

Conférence « Image et 3D »

Module Image, École doctorale I2S, Montpellier

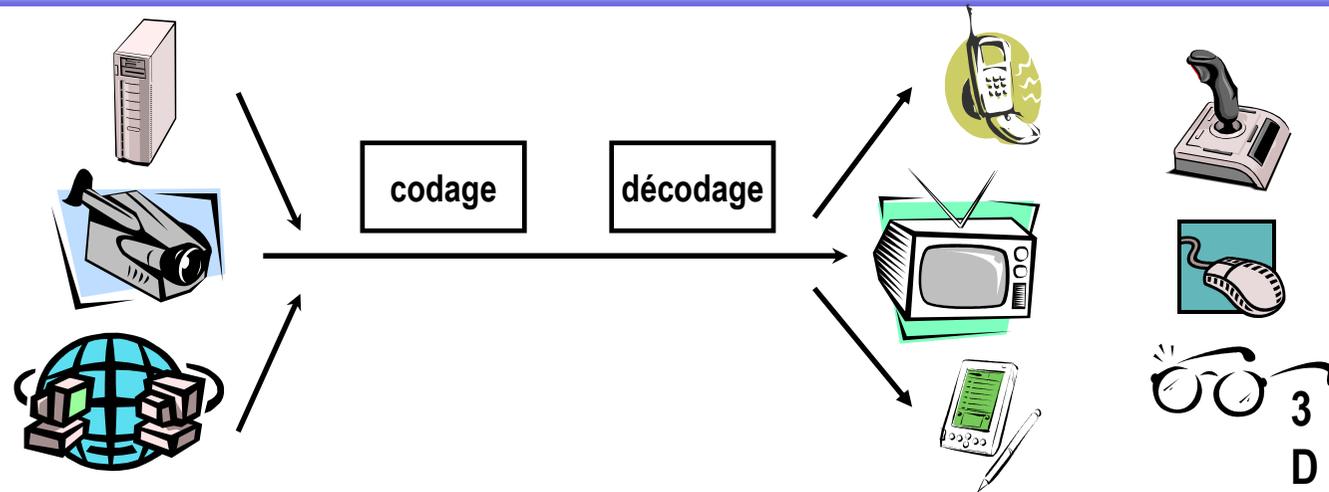
Luce Morin

MCF Univ. Rennes 1, laboratoire IRISA – projet Temics

(transparents aussi de Stéphane Pateux, Vivien Chappelier,
Marc Chaumont, Andrew Zisserman, Eric Marchand)



11 mai 2006



Evolution du monde de la vidéo numérique :

- Nouvelles sources/terminaux : dv, ordinateurs et terminaux plus puissants, terminaux mobiles, écrans 3D,...
- Réseaux hétérogènes : internet, téléphone RTC, GSM, ADSL, WIFI...
- Nouvelles fonctionnalités : interactivité, jeux 3D, TV3D, DVD, , TV sur internet, visiophonie mobile, visioconférence 3D)...

=> Nouvelles orientations du codage vidéo

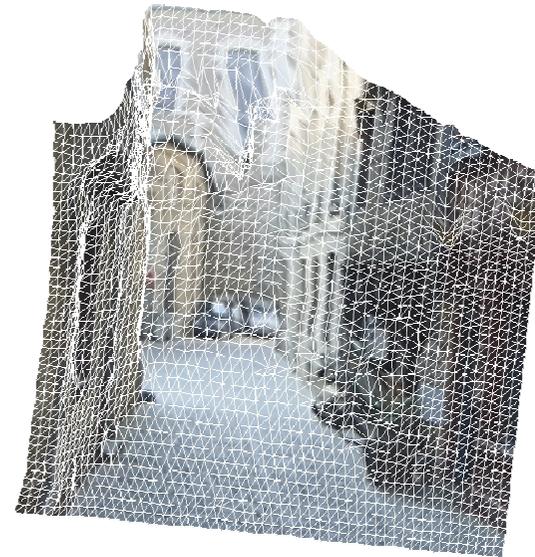
- Compression ET fonctionnalités additionnelles
- Codage adapté contenus
- Scalabilité

Motivations pour une représentation 3D de vidéos

- Acquisition vidéo
- Restitution par rendu 3D



Video sequence



3D reconstruction



Acquérir une vidéo et ...



... l'explorer en 3D,



... rajouter des objets,



... changer l'éclairage,

Fonctionnalités 3D

- Navigation virtuelle (évtnt interactive)
- Manipulation vidéo (ajout d'objets, sources de lumières,...)
- Rendu stéréoscopique

Applications

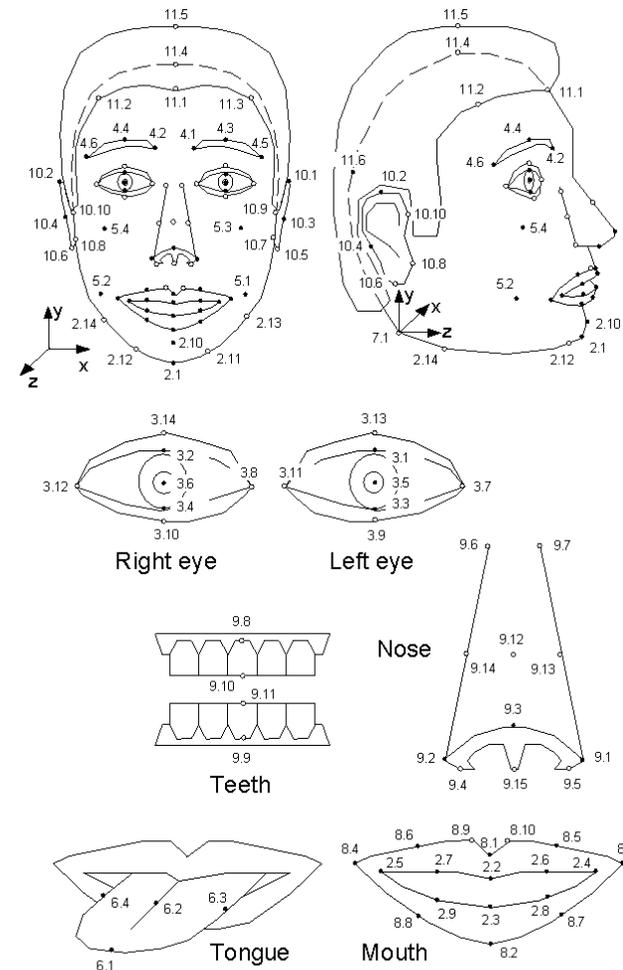
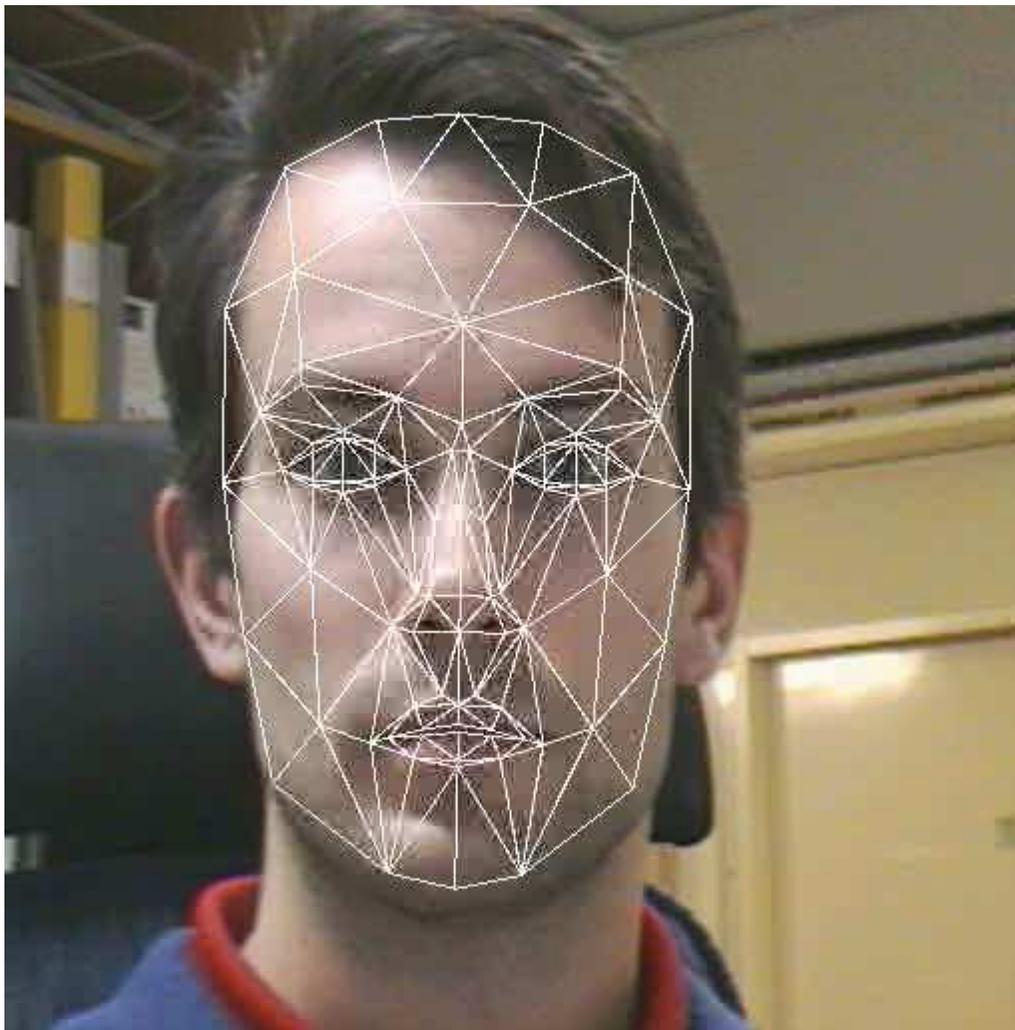
- Réalité virtuelle : aide au géo positionnement, visites virtuelles
- Réalité augmentée : étude d'impact, visioconférence
- Compression vidéo : très bas débits pour les applications distantes



... la regarder en stéréo,

Codage vidéo basé modèles 3D

Codage basé modèles 3D classique (*Mpeg4-faps, SNHC...*)

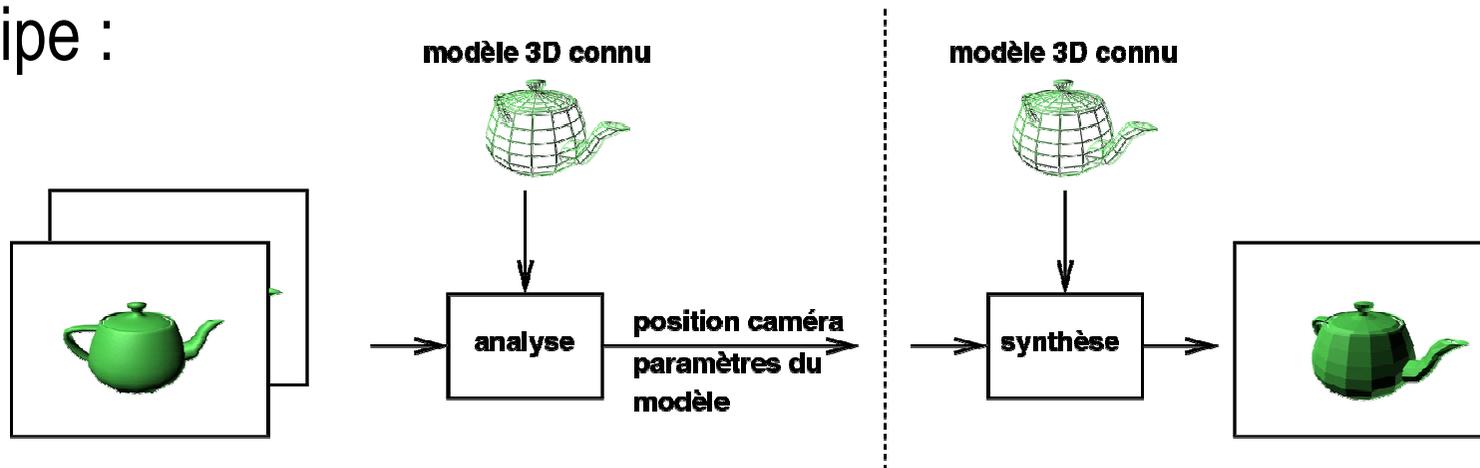


• Feature points affected by FAPs
 ○ Other feature points

Introduction

Codage vidéo basé modèles 3D

Principe :

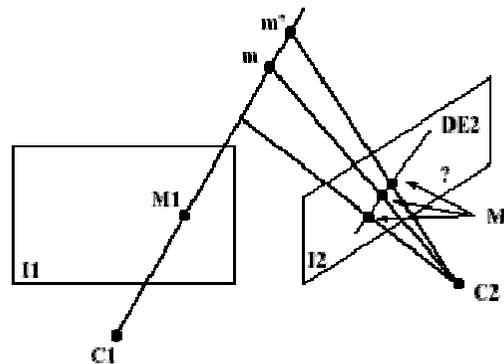


Avantages / inconvénients :

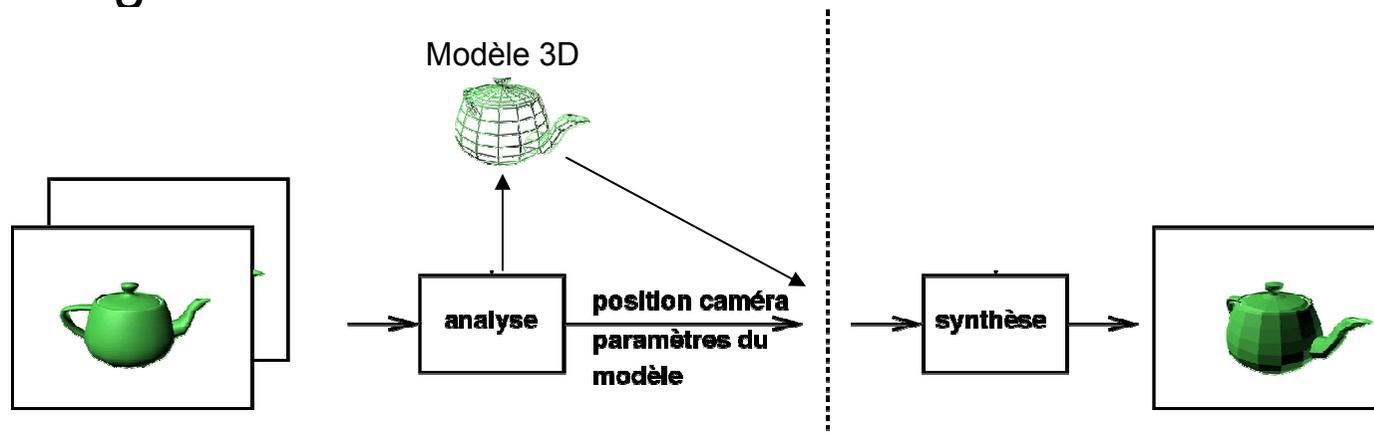
- + bonne compacité : données à transmettre =
 - paramètres du modèle 3D (position, forme, mouvement, texture)
 - le point de vue caméra pour chaque image
 - + manipulations 3D, interactivité
 - modèles 3D connus a priori
- ⇒ est-ce possible sans modèle 3D a priori ?

Vision par ordinateur :

reconstruction de modèles 3D inconnus à partir d'images : stéréovision



⇒ Codage basé reconstruction 3D



Contraintes et hypothèses de travail

Contraintes de la vision par ordinateur :

- mouvement caméra non dégénéré
- scène statique ou segmentée
- instabilité numérique (problème inverse)

Contraintes du codage :

- qualité de reconstruction des images
- pas d'a priori sur l'acquisition (caméra monoculaire non étalonnée)
- pas d'a priori sur le contenu
- "tout automatique"

Introduction

I - Reconstruction 3D

II - Compression Vidéo

III - Modélisation 3D de séquences vidéo

Conclusion

Vision par ordinateur

- Réaliser automatiquement les tâches de vision à partir d'images:
 - Percevoir le relief
 - Déterminer la position des objets observés / du capteur
 - Détecter et analyser le mouvement des objets
 - Suivre des yeux les objets
 - Identifier les objets
- Applications
 - Neurobiologie : étude du SVH
 - Imagerie médicale,
 - Robotique, Métrologie
 - Imagerie satellitaire, Cartographie
 - Télé-surveillance, militaire
 - Réalité virtuelle et augmentée, Jeux, IHM multimodales

passage 2D → 3D : un problème difficile

retrouver une dimension « perdue » :

⇒ classe des problèmes inverses

⇒ problème mal posé

⇒ plusieurs solutions compatibles avec les données

⇒ sensibilité aux erreurs, instabilité numérique

...et pourtant le système visuel humain réussit très bien cette tâche, comment?

1 - Reconstruction 3D par le SVH

- a) Illusions visuelles
- b) Principe de la vision du relief
- c) Principe de la visualisation stéréo

2 - Reconstruction 3D par vision par ordinateur

- a) Reconstruction calibrée
- b) Reconstruction non calibrée
- c) Conclusion

1 - Reconstruction 3D par le SVH

- a) Illusions visuelles
- b) Principe de la vision du relief
- c) Principe de la visualisation stéréo

2 - Reconstruction 3D par vision par ordinateur

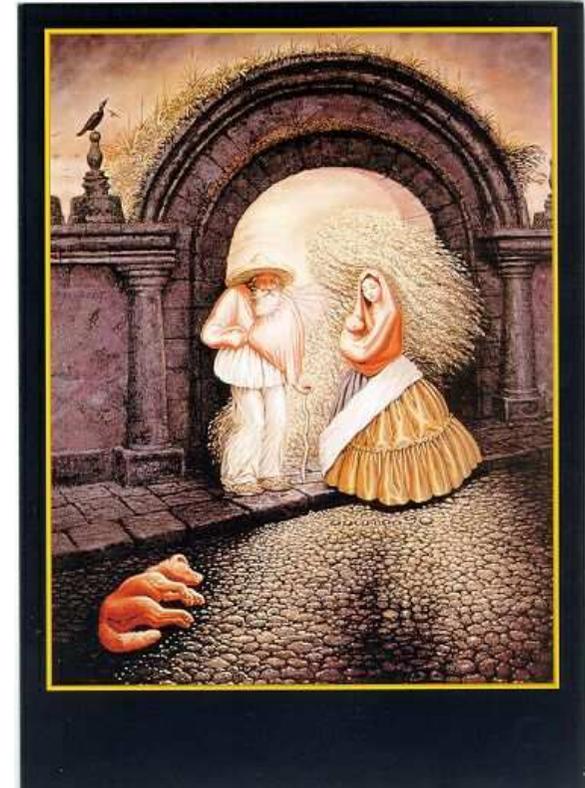
- a) Reconstruction calibrée
- b) Reconstruction non calibrée
- c) Conclusion

1 - Reconstruction 3D par le SVH

a) Illusions visuelles : un fort a priori 3D



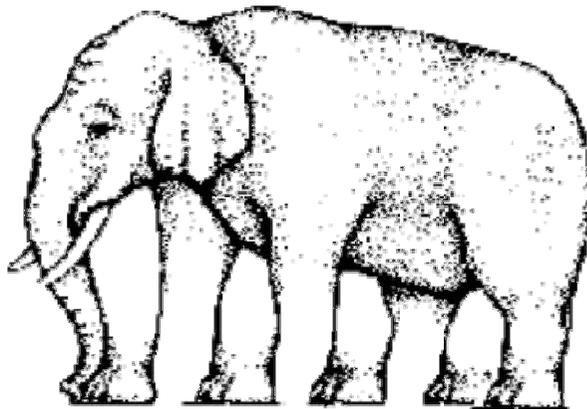
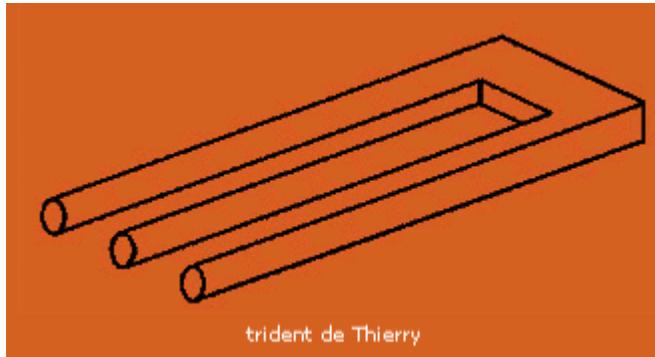
Do you see the face? Or an Eskimo?



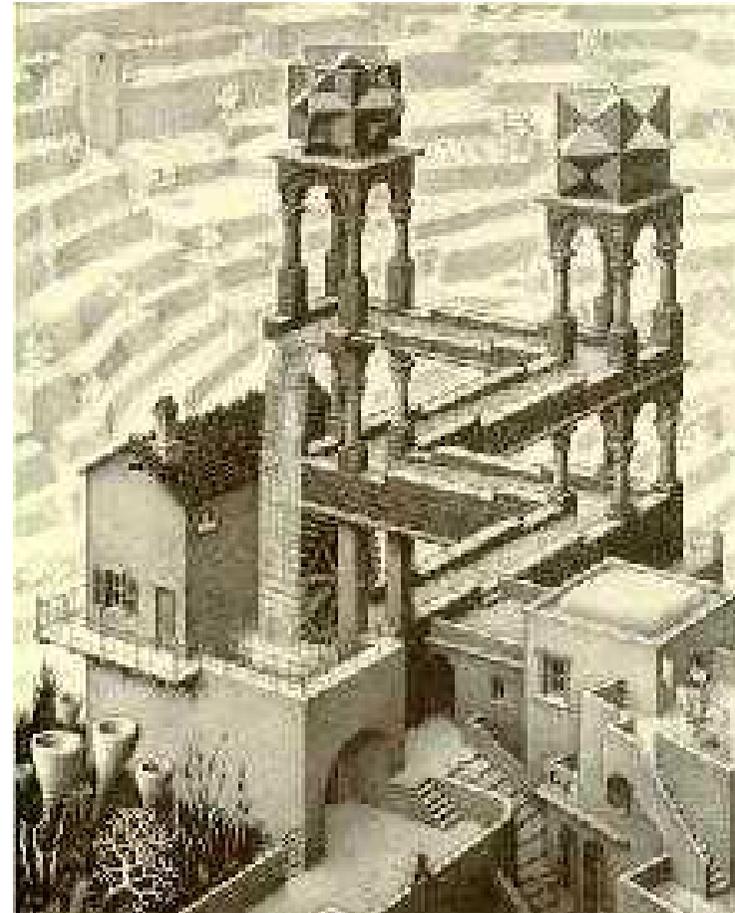
images ambiguës

1 - Reconstruction 3D par le SVH

a) Illusions visuelles : un fort a priori 3D



How many legs does this elephant have?



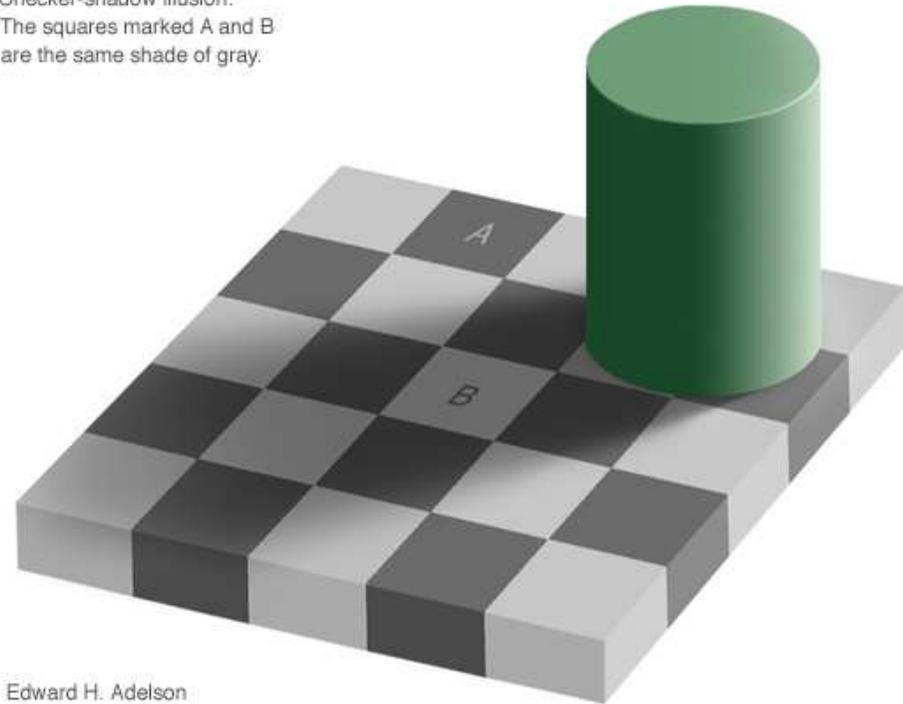
images « impossibles »

1 - Reconstruction 3D par le SVH

a) Illusions visuelles : un fort a priori 3D



Checker-shadow illusion:
The squares marked A and B
are the same shade of gray.



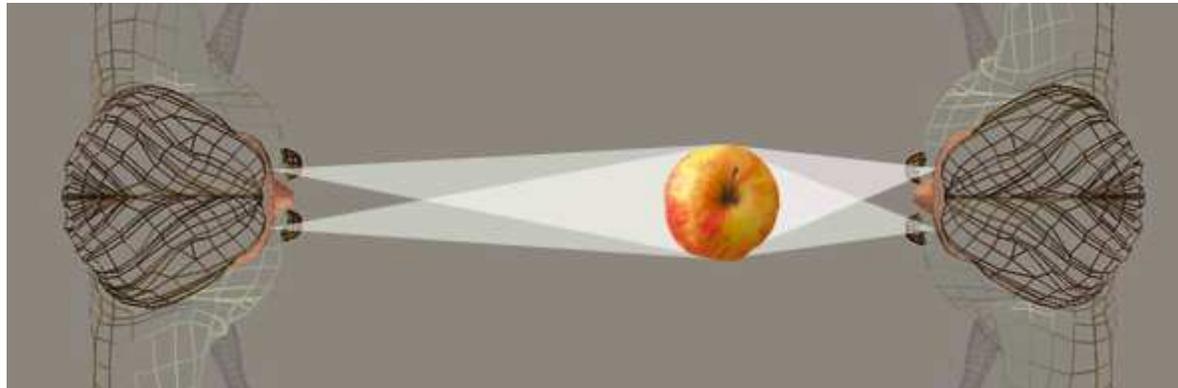
Edward H. Adelson

image = scène 3D éclairée par le haut

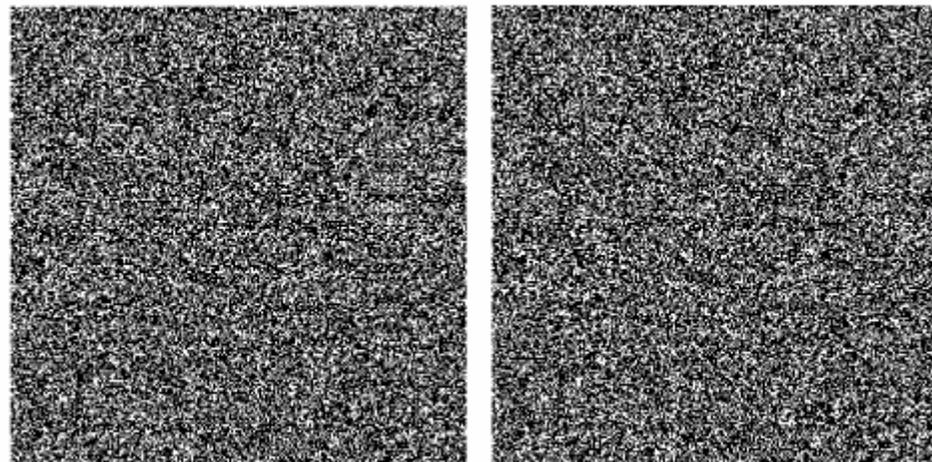
1 - Reconstruction 3D par le SVH

b) Principe de la vision du relief

- 2 yeux, avec des points de vues différents
=> la différence d'angle (**parallaxe**) indique la profondeur.



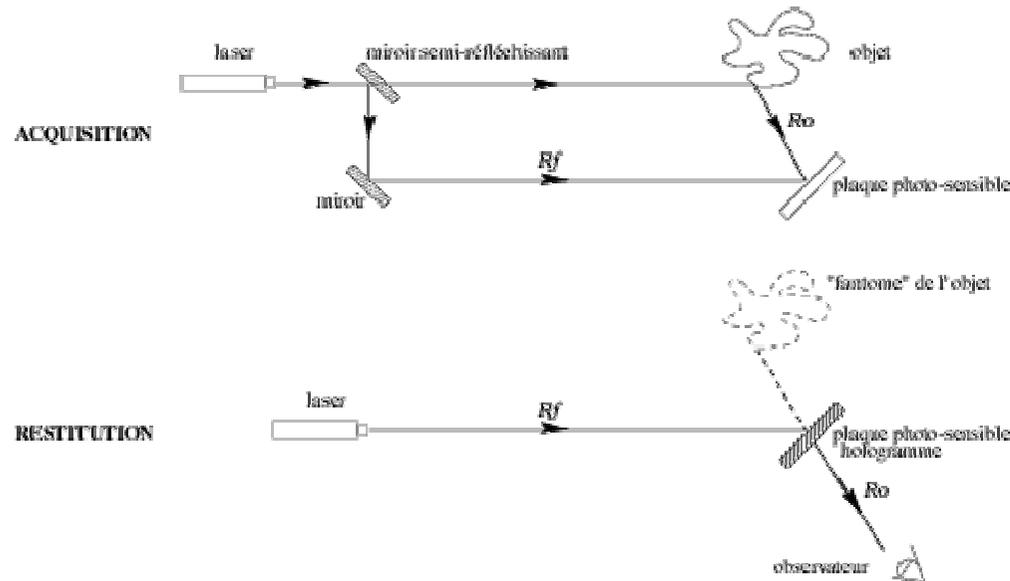
- un procédé « bas-niveau » : **stéréogrammes aléatoires**



1 - Reconstruction 3D par le SVH

c) Principe de la visualisation stéréo

- Reconstitution de l'onde lumineuse : hologrammes



- Envoi d'images décalées sur chaque œil : films en 3D
 - lunettes et écrans polarisés
 - lunettes à obturation synchronisées
 - écrans auto-stéréo (sans lunettes) : réseaux lenticulaires

2 - Reconstruction 3D par vision par ordinateur

1 - Reconstruction 3D par le SVH

- a) Illusions visuelles
- b) Principe de la vision du relief
- c) Principe de la visualisation stéréo

2 - Reconstruction 3D par vision par ordinateur

- a) Reconstruction calibrée
 - Cadre : stéréovision binoculaire
 - Modèle de caméra
 - Étalonnage
 - Géométrie épipolaire
 - Mise en correspondance
 - Triangulation
 - Exemples
- b) Reconstruction non calibrée
- c) Conclusion

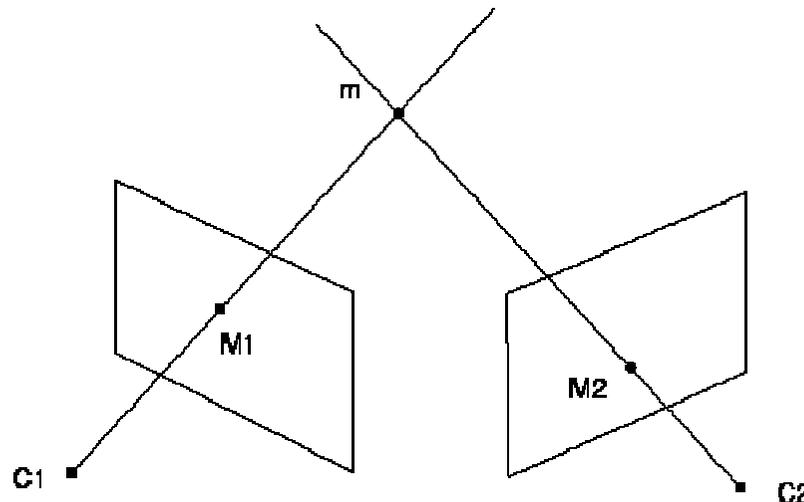
but : à partir de 2 images d'une même scène, restituer le relief

hyp : on connaît les positions des caméras (caméras étalonnées)

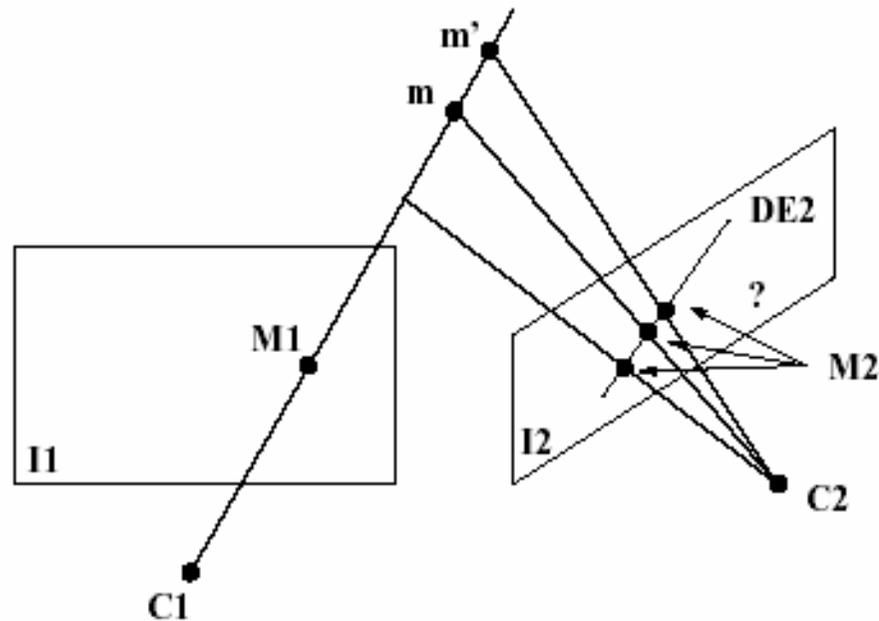


Principe

- Hypothèse : projection perspective
- Soient $M1$, $M2$ projetés du même point 3D m
- Reconstruction par triangulation : $m = (C1M1) \cap (C2M2)$

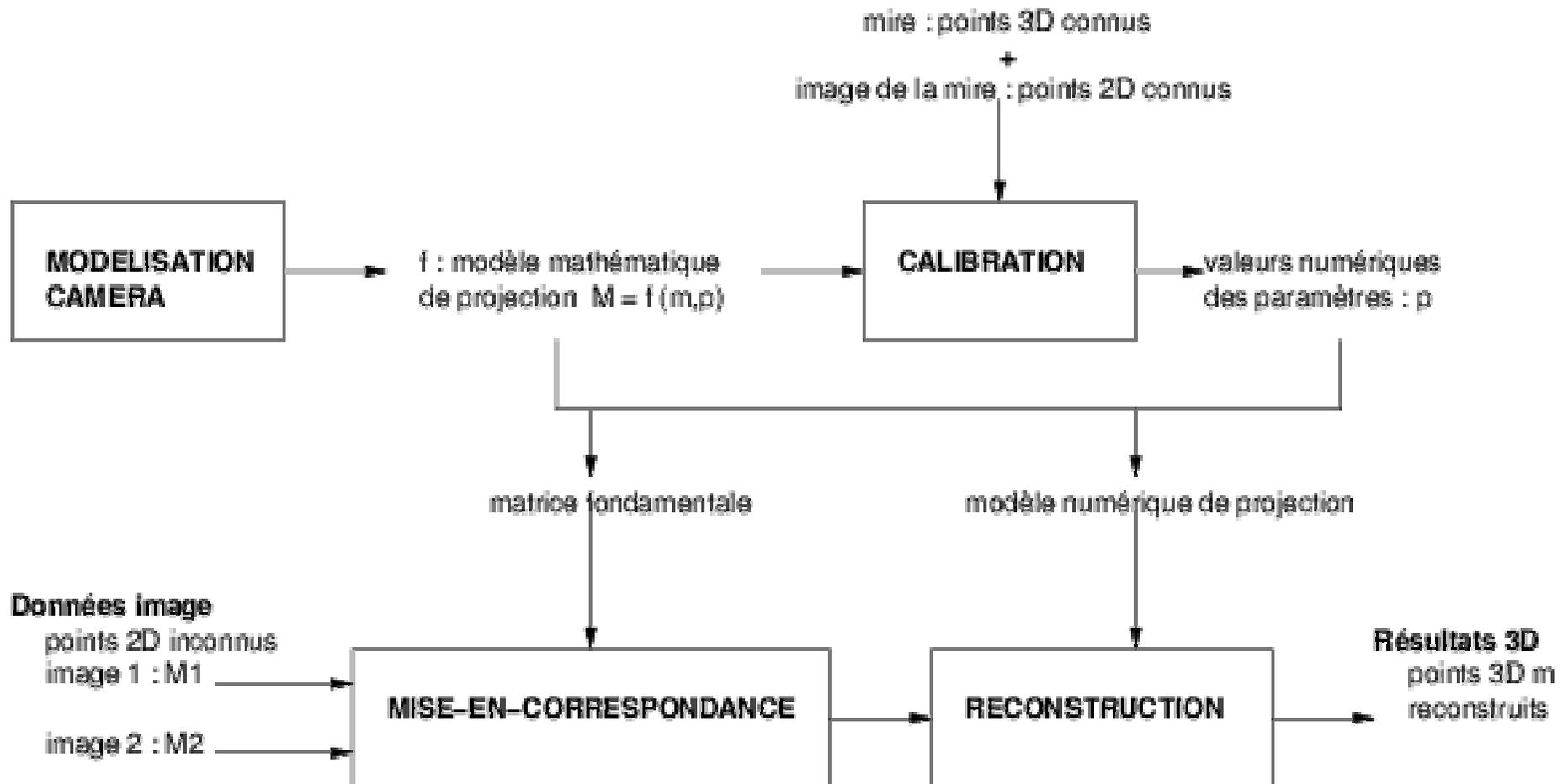


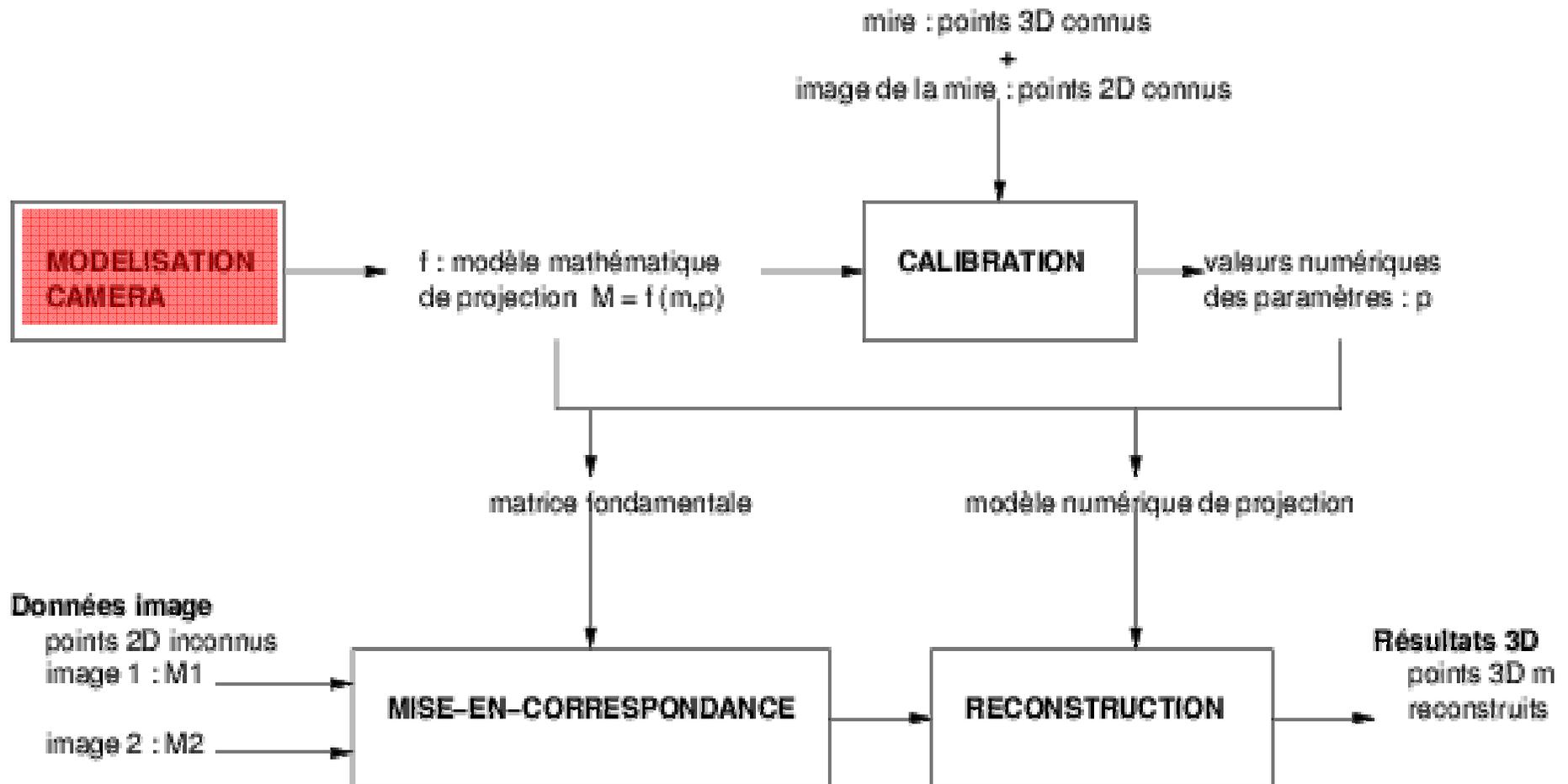
Problèmes à résoudre



- Associer M1 et M2 → mise-en-correspondance
- Position relative de C1 et C2 → étalonnage
- Intersecter des droites 3D → reconstruction par triangulation

Étapes





équations de projection dans le repère caméra :

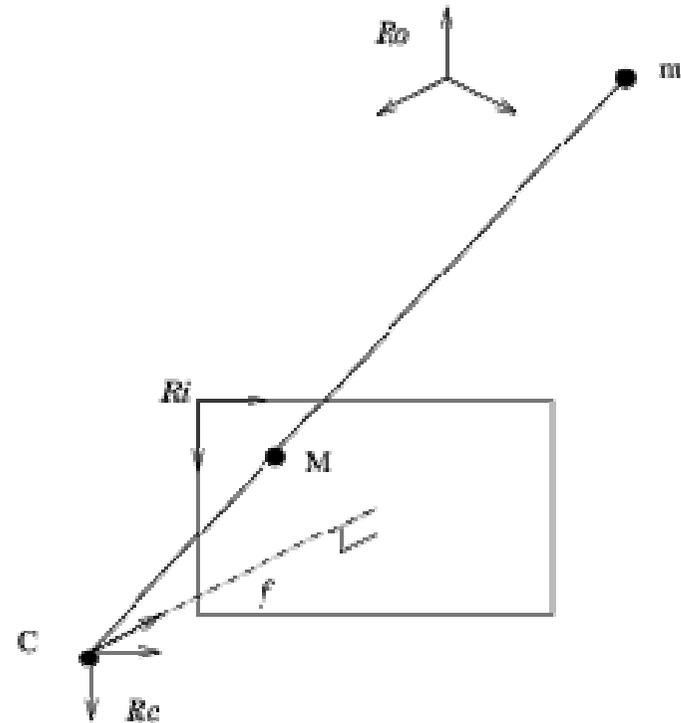
- en coordonnées cartésiennes:

$$X = \frac{x}{z}$$

$$Y = \frac{y}{z}$$

- en coordonnées homogènes :

$$\begin{bmatrix} X \\ Y \\ W \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix}$$



équations de projection du repère objet vers le repère image

- paramètres intrinsèques : (f, X_c, Y_c, l_x, l_y)
- paramètres extrinsèques :
 position de la caméra : vecteur de translation $t = (X_o, Y_o, Z_o)$
 orientation de la caméra : matrice de rotation $R = (r_{ij})_{i,j}$

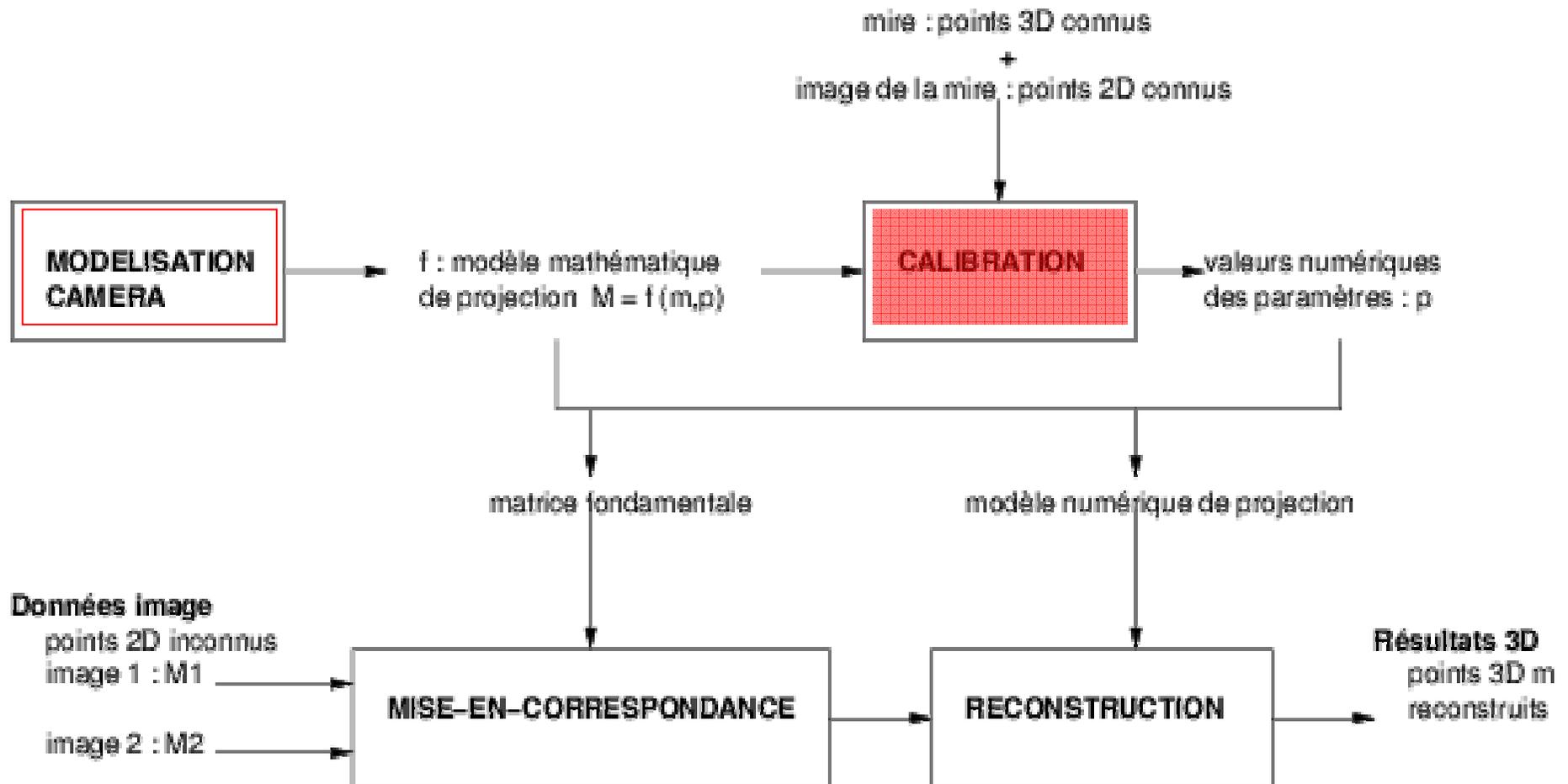
modèle « linéaire » (linéaire en coordonnées homogènes)

$$\begin{bmatrix} X_p \\ Y_p \\ W_p \end{bmatrix} = \begin{bmatrix} \frac{f}{l_x} & 0 & X_c \\ 0 & \frac{f}{l_y} & Y_c \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} r_{11} & r_{12} & r_{13} & X_o \\ r_{21} & r_{22} & r_{23} & Y_o \\ r_{31} & r_{32} & r_{33} & Z_o \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_o \\ y_o \\ z_o \\ w_o \end{bmatrix}$$

$$M = K [R|t] m$$

$$M = P m$$

Remarque : c'est l'équation de la ligne de vue associée au pixel M, exprimée dans le repère objet => ce qu'il nous faut pour intersecter les lignes de vue



But : estimation des paramètres caméra à partir d'une mire

$$\text{Équations de projection : (1) } \begin{cases} X_p z - X_c z - F_x x = 0 \\ Y_p z - Y_c z - F_y y = 0 \end{cases}$$

$$\text{avec } \begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} = {}^c T_o \begin{bmatrix} x_o \\ y_o \\ z_o \\ 1 \end{bmatrix} = \begin{bmatrix} r_{11} & r_{12} & r_{13} & X_o \\ r_{21} & r_{22} & r_{23} & Y_o \\ r_{31} & r_{32} & r_{33} & Z_o \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_o \\ y_o \\ z_o \\ 1 \end{bmatrix}$$

$$(1) \text{ équivaut à } A I = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ avec } \begin{cases} A \text{ fonction des données} \\ I \text{ fonction des inconnues} \end{cases}$$

$$A^t = \begin{bmatrix} X_p x_o & Y_p x_o \\ X_p y_o & Y_p y_o \\ X_p z_o & Y_p z_o \\ -x_o & 0 \\ -y_o & 0 \\ -z_o & 0 \\ 0 & -x_o \\ 0 & -y_o \\ 0 & -z_o \\ -1 & 0 \\ 0 & -1 \\ X_p & Y_p \end{bmatrix} \text{ et } I = \begin{bmatrix} r_{31} \\ r_{32} \\ r_{33} \\ r_{31} X_c + r_{11} F_x \\ r_{32} X_c + r_{12} F_x \\ r_{33} X_c + r_{13} F_x \\ r_{31} Y_c + r_{21} F_y \\ r_{32} Y_c + r_{22} F_y \\ r_{33} Y_c + r_{23} F_y \\ X_c Z_o + F_x X_o \\ Y_c Z_o + F_y Y_o \\ Z_o \end{bmatrix}$$

Chaque points de la mire donne 2 équations $A_i X = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$

Pour tous les points $i=1..n$

$$A = \underbrace{\begin{bmatrix} A_1 \\ A_2 \\ \dots \\ A_n \end{bmatrix}}_{12 \text{ colonnes}} \} 2n \text{ lignes}$$

Systeme à résoudre $A I = Null$

où $Null$ est un vecteur nul de taille $2n$

Solution 1

recherche de I comme vecteur du noyau de A (SVD)

Solution 2

résolution sous la contrainte $[r_{31}, r_{32}, r_{33}]$ unitaire
(méthode de Lagrange)

Détermination des paramètres intrinsèques et extrinsèques
à partir de I

$$X_c = I_1 I_4 + I_2 I_5 + I_3 I_6$$

$$Y_c = I_1 I_7 + I_2 I_8 + I_3 I_9$$

$$F_x = \sqrt{I_4^2 + I_5^2 + I_6^2 - X_c^2}$$

$$F_y = \sqrt{I_7^2 + I_8^2 + I_9^2 - Y_c^2}$$

$$r_{31} = I_1$$

$$r_{32} = I_2$$

$$r_{33} = I_3$$

$$r_{11} = (I_4 - r_{31} X_c) / F_x$$

$$r_{12} = (I_5 - r_{32} X_c) / F_x$$

$$r_{13} = (I_6 - r_{33} X_c) / F_x$$

$$r_{21} = (I_7 - r_{31} Y_c) / F_y$$

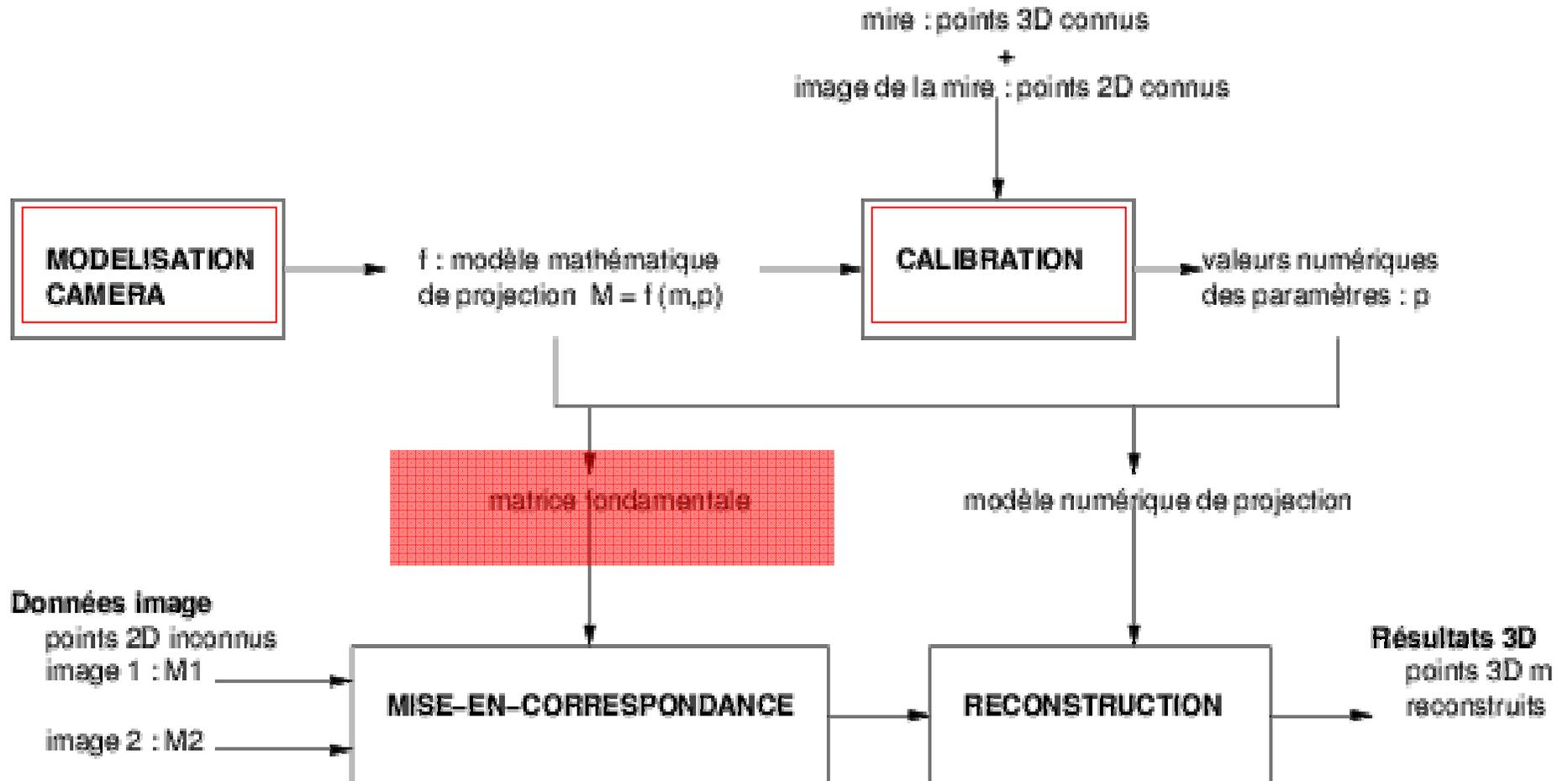
$$r_{22} = (I_8 - r_{32} Y_c) / F_y$$

$$r_{23} = (I_9 - r_{33} Y_c) / F_y$$

$$X_o = (I_{10} - X_c I_{12}) / F_x$$

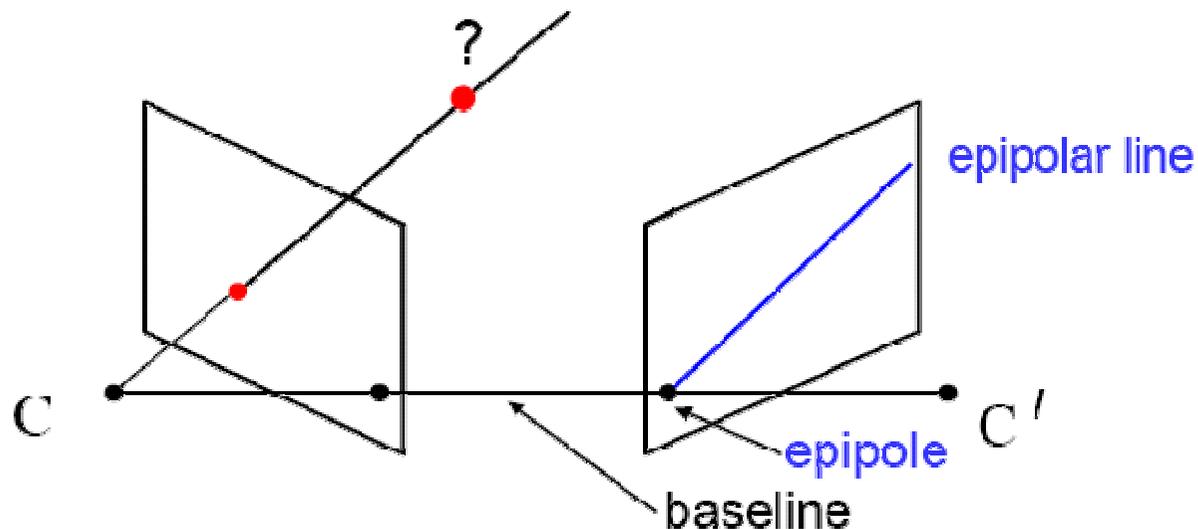
$$X_o = (I_{11} - Y_c I_{12}) / F_y$$

$$Z_o = I_{12}$$



Problème posé:

Étant donné un point dans une image, où est son correspondant dans l'autre image?



- Chaque point a une **droite épipolaire associée** dans l'autre image
- Cette droite contient son correspondant

Formulation matricielle

C_1, C_2, M_1, M_2 coplanaires
 $\Leftrightarrow C_1 \vec{M}_1 \cdot (C_1 \vec{C}_2 \wedge C_2 \vec{M}_2) = 0$

d'où $M_1^t [t]_{\times} R M_2 = 0$ (1)

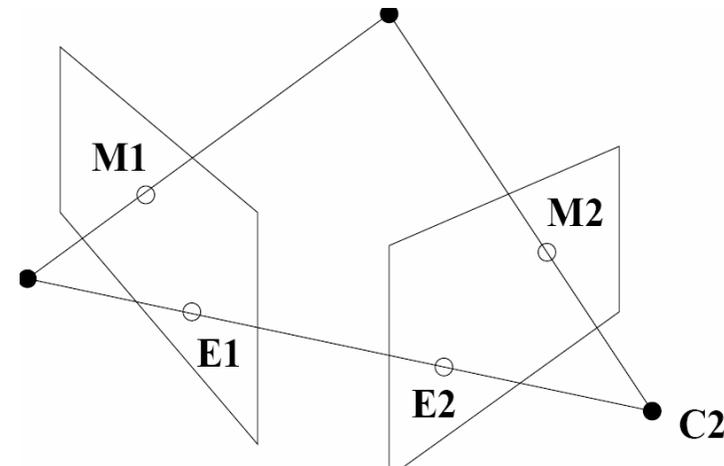
avec

M_1 : coordonnées de M_1 dans \mathcal{R}_{C_1}

M_2 : coordonnées de M_2 dans \mathcal{R}_{C_2}

(R, t) mouvement relatif entre \mathcal{R}_{C_1} et \mathcal{R}_{C_2}

$[t]_{\times}$: matrice de produit vectoriel par t



Formulation matricielle (suite)

- on définit :

la **matrice essentielle** $E = [t]_{\times} R$

la **matrice fondamentale** $F = K^{-t} [t]_{\times} R K^{-1}$

- la contrainte épipolaire (1) s'écrit alors :

pour tous pixels $M_1 (X_{1p}, Y_{1p})$ et $M_2 (X_{2p}, Y_{2p})$ homologues

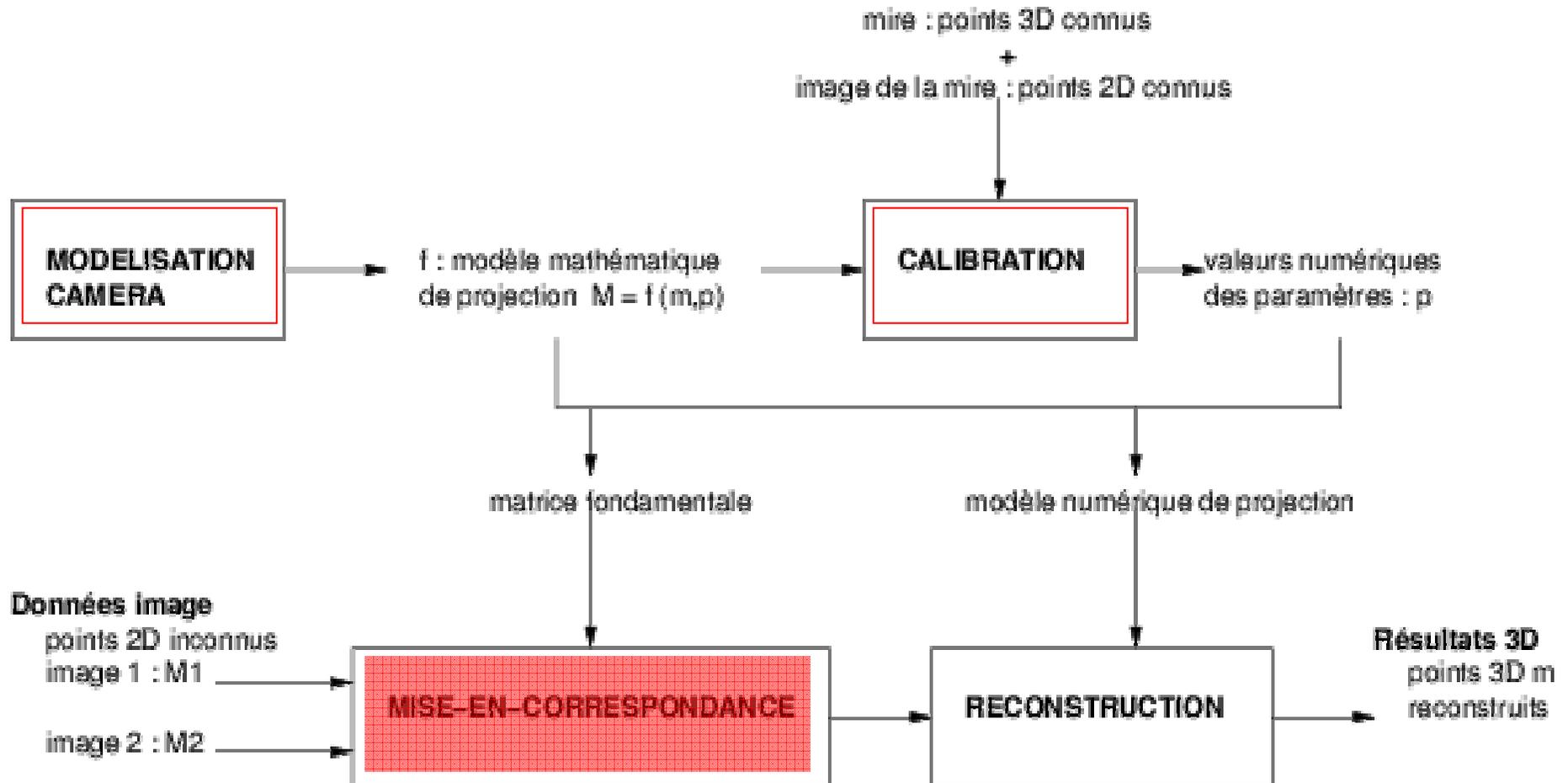
$$\boxed{M_1^t F M_2} = \begin{bmatrix} X_{1p} \\ Y_{1p} \\ 1 \end{bmatrix}^t F \begin{bmatrix} X_{2p} \\ Y_{2p} \\ 1 \end{bmatrix} = 0$$

- F définit l'équation de la droite épipolaire associée à M_1 par $F^t M_1 = (a_2, b_2, c_2)$
La contrainte épipolaire exprime que M_2 appartient à la droite épipolaire

$$M_1^t F M_2 = 0 \Leftrightarrow a_2 X_2 + b_2 Y_2 + c_2 = 0$$

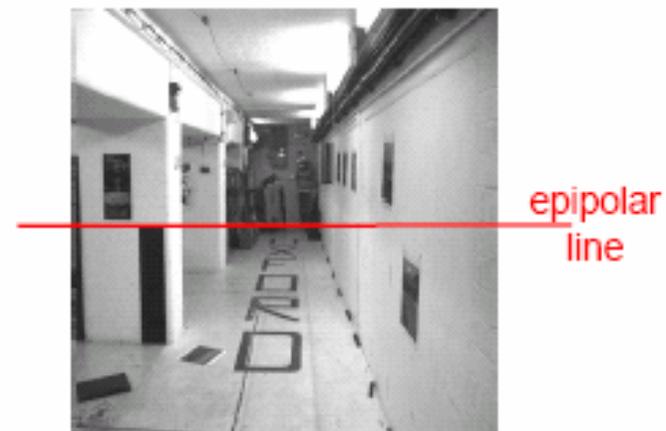
Exemple de faisceaux de droites épipolaires





Problème posé

- Problème : étant donné 2 images et les paramètres des caméras, trouver les correspondants:
 - de tous les points de l'image 1 (m-e-c **dense**)
 - de certains points de l'image 1 (m-e-c **éparse**)
- Solution : recherche du « meilleur point » sur la droite épipolaire



Principe

- Hypothèse : le correspondant a un voisinage similaire
- Utiliser une mesure de similarité entre voisinages :
exemple : NCC : coefficient de corrélation normalisé

subtract mean: $A \leftarrow A - \langle A \rangle, B \leftarrow B - \langle B \rangle$

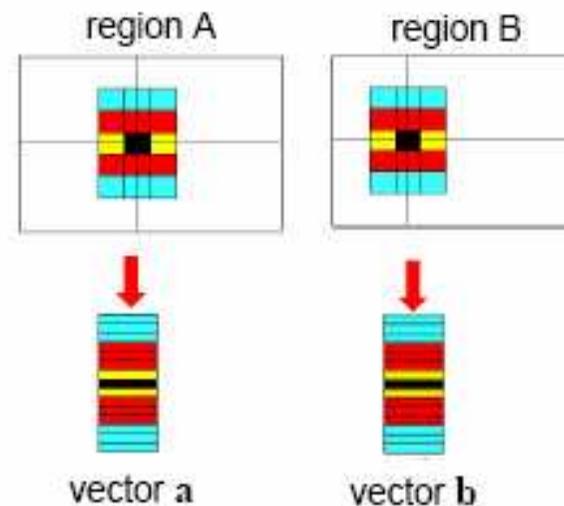
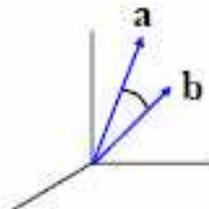
$$NCC = \frac{\sum_i \sum_j A(i, j) B(i, j)}{\sqrt{\sum_i \sum_j A(i, j)^2} \sqrt{\sum_i \sum_j B(i, j)^2}}$$

Write regions as vectors

$$A \rightarrow \mathbf{a}, B \rightarrow \mathbf{b}$$

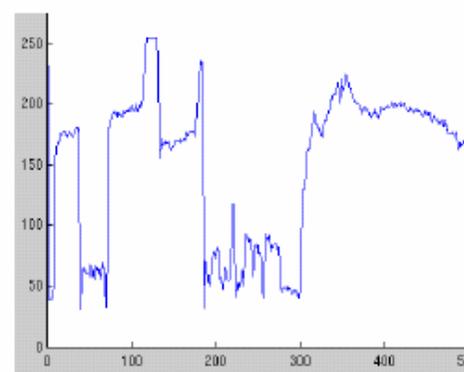
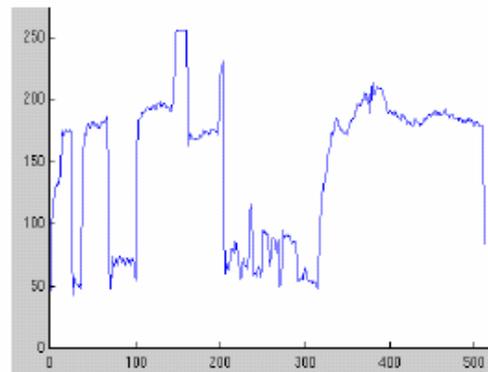
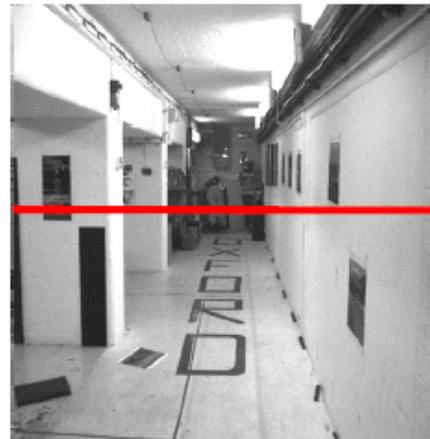
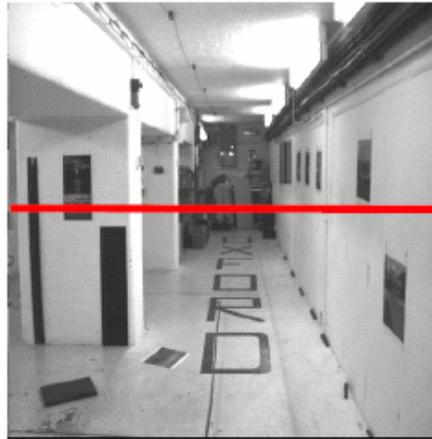
$$NCC = \frac{\mathbf{a} \cdot \mathbf{b}}{|\mathbf{a}| |\mathbf{b}|}$$

$$-1 \leq NCC \leq 1$$



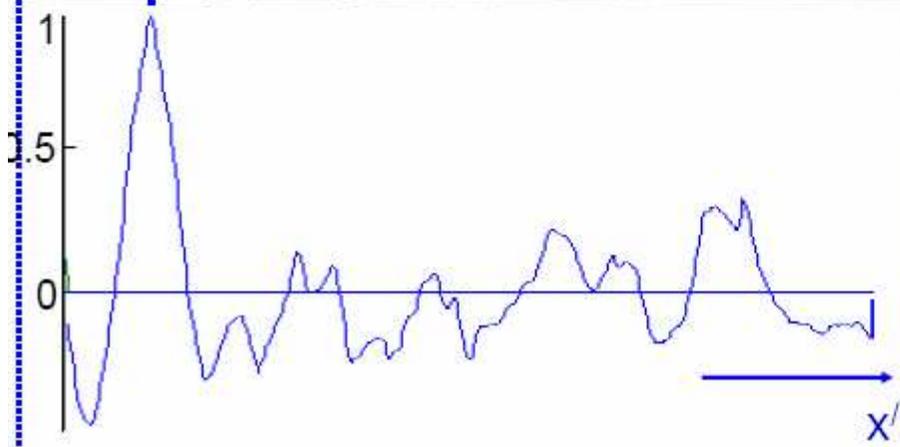
Exemple

caméra fronto-parallèles => droites épipolaires horizontales



nette correspondance entre les profils

Exemple : cas 1



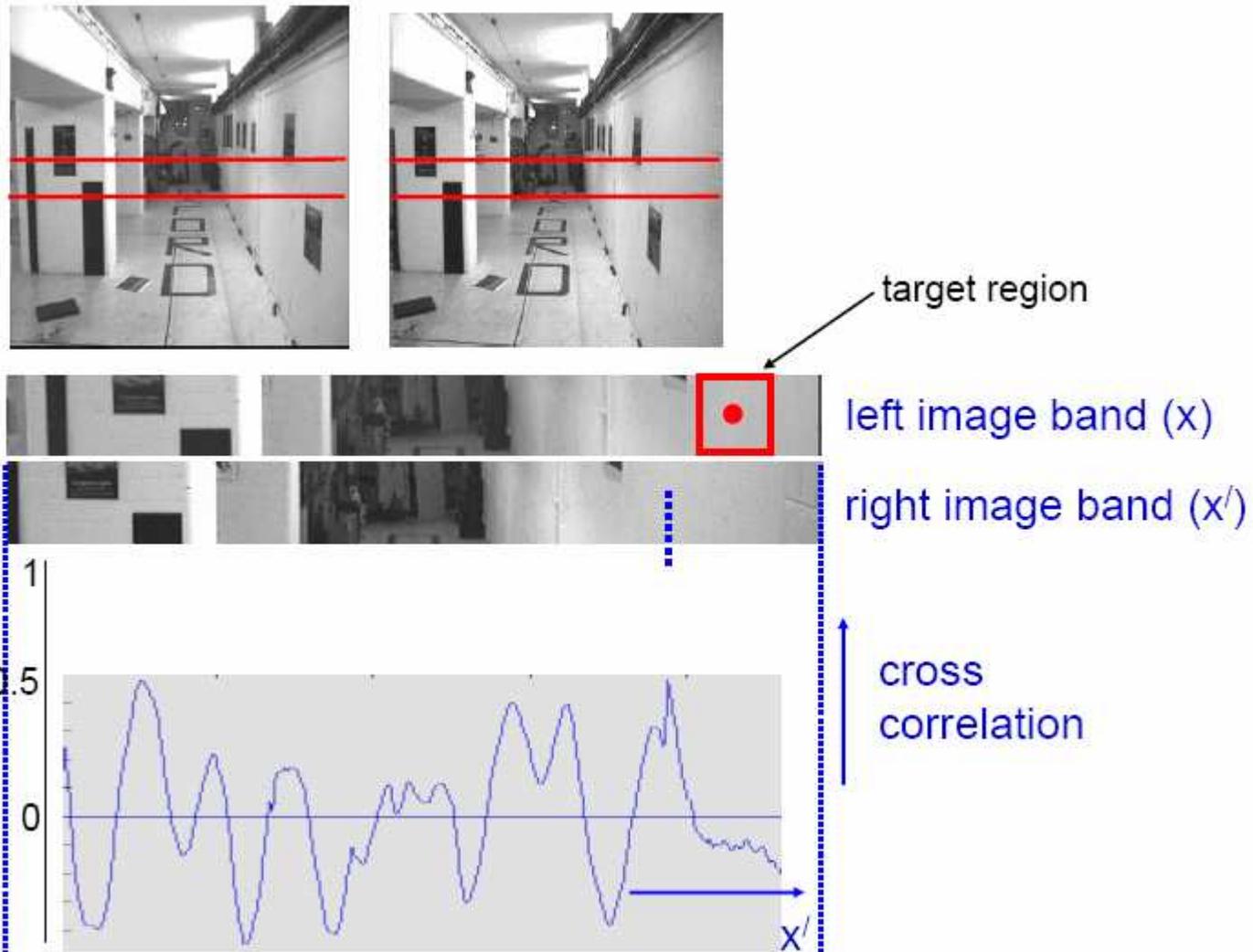
left image band (x)

right image band (x')

cross correlation

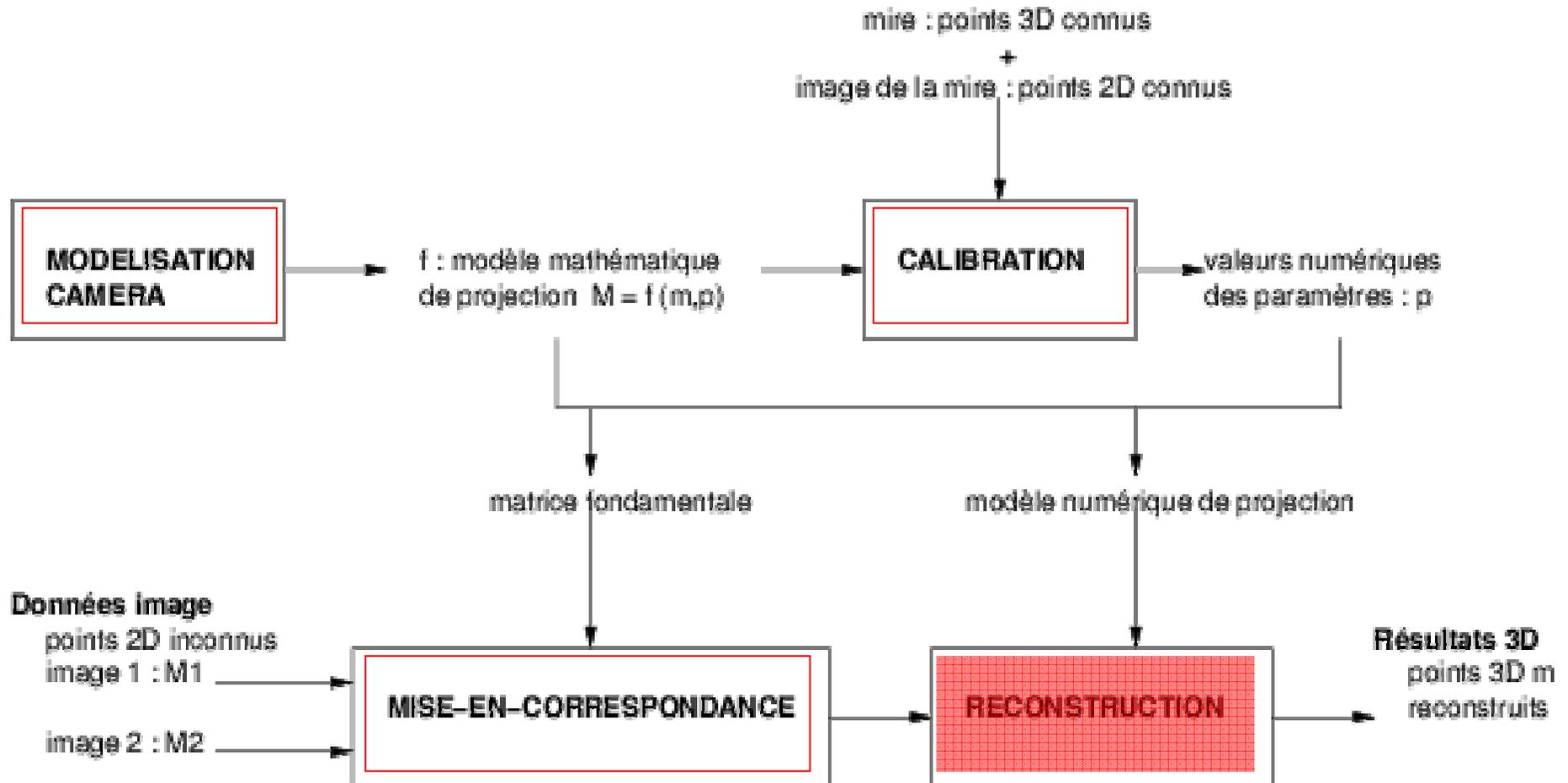
disparity = $x' - x$

Exemple : cas 2



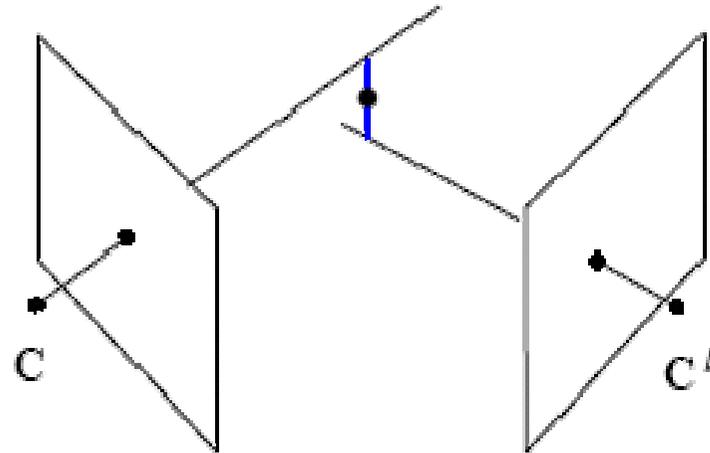
2.a Reconstruction calibrée

Reconstruction par triangulation



Problème posé

- mesures réelles bruitées \Rightarrow les droites 3D ne s'intersectent pas!
- Solution géométrique :
point le plus proche des droites



- Solution algébrique :
N lignes de vues \Rightarrow $2N$ équations linéaires en x, y, z
+ simple
+ généralisable à plus de vues,
- minimise une grandeur non géométrique

2.a Reconstruction calibrée

Exemples de scènes reconstruites

(source A. Zisserman, Univ of Oxford)

Example dense correspondence algorithm



left image



right image



depth map
intensity = depth

Views of a texture mapped 3D triangulation



2.a Reconstruction calibrée

Exemples de scènes reconstruites

(source A. Zisserman, Univ of Oxford)

Pentagon example

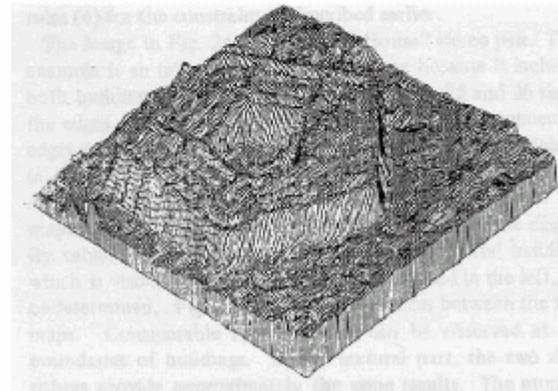
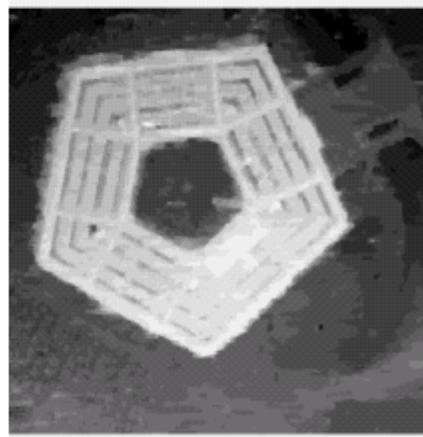
left image



right image



range map



2.b Reconstruction non calibrée

Problème posé:

- Étant données 2 (ou plus) images d'une scène, calculer le relief de la scène **sans connaître les positions des caméras** : il faut les estimer aussi!
- Structure from Motion → Structure **and** Motion from video



- Hypothèse : les paramètres intrinsèques sont connus
- Hypothèse : caméra en mouvement + scène rigide
- Attention : on ne connaît pas la géométrie épipolaire!

2.b Reconstruction non calibrée

- Principe pour 2 images :
 - Trouver des points en correspondance
 - Utiliser la contrainte épipolaire pour estimer la matrice fondamentale F
 - Estimer le mouvement caméra (R,t) par décomposition de F

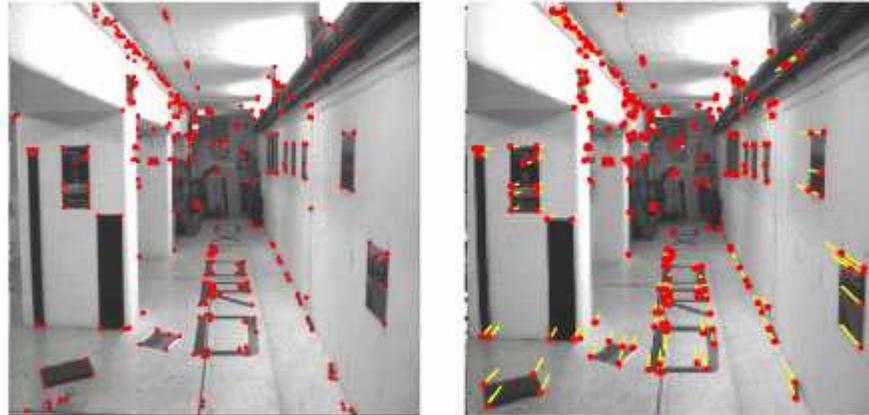
$$F = K^{-t} [t]_{\times} R K^{-1}$$

- Principe pour plus de 2 images
 - 3 images : tenseur trifocal => R,t,K (en théorie)
 - N images : alignement de faisceaux (*bundle adjustment*) => min. non linéaire
- Limitations
 - Reconstruction à un facteur d'échelle près ($\lambda m, \lambda t$) est aussi solution
 - Contrainte sur les mouvements de caméra permettant la reconstruction 3D

2.b Reconstruction non calibrée

Exemple d'estimation
de la géométrie épipolaire

Correspondences consistentes avec la géométrie épipolaire



Computed epipolar geometry



Conclusion

- Un aperçu des méthodes de base de la reconstruction 3D....
- Il existe beaucoup de travaux sur le sujet
- Modélisation assistée :
 - Intervention humaine [debevec96][debevec00]
 - Système d'acquisition spécifique
 - Table tournante [niem94] [debevec96][gibson98]
 - Robot [mellor00] [zisserman01]
 - A priori sur le contenu de la scène
 - Visages [preteux00] [girod02]
 - Scènes architecturales [faugeras95][hartley00][dick00] [bazin01] [werner02][sturm02]
- Modélisation non assistée :
 - Reconstruction d'un modèle 3D à partir d'images ou vidéos
 - [fitzgibbon99] [roning99][pollefeys00][yao02] [Nis03][yu04]

2 - Reconstruction 3D par vision par ordinateur

Tendances actuelles de la vision par ordinateur

- Convergence synthèse/vision : recherche du photoréalisme
- Séquences multivues (MVV) : plusieurs caméras (fixes), une scène en mouvement
 - Intérêt : capture et modélisation du mouvement, « effet matrix » (Free-view-point vidéo FVV)
 - Applications : TV3D, interfaces interactives 3D (visio-conférence 3D, jeux, interfaces intelligentes)
- Temps réel
 - Indispensable pour les applications interactives
 - Rendu possible par les processeurs tels que les GPU
- Reconstruire ... et transmettre? visualiser à distance? (cf CitéVision)

Introduction

I - Reconstruction 3D

II - Compression Vidéo

III - Modélisation 3D de séquences vidéo

Conclusion

- 1 – Schémas de codage
 - a) Image
 - b) Vidéo
- 2 – Compression de l'information : codage entropique
 - a) Principe du codage statistique
 - b) Bases théoriques : théorie de l'information
 - c) Codage de Huffman
 - d) Codage arithmétique
 - e) Limitations du codage entropique
- 3 – Réduction de la dépendance spatio-temporelle
 - a) Dépendance spatiale
 - b) Dépendance temporelle
- 4 – Obtention du débit cible : quantification
 - a) Principe de la quantification
 - b) Mesure de la distorsion
 - c) Compromis débit-distortion
 - d) Courbe R/D
- 5 -- Quelques exemples
- 6 - Principaux standards de compression image et vidéo

Débit

- Débit d'un signal vidéo sans compression :
237 Mb/s (télévision SDTV (images 720x576))
- Débits disponibles :
RTC : 56Kb/s → **facteur 4000**, ADSL : 15 Mb/s → **facteur 15**

Place mémoire

- Volume d'un signal vidéo sans compression :
1,22 Térabits (1h1/2 de film)
- Capacité de stockage disponible :
DVD R/W capacité : 4,7 Gb → **facteur 250**

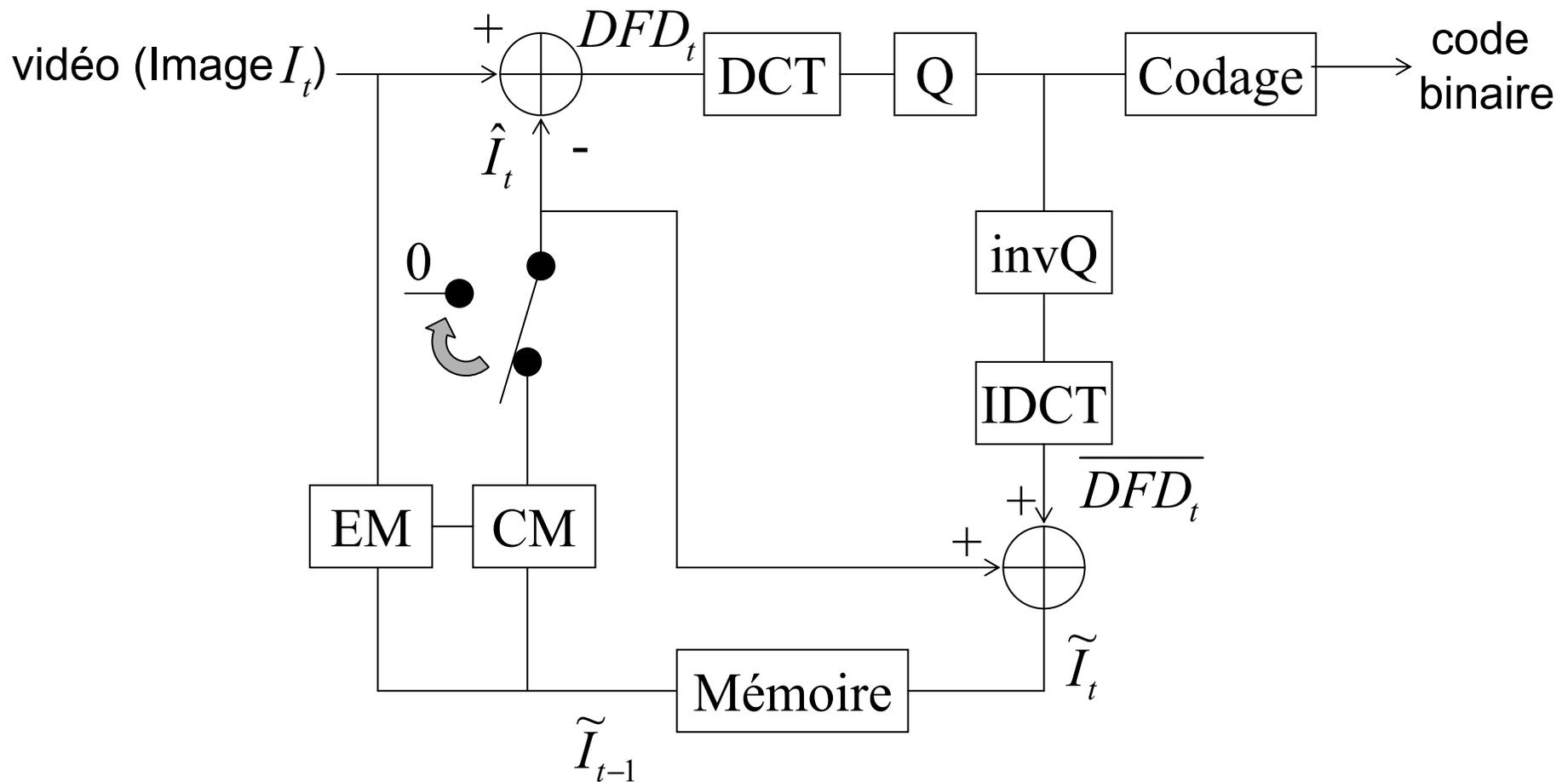
1 - Schémas de codage : image

Image : schéma classique (JPEG, JPEG2000)

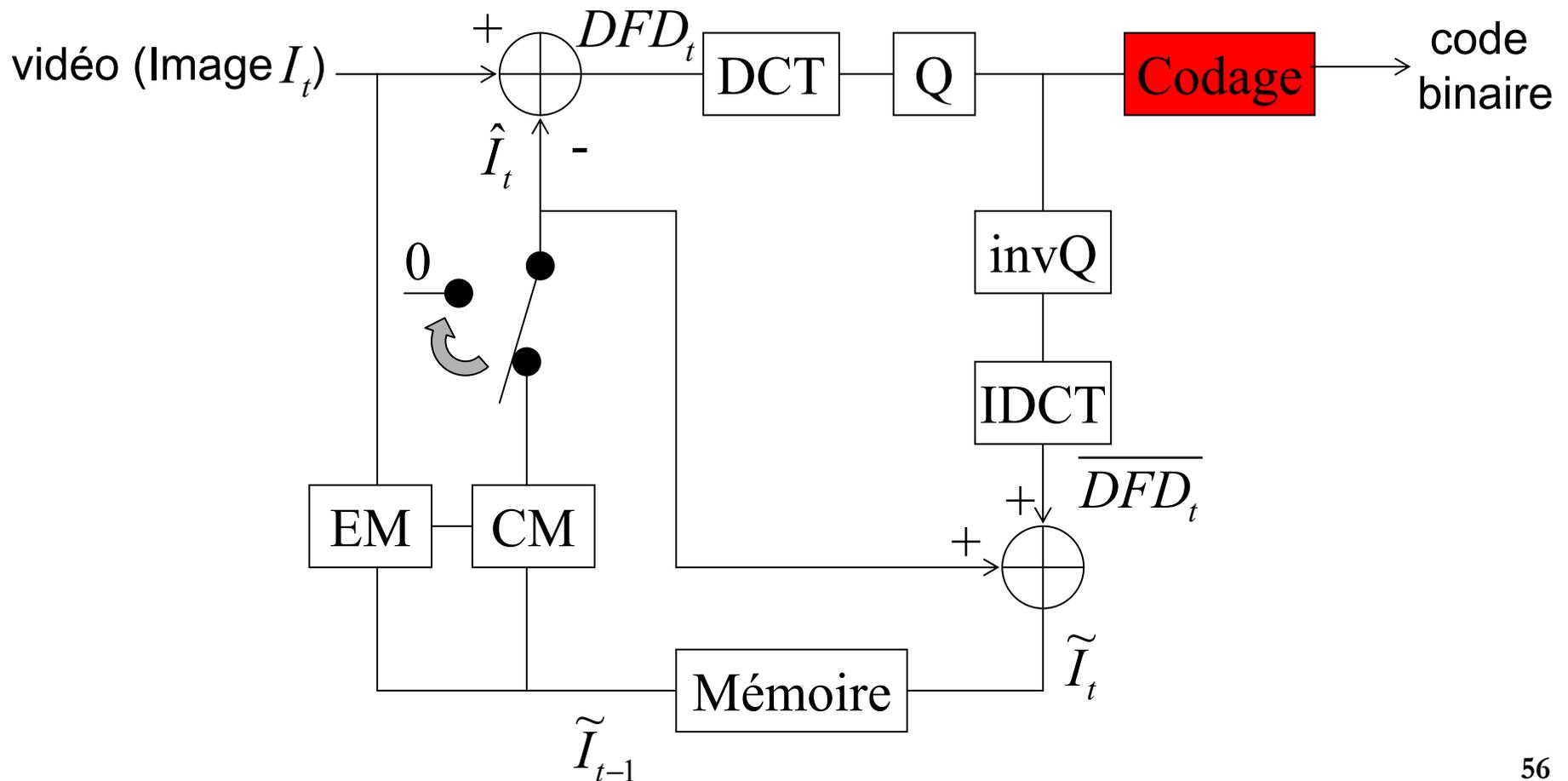


1 - Schémas de codage : vidéo

Codeur vidéo hybride (ex MPEG2, H264)



2 – Compression de l'information : codage entropique



2 – Compression de l'information : codage entropique

a) Principe du codage statistique

- But de la compression : si un message (image) est codé sur une longueur de L bits, faire en sorte que L soit le plus petit possible

- Taux de compression : $\tau = \frac{\text{nb bits représentation initiale}}{\text{nb bits représentation compressée}}$

- Codage statistique : VLC (Variable Length Codes)

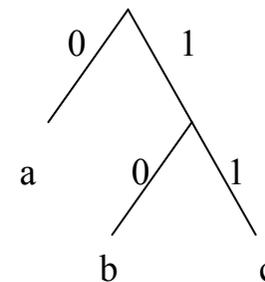
- Valeurs fréquentes : code court
- Valeurs peu fréquentes : code long

$$\tau = \frac{\log_2 N}{L}$$

- Comment découper au décodage?

- Règle du préfixe → codes en arbre

- Comment choisir les codes des valeurs?



ex: décoder 11010011

2 – Compression de l'information : codage entropique

b) Bases théoriques : théorie de l'information (Shannon, 1948)

message = suite de **symboles**, de valeur **codée en binaire**

ex : **image** = suite de **pixels**, de niveau de gris codé sur **8 bits**

chaque symbole = réalisation d'une variable **aléatoire**

source = générateur de messages suivant une **loi de probabilité**

- source de toutes les images plausibles (loi ????)
- source des images issues d'un fax, d'un scanner IRM, d'un photomaton
- source d'images de « neige » sur la TV

2 – Compression de l'information : codage entropique

Notion d'entropie (définition de Shannon) :

- mesure la quantité d'information apportée par un message A
- unité : le bit => fournit le nombre de bits nec et suf pour coder A
- Quantité d'information : $I(A) = -\log_2 p(A)$
- Entropie d'une source : quantité d'information moyenne délivrée source S avec N symboles possibles = $\{x_1, \dots, x_N\}$

$$H(S) = E\{I(x_i)\} = \sum_{i=1}^N -p_i \log_2 p_i$$

2 – Compression de l'information : codage entropique

Exemples

- loi uniforme : $H(S) = \log_2 N$
- autres exemples

P(0)	P(1)	P(2)	P(3)	P(4)	P(5)	P(6)	P(7)	Entropie
1	0	0	0	0	0	0	0	0
0	0	0,5	0,5	0	0	0	0	1
0	0	0,25	0,25	0,25	0,25	0	0	2
0,06	0,23	0,3	0,15	0,08	0,06	0,06	0,06	2,68
0,125	0,125	0,125	0,125	0,125	0,125	0,125	0,125	3

- H augmente avec le nombre de symboles présents
- si loi plus « étalée », entropie plus élevée

Propriété importante

$$H(S) \leq \bar{L} \leq \log_2 N$$

Théorème de Shannon : $\forall \varepsilon, \exists$ un codage tel que $\bar{L} - H(S) < \varepsilon$

2 – Compression de l'information : codage entropique

c) Codage de Huffman (JPEG, MPEG)

- optimal pour les codes entiers (le plus proche de l'entropie)
- version adaptative : maj au fil de l'eau de la loi de probabilité

d) Codage arithmétique (JPEG2000, H264)

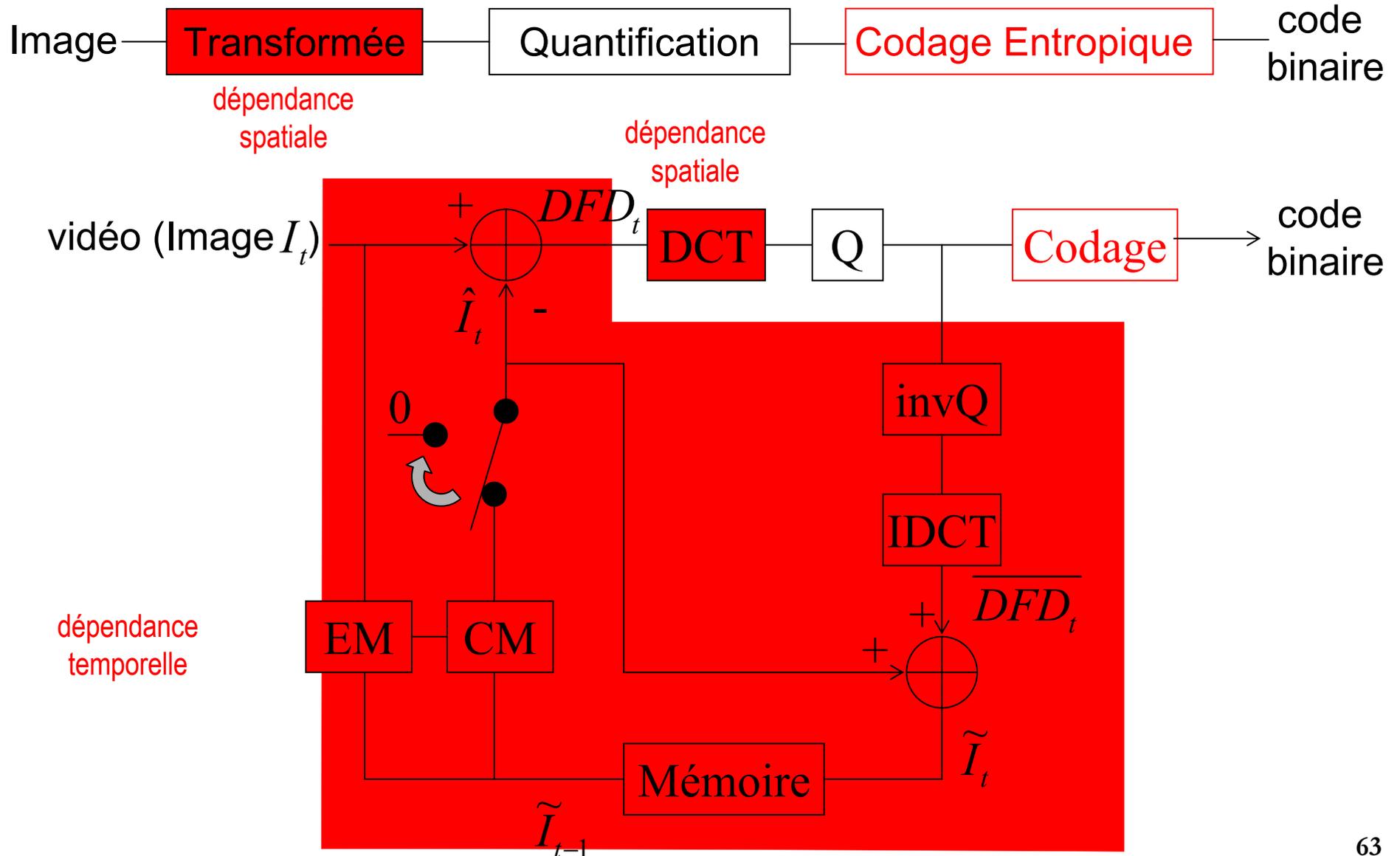
- 1 code pour plusieurs symboles : $L < 1$ bit/symbole possible
- permet d'atteindre l'entropie du message (+2 bits)
- construction incrémentale du code
- version adaptative (ex: CABAC dans H264)

2 – Compression de l'information : codage entropique

e) Limitations du codage entropique :

- Hypothèse non valide
 - symboles indépendants (source sans mémoire)
 - les pixels ne sont pas indépendants!
 - => **but : rendre les données statistiquement indépendantes** en entrée du codeur entropique
- Taux de compression limité :
 - débit le plus bas atteignable = entropie de la source
 - Comment faire si débit (volume) en sortie imposé ?
 - => **but : atteindre des débits aussi bas que l'on veut**

3 - Réduction de la dépendance spatio-temporelle



3 - Réduction de la dépendance spatio-temporelle

a) Dépendance spatiale :

= les pixels voisins sont statistiquement dépendants

Solution : **transformation décorrélante** (changement de base)

- Karhunen-Loeve : décorrélation totale, mais inapplicable
- Transformations fréquentielles :
 - DCT** : bonne décorrélation, rapide (si par blocs)
 - DWT** (ondelettes) : bonne décorrélation : meilleure qualité visuelle, scalabilité
- rem : décorréler \Leftrightarrow rendre indépendant, sous l'hypothèse de données gaussiennes (hypothèse souvent valide)

En sortie : coefficients statistiquement indépendants

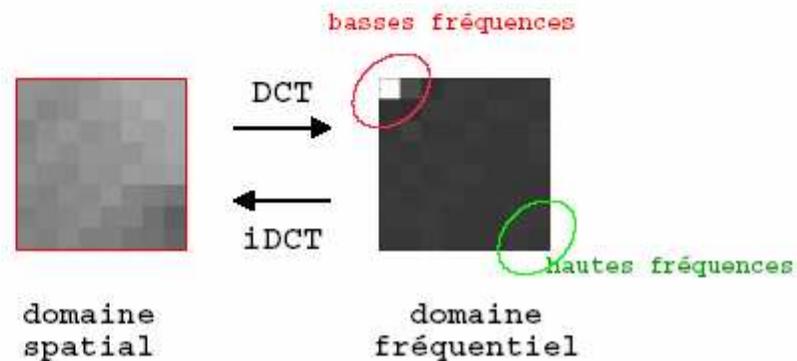
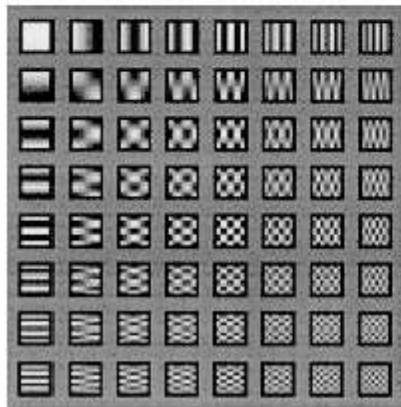
Transformée DCT par blocs 8x8

$$DCT[\mathcal{I}](u, v) = \frac{1}{\sqrt{2N}} c_u c_v \sum_{i=0}^{N-1} \sum_{j=0}^{N-1} \cos\left(\frac{(2i+1)u\pi}{2N}\right) \cos\left(\frac{(2j+1)v\pi}{2N}\right) \mathcal{I}(i, j)$$

$$iDCT[\mathcal{D}](i, j) = \frac{1}{\sqrt{2N}} \sum_{u=0}^{N-1} \sum_{v=0}^{N-1} c_u c_v \cos\left(\frac{(2u+1)i\pi}{2N}\right) \cos\left(\frac{(2v+1)j\pi}{2N}\right) \mathcal{D}(u, v)$$

$$\text{avec } c_k = \begin{cases} \frac{1}{\sqrt{2}} & \text{si } k = 0 \\ 1 & \text{sinon} \end{cases}$$

Décomposition sur la base suivante :



3 - Réduction de la dépendance spatio-temporelle

b) Dépendance temporelle :

Pb: images successives d'une vidéo statistiquement dépendantes

Solution : estimation et compensation de mouvement

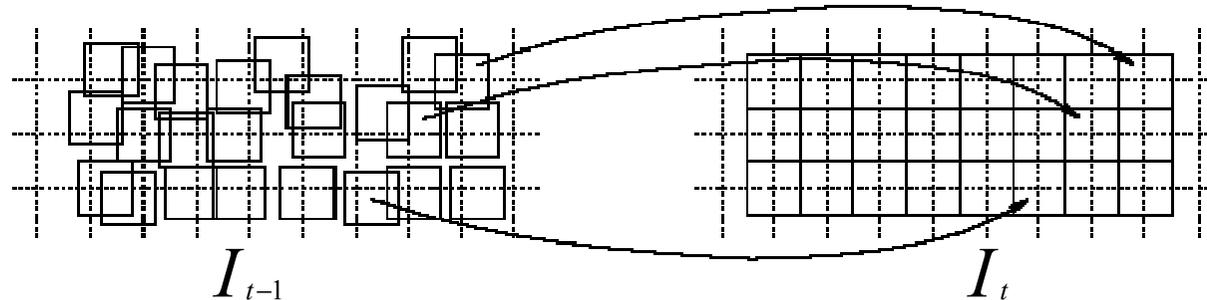
Principe :

- Prédire I_t à partir de I_{t-1} \rightarrow image prédite \tilde{I}_t
- Encoder le résidu de prédiction $I_t - \tilde{I}_t$

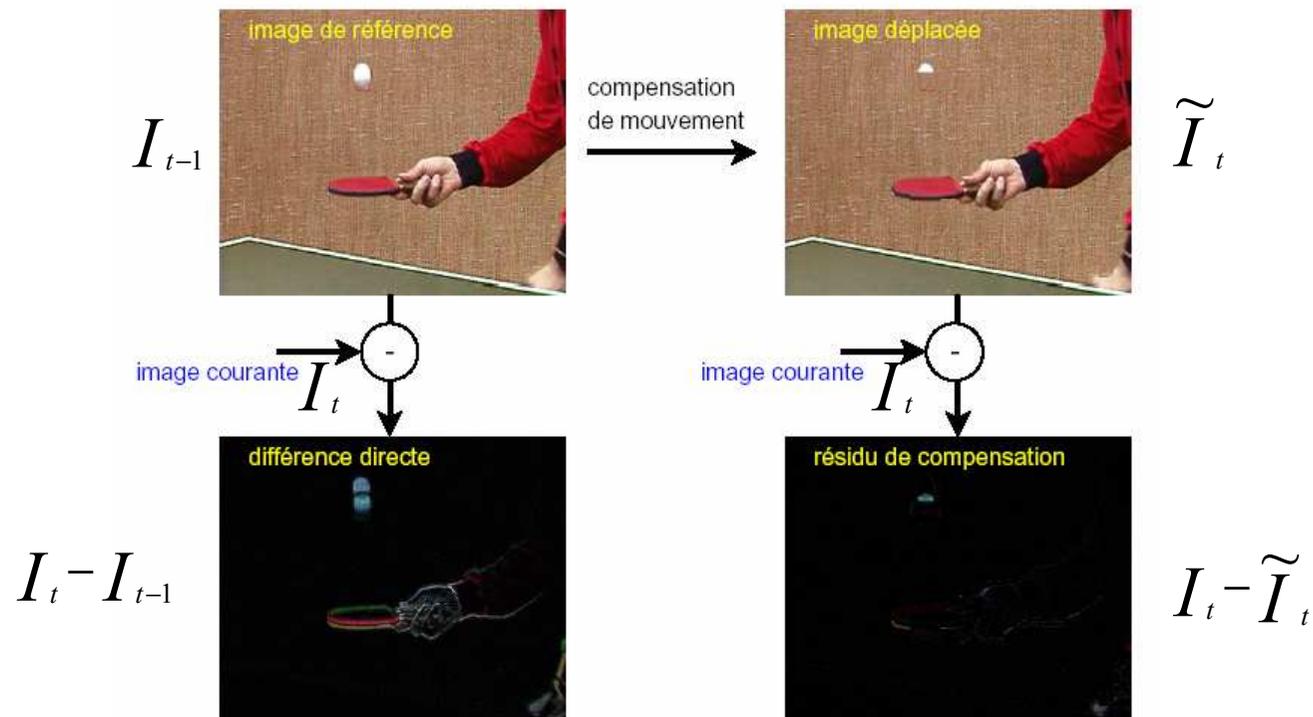
3 - Réduction de la dépendance spatio-temporelle

Prédiction par blocs:

arrière (*backward*)



Exemple



3 - Réduction de la dépendance spatio-temporelle

Estimation de mouvement :

↪ On cherche à minimiser pour chaque macrobloc l'erreur de prédiction :

$$\sum_{(x,y) \in \text{bloc}} \|I(x, y, t) - I(x + dx, y + dy, t - 1)\|$$

par rapport au *vecteur de mouvement* (dx, dy) .

On utilise comme *critère d'erreur* $\|\cdot\|$ soit

- le carré de l'erreur (norme 2)
- la valeur absolue de l'erreur (norme 1)

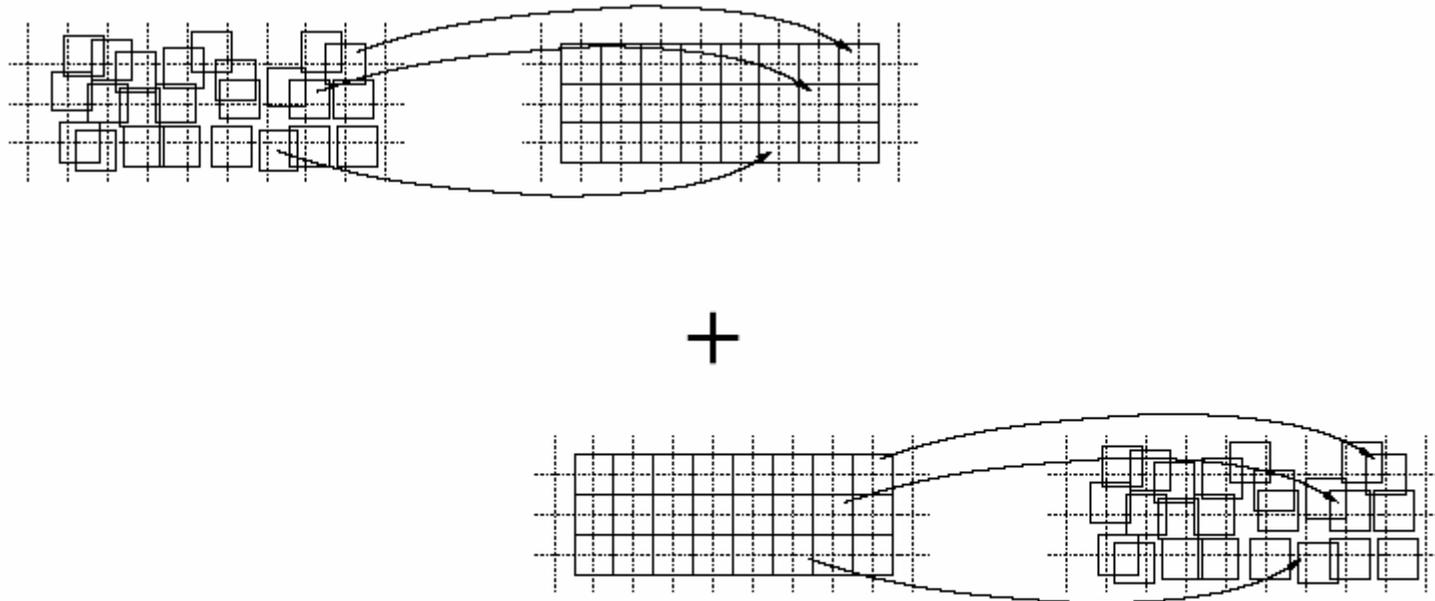
On peut effectuer sur la zone de recherche :

- Une minimisation globale par recherche exhaustive (lent,optimal)
- Une minimisation locale selon diverses heuristiques (rapide,sous-optimal)

3 - Réduction de la dépendance spatio-temporelle

Amélioration de la prédiction : compensation de mouvement bidirectionnelle :

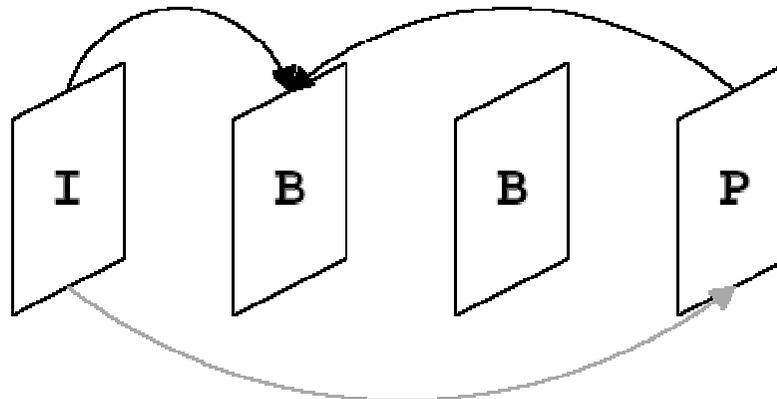
3. Bidirectionnelle



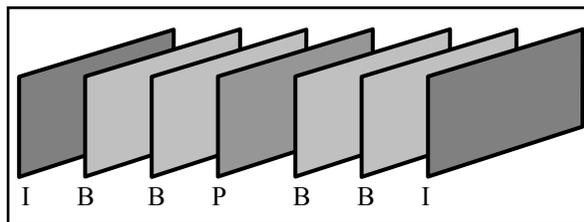
$$\text{prédicteur} = \frac{\text{avant} + \text{arrière}}{2}$$

GOP : Group Of Pictures

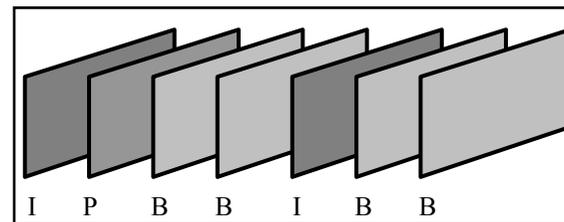
- Un GOP contient différents types d'images :
 - **I** (*intra* ou *clé*) codée indépendamment des autres images
 - **P** (*prédite*) codée par rapport à l'image précédente
 - **B** (*bidirectionnelle*) codée par rapport à une image précédente et suivante



- L'ordre de transmission n'est pas l'ordre dans la séquence



Ordre de Présentation



Ordre de transmission

3 - Réduction de la dépendance spatio-temporelle

Information à transmettre :

- résidu de prédiction

traité comme un image fixe

(décorrélation spatiale par transformation fréquentielle)

- information de mouvement

un vecteur de mouvement par bloc

remarque : nouvelles approches : décorrélation 2D+t

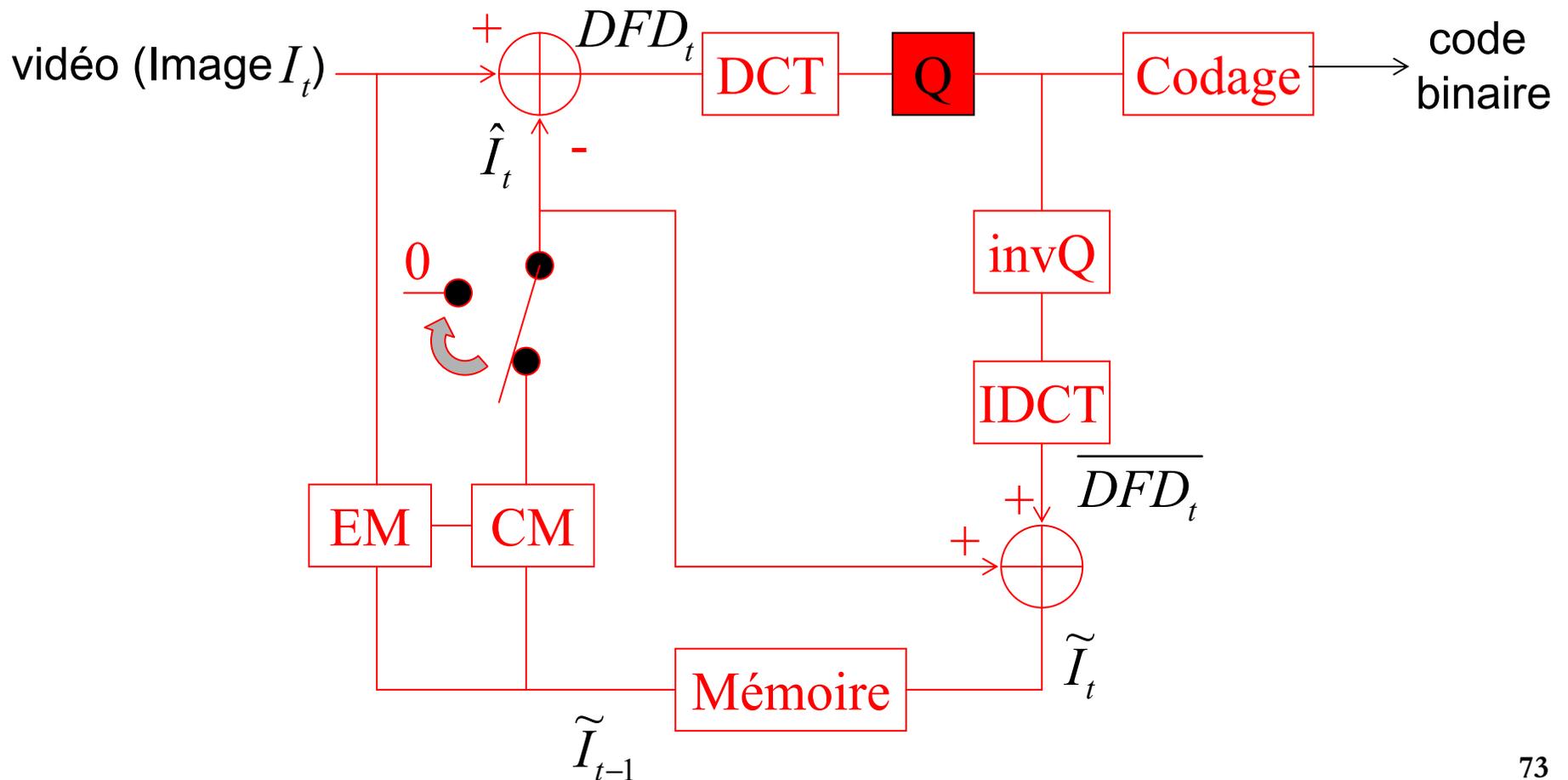
- transfo fréquentielle 3D(x,y,t) de la séquence compensée en mvt
- ondelettes de seconde génération : transformation le long des lignes du mouvement

4 - Obtention du débit cible : quantification

Rappel :

- Taux de compression nécessaires pour transmettre / stocker la vidéo sur les supports actuels : 10, 100, 500
(en bits par pixel : bpp :
 - Codage statistique : taux inférieurs à 5 pour la vidéo
- ⇒ Il faut réduire le débit, quitte à supprimer de l'information
- réduire la taille de l'image
 - réduire la fréquence temporelle
 - réduire le nombre de couleurs
 - dans les zones unies, envoyer (valeur, nombre de répétitions)

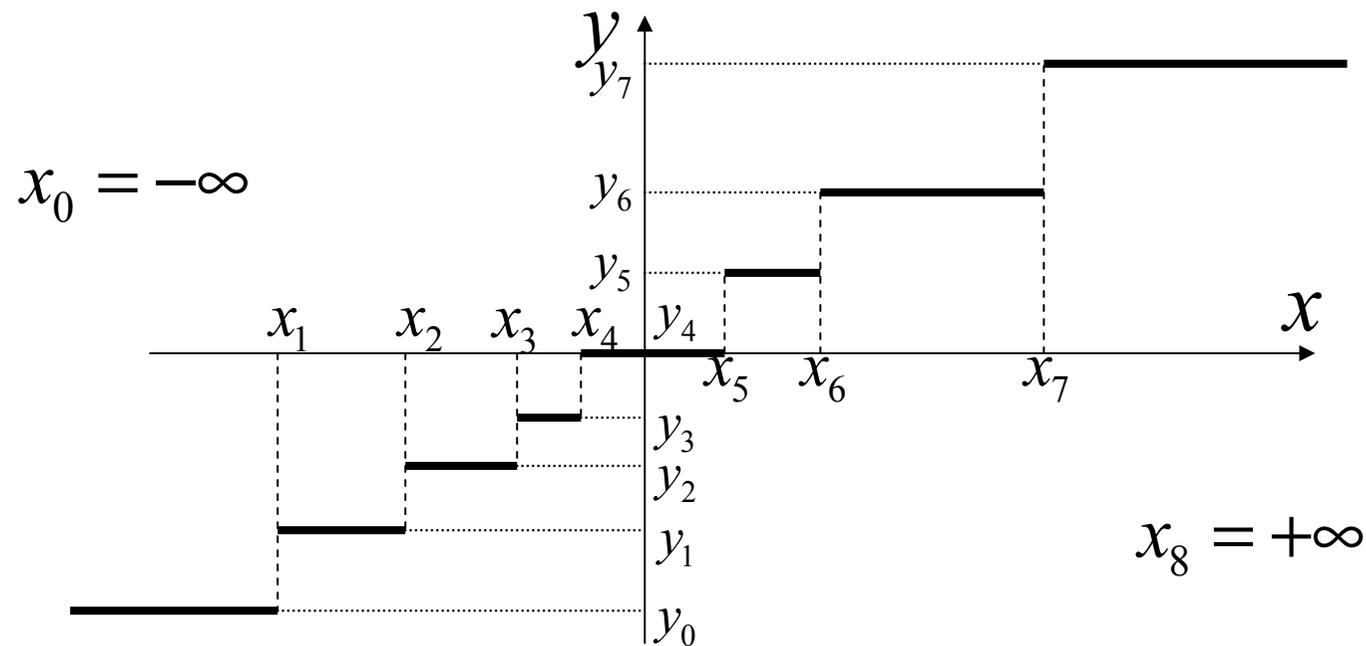
4 - Obtention du débit cible : quantification



4 - Obtention du débit cible : quantification

a) Principe de la quantification

représentation sur moins de bits => approximation



- JPEG : quantification des coefficients DCT,
- JP2K : troncature du train binaire issu de SPIHT

4 - Obtention du débit cible : quantification

b) Mesure de la distorsion

- mesures subjectives : faites par des observateurs humains
- mesures objectives

- Erreur Quadratique Moyenne **EQM** (MSE)

$$EQM = \frac{1}{N} \sum_{i=1}^N (x_i - y_i)^2 = \frac{1}{N \times M} \sum_{i=1}^N \sum_{j=1}^M (I_1(i, j) - I_2(i, j))^2$$

- Peak Signal over Noise Ratio **PSNR**

$$PSNR = 10 \cdot \log_{10} \frac{255^2}{EQM}$$

- mesures objectives de qualité subjective

Prise en compte de phénomènes du SVH

- masquage
- sensibilité nombre de couleurs
- sensibilité sur la teinte

4 - Obtention du débit cible : quantification

c) Compromis débit-distorsion

But:

- pour un débit donné, avoir la plus faible distorsion possible
- pour une distorsion acceptable, avoir le plus faible débit possible

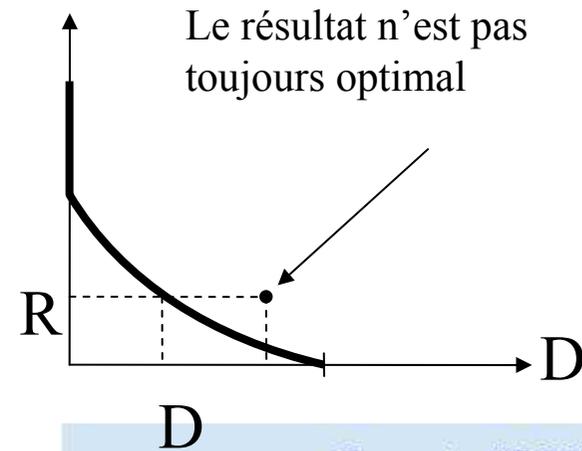
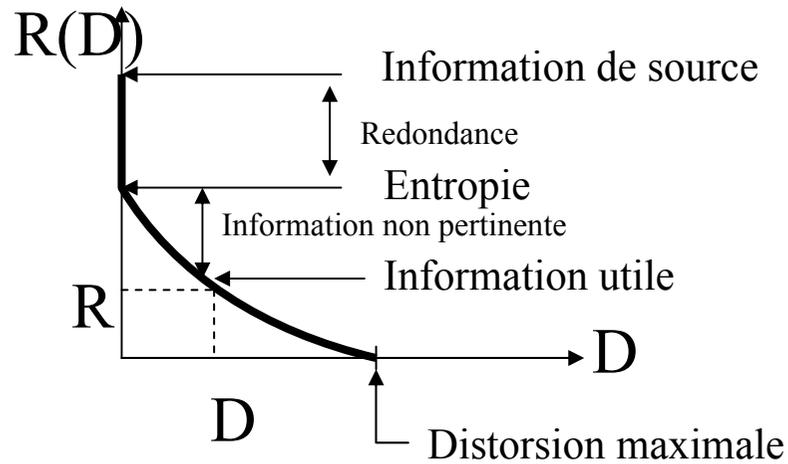
$$\begin{cases} \min_{\theta} D_{\theta}(I) \\ R_{\theta} \leq R_{\max} \end{cases} \quad \begin{cases} \min_{\theta} R_{\theta}(I) \\ D_{\theta} \leq D_{\max} \end{cases}$$

Approche classique : multiplicateurs de Lagrange

- $\min_{\theta} R_{\theta}(I) + \lambda \cdot D_{\theta}(I)$
- recherche du λ afin de vérifier la condition limite

4 - Obtention du débit cible : quantification

d) Courbe R/D



- Permet de comparer différents codeurs

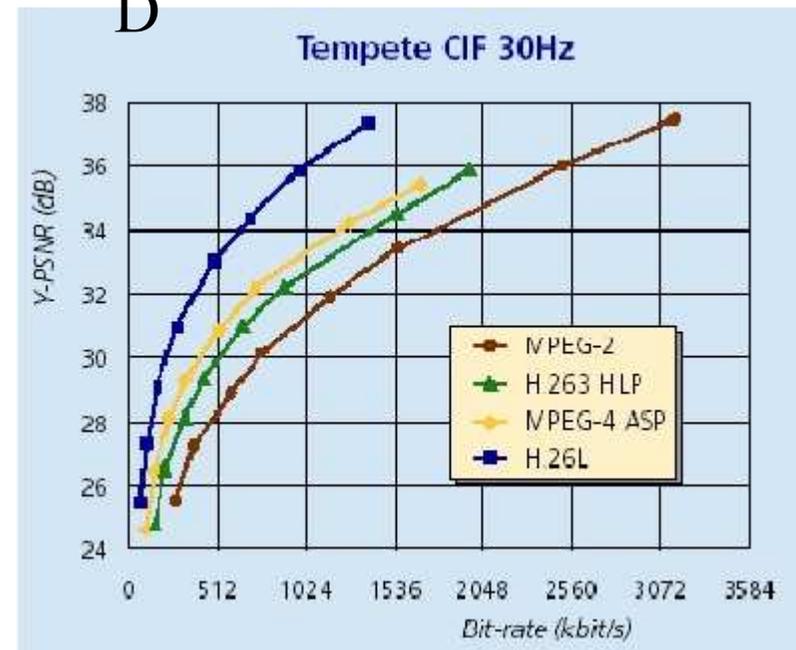




Image originale « Lena »
8 bits par point (bpp)



Taux Compression = 6,03:1
1,32 bpp

$$\text{TauxCompression} = \text{VolumeOriginal} / \text{VolumeCompressé}$$



Image originale « Lena »



Taux Compression = 21,65:1
0,37 bpp

$$\text{TauxCompression} = \text{VolumeOriginal} / \text{VolumeCompressé}$$



Image originale « Lena »



Taux Compression = 38,83:1
0,21 bpp

$$\text{TauxCompression} = \text{VolumeOriginal} / \text{VolumeCompressé}$$



Image originale « Lena »



Taux Compression = 41,37:1
0,19 bpp

$$\text{TauxCompression} = \text{VolumeOriginal} / \text{VolumeCompressé}$$



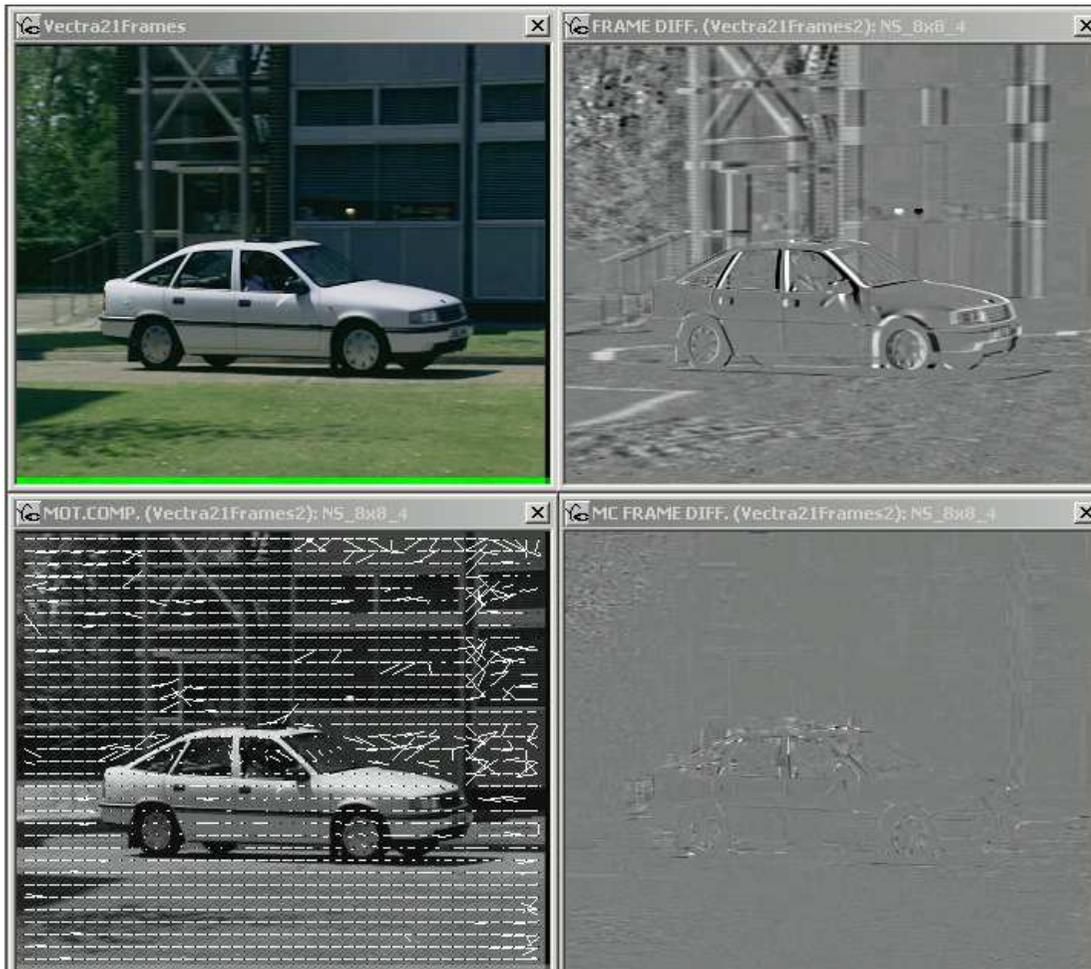
Image originale « Lena »



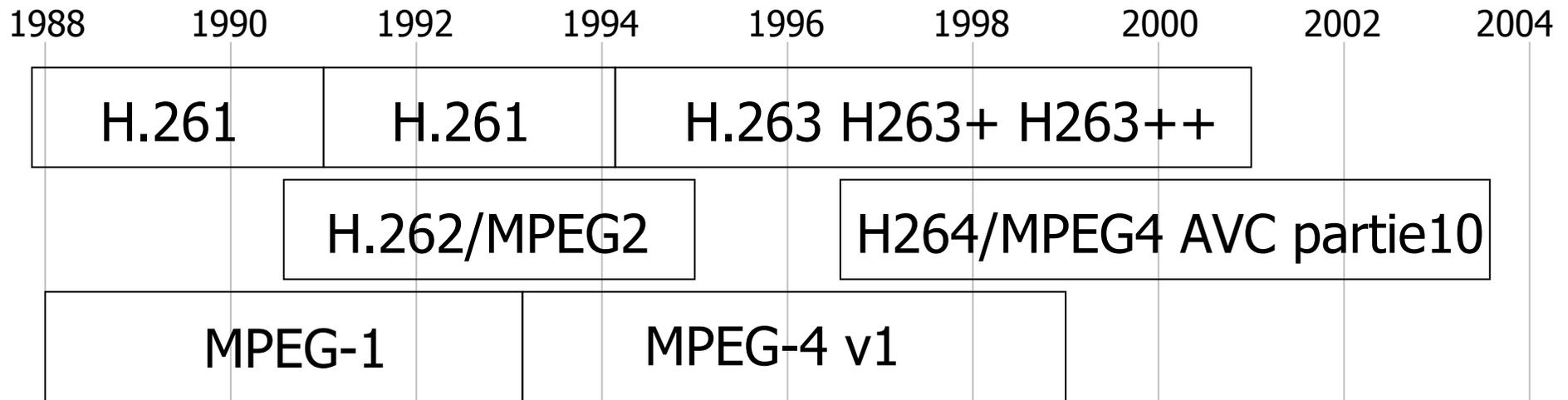
Taux Compression = 80,12:1
0,1 bpp

$$\text{TauxCompression} = \text{VolumeOriginal} / \text{VolumeCompressé}$$

5 - Quelques exemples MPEG1



6 – Principaux standards de compression vidéo



MPEG1 : Débit inférieur à 1,5 Mbit/s;
stockage sur CD d'un film de qualité comparable aux cassettes VHS

MPEG2 : Pour la télévision numérique et les DVDs;
Diffusion à moyens débit (4 à 8Mbit/s);
Qualité studio (30Mbit/s)

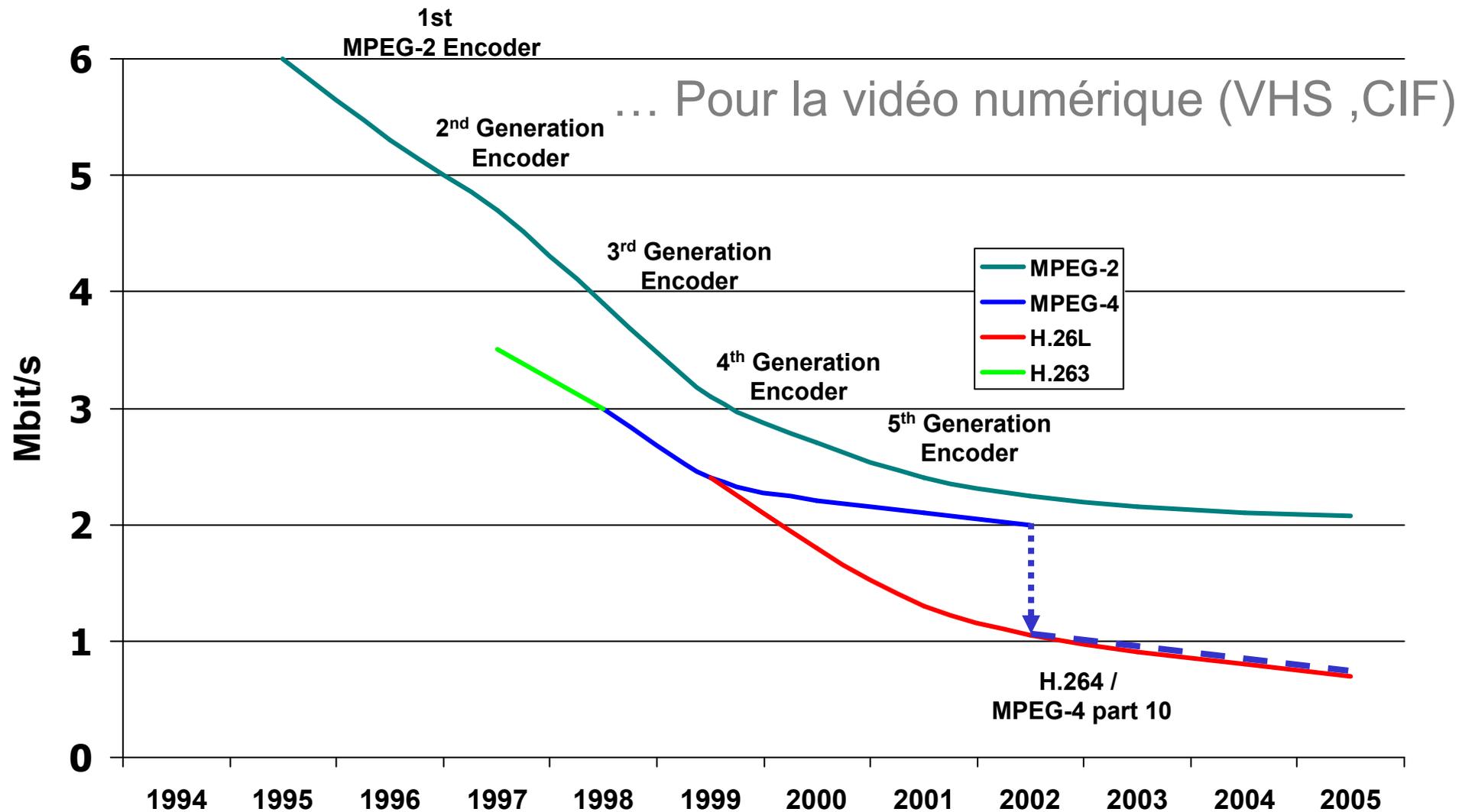
MPEG4 : Reprend les possibilités des anciens standards;
Permet de faire du codage basé objet; DiVx

H264/MPEG4 partie10/AVC : prochain standard « universel »

Performances MPEG2 x 2

Diffusion numérique TVnum (2 à 4 Mbit/s) ;

Visioconf (10kbit/s à 1 Mbit/s) ; DVD-HD ; TNT ; TV-HD sur ADSL



plan de l'exposé

Introduction

I - Reconstruction 3D

II - Compression Vidéo

III - Modélisation 3D de séquences vidéo

Conclusion

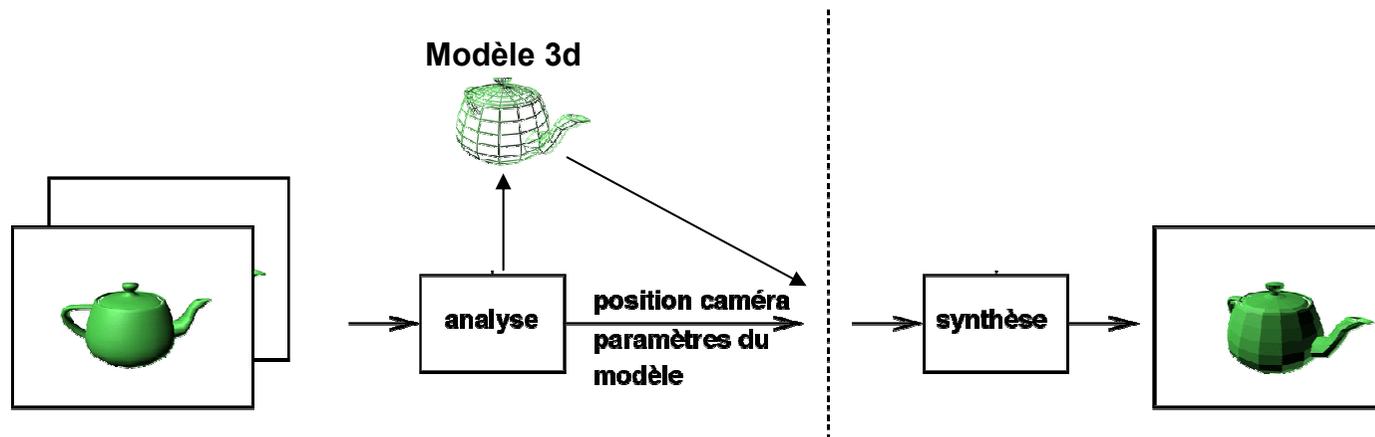
III - Modélisation 3D de séquences vidéo

plan de la partie

III - Modélisation 3D de séquences vidéo

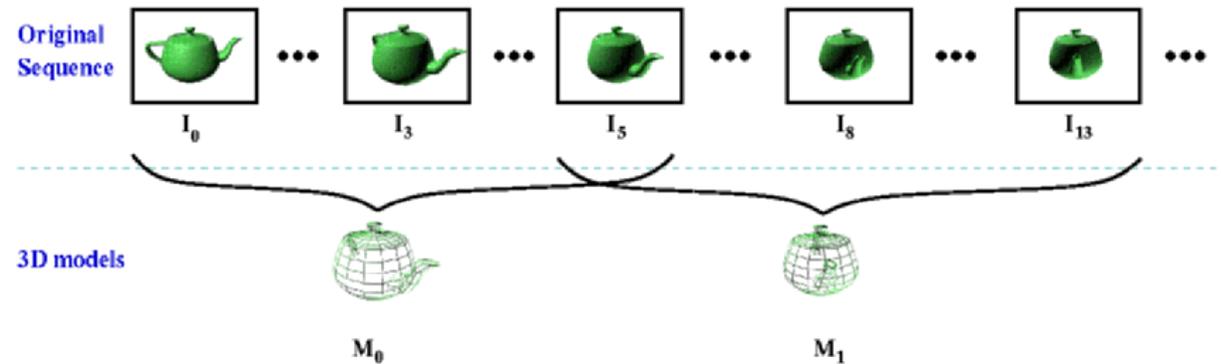
1. Représentation proposée
2. Estimation des modèles 3D
3. Amélioration du rendu visuel
4. Application à la compression vidéo
5. Application à la réalité virtuelle
6. Conclusion et perspectives

Rappel : objectif



1- Représentation proposée

Flux de modèles 3D



- 1 modèle 3D valide pour 1 groupe d'images (GOP)
- 2 GOPs successifs partagent 1 image clef K
- + : longues séquences, traitement au flot, gestion des dérives
- - : modèles 3D moins réalistes

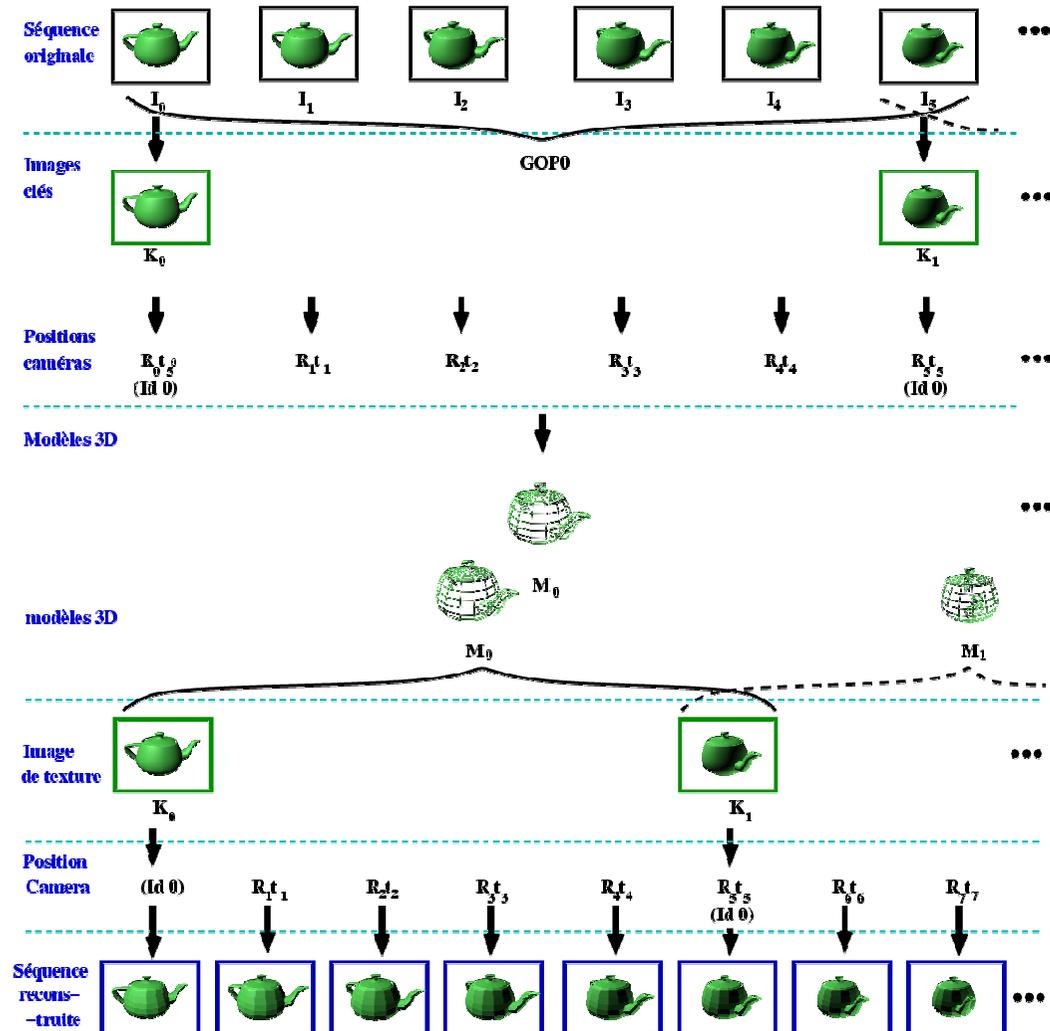
1- Représentation proposée

Schéma de type Analyse-Synthèse

- Analyse :
vidéo → modèles 3D

- Transmission :
 - modèles 3D
 - images clés
 - positions caméra

- Synthèse :
modèles 3D → vidéo reconstruite



1- Représentation proposée

Contraintes et hypothèses de travail

Contraintes du codage :

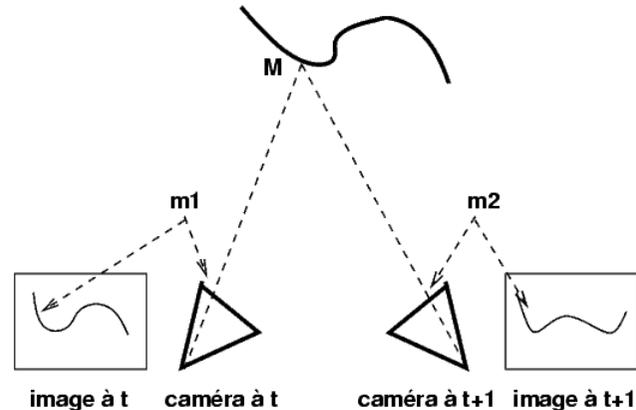
- qualité de reconstruction des images
- pas d'a priori sur l'acquisition (caméra monoculaire non étalonnée)
- pas d'a priori sur le contenu
- "tout automatique"

Contraintes de la vision par ordinateur :

- mouvement caméra non dégénéré
- scène statique ou segmentée
- instabilité numérique (problème inverse)

2- Estimation des modèles 3D

Contrainte de codage \Rightarrow stéréovision avec caméra monoculaire



- modèle sténopé :
 - $m_1 = K_1(l_3|0)M$
 - $m_2 = K_2(R|t)M$
- étapes de l'approche classique :
 - \Rightarrow mise en correspondance de m_1 et m_2
 - \Rightarrow estimation de K_i (auto-étalonnage)
 - \Rightarrow estimation (R,t) (calibration)
 - \Rightarrow estimation de M (rétroprojection)

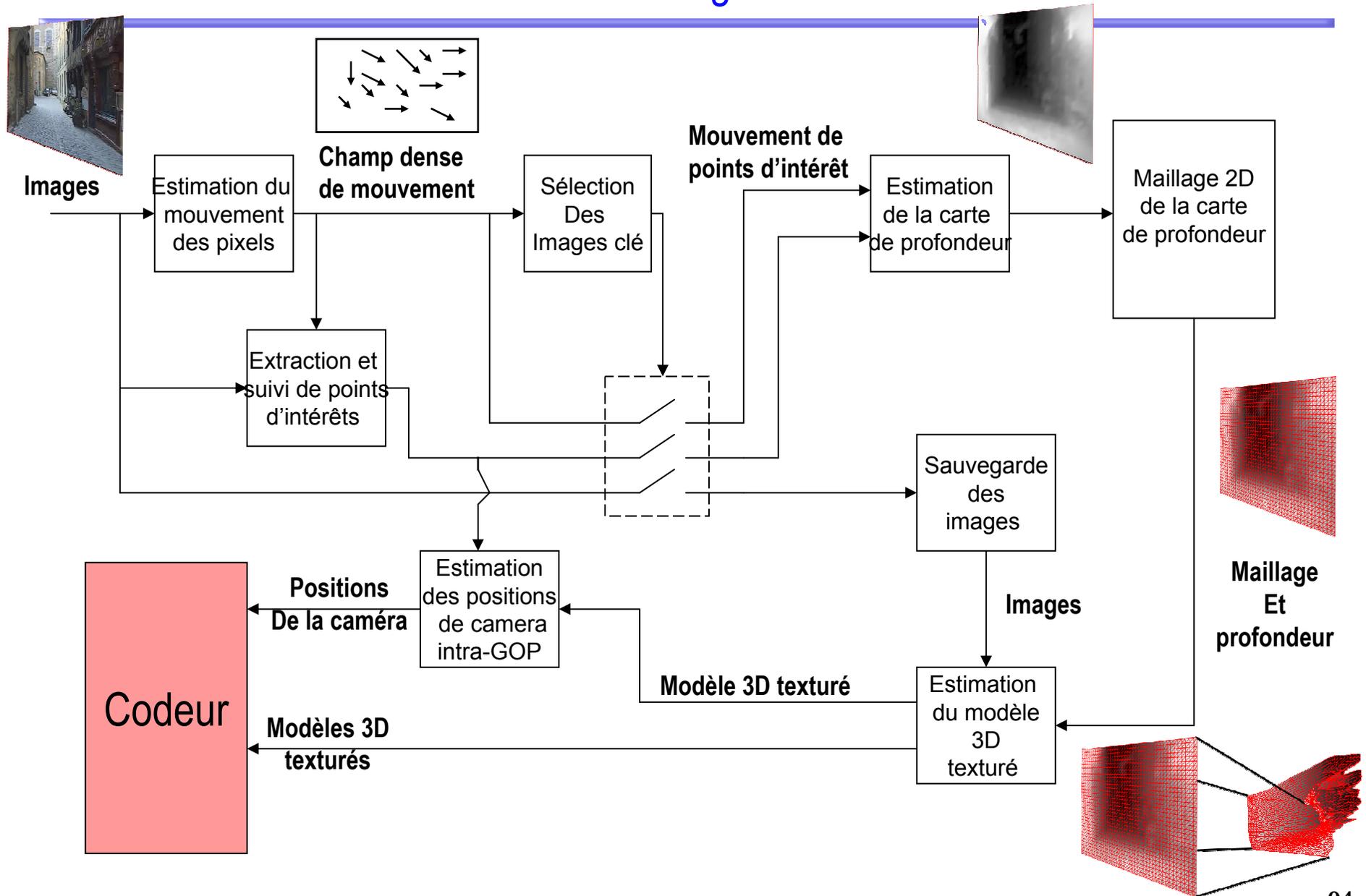
2- Estimation des modèles 3D

Outils algorithmiques

- Mise-en-correspondance dense:
 - estimation de mouvement par maillage déformable [Marquant, Pateux]
- Estimation du mouvement caméra :
 - paramètres intrinsèques : valeurs fixées
 - images clés : suivi de points + décomposition de la matrice essentielle
 - images intra-gop : estimation de pose
- Modèle 3D : carte de profondeur + maillage triangulaire uniforme
- Cohérence entre modèles successifs : ajustement glissant
- Sélection des images clés : critères 2D et 3D

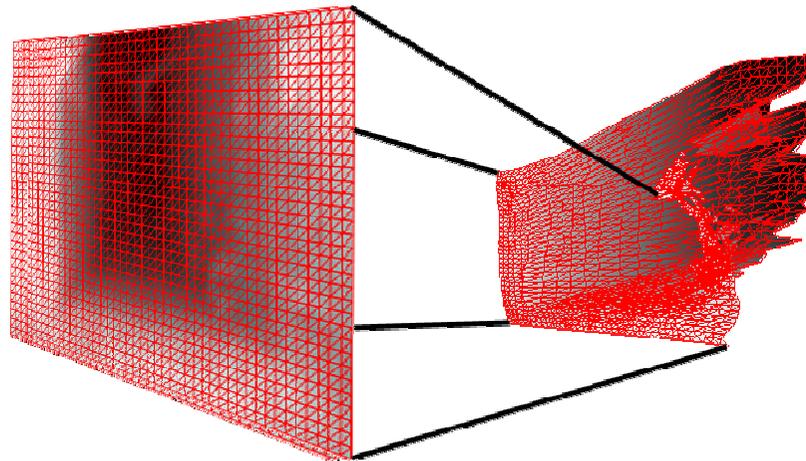
2- Estimation des modèles 3D

Schéma global



2- Estimation des modèles 3D

Obtention du maillage 3D



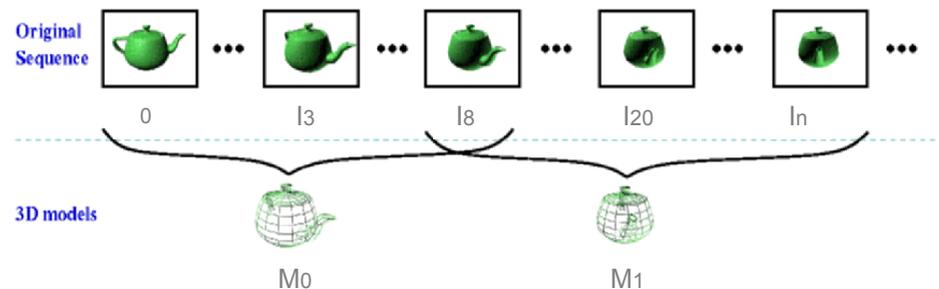
Principe :

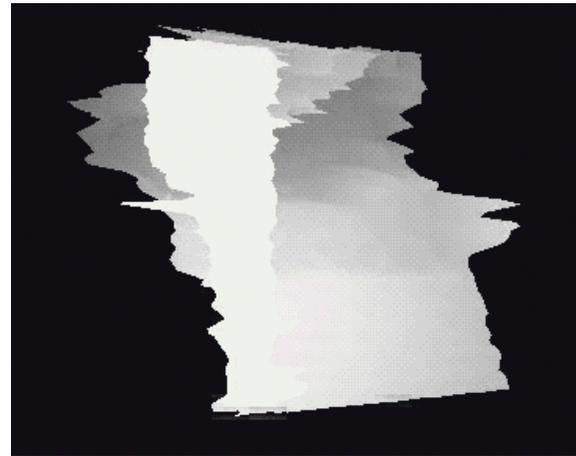
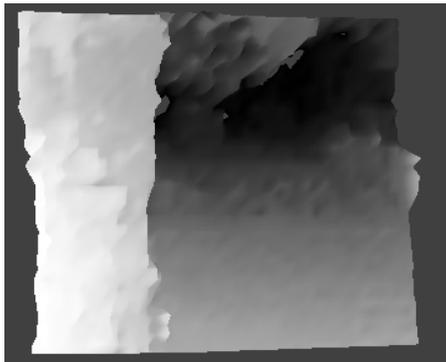
- champ dense $C_{n,n+1} + (A, R_n, t_n, R_{n+1}, t_{n+1}) \Rightarrow$ carte de profondeur Z_n
- maillage uniforme de Z_n
- maillage 3D par facettes triangulaires
- \Rightarrow compact (codage)
- \Rightarrow visualisation par moteur de rendu 3D

2- Estimation des modèles 3D

Sélection des images clés

- Découpage de la vidéo en GOPs de taille variable, délimités par des images clés
 - Sélection des images clés
 - automatique
 - adaptée au contenu vidéo
 - Critères de sélection
 - amplitude mouvement 2D global : disparité suffisante pour permettre la reconstruction 3D
 - pourcentage de points sortants : portion de champ de vue commun
 - résidu épipolaire : validité du mouvement 3D estimé
- Un groupe d'images (GOP) est représenté par :
 - un maillage 3D
 - une image de texture (l'image clé de début de GOP)
 - un couple (R,t) : orientation, position de la caméra pour chaque image





modèle 3D n°0



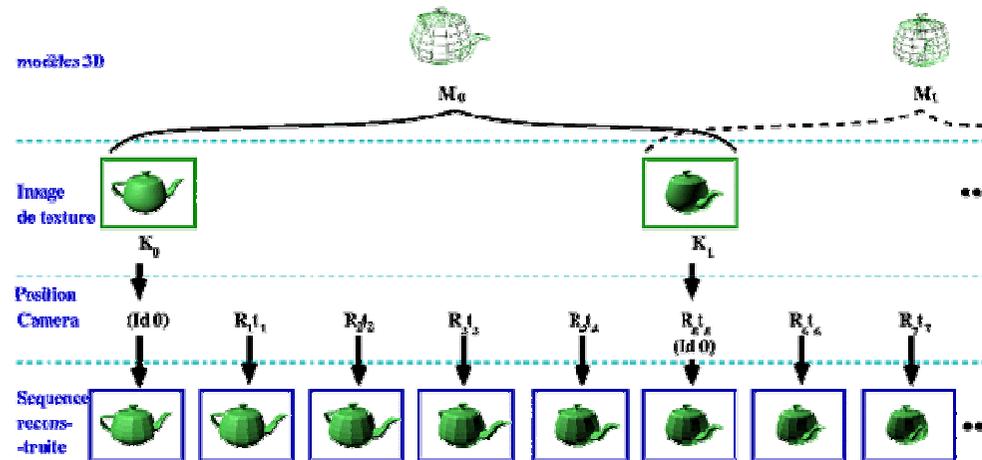
modèle 3D texturé n°0

Modèle 3D extrait de la séquence *escalier*

- ✓ géométrie générale bonne
- ✓ modèle 3D satisfaisant
- ✗ distortions visuelles (étirement des textures etc.)

3 – Amélioration du rendu visuel

Reconstruction de la séquence originale



Principe

- synthèse : projection des modèles 3D texturés successifs

Problèmes :

- saut visuel d'un GOP à l'autre
- problèmes des zones entrantes

Solutions proposées :

- *morphing* approché : *fading 3D* des modèles 3D
- mosaïque 3D de modèles successifs

3 – Amélioration du rendu visuel

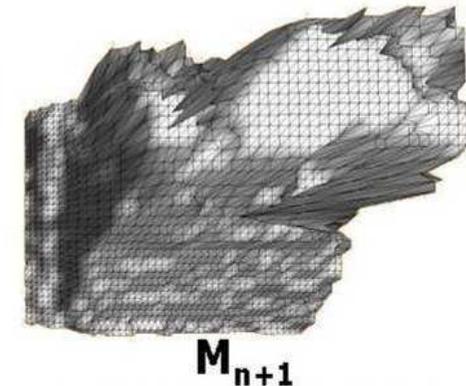
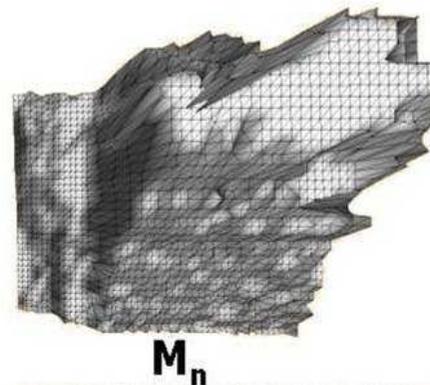
Problème : sauts visuels aux changements de GOP

En théorie :

- image clé commune assure continuité visuelle

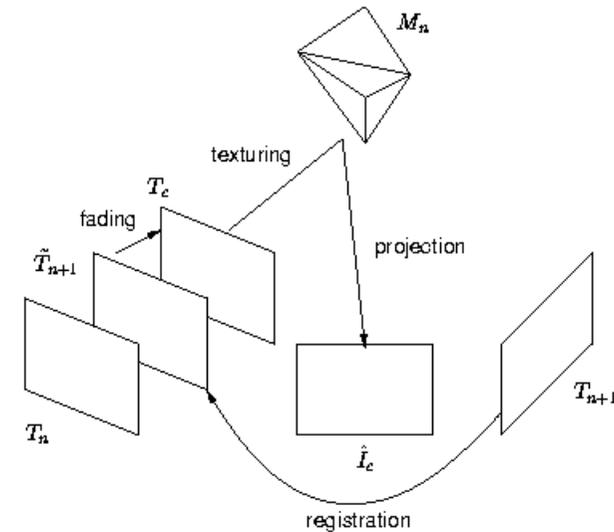
En pratique :

- qualité décroissante et transition abrupte
- changements de texture, géométrie et connectivité
- zones entrantes et zones découvertes



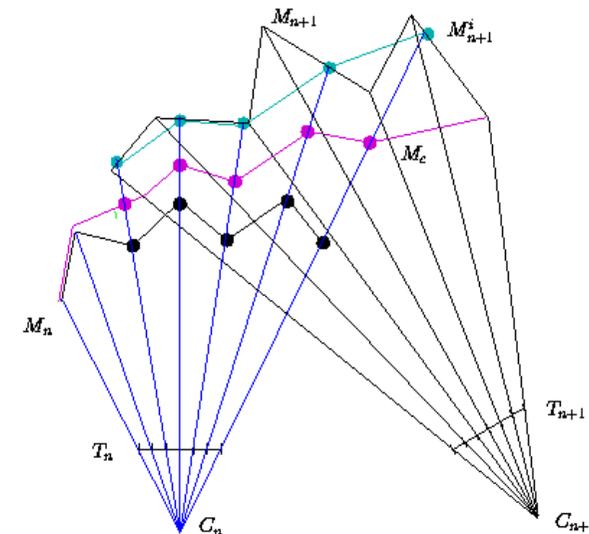
- Fondu de texture

- Création d'une image de texture spécifique pour chaque image du GOP
- Combinaison linéaire des images clés recalées
- Pondération par le rapport des translations caméra



- Métamorphose 3D (morphing)

- Création d'une carte de profondeur spécifique pour chaque image du GOP
- Interpolation linéaire de la profondeur le long des lignes de vue
- Remaillage aux changements de GOP



3 – Amélioration du rendu visuel



sans fondu de texture



avec fondu de texture



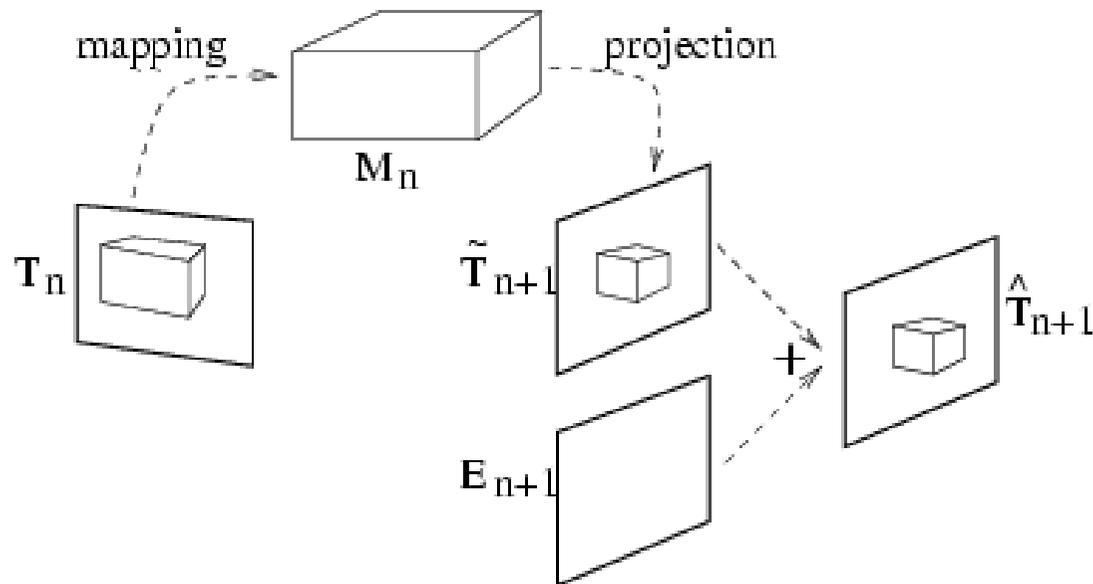
2 modèles 3D superposés pour une vue virtuelle
avec / sans métamorphose

Principe

- analogie avec les codeurs standard : schéma IPP
- pour chaque GOP :
 - un modèle 3D (géométrie)
 - une image d'erreur de prédiction
 - une position/orientation caméra pour chaque image

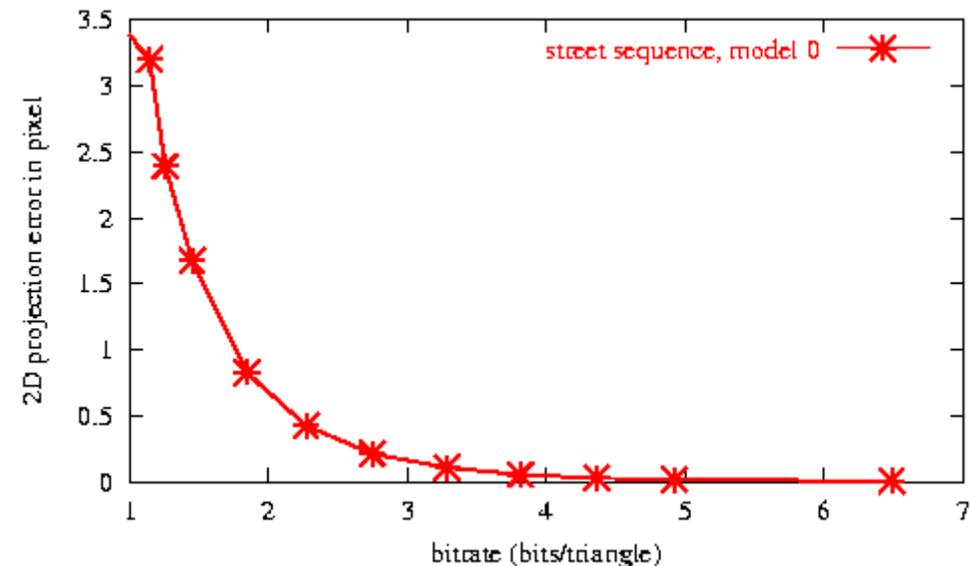
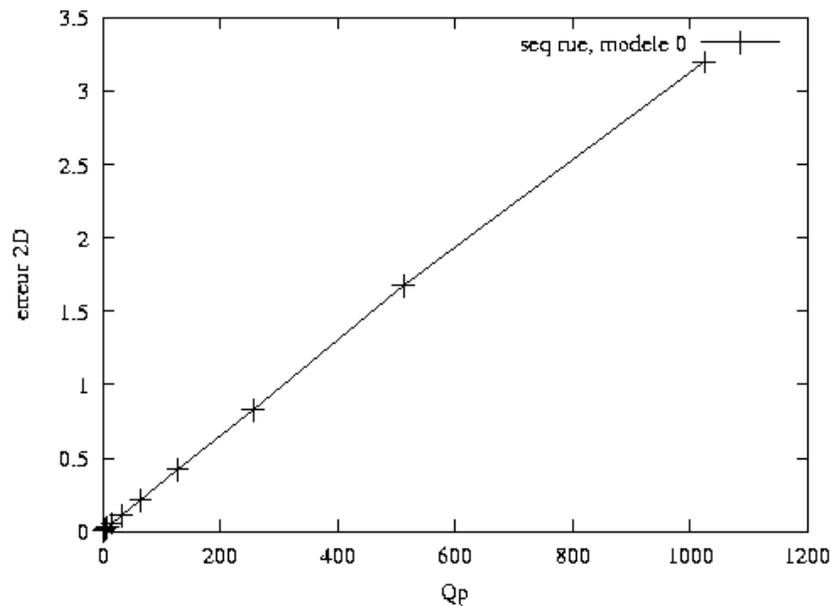
Image	I_0 =image clef 0	I_1	...	I_{40}	I_{41} =image clef 1	I_{42} ...
Modèle 3D						
Texture	 image 0				 image erreur	
Position caméra	R,t	dR,dt	...	dR,dt	dR,dt	$dR,dt \dots$
Type d'image	I	B	...	B	P	B ...

- Compression des images de texture
 - prédiction via le modèle 3D précédent
 - codage scalable du résidu de prédiction (JPEG2000)

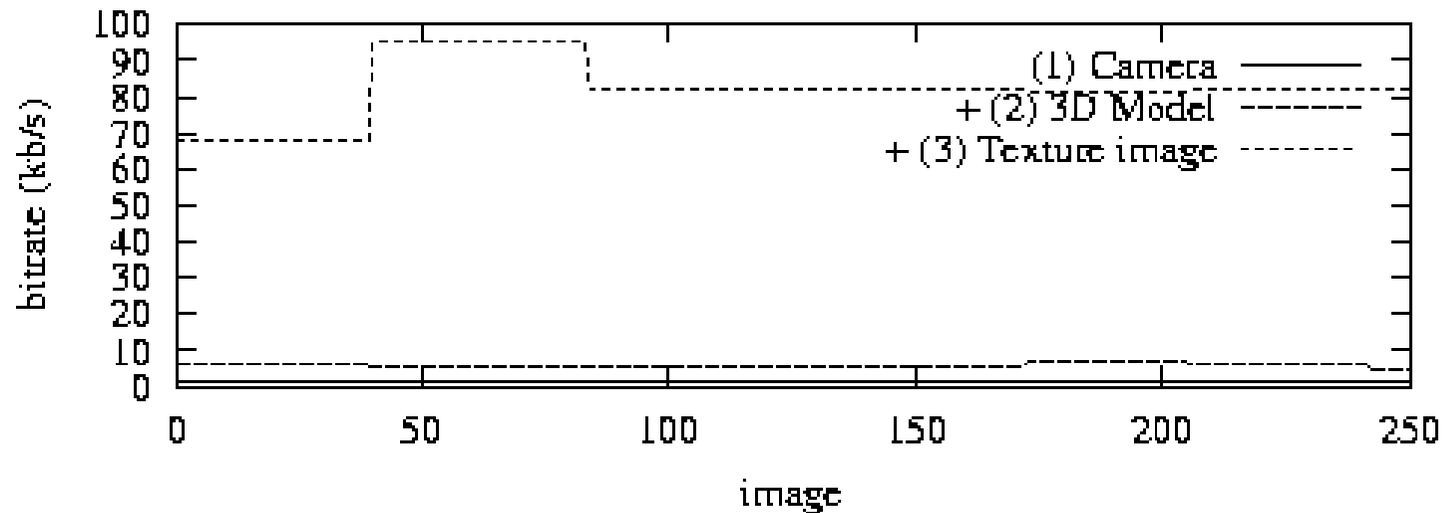


5 - Application à la compression vidéo

- Compression des modèles 3D
 - Triangulation uniforme de la carte de profondeur
 - Quantification uniforme de la profondeur inverse de chaque sommet
=> erreur de reprojection linéaire en fonction du pas de quantif.
 - Codage scalable par ondelettes (JPEG2000)



Répartition du débit



- Texture : information la plus volumineuse
- Adaptation au débit cible :
 - Sélection d'un pas d'échantillonnage pour le maillage uniforme
 - Quantification des profondeurs tq erreur de projection subpixel
 - Codage progressif FGS de la texture

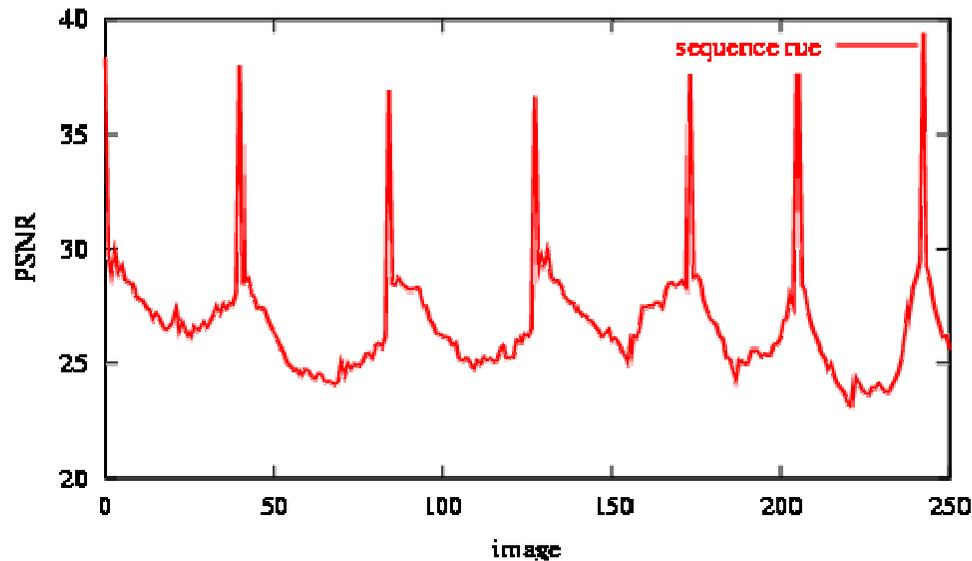


image 59, PSNR=24.5dB

Résultats obtenus :

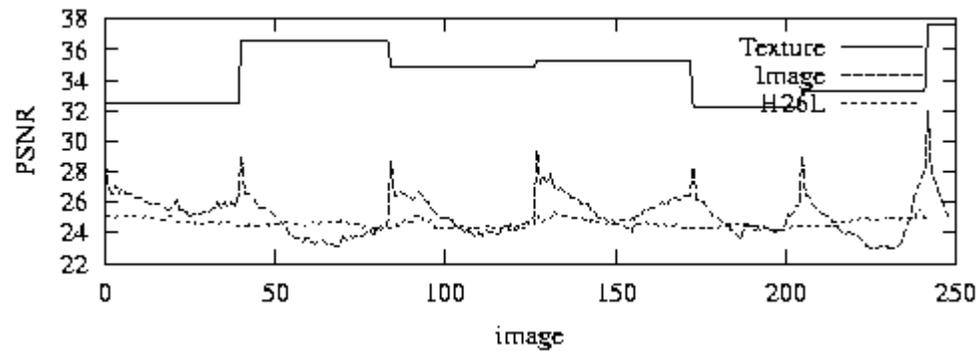
- ✓ qualité objective (PSNR) en dents de scie :
 - ✓ PSNR élevé pour les images clés
 - ✓ PSNR faible pour les images intermédiaires
- ✓ qualité visuelle : bonne pour toutes les images de la séquence



H264 82kb/s, CIF, 25Hz



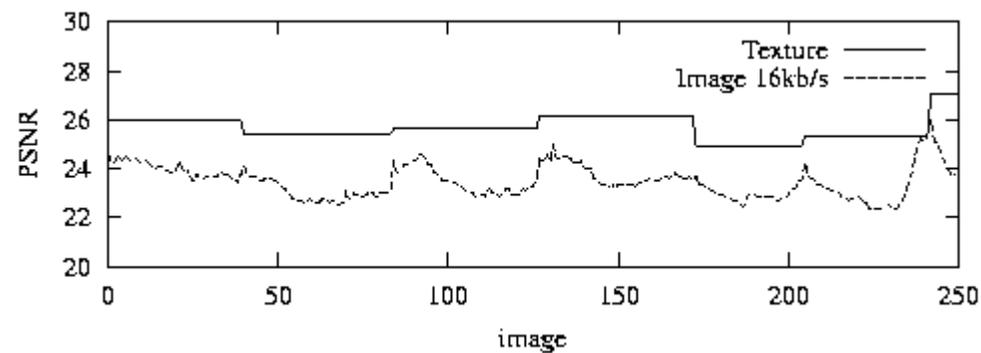
Rec3d 82kb/s, CIF, 25Hz



PSNR , 82kb/s



Rec3d 16kb/s, CIF, 25Hz



PSNR , 16kb/s

Conclusion

Compression à très (très) bas débits (<64kb/s) :

- Débit minimum atteint = 16kb/s pour du CIF 25Hz
- Bonne qualité objective et subjective (pour ce débit)

Comparaison avec un codeur de l'état de l'art (*H264 v6.0*) :

- Meilleure qualité visuelle (plus pour longtemps?)
- PSNR inférieur pour les débits moyens et élevés (>128kb/s)
- PSNR similaire pour les bas débits (<128kb/s)

6 – Application à la réalité virtuelle



séquence *escalier* originale



séquence *escalier* modifiée

- fonctionnalités :
 - stabilisation de trajectoire + manipulation vidéo (incrustation, changement d'illumination...)
 - facilité des manipulations (prise en compte de la géométrie, du mouvement etc.)
 - qualité visuelle satisfaisante en approche automatique

6 – Application à la réalité virtuelle



trajet virtuel



modification d'éclairage



séquence stéréoscopique

- **Autres applications :**
 - visualisation stéréoscopique
 - navigation interactive contrainte/libre
 - player interactif porté en salle Immersia de l'Irisa
 - navigation à distance

Conclusion

- La modélisation 3D : une représentation intéressante des vidéos ...
 - Nouvelles fonctionnalités (réalité augmentée...)
 - Gain en compression à bas débit
 - Permet d'atteindre des débits extrêmement bas
- ... mais qui présente encore des limitations
 - Scènes statiques seulement
 - Validité locale : limite la navigation libre
 - Ensemble discret de modèles : pas satisfaisant pour représenter un scène évoluant continûment
 - Scalabilité limitée : sur la texture et la profondeur, scalabilité au codeur

Conclusion

Travaux en cours et futurs

- Lever les limitations sur le mouvement caméra et la scène
- Passer à une représentation continue : modèle évolutif
- Encodage progressif des modèles par ondelettes,
- Utilisation de données complémentaires / multimodales (SIG pour la modélisation de scènes urbaines)



Reconstruction 3D

- Reference book regarding 3D reconstruction
[Multiple View Geometry in Computer Vision](#)
 Richard Hartley and Andrew Zisserman
<http://www.robots.ox.ac.uk/~vgg/hzbook/>
 Andrew Zisserman 's slides
<http://www.robots.ox.ac.uk/~az/lectures/cv/index.html>
- Many vision algorithms and their implementation
 An invitation of 3-D vision
 Ma et al.
<http://vision.ucla.edu/MASKS/>
- For the math lovers
 The geometry of multiple images
 Faugeras & Luong
<http://mitpress.mit.edu/catalog/item/default.asp?tid=4147&ttype=2>
- Many ressources about computer vision (links, on-line books, software...)
 CV-ONLINE
<http://homepages.inf.ed.ac.uk/rbf/CVonline/CVentry.htm>
- CVHP (The computer vision home page)
<http://www.cs.cmu.edu/~cil/vision.html>

Compression

- Techniques de compression des images
 Jean-Paul Gullois, Hermes, 1996
- VcDemo : logiciel de démonstration pédagogique de codeurs vidéo
<http://ict.ewi.tudelft.nl/index.php?Itemid=124>
- Multivue, TV3D
<https://www.3dtv-research.org/>
- Navigation interactive 3D à distance dans la ville de Rennes (société Archividéo)
<http://www.citevisions.rennes.fr>
<http://www.archivideo.fr/>

Remerciements

Doctorants:

- **Lionel Oisel, Franck Galpin, Raphaèle Balter, Gaël Sourimant**

Stagiaires:

- **Philippe Coval** (interactive player for 3D representation)
- **Eric Morillon, Gaël Sourimant** (hybrid 2D/3D representation)
- **Benjamin Leguen** (3D morphing and representation via surface parameterization)
- **Jonathan Delhumeau** (rigid motion segmentation according to epipolar geometry)
- **Matthieu Maitre** (using 3D modelling for Distributed Source Coding)

Collaborations :

- **INRIA ACI Télégéo**
- **RNRT project V2NET**
- **France Telecom R&D**
- **European NOE : SIMILAR**