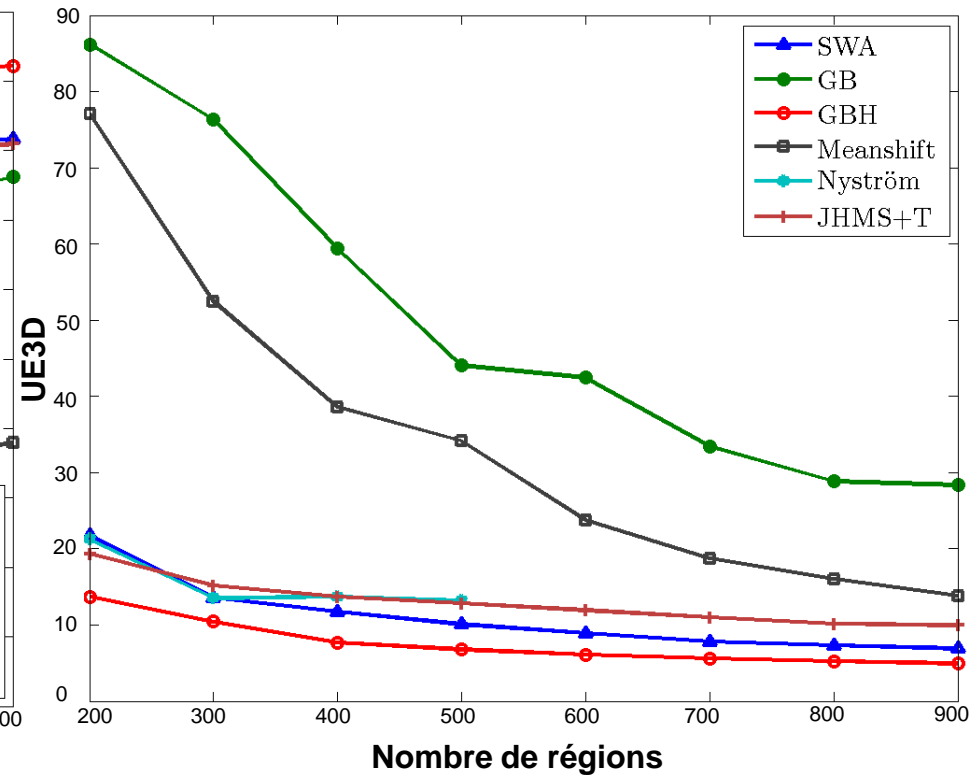
- Pas de benchmark intégrant les techniques 2D+T
- **Benchmark Xui.org [Xu13]** :
 - Techniques de segmentation 3D
 - 8 vidéos de 80 images
 - 200 – 900 régions par séquence
 - Plusieurs métriques sont proposées
 - Précision de la segmentation
 - Rappel des contours
 - Erreur de sous segmentation
 - Durée de vie des régions

Comparaison avec l'état de l'art

- Bon taux de rappel de contours
- Faible erreur de sous-segmentation



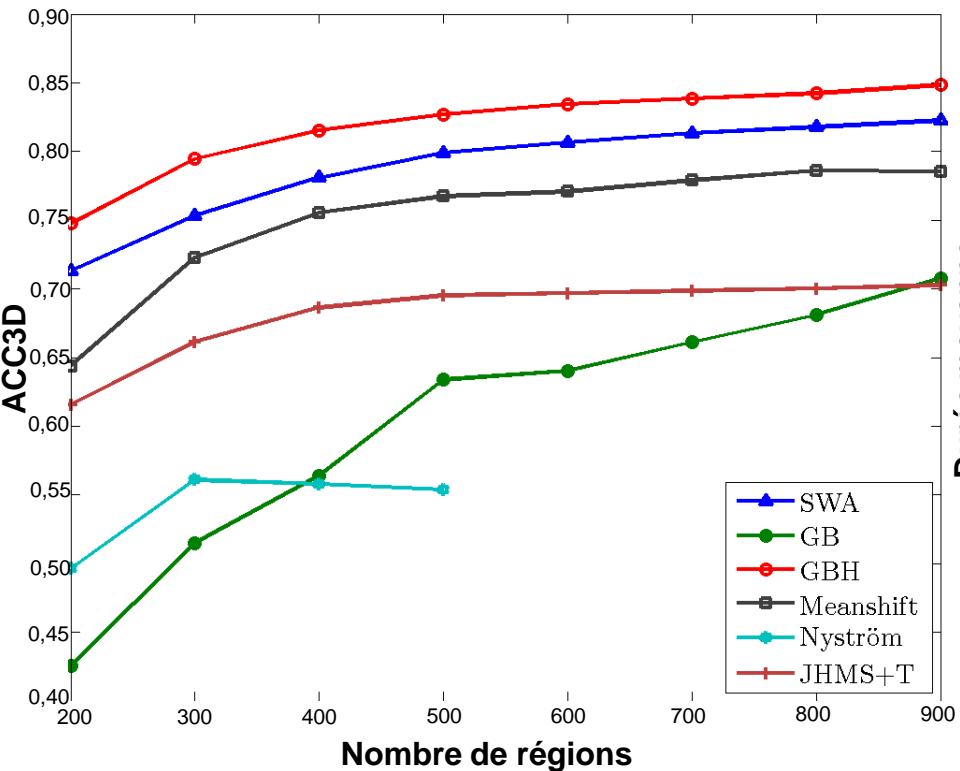
Recall3D: Rappel de contours



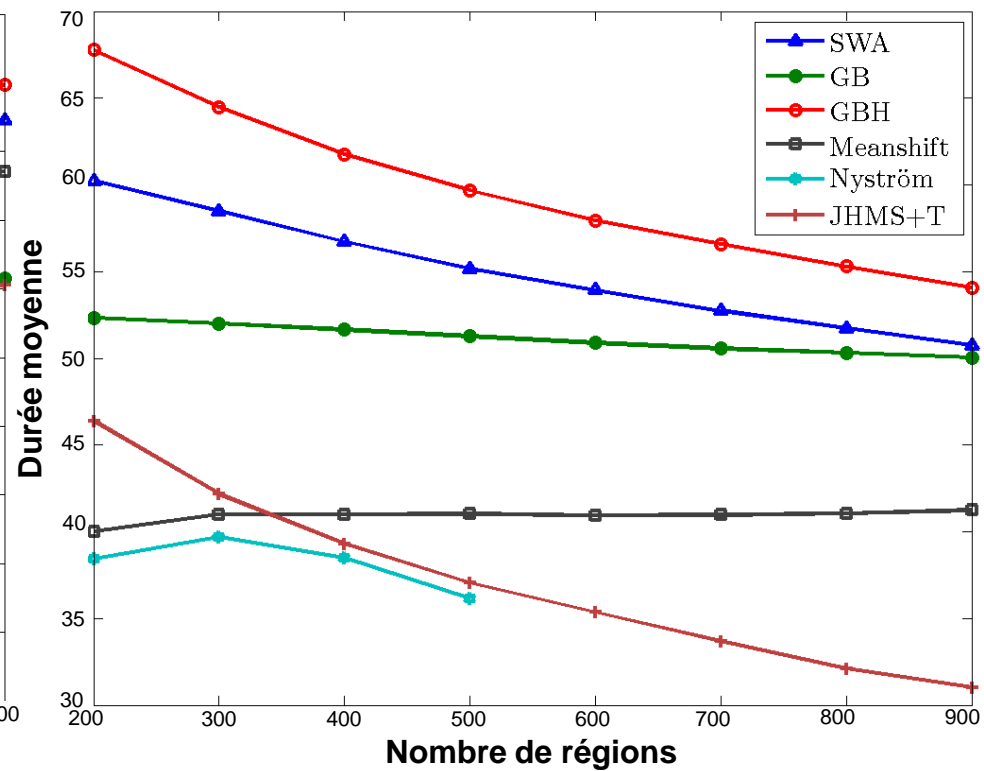
UE3D: Erreur de sous-segmentation

Comparaison avec l'état de l'art

- Bonne précision des régions
- Durée moyenne des régions moins importante



ACC3D: Précision de segmentation



MD: Durée moyenne

Résultats visuels

- Vidéos benchmark



- Vidéos de couloir



Conclusion : Segmentation 2D+T

- **JHMS+T**
 - Adapté au vidéo online
 - Suivi des régions par le suivi de leur contours
 - Projection correction des contours
 - Détection de nouvelles régions
- **Résultats objectifs :**
 - Comparables aux techniques 3D
 - Cohérence temporelle:
 - Bon suivi des étiquettes
- Moins coûteux en termes de consommation (seulement image courante et le RAG)

Conclusion et perspectives

Conclusion

- Techniques de suivi de détection et suivi d'objets 2D
- **Détection et suivi de portes dans un couloir**
 - **Localisation** de la caméra dans le couloir
 - Estimation de la forme de la porte dans l'image
 - **Reconnaissance** de plus de **82%** des portes
 - **Suivi** des montants de portes
- **Segmentation d'images JHMS**
 - Représentation quadtree
 - **RAG multirésolution**
 - **Segmentation hiérarchique**
 - Bons résultats objectifs et visuels

Conclusion

- Suivi de régions dans séquence d'images 2D+t :
 - Représentation des régions par leurs contours
 - Suivi de contours en utilisant KLT
 - Projection et correction des déplacements de contours
 - Détection des zones de mouvement et les nouvelles régions
 - Raffinement de la segmentation par croissance de région
 - Résultats objectifs comparables aux techniques 3D
 - Résultats visuels pertinents

Perspectives

- **Détection et suivi de portes :**
 - Etendre la détection de porte pour d'autres environnements (hall, chambre..)
 - Point de fuite non-visible
 - Robustesse dans le suivi : filtre de Kalman, odométrie
- **Segmentation d'images :**
 - Adaptation des seuils de fusion au contenu de l'image
 - Utilisation des lignes de fuite pour améliorer la segmentation
- **Segmentation vidéos :**
 - Utiliser un suivi de points autour de la normale
 - Ajouter le contrôle du nombre de régions