

Action Spécifique Geste et Action : Compte-rendu de la réunion du 25/03/2008

Sur la demande du Comité de Direction du GDR ISIS, la réunion de l'action spécifique "Geste et Action" du GDR ISIS, initialement programmée le 31 janvier 2008, est reportée au mardi 25 mars 2008 dans les locaux de TELECOM ParisTech (ex ENST). Elle fait suite à deux réunions en novembre 2006 et juin 2007 dont les actes sont disponibles en ligne sur http://imtr.ircam.fr/index.php/Geste_et_Action.

Elle a donné lieu à huit exposés sur le visage et geste, dont une présentation invitée par A. Gagalowicz sur les travaux de l'équipe INRIA/MIRAGES.
La journée a été suivie par plus de 48 participants.

Les présentations sont en ligne à l'adresse suivante :
http://imtr.ircam.fr/index.php/AS_Geste_25_mars_2008

Organisateurs : Catherine Achard, Frederic Bevilacqua, Patrick Horain

Programme et résumés

1) Le suivi en temps réel des caractéristiques faciales avec l'algorithme de Lucas-Kanade

F.Abdat, C.Maaoui, and A.Pruski

Laboratoire d'automatique et des systèmes coopératifs, Université de Metz,

Notre étude a pour but de développer un système de reconnaissance d'émotion à partir des expressions faciales. Il est constitué d'une caméra et d'un ordinateur avec un logiciel chargé de reconnaître les émotions. Ce logiciel analyse les mouvements des caractéristiques faciales ; sourcils, lèvres et yeux pour déterminer si une personne est contente, triste, dégoûtée, ou a peur.

Un système d'analyse automatique des expressions faciales s'effectue généralement selon les étapes suivantes :

- Localisation des caractéristiques faciales dans la première image de la séquence vidéo.
- Suivi des caractéristiques dans le reste des images de la séquence,
- Codage et classifications des expressions faciales pour la reconnaissance des émotions.

Dans cette présentation, on va se concentrer sur la partie du suivi des caractéristiques faciales. Pour réaliser cette tâche, on a :

- A. détecté le visage dans la première image de la séquence, en utilisant le détecteur de Viola & Jones qui se base sur les descripteurs de HAAR et des classifieurs en cascade.
- B. localisé les caractéristiques faciales, en localisant d'abord leurs axes horizontaux, puis on délimite la zone de chacun en appliquant un modèle géométrique. Ce dernier suppose que les distances verticales entre les yeux et le nez et entre les yeux et la bouche sont proportionnelles à la distance horizontale entre les deux centres des yeux.
- C. sélectionné les points faciaux caractéristiques selon deux méthodes:
 1. Les points caractéristiques sont distribués d'une façon uniforme.
 2. Les points caractéristiques sont extraits en se basant sur les coefficients d'ondelettes.
- D. Une fois on a sélectionné les points dans la première image, leur position dans les images suivantes est obtenue par l'algorithme de Kanade-Lucas.

2) Suivi de visage temps réel pour l'interaction avec un logiciel ludique / pédagogique

Vincent Courboulay

L3i, laboratoire informatique image interaction, La Rochelle

Introduction

De nombreux algorithmes permettent la détection le suivi et/ou l'estimation de l'orientation d'un visage dans un flux vidéo. Cependant la majorité de ces solutions se concentrent sur une sous partie spécifique du problème (détection, suivi ou estimation de pose) et utilisent la majorité des ressources processeur à la résolution de celui-ci. Dans le cadre du développement de nouvelles méthodes d'interaction avec l'utilisateur d'une application interactive, nous avons été amenés à développer un système de détection / suivi / estimation de l'orientation du visage temps réel économe en ressources processeur. En effet, ce système se devait de pouvoir fonctionner en parallèle avec l'exécution d'une application ludique sans perturber son bon déroulement.

Apport

Le système proposé présente un bon compromis entre robustesse, vitesse d'exécution et précision de l'estimation, permettant son intégration dans des jeux pour une nouvelle forme d'interaction (conduite de voiture, jeux de Tetris) ou pour la détermination de l'attention de l'utilisateur. Ce système ne nécessite pas d'équipement spécifique et effectue ses traitements sur des images issue d'une simple webcam en niveaux de gris. Nous présenterons notre architecture de vision basée sur des algorithmes de traitement d'images simples et rapides (différences de gaussiennes, opérateur de symétrie de Loy-Zelinsky, images intégrales, etc.) ainsi que ses applications.

3) Modèles Actifs d'Apparence Adaptés

R. Segurier

SUPELEC, Campus de Rennes

Nous nous intéressons à la localisation précise des éléments du visage tels que les yeux, le nez et la bouche. Nous nous plaçons dans le cadre d'une acquisition de faible qualité avec une caméra de type webcam. L'alignement de visage dans un environnement non contraint est un enjeu majeur. En effet, pouvoir interagir avec les machines ou pouvoir reconnaître un visage dans un environnement avec un éclairage et des variations d'illumination non contrôlés, des expressions et poses (orientation 3D) du visage variées, est une problématique ayant de multiples applications y compris pour les systèmes embarqués.

Afin d'extraire les traits caractéristiques du visage, nous utilisons les Modèles Actifs d'Apparence (AAM), modèles déformables permettant à la forme et à la texture d'être conjointement synthétisées. Nos travaux consistent à rendre les AAM plus robustes.

Nous proposons, dans un premier temps, un prétraitement basé sur des cartes orientées pour s'affranchir des variations d'éclairage qui rendent instable l'utilisation des AAM lorsque l'illumination est quelconque.

Nous proposons, dans un second temps, de rendre les modèles d'apparence déformable robustes à la variabilité de la base de données. Afin d'obtenir un modèle déformable s'adaptant à la variabilité de l'objet (ses déformations internes, sa variabilité extrinsèque, ses changements d'apparence sous des prises de vues différentes et sous des illuminations différentes), il faut constituer une base de données d'exemples de l'objet contenant cette

variabilité. Malheureusement, plus la variabilité d'une base de données augmente, moins le modèle déformable correspondant a de chance de synthétiser correctement l'objet. C'est pourquoi nous proposons d'adapter le modèle déformable au nouvel objet à analyser grâce à des bases de données et des modèles pré-appris. Notre proposition est à la fois rapide et efficace et permet de traiter un visage inconnu dans un flux vidéo, même lorsque ce visage est animé (variation en pose et en expression).

4) Tracking and Analysis of subtle 3D Facial Expressions

Dimitris Samaras

Stony Brook University

<http://www.cs.sunysb.edu/~ial/>

Facial expressions have been up to recently only studied in the context of basic facial expressions. However a lot of mean is often communicated by subtle variations on expressions, which up to now were hard to capture and analyze. In this talk I will present our work in the last few years on accurately capturing, analyzing and tracking 3D facial expression data, in high temporal and spatial resolution. I will present the 3D capture pipeline as well as an example of facial expression synthesis, using a style/content decomposition. A significant difficulty in building accurate models of subtle expression is caused by the difficulty in registering/ tracking precisely a human face, due to its deformable motions and relative lack of features. I will describe a family of registration methods that we have been exploring based on conformal geometry which allow us to convert the 3D registration problem into a 2D problem and achieve high accuracy registration results.

5) Détermination de la pose du visage à partir d'images de synthèse

K. Bailly, M. Milgram

ISIR, Université Pierre et Marie Curie

Dans cette présentation, nous abordons l'estimation de la pose d'un visage dans une image. La première étape consiste à estimer les paramètres de pose et de forme du visage dans une image de référence. Cette estimation s'effectue en modifiant alternativement ces paramètres pour minimiser l'erreur de reprojection entre les points image et ceux du modèle déformable 3D.

Nous pouvons ainsi générer un grand nombre d'images du modèle de visage obtenu pour différentes orientations et expressions faciales.

Dans une seconde étape en ligne, nous estimons la pose d'un visage dans une image en la comparant aux images de synthèse générées précédemment. Cette mesure de ressemblance s'appuie sur une mesure de distance entre les contours orientés.

Nous présentons d'abord les résultats obtenus pour l'estimation de la pose et de la forme d'un modèle déformable, puis les premiers résultats sur la base de données Pointing'04.

6) Modélisation interactive 3D sans marqueurs de visages à partir de quelques images.

Application à la rotoscopie 3D.

André Galalowicz

INRIA/Mirage

Nous exposerons tout d'abord le principe de la technologie sous-tendant ces travaux (approches à base de modèles et avec feedback). Nous exposerons ensuite notre technique de modélisation utilisant un modèle générique de visage (model-based!) qui est déformé en deux étapes.

Dans la première étape, nous utilisons un ensemble de points caractéristiques du visage pointés manuellement sur le modèle générique ainsi que sur les images. Un algorithme coopératif calibration/reconstruction 3D détermine la position 3D de ces points caractéristiques du visage réel et un algorithme d'interpolation propage ces modifications sur tous les sommets du visage générique.

Dans la deuxième étape, on utilise la mise en correspondance des silhouettes image et modèle pour raffiner la reconstruction du visage avec des courbes, plus riches que les points, mais plus difficile à piloter.

Nous exposerons ensuite brièvement le modèle d'animation du visage et son utilisation dans la technique de poursuite 3D de visage sans marqueurs à partir d'une séquence monoculaire..

7) Suivi d'un corps articulé en 2D avec gestion des auto-occultations

Eric Para¹, Olivier Bernier¹ and Catherine Achard²

1 Orange Labs, France Telecom R&D,

2 ISIR, Université Pierre et Marie Curie

Cette présentation concerne le suivi en temps réel de chaque membre supérieur du corps d'une personne avec gestion de leurs occultations respectives. Pour réaliser ce suivi et initialiser automatiquement la cible, nous utilisons un modèle 2D articulé permettant de prendre en compte la déformabilité du corps humain. Chaque membre rigide est ainsi relié à ses voisins pour créer une chaîne contrainte. Notre méthode est fondée sur les "Pictorial Structures" et décompose la recherche de la meilleure solution en 3 étapes :

- La première consiste à calculer les scores de toutes les hypothèses de position de chaque objet. Une sélection des meilleures d'entre elles est ensuite faite pour limiter les couts de calcul.
- La seconde effectue le calcul des scores de liaison pour chaque paires d'objets liés.
- Enfin, la dernière étape consiste à trouver la meilleure configuration maximisant les scores des objets et de ceux des liens afin de tenir compte de la déformation du modèle.

La description de la méthode est tout d'abord illustrée à travers l'initialisation automatique de notre modèle. Les indices à observer dans l'image doivent être génériques afin de pouvoir traiter n'importe quel utilisateur. Une soustraction du fond, la détection des contours et de la géométrie elliptique du visage sont ainsi utilisés. un algorithme de programmation dynamique est utilisé pour obtenir la meilleure configuration en tenant compte des liens entre les différents membres. Le suivi des membres de la personne est ensuite effectué au long d'une séquence. Elle se fait cette fois-ci par une corrélation SAD des patchs 2D correspondant aux membres extrait lors de l'initialisation automatique. Pour gérer les auto-occultations, les membres occultant sont recherchés en premier et les masques d'occultation alors créés utilisés pour calculer les scores des membres occultés.

La gestion des auto-occultations entre les membres d'une personne permet d'améliorer le suivi dans les cas difficiles d'interactions face à une caméra ou de telles occultations sont nombreuses. Les principales limitations du système actuel portent sur la connaissance à priori de l'ordre de visibilité des membres occultant. Pour réaliser une gestion complète des occultations, l'extension des algorithmes actuels à une gestion de graphes multiples est nécessaire tout en permettant le suivi multi-personnes avec gestion des occultations entre ces

différentes personnes. Enfin, le modèle 2D utilisé limite l'interaction dans le plan. Les perspectives portent donc sur l'élaboration d'un modèle 3D avec utilisation d'une caméra binoculaire pour le calcul de la disparité et donc de la position 3D de chaque membre de la personne [3].

8) Gaussian Process Dynamical Models (GPDM) for Motion Analysis

Zhenbo LI & Patrick HORAIN

Institut TELECOM ; TELECOM & Management SudParis

In this talk, I will present the Gaussian process dynamical models [1] and their applications in human motion analysis. GPDM is a latent variable model for nonlinear time series analysis. It comprises a low-dimensional latent space with associated dynamics, and a map from the latent space to an observation space. GPDM can learn an effective representation of high dimensional nonlinear dynamics data even using small data sets. This is especially useful for human motion analysis as human motions are time related and nonlinear movements in high space. GPDM have been used to define motion styles, thus providing prior models for motion tracking [2]. The probability density function in latent space can help to generate same style motion with variances. The trajectories learned by GPDM in latent space can be regarded as the representation of motions, same type of motions have same kind of trajectory shapes. This can be used for motion recognition.

References

- [1] Wang, J. M., Fleet, D. J., Hertzmann, A. Gaussian Process Dynamical Models for Human Motion. In IEEE Trans. PAMI. February, 2008. pp. 283-298.
- [2] R. Urtasun, D. J. Fleet and P. Fua. Gaussian Process Dynamical Models for 3D people tracking. In Conference on Computer Vision and Pattern Recognition (CVPR) New York, June 2006.