Nouvelles de l'AS fouille d'images

Émergence de caractéristiques sémantiques

Patrick GROS
Projet TEXMEX
IRISA - UMR 6074, CNRS, université de Rennes 1, INSA Rennes, INRIA

L'AS fouille d'images

- Qu'est-ce qu'une AS ?
  - travail prospectif sur un thème
  - émergence de travaux nouveaux
  - lobbying auprès du DSTIC
  - complémentarité avec les GDR (ISIS)
- Qu'est ce que la fouille d'images ?
  - un équivalent de « indexation d'images »
  - tout technique d'utilisation d'un grand stock d'images
    - où l'on s'intéresse au contenu des images
    - par ex : classification, navigation, recherche...

Mode de travail

- Travail en comité restreint
  - choix de la méthode de travail
  - organisation des réunions plénières
  - organisation d'un workshop à RFIA'04
- Travail public
  - des réunions
  - annoncées sur le site du GDR ISIS
  - ouverte à la participation de tous
- L'idée maîtresse :
  - aux limites de la fouille d'images
  - interaction avec d'autres disciplines

Fouille d'images et perception visuelle

- Rencontre avec la perception visuelle, la psychophysique, la biologie
  - Comment voit-on les images ?
  - ce que l'oeil leur fait subir
  - ce que le cerveau leur fait subir
  - comment modéliser tout cela ?
- Conclusion : comment utiliser ces travaux ?

Fouille d'images, statistiques et apprentissage

- Quelques problèmes rencontres par les traiteurs d'images
  - classifier (clusteriser ?), calculer des distributions, apprendre
- Quelques travaux pour y répondre
  - comment classifier
  - comment apprendre
- Conclusion : il y a encore du travail...

Fouille d'images et bases de données

- Qu'est ce que l'indexation ?
  - pourquoi est-ce dur ?
  - pourquoi ça ne marche pas ?
  - quelles sont les pistes ?
- Et MPEG'7 ?
  - tout ce que MPEG'7 ne fait pas pour vous.
- Indexer du XML ?
  - pourquoi ça ne rentre pas dans les SGBD-R ?
  - pourquoi les SGBD-O ne marchent pas non-plus ?
- Il reste du travail... (ouf !)

Autres journées

- On s'est fait griller (par Bruno B.) sur les usages et les sciences humaines
- Journée état de la recherche le 7/7...
  - sur appel à exposé

Bilan

- Ouverture large indispensable
- Intérêt et difficulté de la pluridisciplinarité
- Les thèmes émergents :
  - comprendre les espaces de grande dimension
  - méthodes statistiques pour grands volumes de données
  - indexation : tout reste à faire
  - finis les histogrammes, vive les applications...

Émergence de caractéristiques sémantiques

- Avertissement :
  - discutons un peu...
  - regardons nous en face,
  - essayons de comprendre et de nous situer...

Qui suis-je ?

- Patrick Gros
  CR CNRS à l'IRISA Rennes (UMR CNRS - INRIA - UR1 - INSA)

- Responsable d'une équipe de recherche (projet INRIA)
  « Techniques d'exploitation des documents multimédias »
  traitement d'images + TAL + bases de données + statistiques

- Des projets en indexation d'images fixes :
  RNRT DIPHONET : vérification du copyright pour les agences de photos
  RIAM ANNAPURNA : gestion de photos personnelles sur magnétoscope numérique
  EUREKA ALIMAGES : indexation de données numériques...

- Des projets en indexation de vidéos :
  RNRT AGIR : Architecture globale pour líndexation et la recherche
  PRIAMM Mediaworks : méthodes hybrides pour l'indexation d'archives télévisuelles
  RIAM FERIA : Framework pour la valorisation des archives télévisuelles...

L'indexation sémantique, ou la quête du Graal

- Ce qu'on cherche : une description des images
  - automatique
  - universelle (vis à vis des utilisateurs, des images, des besoins...)
  - sémantique
- Ce qu'on fait : une description des images
  - automatique
  - pour reconnaître les zèbres et les couchers de soleil
  - dans des bases de 2000 images
- Ce qu'on oublie :
  - décrire ne suffit pas
  - Les bases sont volumineuses
  - l'utilisateur existe, il peut changer d'idées
  - les situations sont variables, multiples, mais précises

Le problème de base : la sémantique

- Essayons nous à un peu de théorie de l'information...
- Ce qui constitue un acte de communication :
  - un émetteur (celui qui prend la photo)
  - un message (la photo)
  - un récepteur (celui qui regarde la photo)
- Ce peut être plus compliqué :
  - le photographe prend la photo
  - la documentaliste la classe
  - la documentaliste la retrouve
  - le journaliste la publie
  - le lecteur la voit

Le problème de base : la sémantique

- La sémantique : le sens échangé entre l'émetteur et le récepteur par le biais du message.
  - Pas de sémantique intrinsèque au message
  - Problème de l'intention de l'émetteur
  - Problème de la subjectivité
  - Problème du contexte de chaque personne
- Si on oublie l'émetteur ?
  - l'utilisateur peut réémettre l'image
    - = réutiliser le message pour une nouvelle communication
  - valeur esthétique de l'image
- Si on oublie aussi le récepteur ?
  - On ne peut que décrire le message (sa syntaxe)
  - On peut injecter des connaissances, de la culture...

Petite caricature du travail en image fixe

- Calcul de descripteurs :
  - variations sur les histogrammes de couleur, la texture et la forme
  - requêtes par l'exemple
  - relevance feedback pour une recherche itérative
- Pas de contexte applicatif :
  - base d'images « généralistes »
  - pas d'utilisateurs
  - pas de contexte de recherche
- Il ne peut y avoir de sémantique
  - des images isolées hors de tout contexte !

D'un point de vue technique

- On décrit :
  - la couleur (dominante, répartition)
  - la texture (est-ce utilisable ?)
  - la forme (mais il faut l'extraire...)
- On détecte et reconnaît :
  - les visages, de face, un peu de profil
  - le corps humain

On extrait :

- des points, des régions, des contours...

Ce qu'on « sait » faire

- Comparer des images et retrouver des copies
  - ◇ recherche de copies sur le web
  - ◇ recherche de fragments
  - ◇ reconnaissance d'objets
    - ⋆ notion de similarité forte
    - ⋆ contexte applicatif bien défini
    - ⋆ pas de sémantique

Ce qu'on « sait » faire

- Classer des images
  - clustering sur critère visuel
  - propagation de mots clés
    - de la sémantique ...
      la forme rejoint le fond en univers fermé (et petit)

Ce qu'on « sait » faire

- Apprentissage de critères pertinents pour une recherche
  - relevance feedback
  - prise en compte d'un « point de vue utilisateur »...
    pour autant que ce point de vue soit visuel

Ce qu'on ne fait pas

- Description de scènes
- Compréhension d'images
- Analyse esthétique

Conclusion

- On est scotché au signal
- Seul le contexte peut le relier à un sens
  - contexte applicatif limité
  - apprentissage (reconnaissance de visages)

Du côté de la vidéo

- Des documents différents
  - plus longs, complets, explicites
  - présence de la parole, de textes
- Ce qu'on fait :
  - de la segmentation : plans, objets mobiles
  - de la segmentation PMB, du locuteur, de la transcription
- Ce qu'on ne fait pas encore :
  - du multimédia
- La sémantique
  - un peu de texte, la parole
  - pas de contexte visuel

La sémantique dans la parole ?

- La sémantique en TAL
  - Approches syntaxiques : ne marchent pas
  - Approches statistiques : sur corpus fermé
- Les grandes questions :
  - utilisation de connaissances, ontologies...
  - apprentissage de ces connaissances

Perspectives

- Images fixes :
  - De très grandes bases
  - Recherches précises ou navigation
  - Qui veut payer pour retrouver les zèbres ?
- Vidéo :
  - du multimédia
  - de la structure
- Des applications :
  - on aide les documentalistes
  - on ne les remplace plus

Perspectives techniques

- Comment prendre en compte les particularités d'un contexte ?
  - intégration de connaissances
  - modélisation de l'utilisateur
  - bases réalistes
- Des applications limitées mais réalistes
  - plutôt que la généralité inutile ?