



# Détection et modélisation de gestes monoculaires par recherche stochastique

**B. Bascle**

**Orange Labs / FT R&D**

**<http://perso.rd.francetelecom.com/bascle/>**



# Motivation



## → Analyse du mouvement 3D du haut du corps / gestes

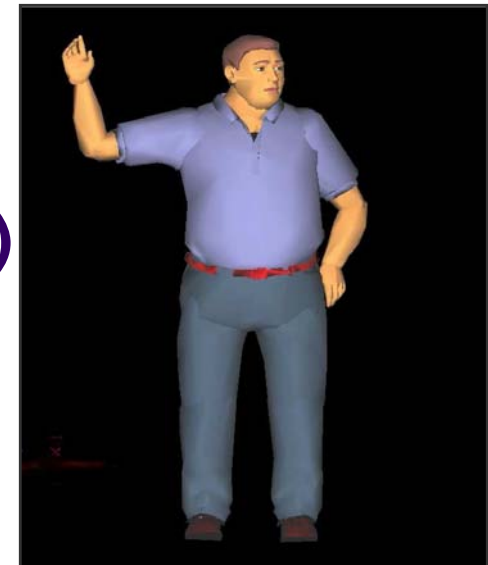
- animation de clones
- interfaces gestuelles
- Réalité augmentée



Images courtesy  
of O. Bernier

## → à partir d'une seule caméra

- grand public (coût, calibration, synchronisation)
- réutilisation de vidéos existantes



# Détection et suivi de corps articulé: un problème complexe



- Difficultés habituelles de la détection et du suivi d'objet:
  - Fonds complexes (clutter),
  - Variations d'illumination,
  - Apparences variables (vêtements, orientation).
  - Occultations.
- Objet articulé
  - Déformable avec un grand nombre de degrés de liberté
    - ~ 30-40 pour un corps complet,
    - ~ 13-20 pour le haut du corps.
  - Auto-occultations.
- Une seule caméra (manque d'info de profondeur, ambiguïtés)

# Etat de l'art monoculaire



## → suivi

- filtrage particulière [Isard & Blake 98]
- modèle "loose-limbed" avec propagation de croyance [Sigal & Black 06]
- apprentissage de l'apparence [Ramanan *et al* 04]
- ...

## → détection

- régression directe par les SVMs [Agarwal & Triggs 06]
- hashing [Shakhnarovich 03]
- proposal maps driven MCMC [Lee & Cohen 04]
- détection de poses "stylisées" [Ramanan *et al* 04]
- ...

# Etat de l'art monoculaire



## → suivi

- exploite la continuité du mouvement, de la pose et de l'apparence d'une image à l'autre
- problème d'initialisation et de ré-initialisation quand le suivi échoue

## → détection

- beaucoup d'approches restreintes à des postures ou à des mouvements limités et appris
- multiples solutions et faux positifs possibles

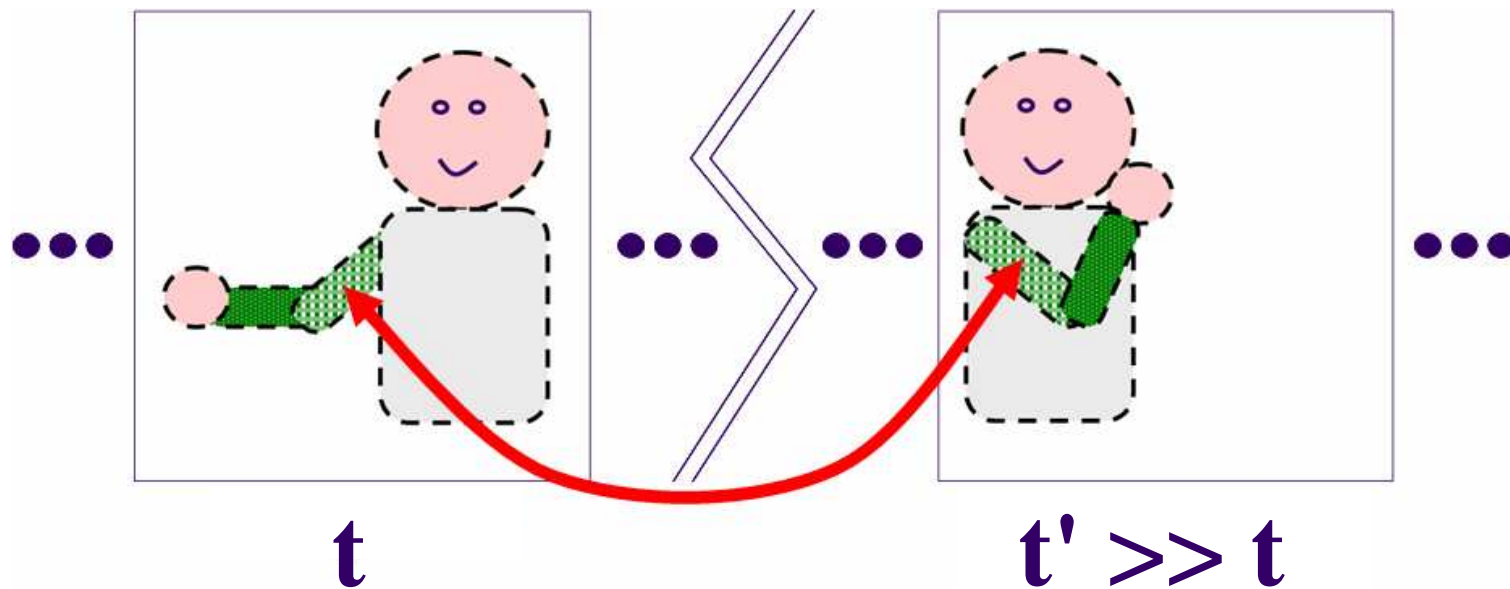
→ idée: exploiter la continuité de l'apparence sur de grands intervalles de temps

# Etat de l'art monoculaire



→ idée: exploiter la continuité de l'apparence (couleur, texture) sur de grands intervalles de temps

- Détection
- Apprentissage en ligne de l'apparence et des mensurations
- Suivi rapide ultérieur



# Etat de l'art monoculaire



## → approche

- génération d'hypothèses à partir des images par inférence sur un modèle graphique
- confrontation d'hypothèses séparées par de grands intervalles de temps (choisis aléatoirement)
- rejet d'hypothèses par mécanisme MCMC

## → intérêt

- robustesse aux interruptions du suivi et aux discontinuités
- robustesse aux conditions adverses à la détection (clutter, occultations, mouvements imprévisibles, lents puis rapides)
- robustesse aux faux positifs
- pas de poses pré-déterminées

# Approche:

## 1. paramétrisation du haut du corps

- paramétrisation 2D

→ reconstruction par la methode de [ Taylor 2000 ]



Images courtesy of C.J. Taylor

- bras détectés séparément pour réduire la dimension de l'espace de recherche

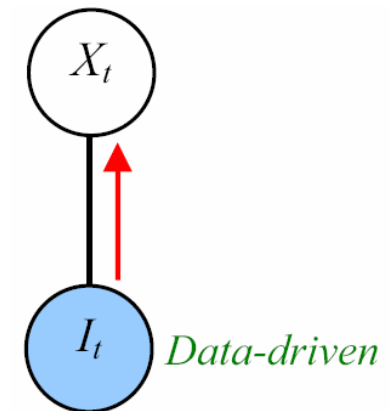
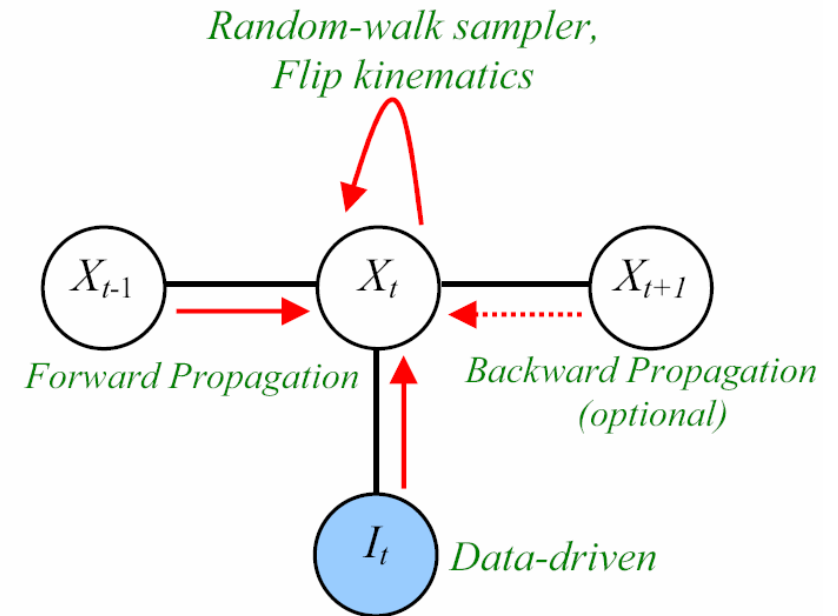


# Approche:

## 2. génération stochastique d'hypothèses



- proposal maps driven MCMC [Lee & Cohen 04] [Lee & Nevatia 05]
  - random walk sampler
  - flip kinematics
  - motion dynamics
  - data driven
- Here: MCMC with independent proposals
  - minimum discrepancy / spacing
  - stratified sampling
  - purely data driven



# Approche:

## 2. génération stochastique d'hypothèses



classic MCMC

$\Omega$

$x_0$

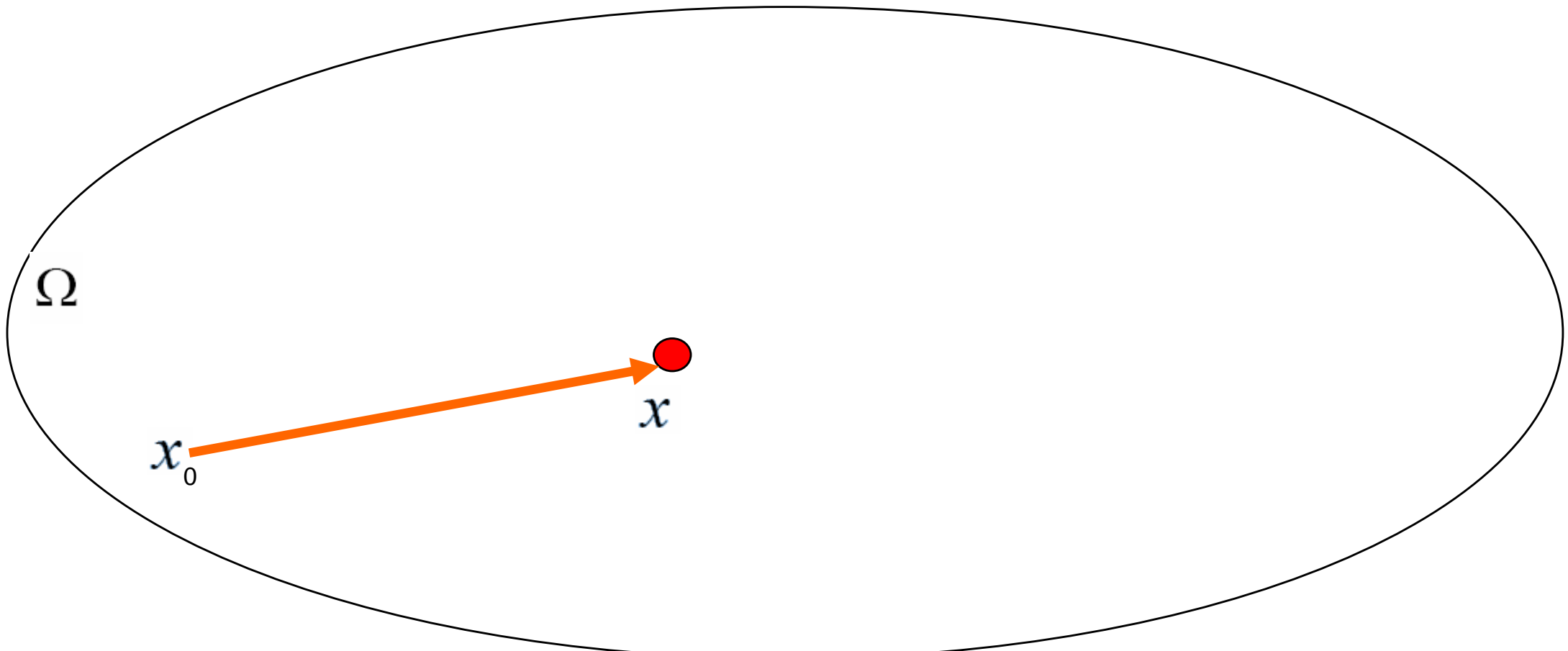


# Approche:

## 2. génération stochastique d'hypothèses



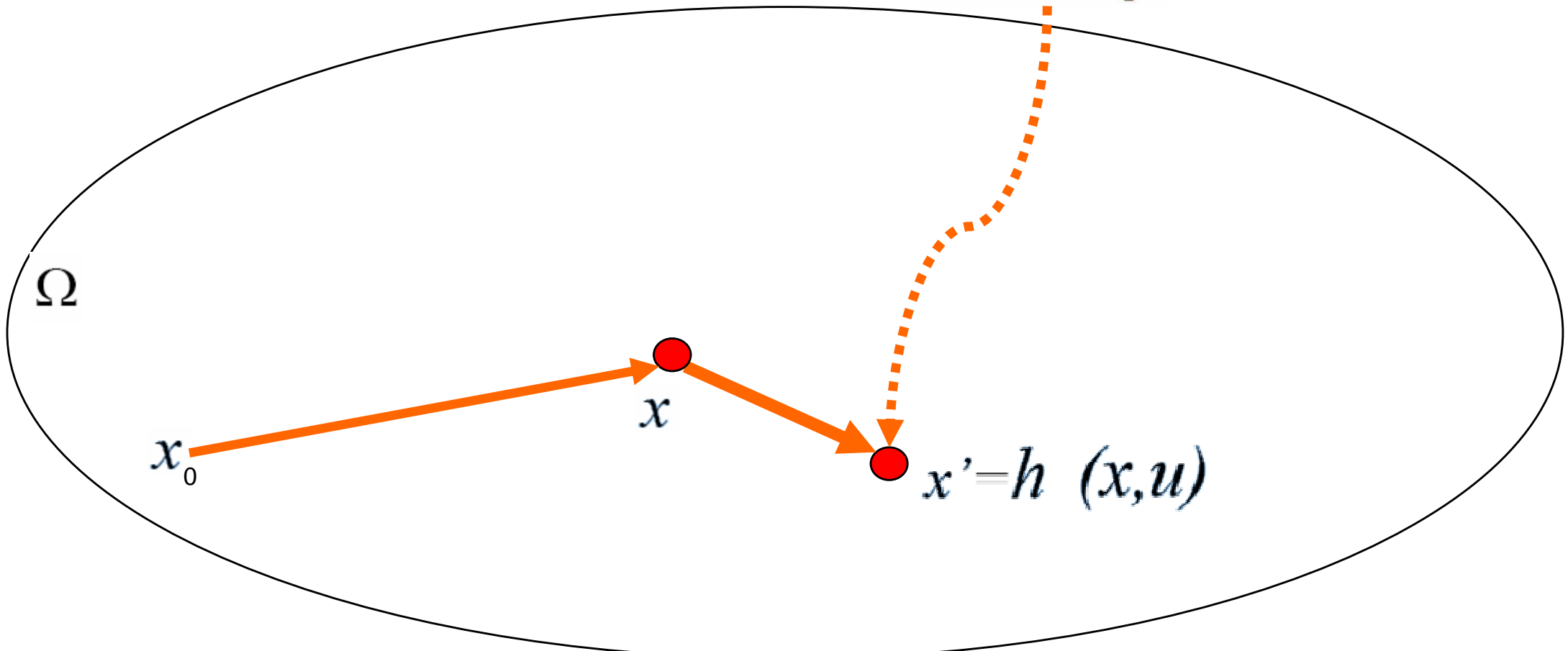
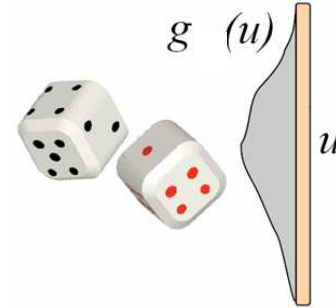
classic MCMC



# Approche:

## 2. génération stochastique d'hypothèses

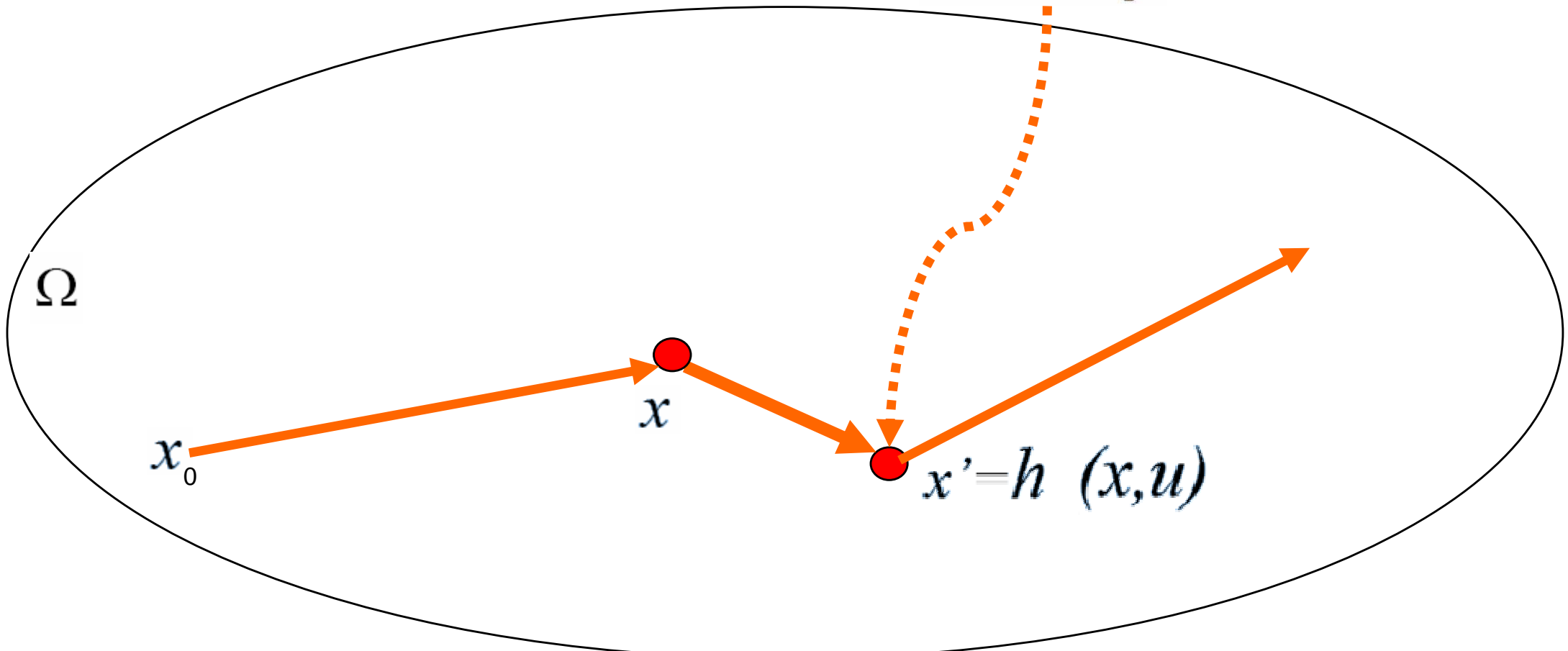
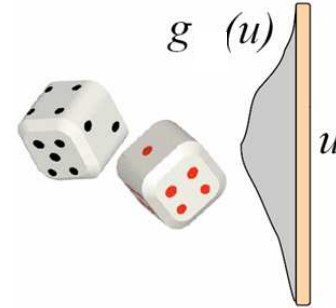
classic MCMC



# Approche:

## 2. génération stochastique d'hypothèses

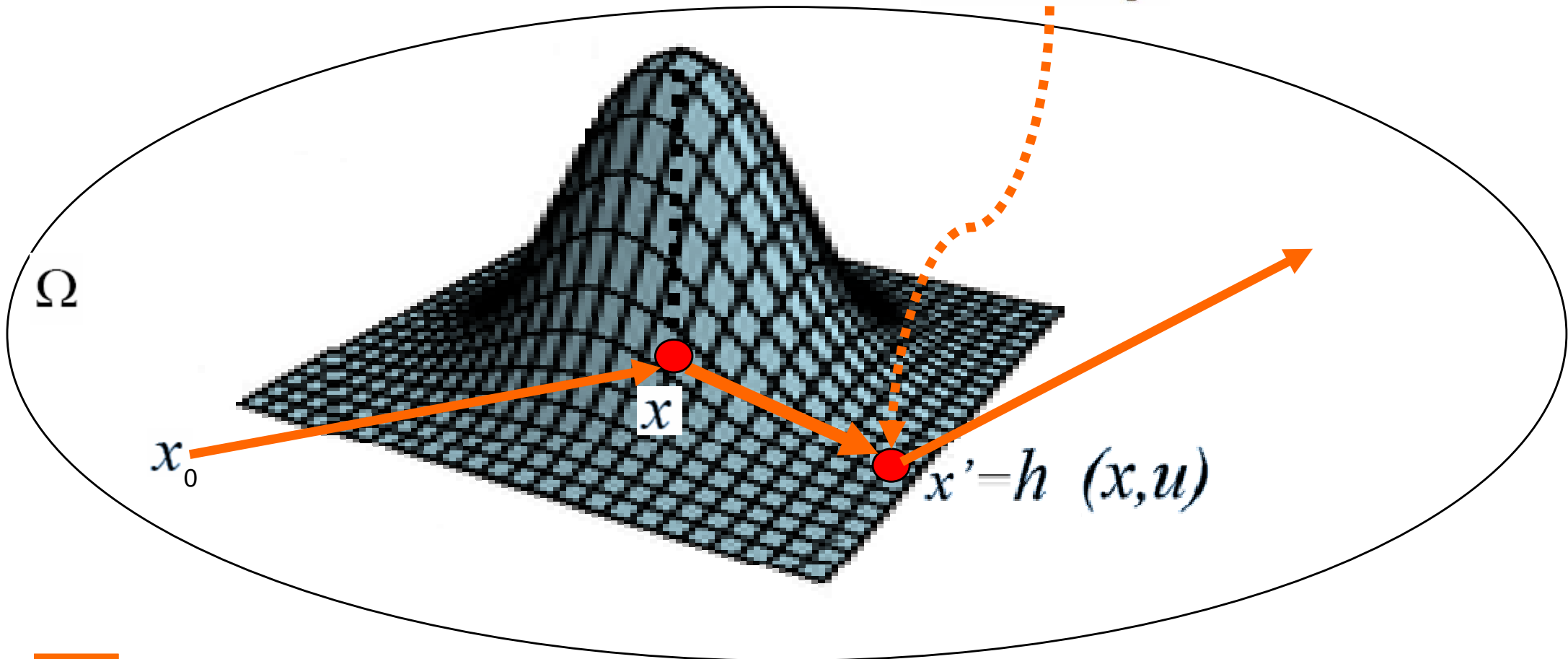
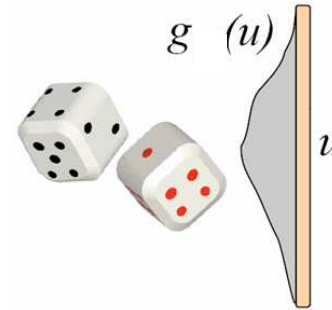
classic MCMC



# Approche:

## 2. génération stochastique d'hypothèses

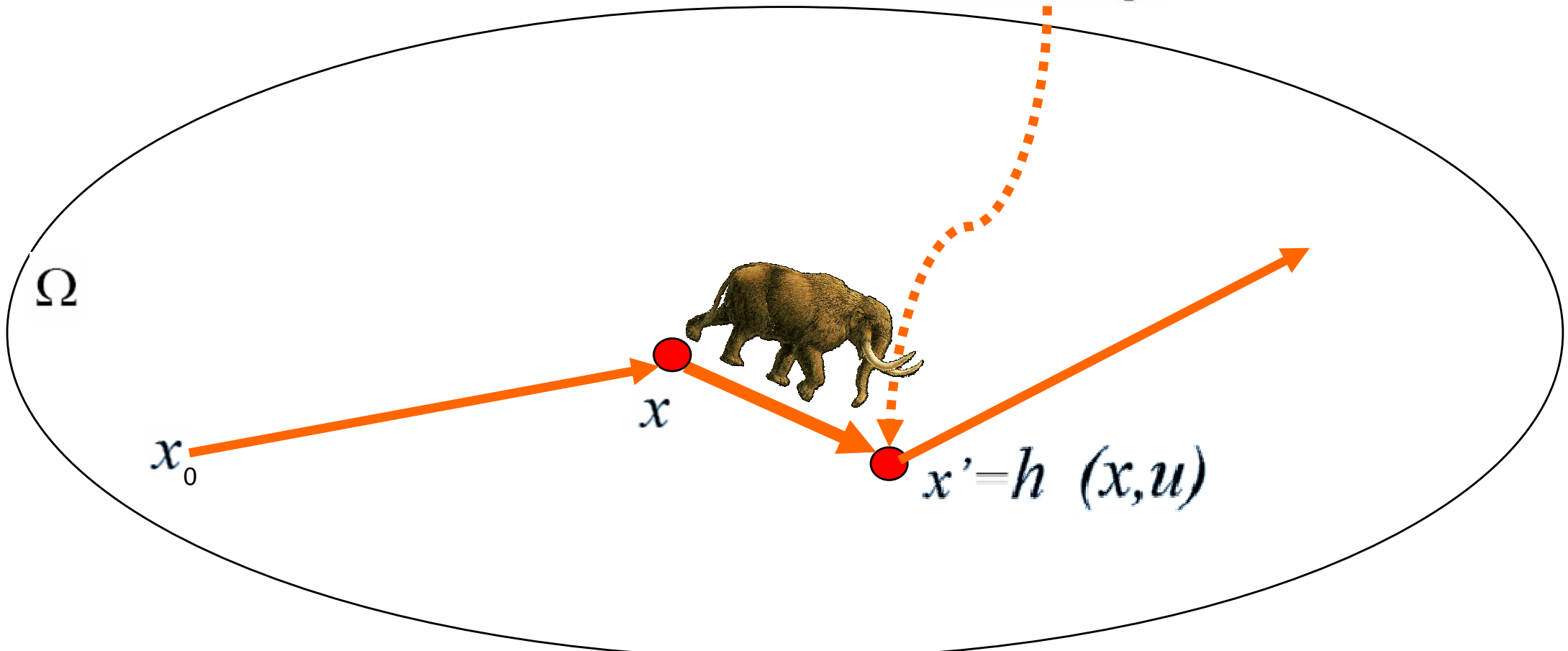
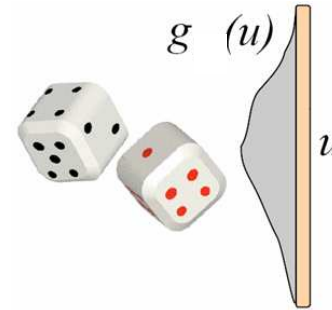
classic MCMC random walk:  
small displacement  
(gaussian kernel)



# Approche:

## 2. génération stochastique d'hypothèses

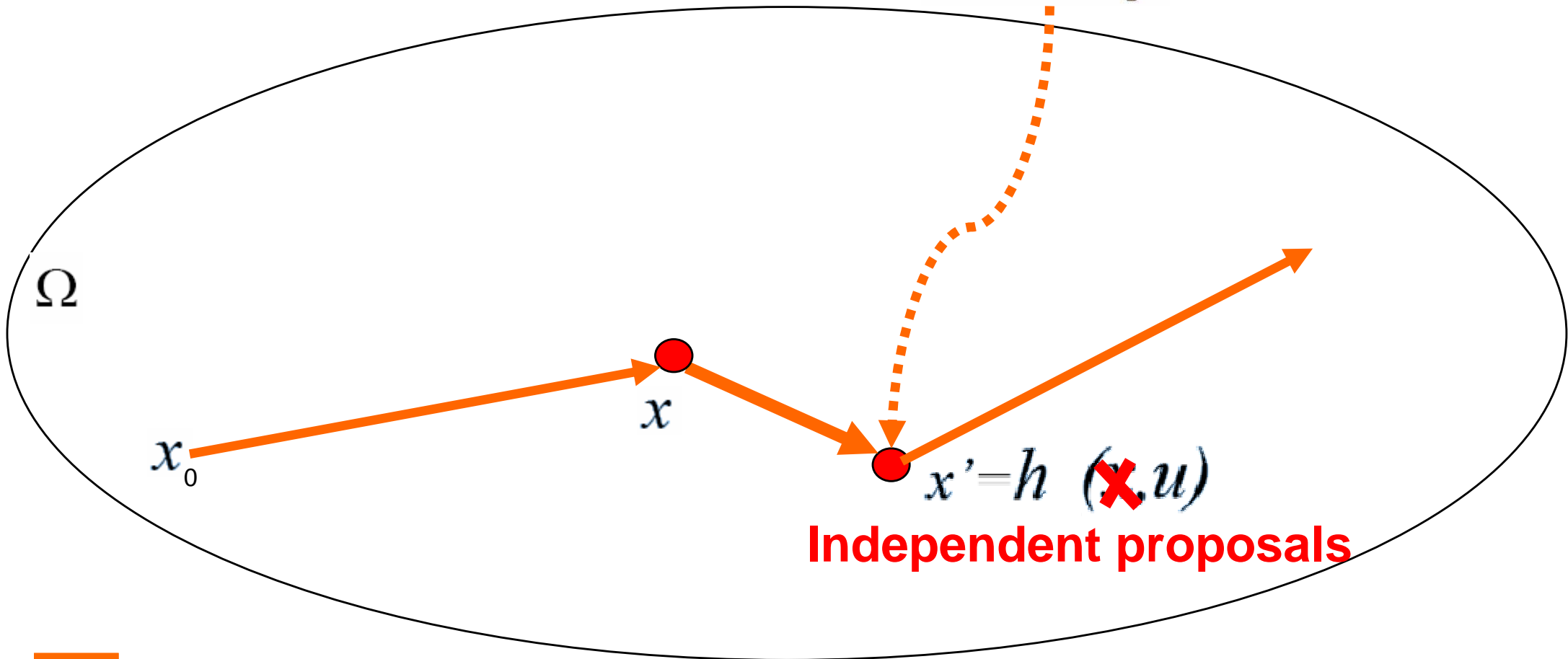
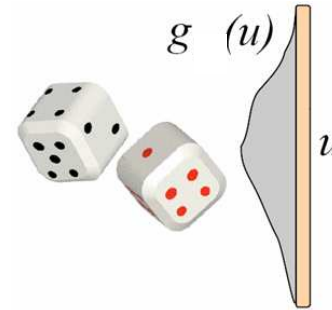
classic MCMC move:  
dynamics-based move



# Approche:

## 2. génération stochastique d'hypothèses

here MCMC move  
to maximise space coverage

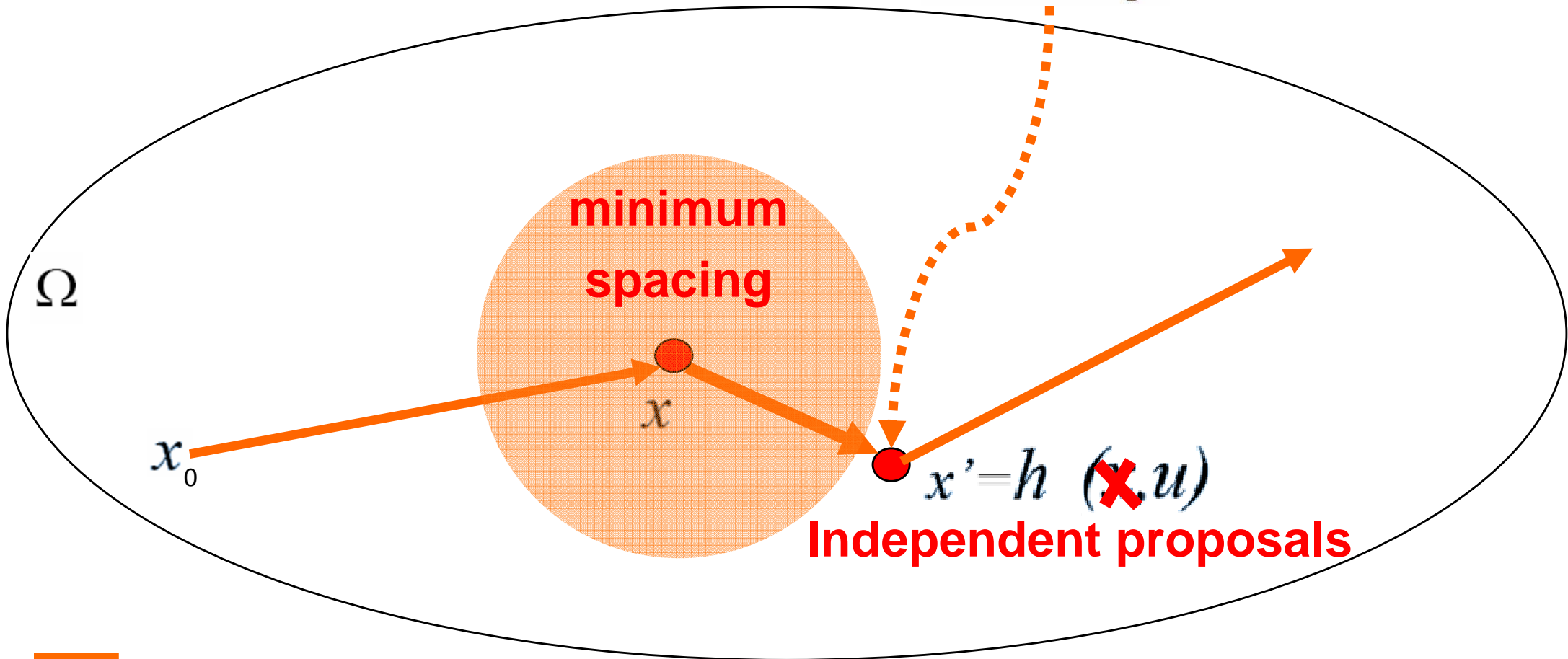
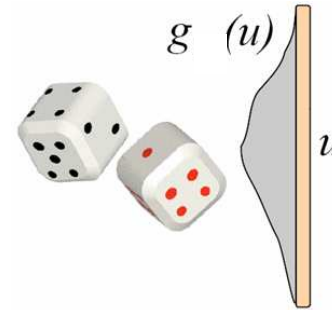




# Approche:

## 2. génération stochastique d'hypothèses

here MCMC move  
to maximise space coverage

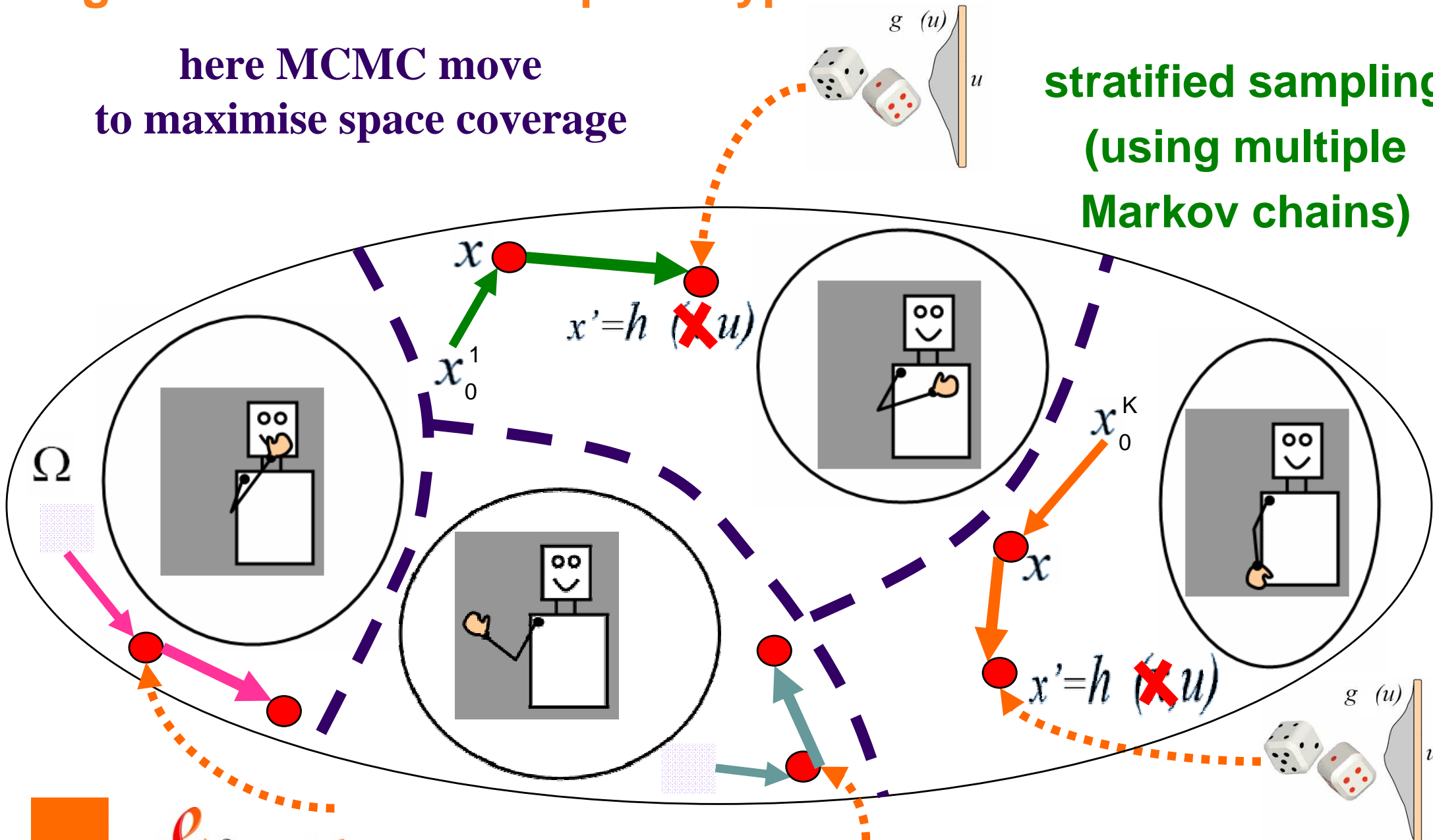


# Approche:

## 2. génération stochastique d'hypothèses

here MCMC move  
to maximise space coverage

stratified sampling  
(using multiple  
Markov chains)



# Approche:

## 2. génération stochastique d'hypothèses

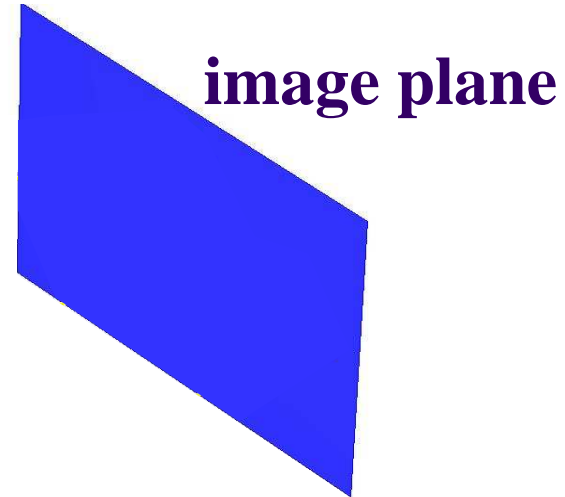


### data driven

- modèle graphique sur les points articulaires 2D
- multi-cues  
(background subtraction, contours, image differencing, skin tone)
- non cinématique
- correspondant aux composantes du bras les plus faciles à détecter  
(nb degrés de liberté, reliability)

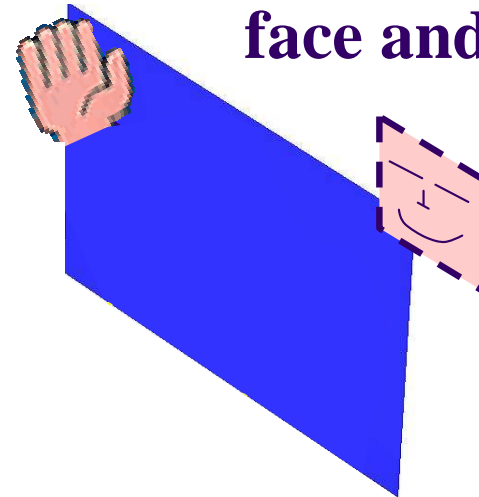
# Approche:

## 2. génération stochastique d'hypothèses



# Approche:

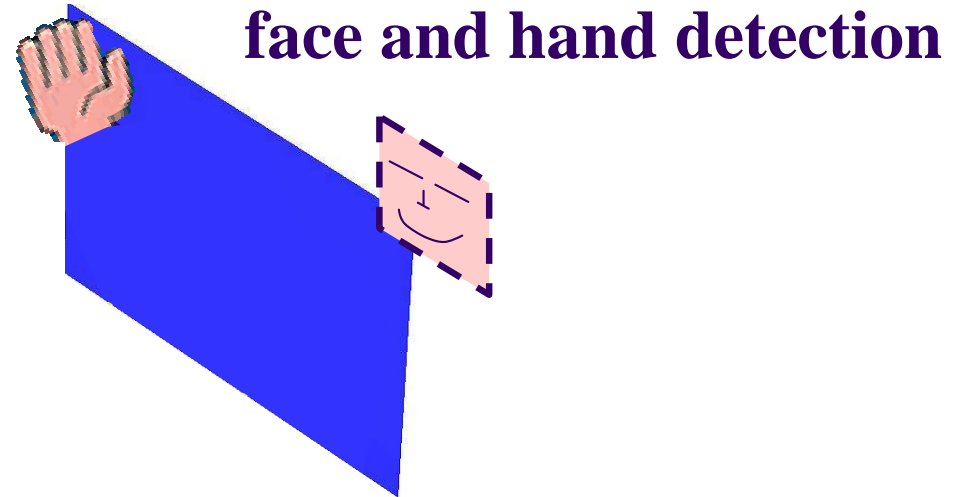
## 2. génération stochastique d'hypothèses



face and hand detection

# Approche:

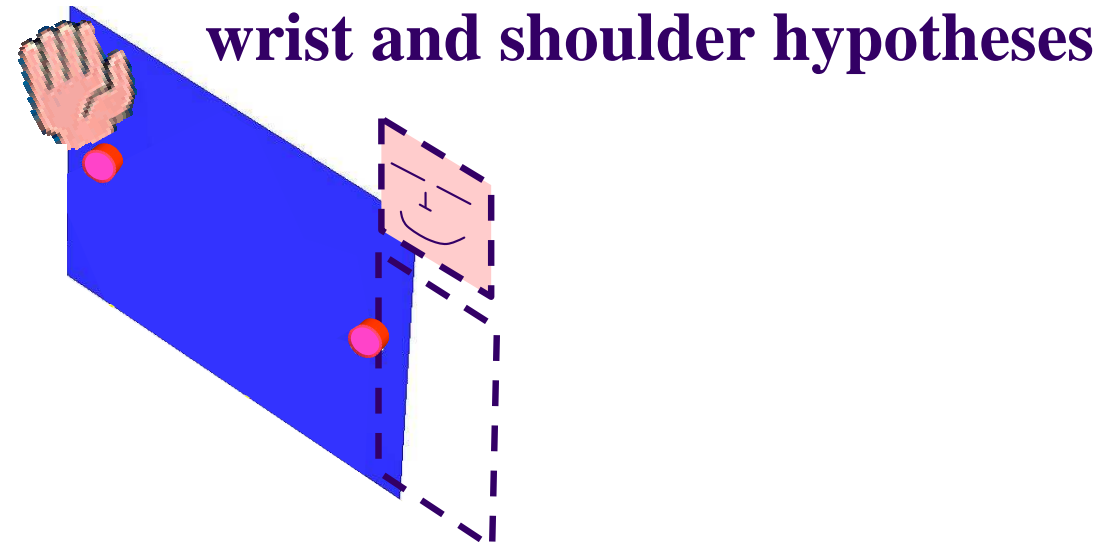
## 2. génération stochastique d'hypothèses



face width and height  $\Rightarrow$  limb length  $L \pm \delta L, L_{\max} = L + \delta L$

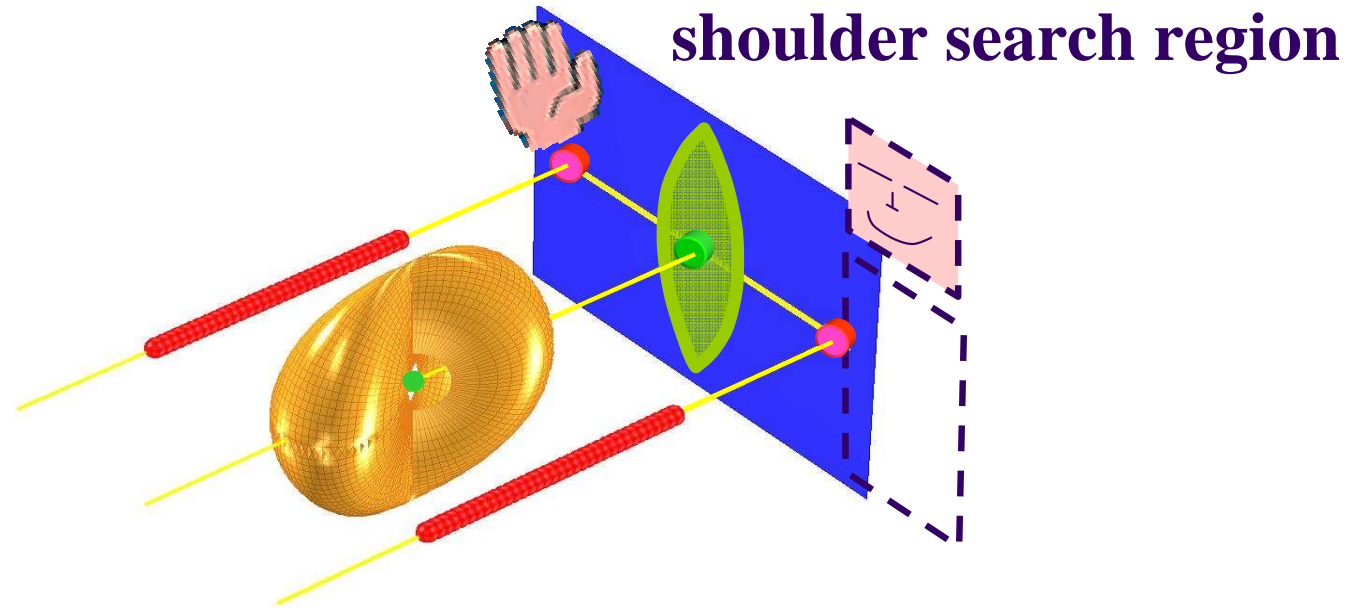
# Approche:

## 2. génération stochastique d'hypothèses



# Approche:

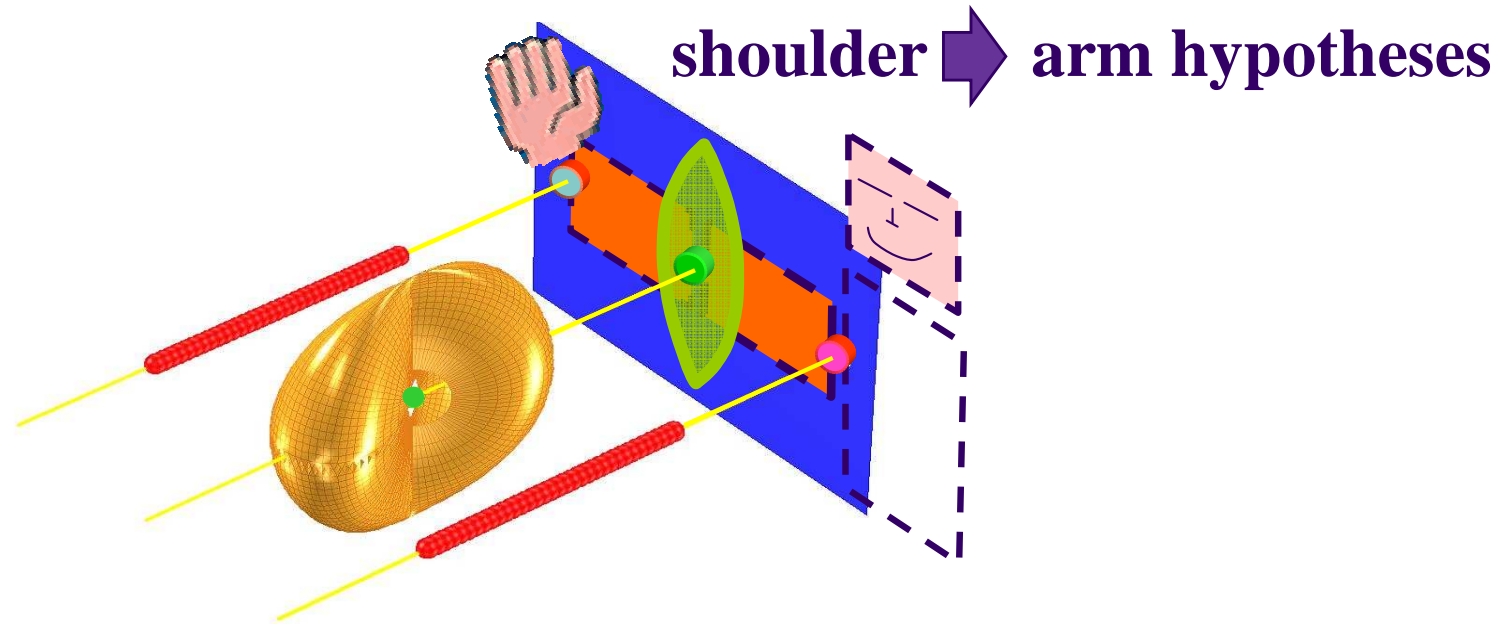
## 2. génération stochastique d'hypothèses





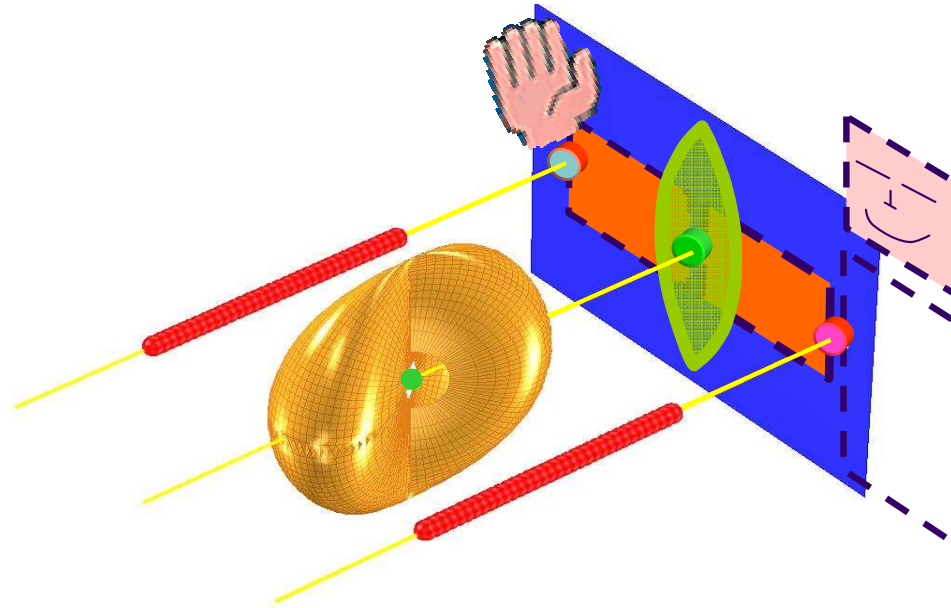
# Approche:

## 2. génération stochastique d'hypothèses



# Approche:

## 2. génération stochastique d'hypothèses



single image arm hypotheses  
with good  
image support

# Approche:

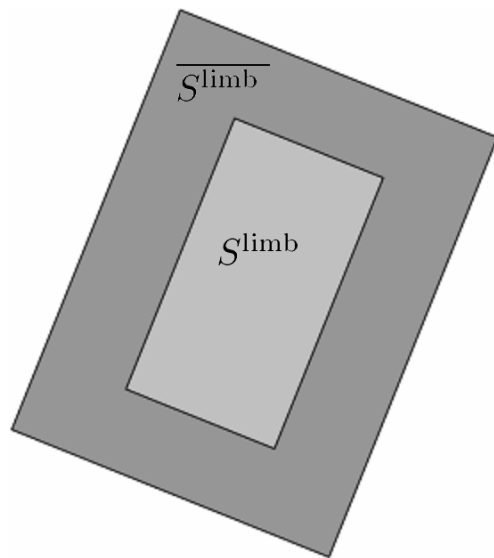
## 2. génération stochastique d'hypothèses



### → hypotheses with good image-based likelihood

- combining edge and background subtraction information
- power weighted by capacity of clue to discriminate between limb (or limb contour) and background (or rest of the image)

$$p(I_t | X'_{m,t}) = p_{\text{bckg}}^{\text{upper arm}} p_{\text{edg}}^{\text{upper arm}} p_{\text{bckg}}^{\text{lower arm}} p_{\text{edg}}^{\text{lower arm}}$$



$$p_{\text{bckg}}^{\text{limb}} = (S_{\text{bckg}}^{\text{limb}})^{\gamma_{\text{bckg}}^{\text{limb}}}$$

$$S_{\text{bckg}}^{\text{limb}} \propto \sum_{\text{all pixels } (x,y) \text{ on limb}} p_{\text{bckg}}(x,y)$$

$$\overline{S}_{\text{bckg}}^{\text{limb}} \propto \sum_{\text{all pixels } (x,y) \text{ not on limb}} p_{\text{bckg}}(x,y)$$

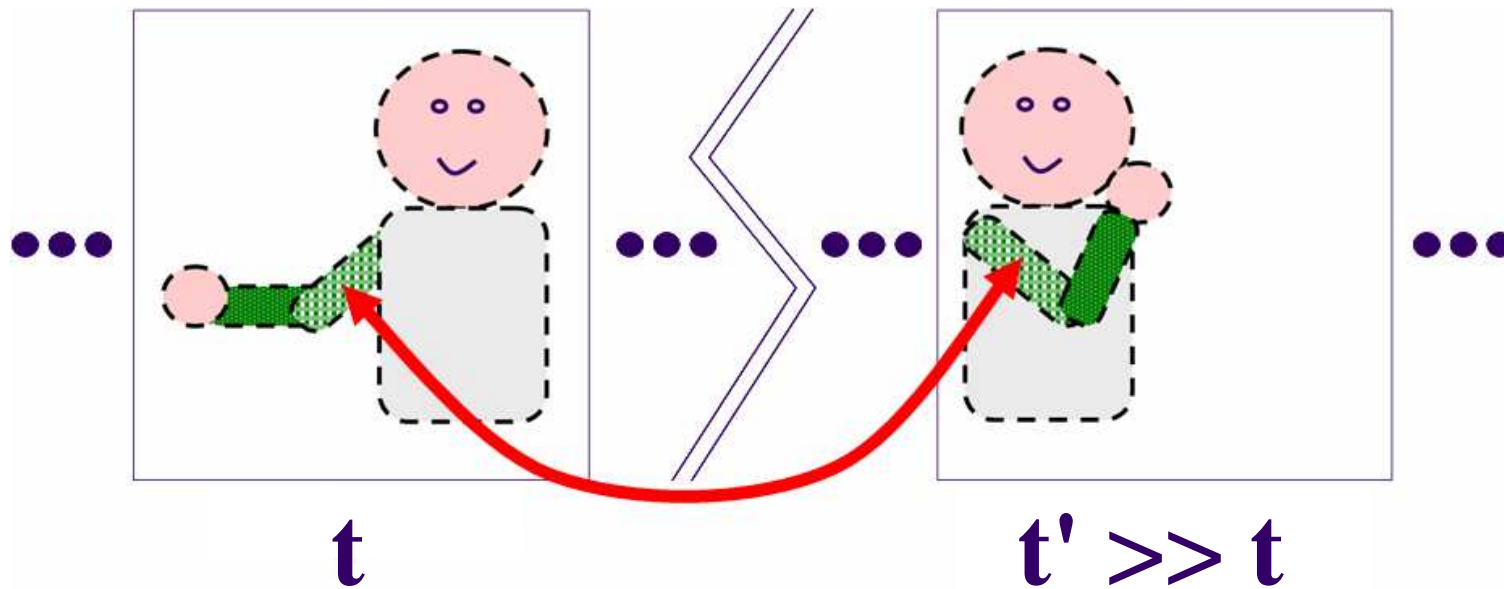
$$\gamma_{\text{bckg}}^{\text{limb}} \propto \left[ \frac{\overline{S}_{\text{bckg}}^{\text{limb}} + \epsilon}{S_{\text{bckg}}^{\text{limb}} + 1} \right]$$

# Approche:



## 3. paires d'hypothèses à grands intervalles de temps

→ continuité de l'apparence (couleur, texture) sur  $(t'-t) \gg 1$



→ mesure de similarité de l'apparence  
(combined power-weighted couleur & gradient clues)

$$\psi(X'_{m,t}, X'_{m',t'}, I_t, I_{t'})$$

# Approche:



## 3. paires d'hypothèses à grands intervalles de temps

- continuité de l'apparence (couleur, texture) sur  $(t'-t) \gg 1$
- $(t'-t)$  de 100 ms à plusieurs secondes



# Approche:



## 3. paires d'hypothèses à grands intervalles de temps

→ continuité de l'apparence (couleur, texture) sur  $(t'-t) \gg 1$

- prunage des faux positifs dûs à la détection sur 1 seule image
- possibilité de leurres non nulle





# Approche:

## 4. rejet d'hypothèses par mécanisme MCMC

acceptance ratio

posterior proba at current iteration (current sample)

proposal density

$$\alpha = \frac{P(X'_t | Z^t) Q(X_t^{(s)}; X'_t)}{P(X_t^{(s)} | Z^t) Q(X'_t; X_t^{(s)})}$$

posterior proba at previous iteration



# Approche:

## 4. rejet d'hypothèses par mécanisme MCMC

$$\alpha = \frac{\overbrace{p(I_t | X'_{m,t}) * p(I_{t'} | X'_{m',t'}) * \psi(X'_{m,t}, X'_{m',t'}, I_t, I_{t'})}^{\text{posterior proba}} Q(X_t^{(s)}; X'_t)}{P(X_t^{(s)} | Z^t) Q(X'_t; X_t^{(s)})} \quad \text{proposal density}$$





# Approche:

## 4. rejet d'hypothèses par mécanisme MCMC

posterior proba

proposal density

$\alpha =$

$$\frac{\cancel{p(I_t | \Lambda_{m,t}^{(s)})} \cdot \cancel{P(I_{t'} | X'_{m,t,t'})} * \psi(X'_{m,t}, X'_{m',t'}, I_t, I_{t'}) \cdot Q(X_t^{(s)}; X'_t)}{P(X_t^{(s)} | Z^t) \cdot \cancel{Q(X'_t; \Lambda_t^{(s)})}}$$



# Approche:

## 4. rejet d'hypothèses par mécanisme MCMC

$$\alpha = \frac{\psi(X'_{m,t}, X'_{m',t'}, I_t, I_{t'})}{\psi_{\text{prev}}}$$

# Résultats

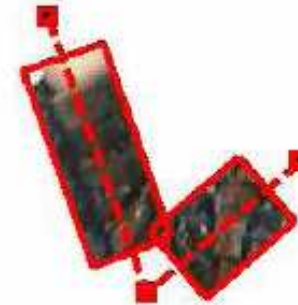
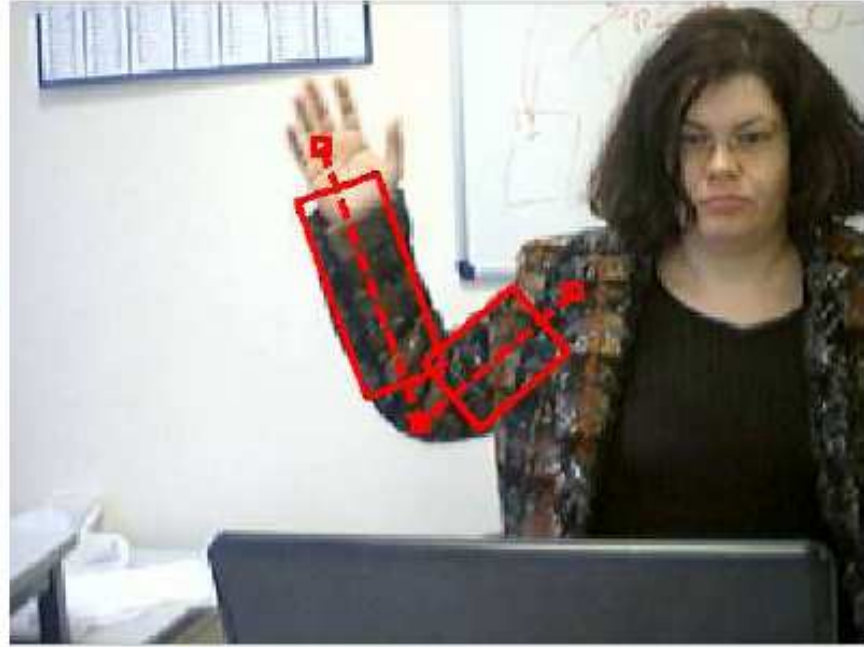
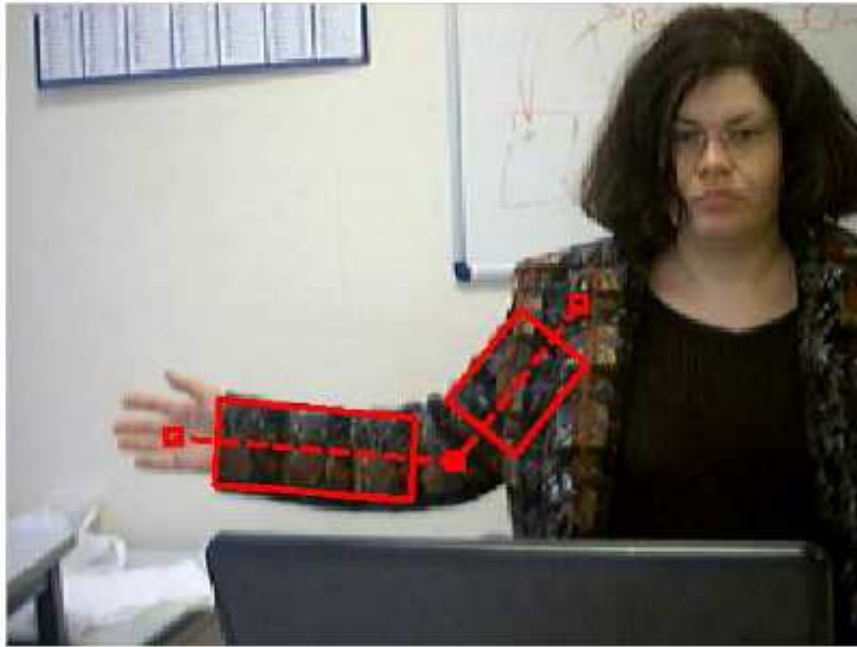


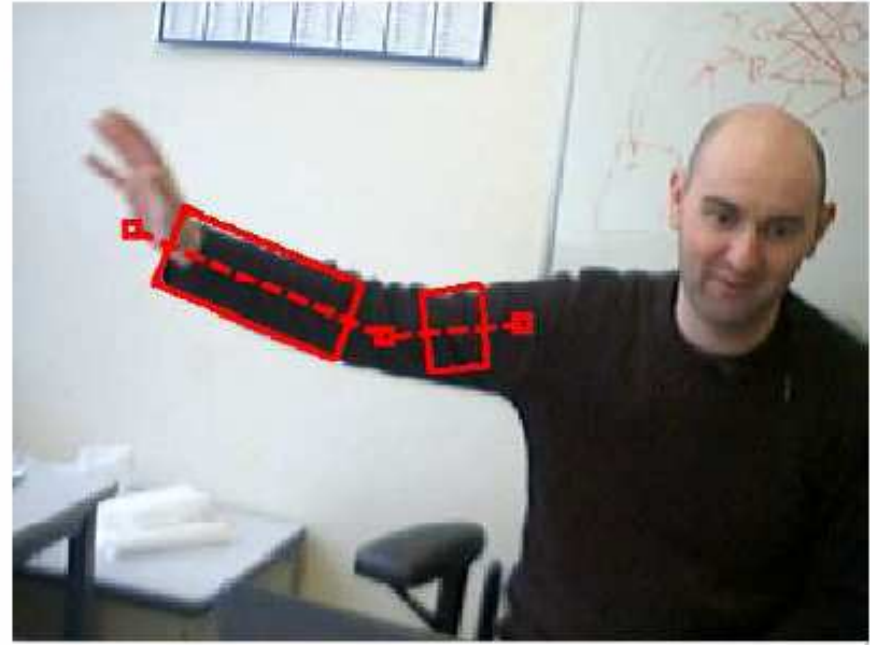
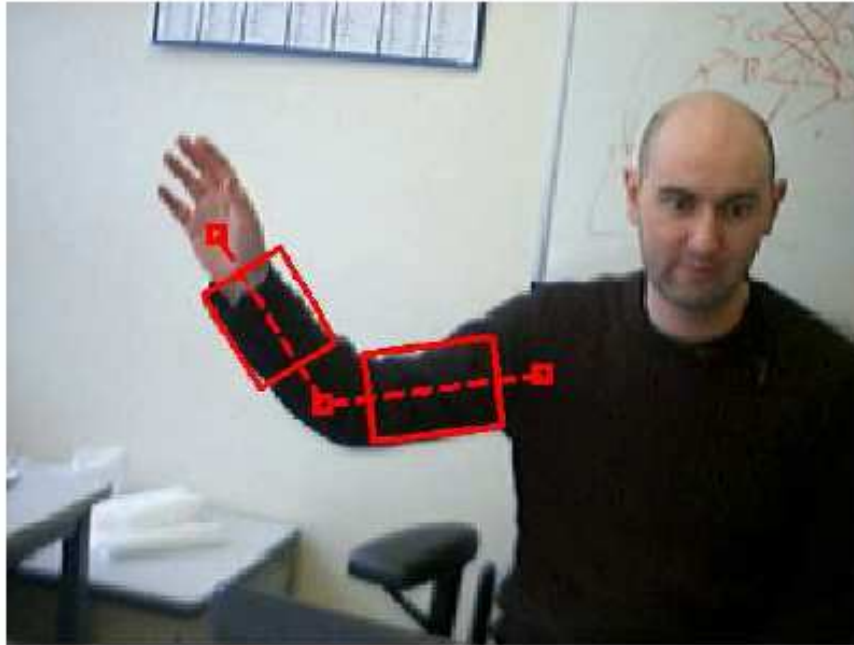
→ (t'-t) de 100 ms à plusieurs secondes

→ variété

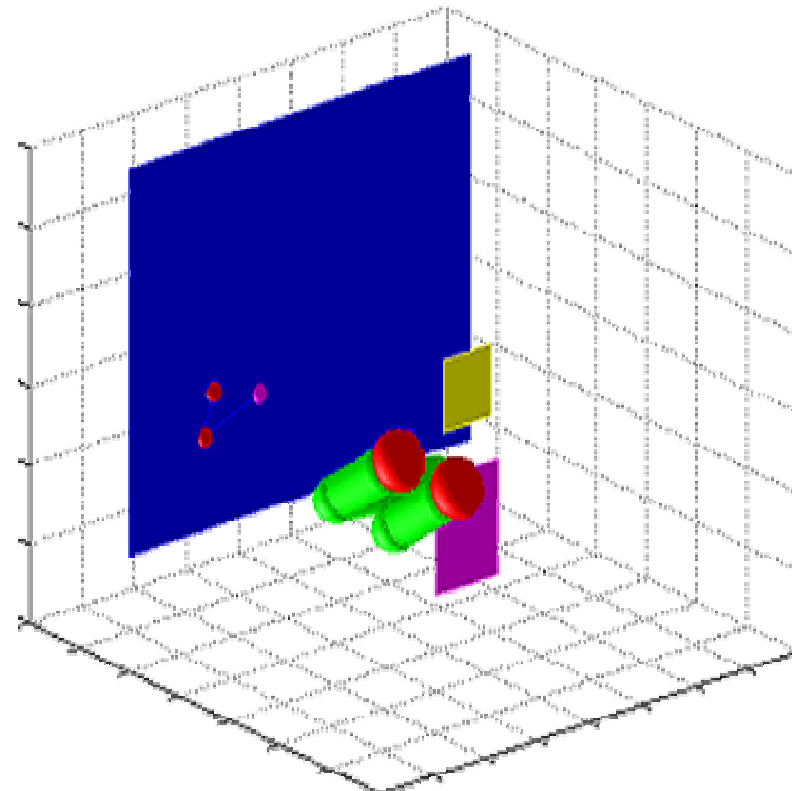
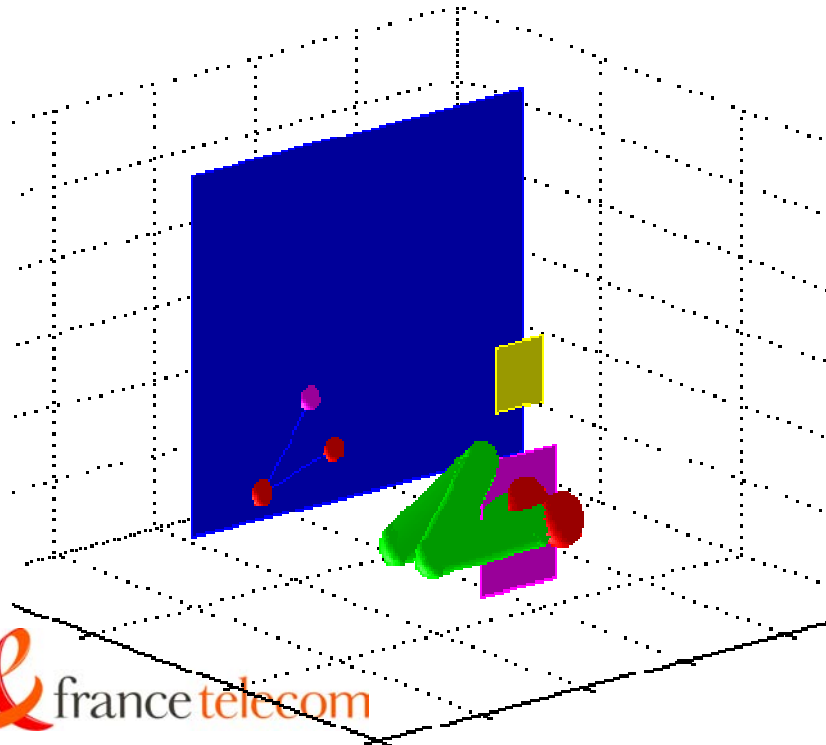
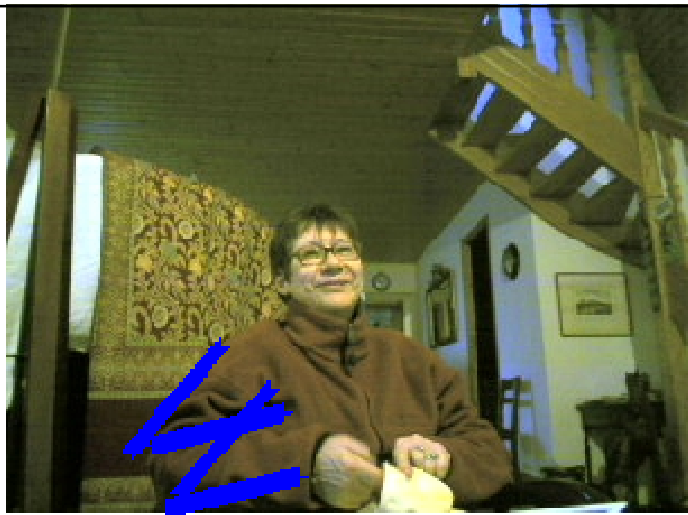
- personnes
- taille, mensurations
- habits (collants ou non)
- mouvements

(rapides mouvements entre des périodes statiques)









# Conclusion



- paires d'images non consécutives
- contrainte de cohérence de l'apparence (couleur, texture) sur de longs intervalles de temps
- opportuniste (configurations favorables trouvées d'abord)
- apprentissage en ligne de l'apparence
- pas besoin d'a priori ou de base d'apprentissage exhaustive sur les poses et mouvements possibles





**MERCI !**

