

Résumé : L'indexation supervisée, nécessaire pour atteindre un document multimédia, se traduit par un **lourd temps de calcul**. Une grande partie des données étant déjà codée au format **MPEG 1&2**, nous allons directement **utiliser ce flux** en nous limitant au **décodage entropique** et à la **quantification inverse**. Dans un premier temps, nous décrivons la méthode de segmentation utilisant les informations de couleur et de mouvement, en nous limitant aux **termes de basse fréquence de la DCT** et aux **vecteurs mouvements** directement donnés dans le flux. Puis, nous enchaînons par le calcul du **mouvement de la caméra** et des objets, ce qui permet l'indexation (**détermination de champs MPEG7**).

PROBLEMATIQUE vers une indexation MPEG7 :

Existant :

- calcul du mouvement de la caméra
- segmentation des objets mobiles
- Opération gourmande en temps**

Calcul des champs MPEG 7, i.e. pour chaque image (ou groupe d'images) :

n° image : i | Durée : n images | zoom - rotation - pan - tilt

Grande quantité d'information au format MPEG1 & 2

Utilisation de l'analyse de la séquence effectuée lors de l'encodage (estimation du mouvement et valeurs DCT)

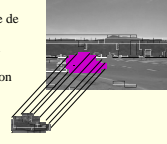
Pouvoir extraire une information :

Pertinente | Non supervisée | Rapide



Notre contribution :

- Utilisation des vecteurs mouvements donnés directement par le flux MPEG 1&2 pour estimer le modèle de mouvement de la caméra
- évaluation de la pertinence des vecteurs mouvement en utilisant les images d'erreur
- Utilisation des moyennes des blocs de luminance et des macroblocs de chrominance pour la segmentation des zones d'intérêts
- vers une détection des objets en mouvement



MPEG1 utilise les redondances :

- Subjectives (luminance / chrominances)
- Temporelles (corrélation entre images)
- Spatiales (plage uniforme - DCT)
- Statistiques (codage entropique)

Rappel norme 1 & 2

4 blocs de Luminance

4 blocs de chrominance bleue (Cb)

4 blocs de chrominance rouge (Cr)

Chaque bloc contient 8x8 pixels

Un macrobloc est composé de quatre blocs adjacents

Décomposition de la séquence en groupe d'images :

- prédic. avant
- prédic. arrière

groupe d'images

Prédiction du mouvement par macrobloc pour les Prédites (P) et les Bidirectionnelles (B)

- découpage en macroblocs 16x16 de l'image à prédire
- recherche d'un macrobloc similaire dans une image déjà codée

P :

P ou I précédente

P à coder

B :

P ou I précédente

B à coder

P ou I suivante

*** Segmentation des zones de couleur uniforme :**

Pour l'axe x, nous avons :

$$L = \{Lum, Cr, Cb\}$$

$$dist_{i,j}^x(f) = \frac{\sum_{l \in L} \lambda_l \left(\frac{f_{i,j}^l - f_{i,j+1}^l}{moy_i^l} \right)^2}{\sum_{l \in L} \lambda_l}$$

- $f_{i,j}^l$ et $f_{i,j+1}^l$: valeur moyenne, de type l, des deux blocs voisins sur l'axe x.
- moy_i^l : valeur moyenne de l'image de type l.
- λ_l : poids fixé empiriquement.

Les deux blocs n'appartiennent pas à la même zone d'intérêt

*** Utilisation des vecteurs mouvement pour recoller les zones de couleur uniforme :**

- MPEG 1&2 contient des informations sur le mouvement des macroblocs
- le mouvement maximum de recherche n'est pas connu lors du décodage
- ce mouvement est cantonné à un voisinage de recherche $\leq \pm 7$ pixels

Si recherche à ± 2 pixels

Si recherche à ± 7 pixels

*** Qualité des résultats proportionnelle à la grandeur du voisinage de la recherche :**

zone_recherche $\leq \pm 2$

zone_recherche $\leq \pm 7$

*** Utilisation de l'image d'erreur pour la pertinence du vecteur mouvement :**

Plus la valeur DC est grande

Plus le nombre des éléments de la DCT sont importants

Plus la valeur de ces dernières est grande

Moins le vecteur mouvement donné dans le flux est pertinent

*** Différence de segmentation entre les B et les P :**

Meilleurs résultats sur les B que sur les P.

Données utilisées

Pour un macrobloc de I :

4 blocs de luminance

4 blocs de chrominance bleue (Cb)

4 blocs de chrominance rouge (Cr)

Moyennes de chaque bloc (valeurs DC de la DCT à une constante multiplicative près).

Travail sur des images "mosaïque" :

Pour les images P et B :

- Vecteurs mouvements pour chaque macrobloc
- Image d'erreur

Il faut reconstruire la valeur des blocs et macroblocs grâce à la prédiction de mouvement et l'image d'erreur données dans le flux.

Pour un macrobloc de P, par exemple :

Macrobloc dans le P ou I précédent

Moyennes du macrobloc à prédire

Prédiction

Erreur

Mouvement de la caméra

Translations possibles :

Rotations possibles :

Plus le Zoom : R_{zoom}

3 translations

3 rotations

le zoom

Utilisation d'une caméra à sténopé $\frac{f}{Z} = cte$

Modèle orthographique

Entre deux images consécutives :

- T_x et R_y ou T_y et R_x jouent le même rôle
- R_{zoom} et T_z jouent le même rôle

$$U_x = -cte(T_x - xT_z) + yR_z$$

$$U_y = -cte(T_y - yT_z) - xR_z$$

En calculant les dérivées, nous trouvons un système de 4 équations à 4 inconnues

On estime donc :

- le pan R_x (équivalent à T_x)
- le tilt R_y (équivalent à T_y)
- le zoom R_{zoom} (équivalent à T_z)
- la rotation R_z

Remplissage du champ : `SegmentedCameraMotion_Info[1]`

n° image : i | Durée : 1 image | zoom - rotation - pan - tilt

Mouvement de la caméra

Translations possibles :

Rotations possibles :

Plus le Zoom : R_{zoom}

3 translations

3 rotations

le zoom

Utilisation d'une caméra à sténopé $\frac{f}{Z} = cte$

Modèle orthographique

Entre deux images consécutives :

- T_x et R_y ou T_y et R_x jouent le même rôle
- R_{zoom} et T_z jouent le même rôle

$$U_x = -cte(T_x - xT_z) + yR_z$$

$$U_y = -cte(T_y - yT_z) - xR_z$$

En calculant les dérivées, nous trouvons un système de 4 équations à 4 inconnues

On estime donc :

- le pan R_x (équivalent à T_x)
- le tilt R_y (équivalent à T_y)
- le zoom R_{zoom} (équivalent à T_z)
- la rotation R_z

Remplissage du champ : `SegmentedCameraMotion_Info[1]`

n° image : i | Durée : 1 image | zoom - rotation - pan - tilt

Minimisation des moindres carrés par rapport aux paramètres du modèle de mouvement :

$$J = \sum \left[(U_x(x_i, y_i) - V_{x_i})^2 + (U_y(x_i, y_i) - V_{y_i})^2 \right]$$

RESULTATS

*** Calcul du mouvement apparent de la caméra (sur la séquence Stephan Eidelberg) :**

Pour MPEG 1&2, la recherche du macrobloc similaire ne peut s'effectuer que dans un voisinage maximal de ± 7 pixels. Remarquons ici que le résultat est bon (± 0.5 pixel) pour du "presque temps réel".

Problème autour des images 90 : le Pan proche de 11, ne peut être trouvé. Cela entraîne un calcul faux pour les autres composantes. Pour le zoom et la rotation, la valeur 0.05 représente un déplacement d'un maximum de 0,88 pixel sur les bords de l'image. Cela explique que les deux courbes ne se juxtaposent pas complètement. A ce niveau d'amplitude, la courbe donne une indication de direction plus qu'une valeur exacte.

*** Segmentation des objets en mouvement :**

Les résultats sont meilleurs lorsque le mouvement de la caméra n'est pas trop grand. L'étude de l'image d'erreur, transmise dans le flux, permet de déterminer la validité du mouvement trouvé.

Lorsque le mouvement de la caméra est supérieur au mouvement autorisé, i.e. ± 7 pixels, le programme est capable de le déterminer et de ne donner aucune réponse. Il faut donc faire une étude précise avec un calcul de prédiction de mouvement par bloc ayant une région de recherche plus grand.

Conclusion

La méthode mise en œuvre évite la **décompression totale** du flux MPEG1-2 et utilise le maximum d'informations déjà présentes dans le flux :

- vecteurs mouvement pour les B et les P.
- coefficients de la DCT : dans les I, pour la reconstruction de l'image ; dans les B et les P pour la pertinence des vecteurs mouvement.

Le fait de rester sur la notation de bloc et macrobloc donne des résultats moins précis mais dans un laps de **temps beaucoup plus court** (proche du temps réel). Pour ce qui est du calcul du mouvement de la caméra, les **résultats trouvés sont proches du mouvement connu** des séquences et ce quel que soit le codeur utilisé (± 0.5 pixel pour les translations).

Une amélioration de la méthode serait d'arriver à différencier les sept paramètres (les trois translations, les trois rotations et le zoom), mais pour cela, nous ne pourrions plus considérer une image et sa suivante, mais une suite d'images.