

Reconstruction et Animation de Visage

Charlotte Ghys

15/06/07

Contexte

- ❑ 3ème année de thèse
- ❑ Thèse CIFRE financée par Orange/France Telecom R&D et supervisée par Nikos Paragios (Ecole Centrale Paris) et Bénédicte Bascle (Orange/France Telecom R&D).
- ❑ Sujet : Reconstruction de visage, animation et analyse.

Le Visage en Vision

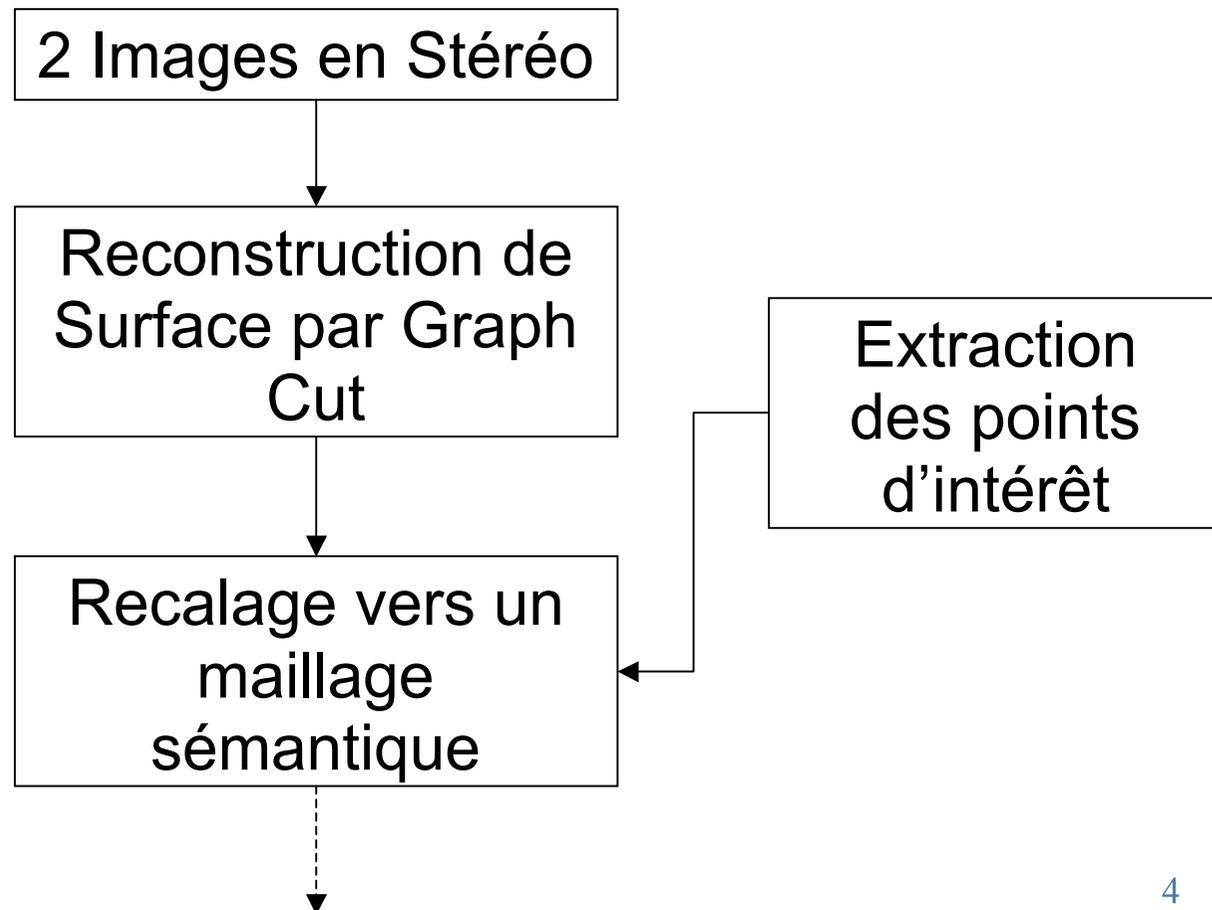
- Pour quoi faire ?
 - ↳ Reconstruction et animation (visio-conférence,...)
 - ↳ Reconnaissance (sécurité,...)
 - ↳ Analyse d'émotions (Interaction Homme Machine, ...)
 - ↳ ...

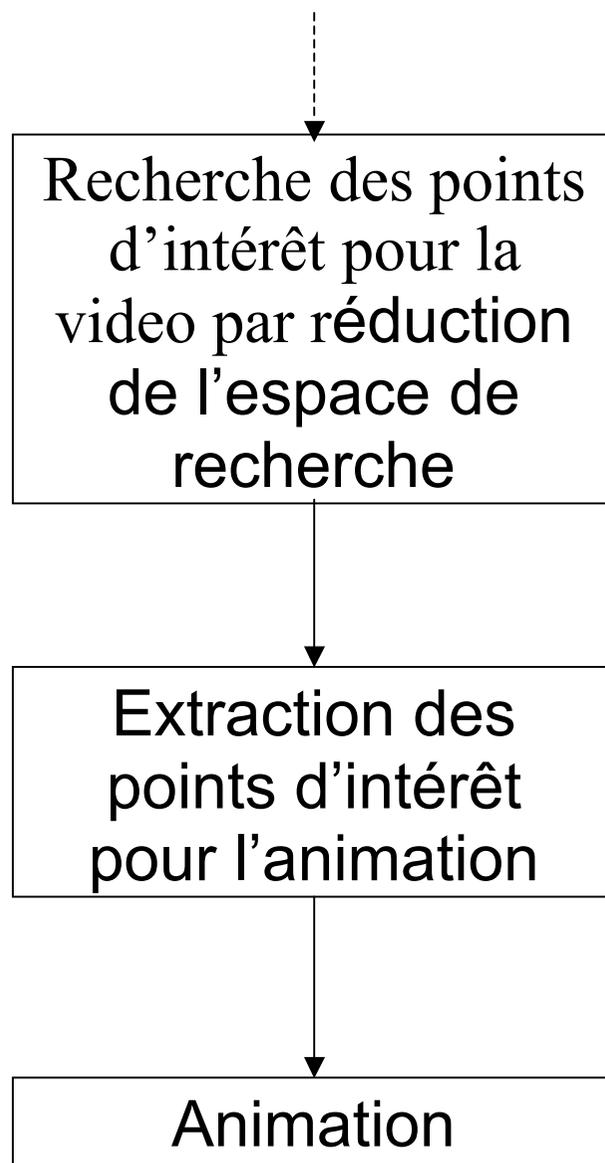
- Comment le faire ?
 - ↳ Stéréoscopie
 - ↳ Apprentissage
 - ↳ Extraction de points d'intérêts.
 - ↳ ...

- Le visage est aussi un sujet très étudié en Graphics.

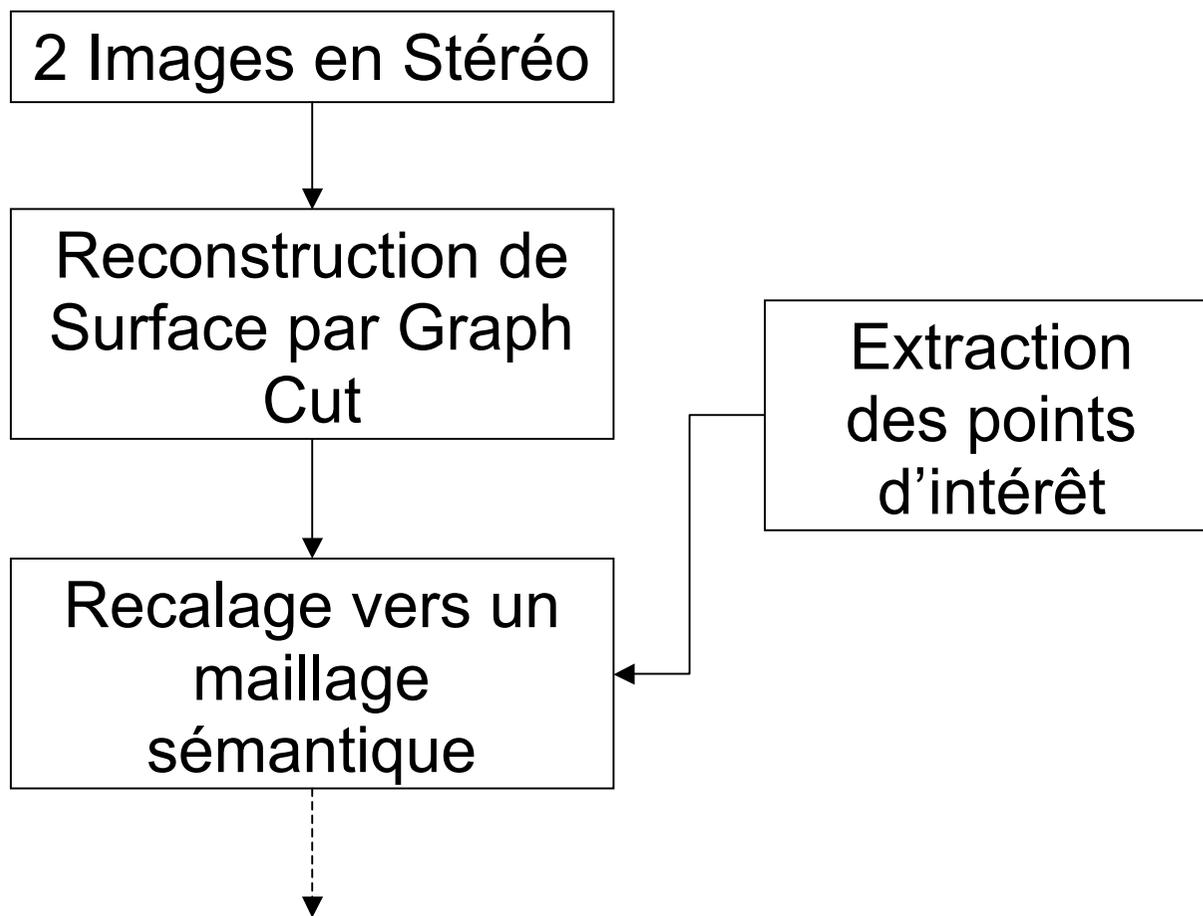
Plan

Objectif : Mimer des expressions en 3D à partir d'une séquence monoculaire.



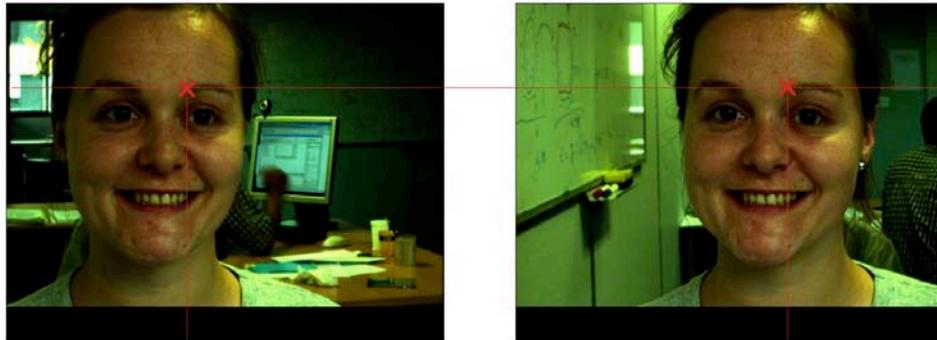


Plan



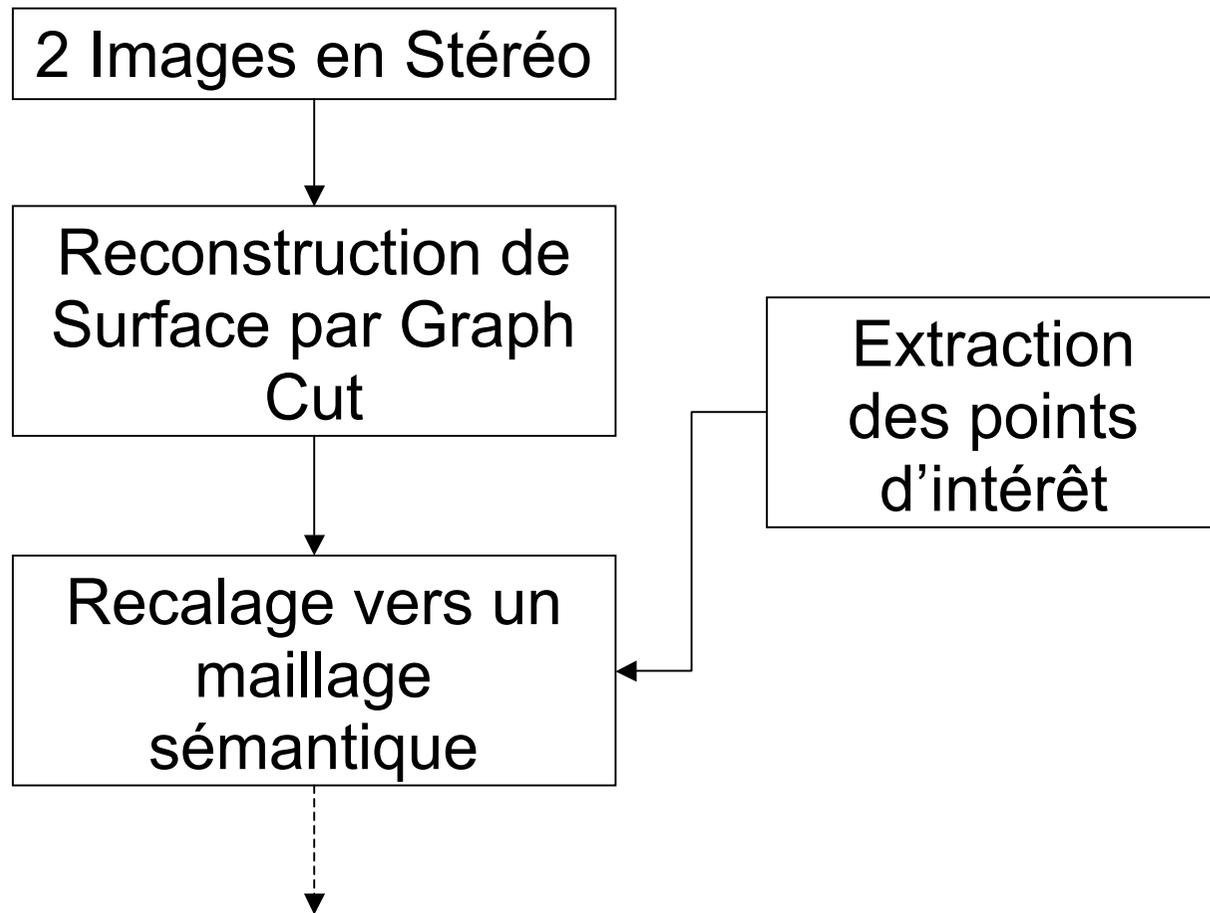
Calibration et Rectification

- Calibration : Calculer les paramètres intrinsèques et extrinsèques des caméras et les relations entre elles (matrice fondamentale).
- Rectification : “Warping” des images telles que 2 points correspondant ont la même coordonnée en ordonnée (en utilisant la matrice fondamentale).
- Trouver les correspondances entre les points devient alors un processus 1D.



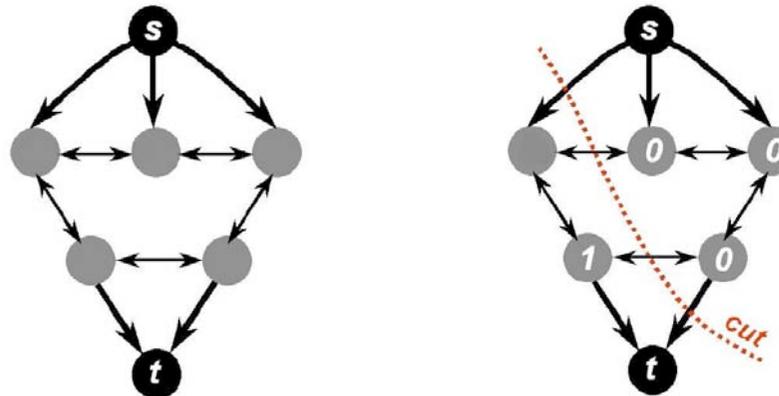
$$\begin{pmatrix} x_2 \\ y_2 \end{pmatrix} = \begin{pmatrix} x_1 + d \\ y_1 \end{pmatrix}$$

Plan



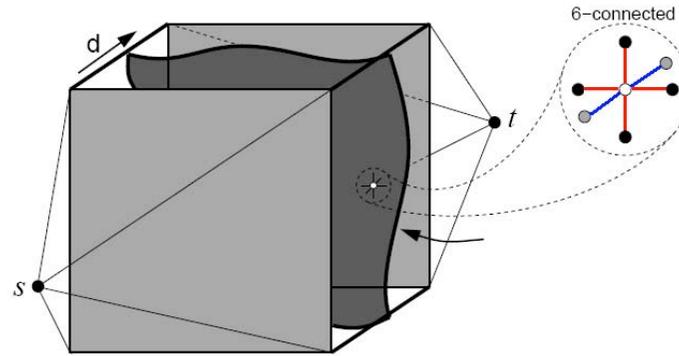
Graph Cut

- Méthode globale et optimale pour minimiser une énergie. [Boykov-Kolmogorov:04]



- Le graph modélise le problème tel que le coût des arcs représente l'énergie à minimiser.
- Une coupe correspond à une configuration parmi toutes les configurations possibles.
- Coupe Minimale=Solution Optimale.

Graph Cut pour la reconstruction 3D

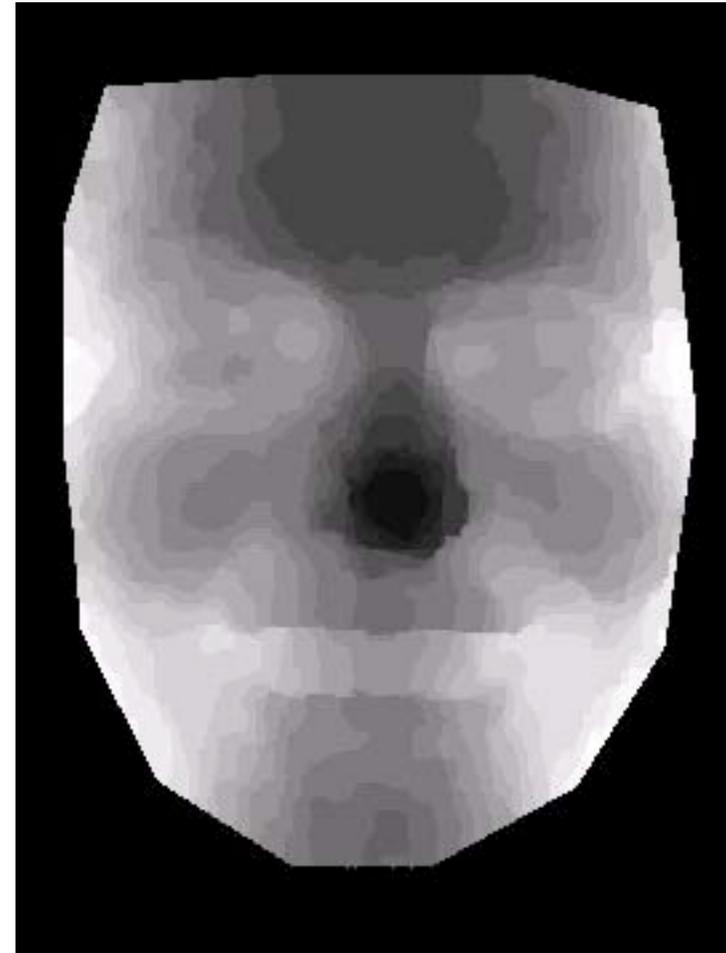


- ❑ Arcs bleus :
 - Terme d'attache aux données basé sur la corrélation entre 2 pixels potentiellement correspondants.
- ❑ Arcs rouges :
 - Terme de lissage, contraignant les pixels voisins à avoir une disparité égale ou proche.

Reconstruction 3D

Pixels clairs -> Disparité élevée.

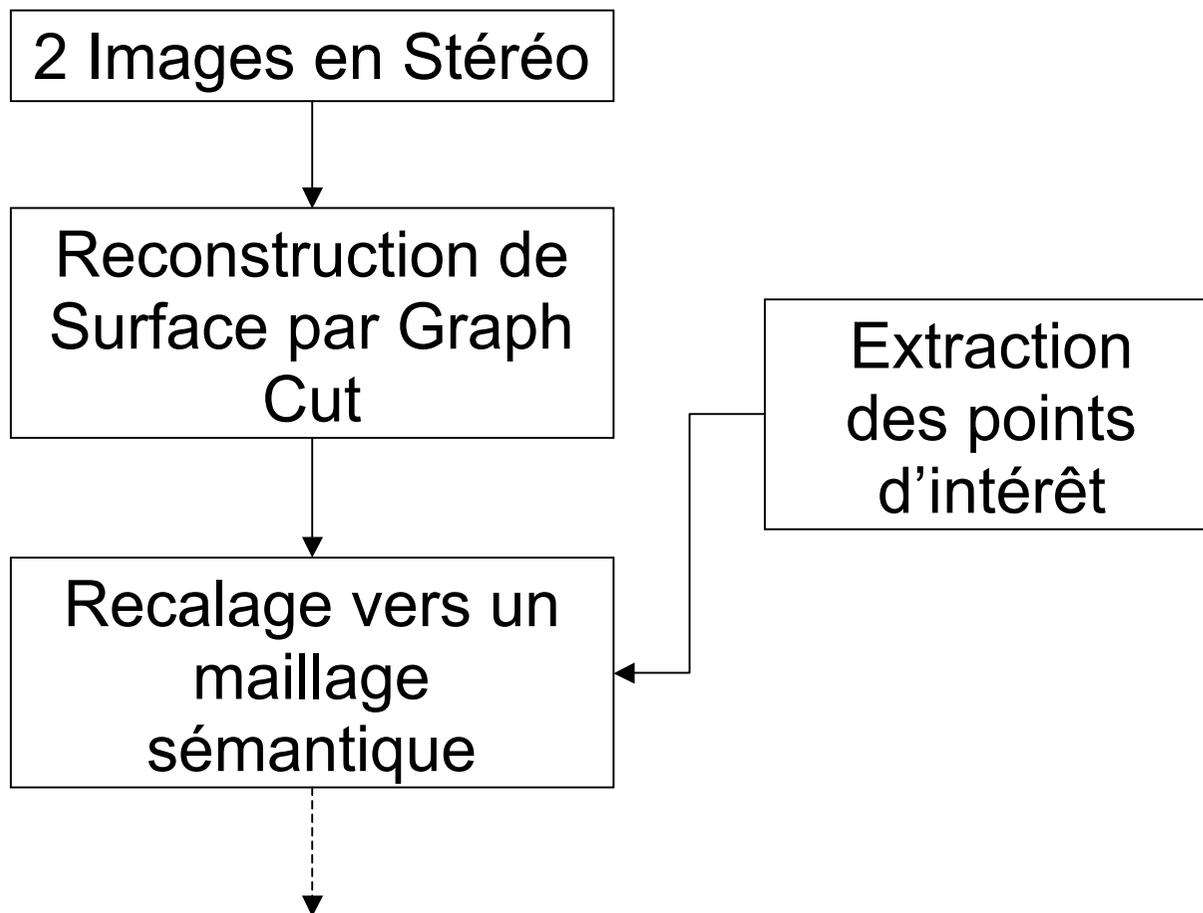
Pixels Foncés -> Disparité faible.



3D Reconstruction



Plan



Adaboost

- AdaBoost est une combinaison linéaire de T classifieurs faibles.

$$h_t : X \rightarrow \{-1, 1\} \text{ with } error < 0.5$$

- L'ensemble d'apprentissage est constitué d'exemples et de contre-exemples de l'objet à détecter. Les données peuvent être des vecteurs de niveau de gris, la réponse à des filtres (Filtre de Gabor, Fonction de Haar,...) auxquels est assigné un poids qui indique l'importance de l'exemple dans la base d'apprentissage.

$$(\mathbf{x}_1, y_1) \cdots (\mathbf{x}_n, y_n) \text{ où } \mathbf{x}_i \in \mathbf{X} \text{ et } y_i \in Y = \{-1, 1\}$$

Adaboost

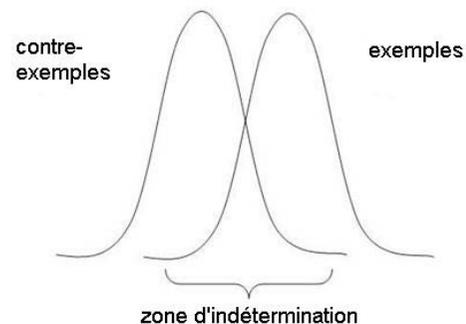
- A chaque itération t , le meilleur classifieur faible et son poids α_t sont calculés. Le $|D_t|$ des exemples qui ne sont pas correctement classés par ce classifieur est augmenté, de manière à ce que le nouveau classifieur soit centré sur ces exemples.
- La classification finale est donnée :

$$H(\mathbf{x}) = \text{sign}\left(\sum_{t=1}^T \alpha_t h_t(\mathbf{x})\right)$$

Adaboost en Cascade [Viola-Jones:01]

Idée : combiner successivement des classifieurs Adaboost avec un petit nombre de classifieurs faibles, supposant qu'il est facile de rejeter certains candidats, comme l'arrière-plan d'une image.

- Apprentissage :
 - Calcul des paramètres du classifieur pour l'ensemble d'apprentissage.
 - Calcul de la distribution des classes.
 - Les exemples négatifs qui sont clairement classés (qui ne sont pas dans la zone d'indétermination) sont retirés de l'ensemble .
 - Répéter jusqu'à ce que les classes soient clairement séparées.



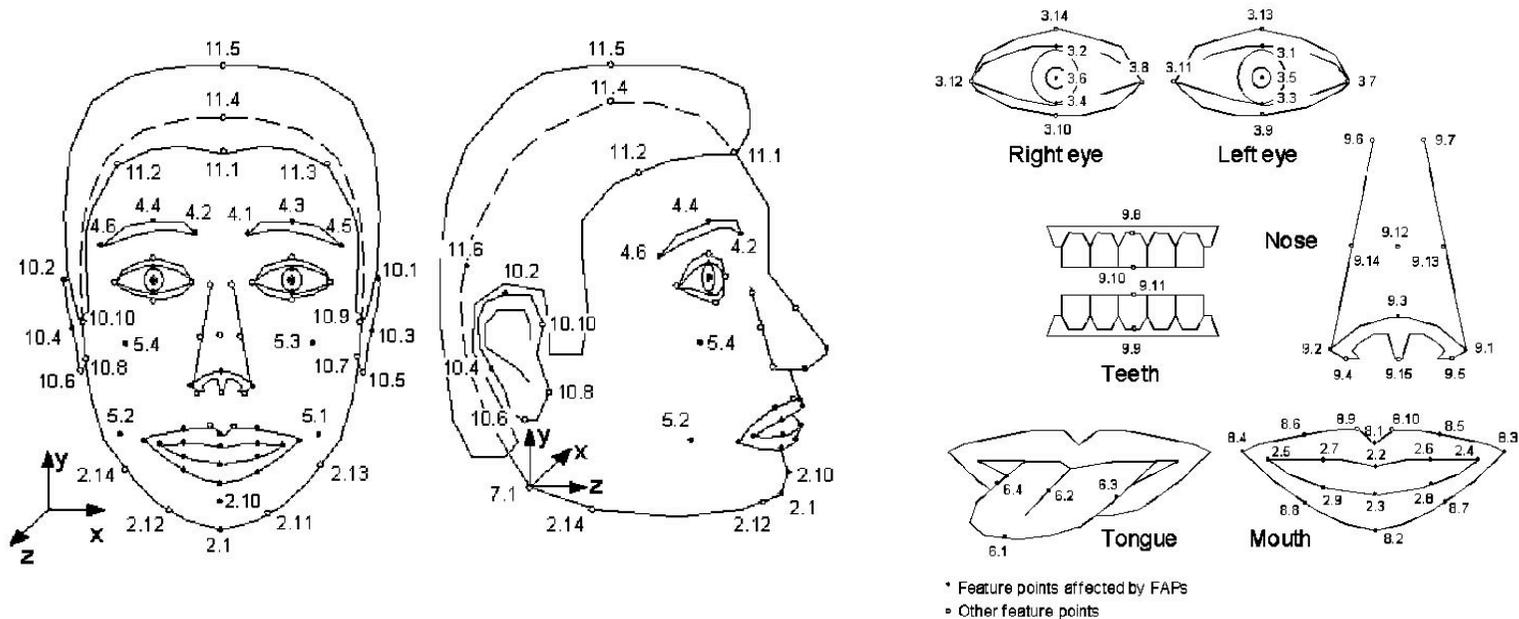
• Lors de la classification, le premier classifieur Adaboost est appliqué à tous les candidats. Pour tous les candidats qui ne sont pas clairement déterminés comme étant l'objet ou non, le deuxième classifieur est appliqué.

Détection du visage



Points d'intérêt du visage

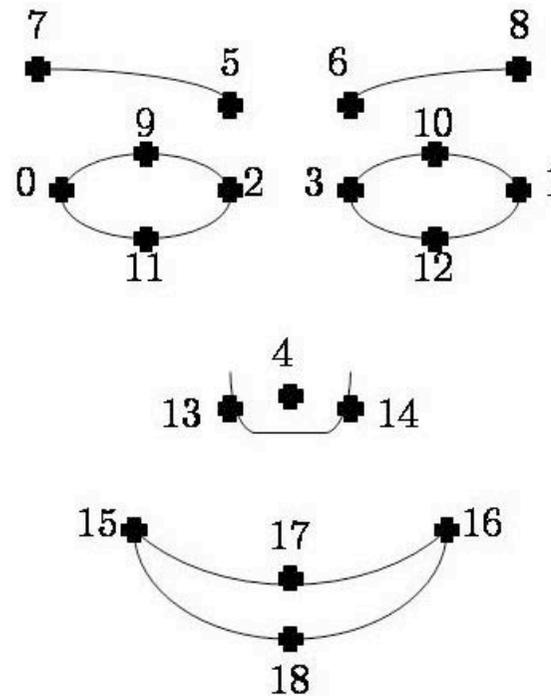
- Points d'intérêt définis par le standard MPEG-4.



205 points

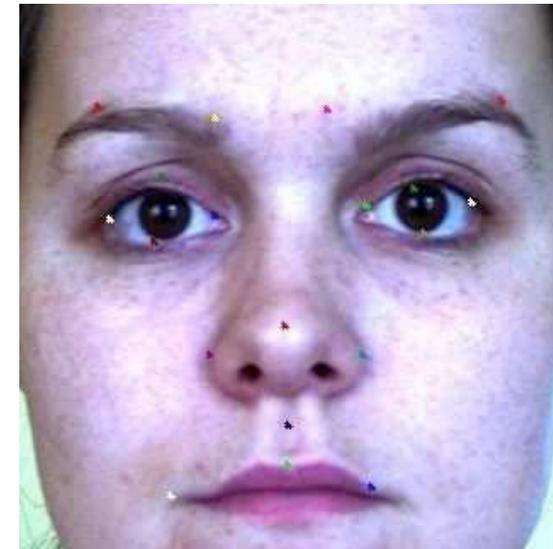
Points d'intérêt du visage

- Nous ne gardons que les 19 points les plus pertinents.



Points d'intérêt du visage

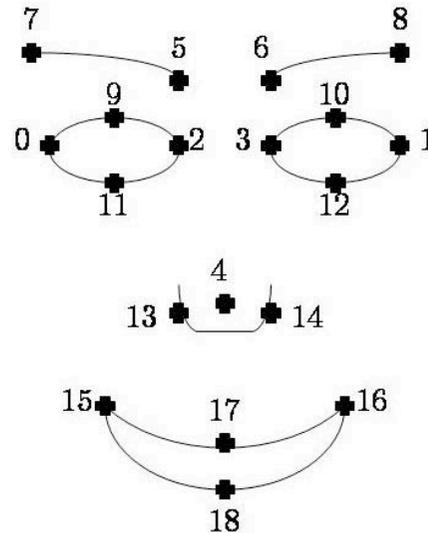
- Adaboost pour le recherche des point d'intérêt [Pantic:00]



- Défaut :
 - ↳ Très dépendant de la base de données.
 - ↳ Ne prend pas en compte la position des points d'intérêts les uns par rapport aux autres.

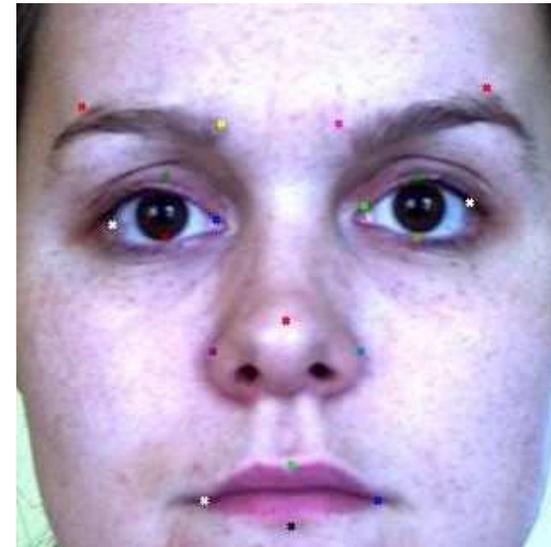
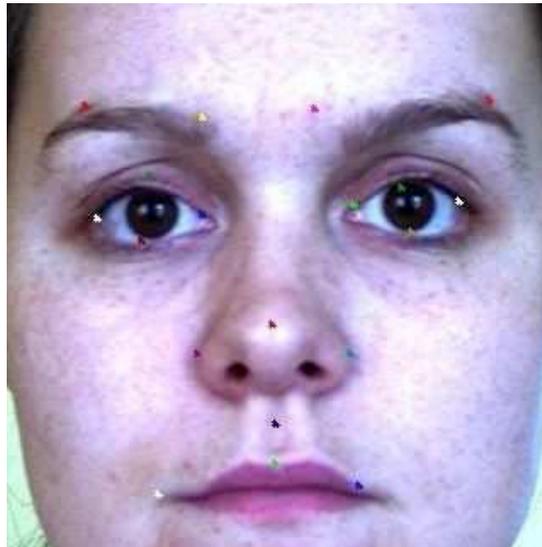
Points d'intérêt du visage

- Des contraintes anthropométriques, basées sur le standard MPEG-4, sont appliquées pour définir la meilleure configuration, parmi les candidats retenus par Adaboost.

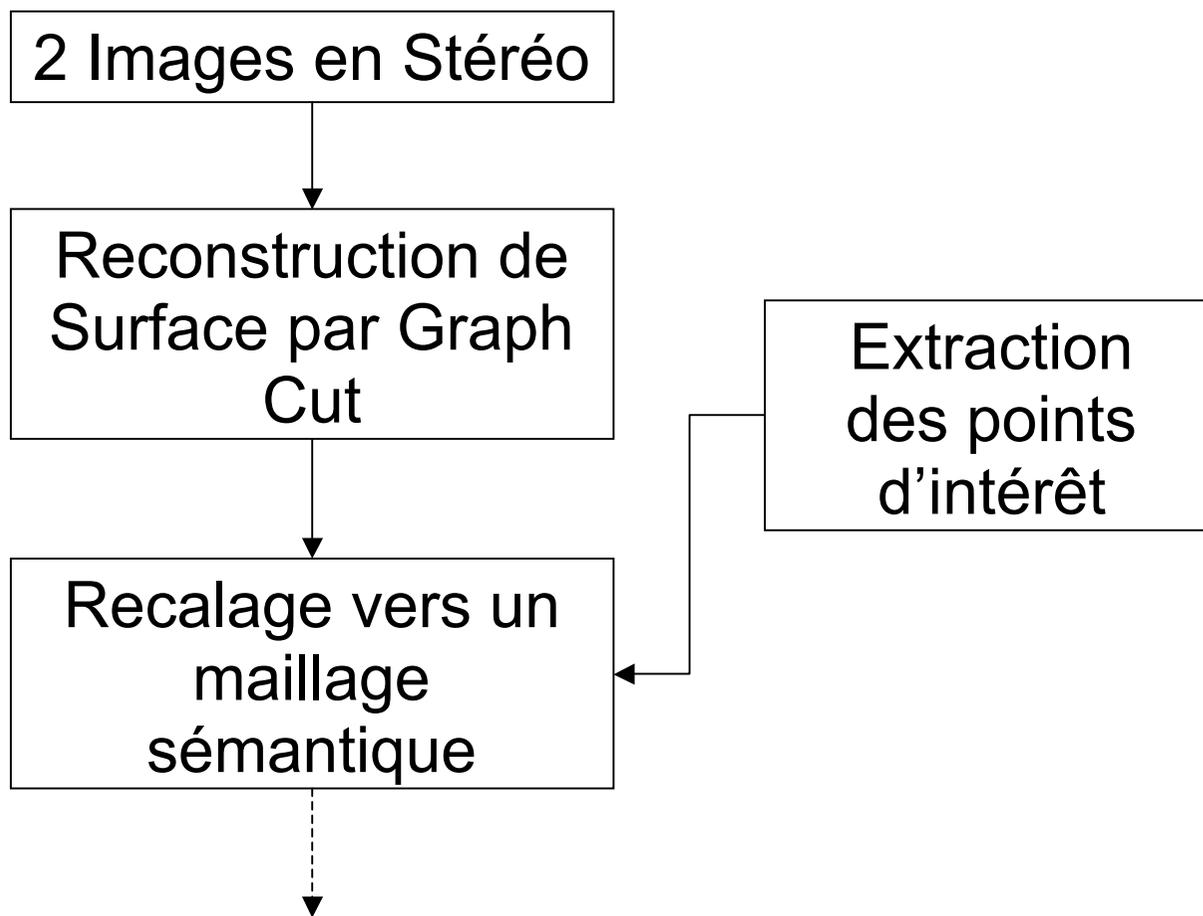


Méthode : en formulant le problème comme un problème d'optimisation discrète [Komodakis-Tziritas:07].

Points d'intérêt du visage

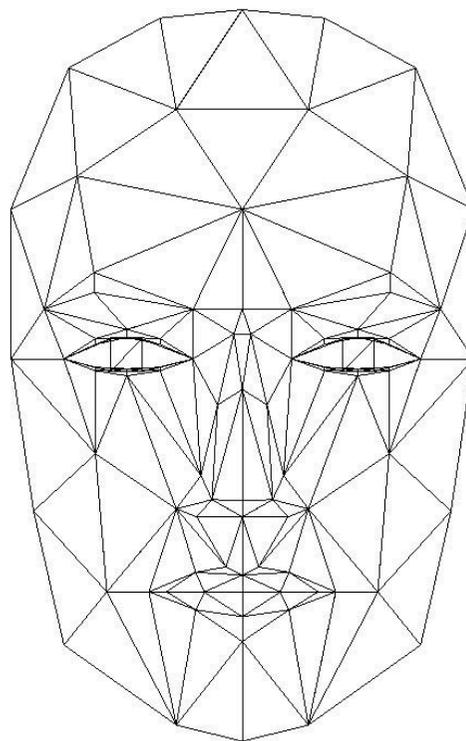


Plan



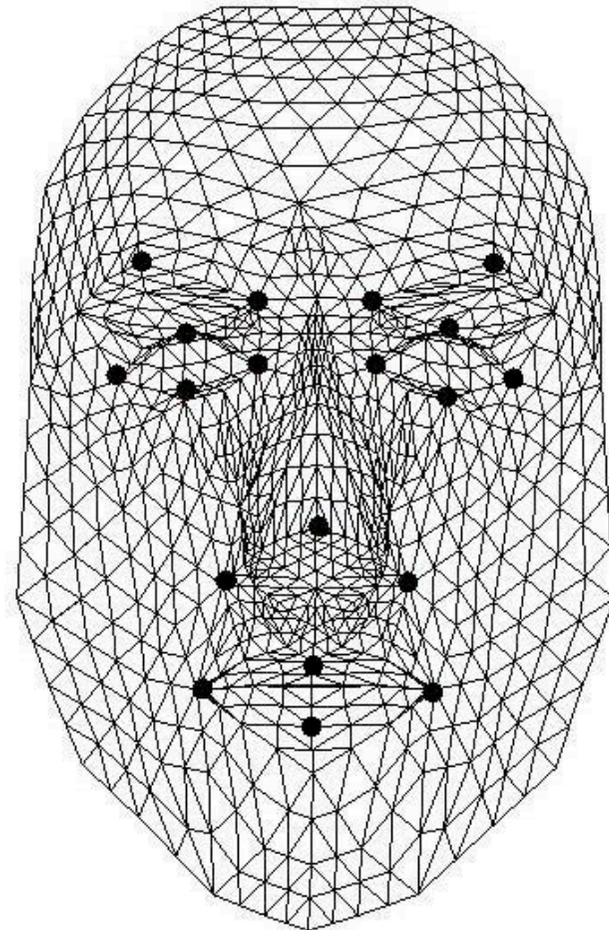
Maillage Générique Sémantique

- Modèle Candide (3ème version) [Ahlberg:01]



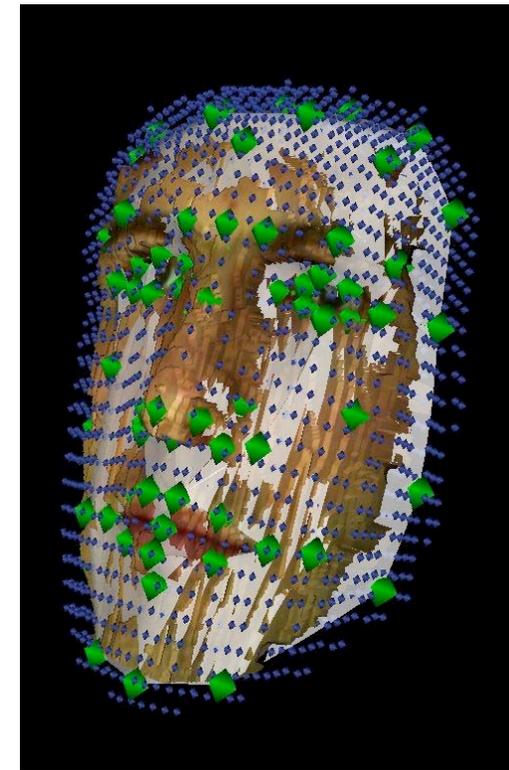
Maillage Générique Sémantique

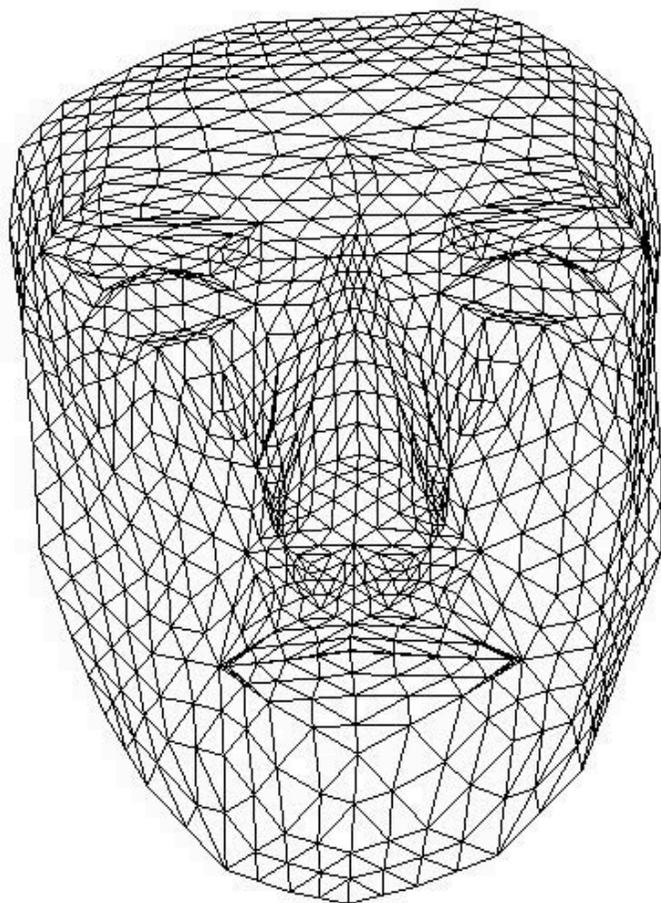
- ❑ Adapté du modèle Candide.
- ❑ Recalé sur une base de reconstructions de visage, pour obtenir un modèle plus humain.
- ❑ Points noirs = Points d'intérêt

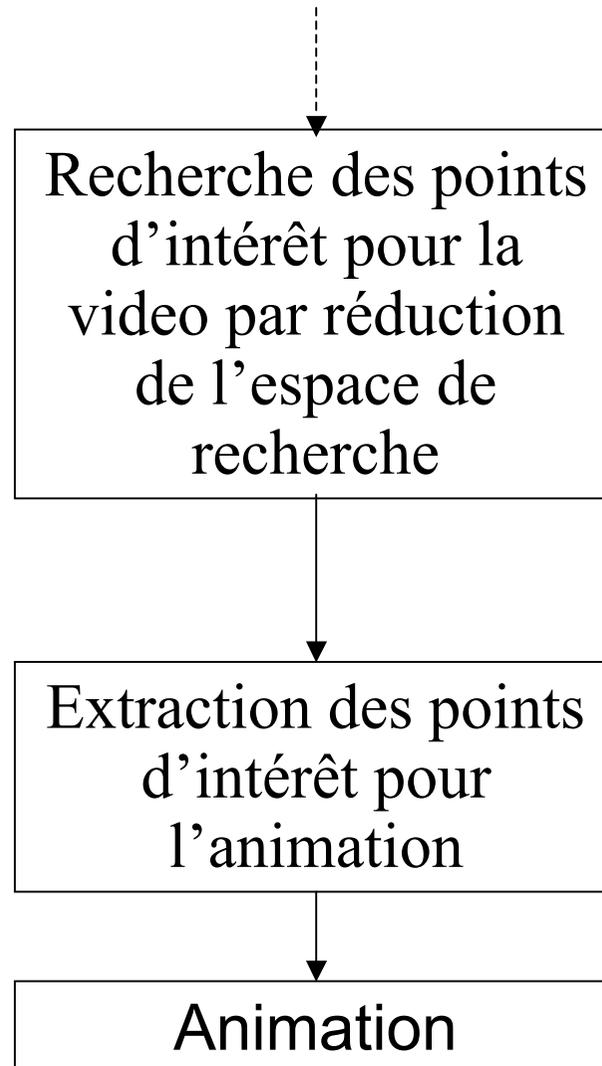


Recalage entre la Reconstruction et le Maillage

- ❑ Travail de Maxime Taron (Ecole Centrale de Paris – Siemens Corporate Research).
- ❑ Surfaces représentées par des fonctions de distance.
- ❑ Recalage global puis local en utilisant les points d'intérêt connus sur les deux surfaces.

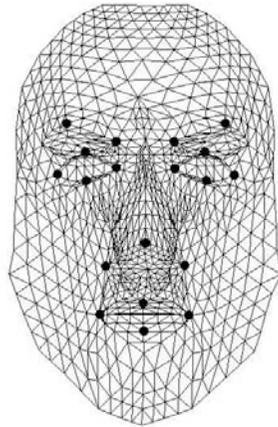




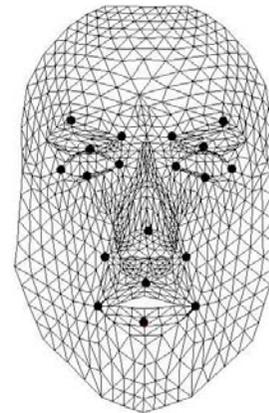


Les 6 Emotions de base

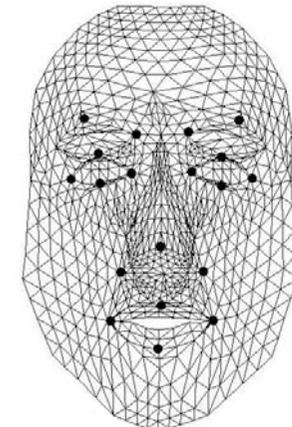
- Définies dans les années 70 par deux psychologues : Ekman et Friesen



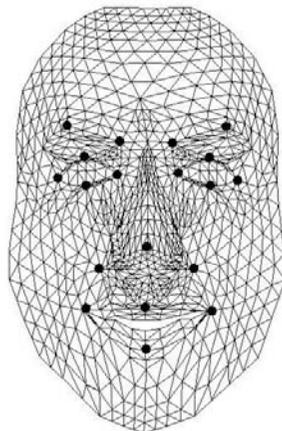
Colère



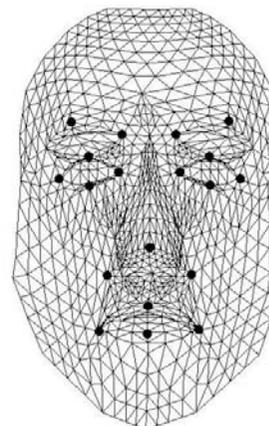
Dégoût



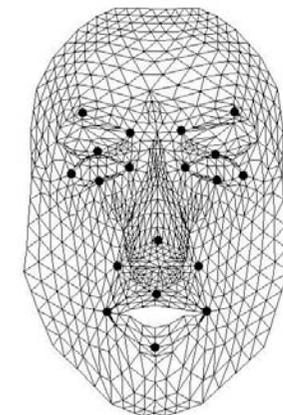
Peur



Joie

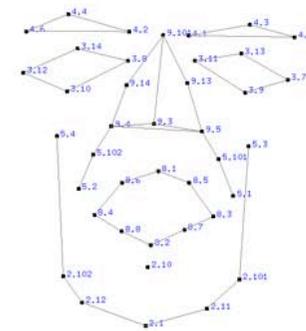
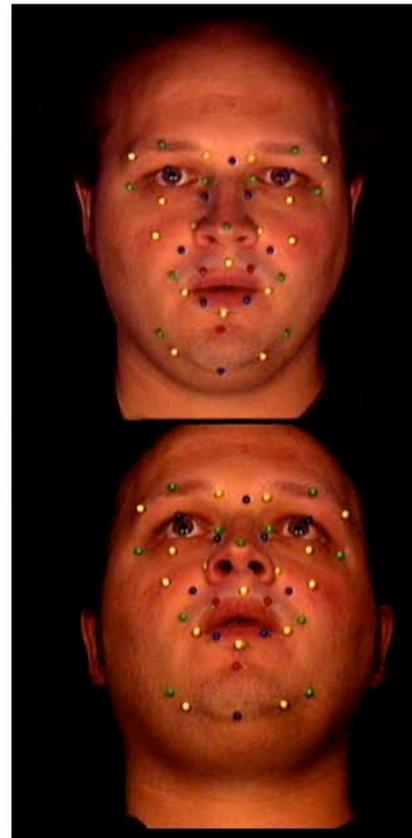
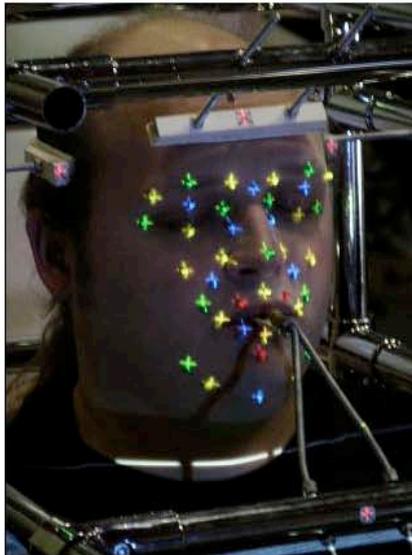


Tristesse



Surprise

Acquisition des données



Modèle d'Expressions

- Idée : Apprendre le mouvement des points d'intérêt du visage entre le temps t et le temps $t + 1$.
- Processus Auto Régressif : l'état d'une variable dans le temps, est calculé grâce aux états précédents en utilisant un modèle linéaire ou non-linéaire.

$$\mathbf{X}(t) = \sum_{k=1}^p H^k \mathbf{X}(t - k) + \eta_k$$

- Où $\mathbf{X}(t)$ est un vecteur représentant les coordonnées des points d'intérêts au temps t et p est l'ordre du processus.

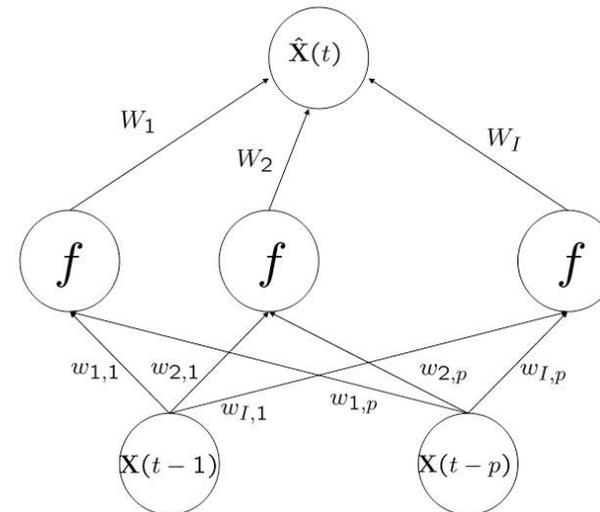
Modèle d'Expressions

- Modèle linéaire : les paramètres sont représentés par une matrice.
- Modèle non-linéaire :

$$\hat{X}(t) = f\left(\sum_{i=1}^I W_i f\left(\sum_{k=1}^p w_{i,k} X(t, k)\right)\right)$$

Où W_i et $w_{i,k}$, $i \in [1; I]$ et $k \in [1; p]$ sont les paramètres du processus auto-regressif, $f(\cdot)$ est une fonction monotone (généralement une sigmoïde).

Le système peut être modélisé par un réseau de neurones :

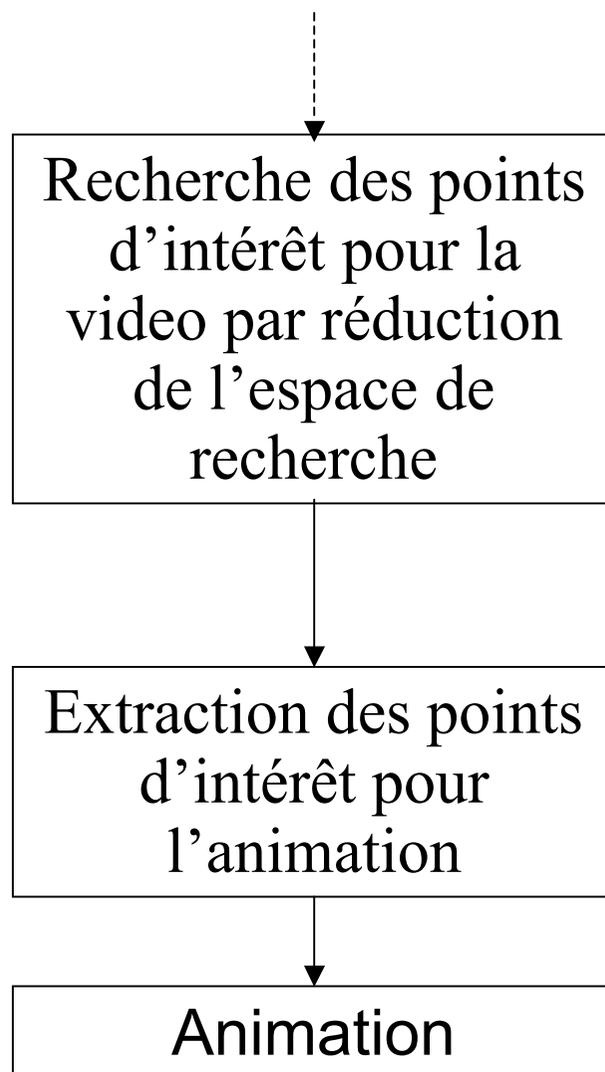


Modèle d'Expressions

- Apprentissage des paramètres
 1. Initialisation aléatoire des poids.
 2. Donner en entrée du réseau de neurones un exemple d'apprentissage et le propager dans le réseau.
 3. Comparer la sortie du réseau avec la véritable sortie et calculer l'erreur .
 4. Retro-propagation de l'erreur dans le réseau et ajustement des poids.
 5. Répéter 1 à 4 pour chaque exemple de l'ensemble d'apprentissage plusieurs fois.

Réduction de l'Espace pour l'Extraction des Points d'intérêt

- ❑ Estimation de la position pour chacune des expressions dans l'image suivante.
- ❑ Réduction de l'espace de recherche dans l'image suivante, basée sur cette réduction.
- ❑ Possibilité d'utiliser le résultat de l'estimation et de l'extraction pour déterminer l'expression.



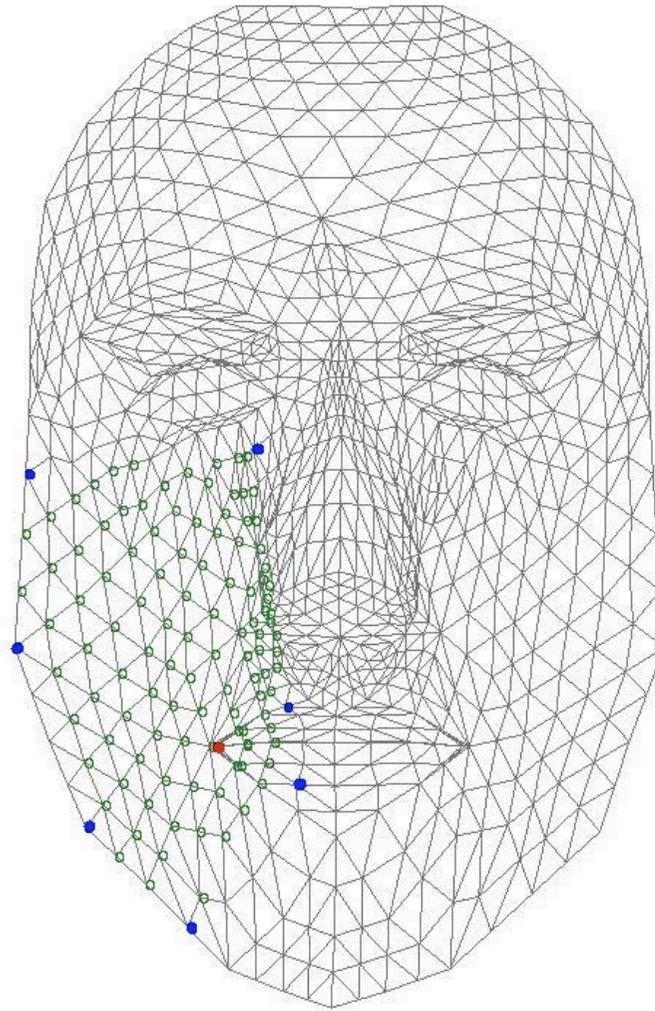
Animation du Modèle

- Utilisation d'une somme pondérée de Fonctions à Base Radiale pour produire des déformations locales.

$$y(x) = \sum_{i=0}^N w_i \rho(\|x - c_i\|)$$

- Où x est la position initiale, $y(x)$ est la nouvelle position du point et c_i est l'ensemble des N points d'ancrage, qui ne bougent pas, et du point de contrôle avec w_i les poids associés.

Animation du Modèle



Point rouge : point de contrôle

Points bleus : Points d'ancrage

Points verts: Points influencés par le mouvement du point de contrôle.

Mimicking Joy

video.mpg

Questions ?