

# IMAGE ET REALITE VIRTUELLE

## Parole et langage

G. Bailly

Sujet d'examen du 9 février 2006 - 2 heures - Cours et documents autorisés

### 1. Compréhension/critique d'article – 10pts

Vous trouverez ci-joint une photocopie de l'article de S. Hamlaoui et al « Suivi des variations de pose et d'apparence de visages dans des séquences vidéo » publié dans *CORESA* en 2004. Cet article présente les résultats obtenus au laboratoire HEUDIASYC sur le suivi de visages par AAM (« Active Appearance Models », modèles actifs d'apparence proposés initialement par Cootes et collègues). Les questions motivent des commentaires sur des sections bien précises de cet article.

NB : Une grande partie de l'évaluation porte sur vos capacités à exploiter l'information déjà présente dans l'article ainsi que vos connaissances. Privilégiez les réponses courtes dans un premier temps puis commentez dans un deuxième temps (s'il vous en reste !).

- Quelle est la problématique du papier ? En quoi réside la contribution des auteurs par rapport à l'existant ?
- Rappelez brièvement le principe des AAM présenté à la section 2.
  - En quoi consiste une transformation « Procrustéenne généralisée », mentionnée à la fin du premier paragraphe de cette section ?
  - Pourquoi introduire une matrice de normalisation dans la formule (3) ?
- La section 3 donne le principe du suivi par analyse-synthèse
  - Rappelez le principe de ce suivi.
  - Pourquoi la matrice  $R$  est considérée comme fixe ? Quels sont les avantages et inconvénients de cette approximation du Jacobien ?
- La section 5 commente quelques résultats expérimentaux
  - Pourquoi le système original autorise-t-il des configurations de la forme du visage telles que celle figurée sur l'image 15 de la figure 2 ?
  - Ne peut-on pas envisager de substituer au filtre particulière des contraintes plus faciles à implémenter ?
  - Comment évalueriez-vous le système pour avoir une vision plus complète de ses performances ?
  - Quelle méthodologie préconisez-vous afin d'accroître en dynamique la compétence du système ... notamment en suivi non seulement d'expressions mais aussi de mouvements de parole ? Comment augmenter le nombre de paramètres  $c$  ?
- Comment peut-on envisager une extension du modèle présenté à un suivi 3D de visage ? Comment acquérir les données d'apprentissage ? Peut-on envisager de suivre ce visage sans calibrer la caméra a priori ?

### 2. Reconnaissance de la parole – 10pts

Soit un système de dictée vocale utilisant un système de reconnaissance des 10 chiffres. Il doit permettre de dicter un numéro de téléphone pour accéder à un annuaire inverse.

- Quels seront les groupes de chiffres qui engendrent le plus de confusions ?
- Comment le modèle de « langage » peut-il améliorer les résultats ?
- Comment l'accès à l'annuaire inverse permet de mieux contraindre l'espace de recherche ?

Le système est utilisé pour identifier un code confidentiel de 4 chiffres, que les usagers doivent chuchoter.

- Pourquoi le système de reconnaissance initial, entraîné sur des signaux prononcés à voix haute, a-t-il des performances qui s'effondrent?
- Le système est entraîné sur de la parole chuchotée mais les performances sont toujours médiocres. Vous êtes chargé de proposer des solutions. Pourquoi un protocole, où l'identité du demandeur est connue (carte insérée par exemple), permet de réaliser une première amélioration du système ? Quelles technologies vocales complémentaires permettraient d'accroître la robustesse ?
- Quelles technologies complémentaires permettraient d'effectuer une détection parole/non parole plus robuste ? Proposez aussi des technologies plus invasives - mais autorisant la mobilité - nécessitant que l'utilisateur porte des capteurs sur lui.
- Aurait-on intérêt à substituer aux chiffres un code en 4 lettres ? Si oui, à quelles conditions ?