

# Documents image et vidéo

- Une image peut être vue d'un point de vue physique comme un signal 2D continu.
- L'image numérique est donc une grille d'éléments appelés pixels.
- La vidéo, quant à elle, est beaucoup plus complexe d'une image.
- D'un point de vue physique ou (informatique), un document (ou un flux) vidéo

- Un document vidéo présente une combinaison de sous-médias ou pistes organisés suivant un axe temporel.
- Chaque piste est présenté sous la forme d'un flux d'éléments et ces flux sont synchronisés entre eux.
- Ces flux peuvent contenir des images animées, du son(un flux ou une composition de plusieurs flux audio émis en parallèle à une fréquence fixe.

- Pour analyser une vidéo, on doit donc structurer une grande quantité d'informations diverses et hétérogènes.
- Une vidéo peut être décomposée selon différents niveaux de détails.

- La liste des trois niveaux cités n'est pas exhaustive.
- Niveau scène: correspond à un groupe de séquences vidéo qui sont homogènes par rapport à un critère sémantique. Une scène doit respecter trois règles de continuité: en espace, temps et action;

- Niveau plan:correspond à une succession d'images d'une vue continue d'une caméra;
- Niveau image/image-clé:correspnd à l'ensemble d'images représentatif de chaque plan pouvant résumer son contenu;

- Description du document vidéo

- Le temps est un facteur clé dans un document vidéo qui synchronise image et son.
- Le but d'une séquence vidéo est d'ordonner des éléments d'un document vidéo (images, son et textes) selon des règles bien définies.
- L'ensemble de ces séquences vidéo constitue le document vidéo.

- Les descripteurs visuels
- Nous pouvons catégoriser les descripteurs d'un document vidéo selon le type de modalité qu'ils représentent, soit des descripteurs visuels, soit des descripteurs audios, soit des descripteurs de mouvement.

- 1. Les descripteurs de couleurs
- 2. Le descripteur de texture
- 3. Les descripteurs de formes
- 3. Les descripteurs basés sur les points d'intérêt

# Les descripteurs audios

- L'audio constitue une source d'information utile et importante dans le contexte de l'analyse sémantique des vidéos.
- En effet, le fait de détecter à titre d'exemple des cris entre personnes, ou des personnes en train de tirer une balle, ou un sifflet, peuvent aider à détecter les concepts : supporters, tirs de ballon, arrêt du match par l'arbitre, etc.
- Dans une approche de recherche sémantique de vidéo l'analyse de l'audio peut être automatiquement catégorisé en classes sémantiques, comme "supporters", "tirs", "arbitrage".
- Généralement la caractéristique audio est regroupée en deux catégories principales : les coefficients temporels qui sont calculés directement à partir du signal audio et des coefficients de fréquence.

# Les descripteurs de mouvements

- Dans les années 2000, plusieurs approches d'annotation du mouvement ont vu le jour.
- Certaines consistent à calculer les paramètres de mouvement d'un modèle prédéfini par une technique d'estimation globale, comme le modèle de translation, qui a été adopté par le standard MPEG-7 (ISO, 2001).
- (Zeng et al., 2002) présentent une approche appelée "Motion Activity Map (MAM)", elle consiste à générer une image synthétisée qui accumule l'activité du mouvement sur les grilles de l'image dans l'axe du temps, cette image représente les informations du mouvement et conduit à une représentation plus intuitive et plus compacte du mouvement.
- D'autres descripteurs de mouvement ont également été proposés dans la littérature : descripteur d'activité de mouvement (Motion Activity), descripteur du mouvement de caméra (Camera Motion), descripteur WPD (Warping Parameter Descriptor), descripteur du mouvement paramétrique (Parametric Motion), etc.

# Les descripteurs sémantiques

- Sont des descriptions textuelles d'une séquence vidéo générant des annotations vidéo de haut niveau.
- En général, elles sont exprimées en langage naturel et peuvent être retrouvées à l'aide d'un moteur standard de recherche pour le texte.

- Nous pouvons être en face de deux problèmes : d'une part, l'écriture des annotations consomme énormément de temps ; et d'autre part, deux annotations différentes peuvent faire référence à une seule vidéo puisque c'est l'être humain qui annote.

- D'un autre point de vue, la vidéo a une structure qui peut être définie en termes de scènes, de plans, d'images et d'objets. D'après (Chan et al., 2001), un objet vidéo est une instance d'un objet physique capturé dans l'espace physique de la vidéo.
- (Mahdi, 2001) considère l'objet vidéo comme « une région de l'image localisée manuellement ou automatiquement par l'utilisation d'un critère d'homogénéité visuelle ou sémantique ».
- Ceci dit, un objet vidéo peut être vu comme une sous-région d'une seule entité sur une ou plusieurs images appartenant à la vidéo.

.

- En utilisant des techniques de détection assez complexes, des objets comme un ballon, un joueur, etc, peuvent être reconnus automatiquement dans une séquence d'images en mouvement.

- Merci