

Reconnaissance de visage par vidéo



Usman Saeed, Jean-Luc Dugelay, Caroline Mallauran
Institut Eurécom

- Introduction.
- Extraction de caractéristique.
- Reconnaissance.
- Résultats.
- Travaux en cours.
- Base de données.
- Reconnaissance Audio Vidéo.
- Conclusion.

Introduction

- Usman Saeed – Pakistanais - 24
- 2000 - 2004
 - BS Computer System Engineering
 - GIK Institute Pakistan
- 2005 – 2006
 - Master IGMMV
 - Université de Nice- Sophia Antipolis
- 2006 – Actuellement
 - Doctorant
 - Eurecom

Introduction

Feature Extraction

Local Features

Mouth

Eye

Global Features

Recognition

Results

Work in Progress

Database

A/V Proposal

Conclusions

Introduction

But du Projet: Reconnaissance de personne par étude comportementale du visage.

Caractéristique: nombreuses variations possibles, de la vitesse avec laquelle une personne parle ou des gestes spécifiques d'œil.

Reconnaissance : utilisation de GMM et classifieurs Bayésiens.

Introduction

Feature Extraction

Local Features

Mouth

Eye

Global Features

Recognition

Results

Work in Progress

Database

A/V Proposal

Conclusions

Introduction

Introduction

Feature Extraction
Local Features
Mouth
Eye
Global Features

Recognition

Results

Work in Progress

Database

A/V Proposal

Conclusions



Extraction de Caractéristique

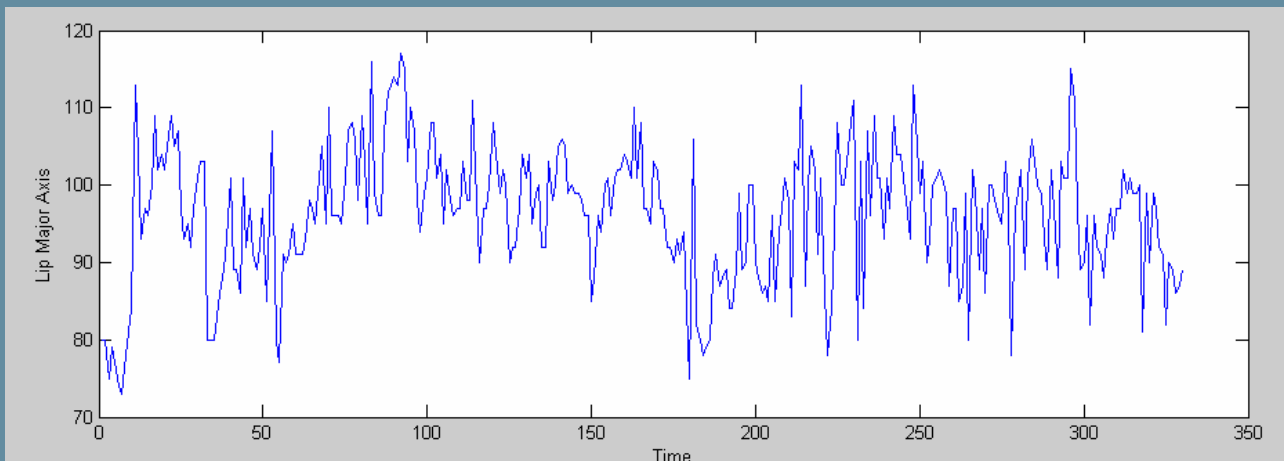
Entrées:

- Séquences vidéos
- Localisation approximative ^[1] des yeux, du nez et de la bouche.

Fonctionnalité: Extraction de caractéristiques

- Caractéristique Local
 - Mouvement de la bouche.
 - Mouvement des yeux.

Sortie: Signal temporel



Introduction

Feature Extraction

Local Features

Mouth

Eye

Global Features

Recognition

Results

Work in Progress

Database

A/V Proposal

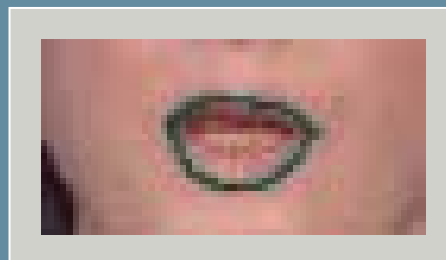
Conclusions

Extraction de Caractéristique

- Mouvement de la bouche
 - Littérature étendue sur la détection et la segmentation des lèvres.
 - Expressions, Reconnaissance Audio Vidéo du locuteur , lecture labiale...
 - Utilisation du bord externe des lèvres comme une mesure de la bouche .
 - Algorithme basé sur la couleur et le contour.

Caractéristique

- Surface
- Axe Majeure
- Axe Mineur
- Excentricité



Introduction

Feature Extraction

Local Features

Mouth

Eye

Global Features

Recognition

Results

Work in Progress

Database

A/V Proposal

Conclusions

- Apprentissage
 - Utilisation des GMM et EM pour enrôler des personnes.
- Reconnaissance
 - Classification Bayésiens
 - Etant donné \mathbf{x}_k le vecteur caractéristique, la probabilité à posteriori pour la classe w_q

$$P(\omega_q | \mathbf{x}_k) = \frac{P(\mathbf{x}_k | \omega_q) P(\omega_q)}{P(\mathbf{x}_k)}$$

Introduction

Feature Extraction

Local Features

Mouth

Eye

Global Features

Recognition

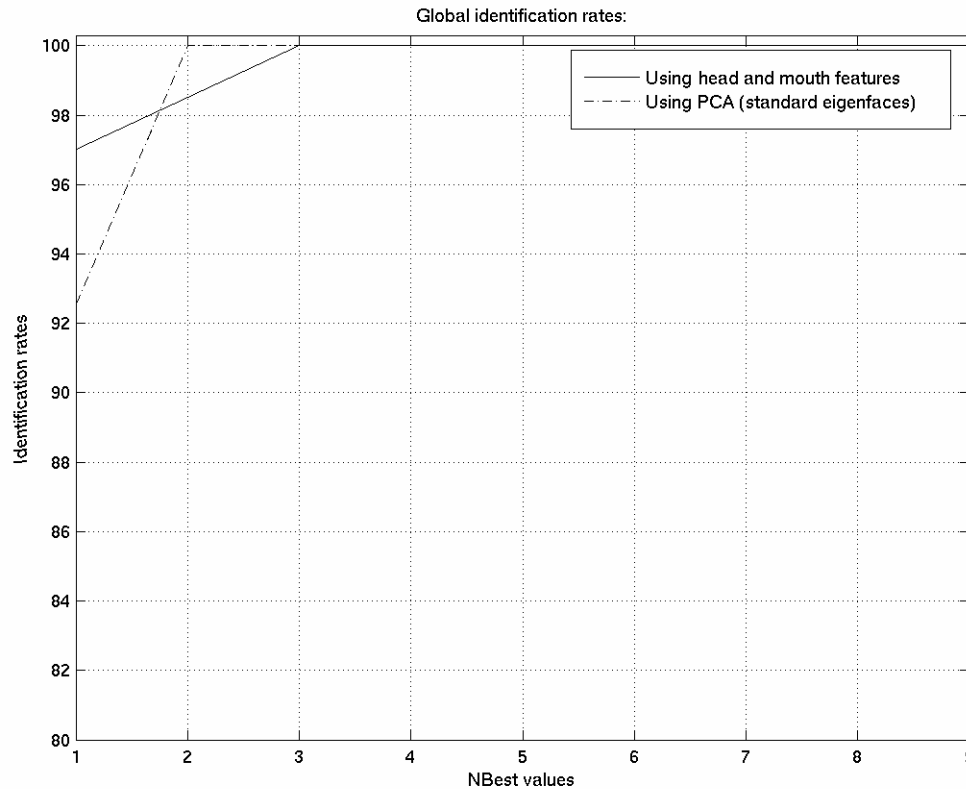
Results

Work in Progress

Database

A/V Proposal

Conclusions



- Par comparaison on a testé sur notre base de données notre algorithme et un algorithme classique, Eigenface.
- Notre Système: Taux d' Identification est 97.0%
- EigenFace: Taux d' Identification est 92.5 %

Introduction

Feature Extraction

Local Features

Mouth

Eye

Global Features

Recognition

Results

Work in Progress

Database

A/V Proposal

Conclusions

- Résultats pas encore finalisés
 - Clignotement des yeux.
 - Mouvement des pupilles.

Introduction

Feature Extraction

Local Features

Mouth

Eye

Global Features

Recognition

Results

Work in Progress

Database

A/V Proposal

Conclusions

Base de données^[1] (Actuelle)

- Pas d' Audio
- 9 présentateurs TV
- 4 mins pour chaque personne.
- 144 séquences de vidéo de durée 14s.

- Résolution Spatiale 352×288 pixels
- Résolution Temporelle 23.97 f/s
- Compression 300 Kb/s
- Sérieux, regardant à la camera



Introduction

Feature Extraction

Local Features

Mouth

Eye

Global Features

Recognition

Results

Work in Progress

Database

A/V Proposal

Conclusions

Base de données (Proposé)

- Contents
 - Enregistrements en mode indépendante du texte.
 - vue frontale du visage.
- Résolution Vidéo -----320X240 ou plus de pixel
- Résolution Temporelle -----25 f/s
- Compression Vidéo -----aucun
- Format Vidéo -----avi
- Couleur -----16 bit ou plus

- Nombre de Personne-----40-60
- Distance entre les yeux-----40-60 pixels
- Longueur requise pour apprentissage -----5 mins
- Longueur requise pour Reconnaissance -----10-15 secs
- Nombre de vidéos par personne-----20

- Camera Spécifique / TV -----?????

Introduction

Feature Extraction

Local Features

Mouth

Eye

Global Features

Recognition

Results

Work in Progress

Database

A/V Proposal

Conclusions

Reconnaissance AV

- Base de données Audio Vidéo .
- Séquences de vidéo de taille fixe .
- Synchronisation de la vidéo et l'audio.
 - Signal qui change avec temps.Vecteur caractéristique extrait à chaque image.
 - Signal Audio ?????
- Conception de Classificateur .
 - Signal d'entrée – Combiné ou séparé.
 - Classificateur
 - si séparé, combiner les résultats au niveau du score ou du rang.
 - Poids de modalité.
- Multi-modalité Vs Détection de Imposteur.
 - Améliorer les résultats avec un système multimodal.
 - Vérification de l'audio par mouvement de la bouche

Introduction

Feature Extraction

Local Features

Mouth

Eye

Global Features

Recognition

Results

Work in Progress

Database

A/V Proposal

Conclusions

Conclusion

- But: Avons-nous assez de variations comportementales entre les personnes pour les identifier.
- Résultats préliminaires prometteurs mais qui valident l'idée, même sur une petite base de données.
- Priorité la plus élevée : Développer une base de données Audio - Vidéo

Introduction

Feature Extraction

Local Features

Mouth

Eye

Global Features

Recognition

Results

Work in Progress

Database

A/V Proposal

Conclusions



Questions /Comments

Bibliography



EURECOM

- [1] F. Matta and J-L. Dugelay, "A behavioural approach to person recognition", to appear in Proceedings of IEEE International Conference on Multimedia and Expo (ICME2006), 9–12 July 2006, Toronto, Canada.
- [2] U. Canzler and T. Dziurzyk, "Extraction of Non Manual Features for Video based Sign Language Recognition", in Proceedings of the IAPR Workshop on Machine Vision Application (MVA2002), 11–13 December 2002 , Nara, Japan, pp. 318–321.