Suivi de la pose 3D du visage en environnement multi-caméras avec un modèle tridimensionnel individualisé

Catherine Herold, Stéphane Gentric, Nicolas Moënne-Loccoz

Suivi de la pose 3D du visage en environnement multi-caméras avec un modèle tridimensionnel individualisé

Multiview 3D Head Pose Tracking with an Individualized Three-dimensional Model

Catherine Herold Stéphane Gentric Nicolas Moënne-Loccoz

Morpho, groupe Safran

11 boulevard Galliéni, 92130 Issy-les-Moulineaux - France
prenom.nom@morpho.com

Résumé

Cet article présente une nouvelle méthode de suivi de la pose du visage dans le cadre d'une application d'authentification faciale. L'acquisition du visage est effectuée pendant l'avancée de l'individu à authentifier dans un sas, ce qui entraîne des variations de son apparence au cours de la séquence. Pour être robuste à ces changements, nous utilisons une modélisation tridimensionnelle de la tête adaptée à chaque personne, bénéficiant ainsi d'une connaissance multi-pose de son apparence. Des vues synthétisées du visage sous de nouvelles poses sont alors utilisées pour effectuer le suivi par filtre particulaire. Outre l'apport du modèle 3D individualisé, nous évaluons différentes variantes du filtrage particulaire, dont des méthodes multi-passes qui améliorent la précision du suivi par le biais d'une exploration de l'espace d'état en plusieurs étapes.

Mots Clef

Suivi du visage, estimation de pose, modèle personnalisé

Abstract

In this paper, we present a new head pose tracking method involved in a facial authentification application. The face acquisition is proceeded while the person is walking in the authentication gate, which induces appearance variations during the sequence. In order to remain robust to these changes, we use a 3D head modelization instantiated at the sequence beginning, thus taking benefit from a multi-pose appareance model. Synthesized views of the head at new poses are then used for the tracking step performed by particle filter. In addition to the contribution of the individualized 3D model, we also evaluate several particle filter methods, especially multi-pass algorithms which improve the tracking accuracy by exploring the state-space in several steps.

FIGURE 1 – Vue globale du système d'authentification

Keywords

Head tracking, pose estimation, personnalized model

1 Introduction

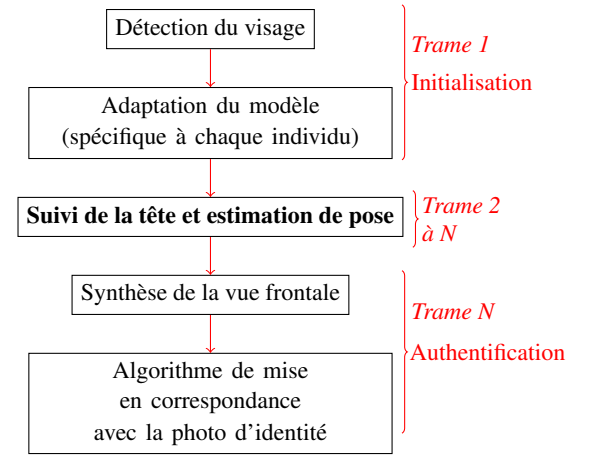
Le suivi d'objets est une thématique largement traitée en vision, dans des applications aussi diverses que la vidéosurveillance (suivi de véhicules ou de personnes), la sécurité routière (suivi de la tête pour le contrôle d'attention du conducteur) ou encore l'interaction humain-ordinateur (suivi des mains pour l'interprétation du langage des signes). Dans cet article, le suivi de la tête et l'estimation de sa pose sont utilisés pour une application de reconnaissance faciale à partir d'un système d'acquisition multi-vues (figure 1).

Une des problématiques majeures rencontrées en suivi est liée au choix des données modèles à utiliser pour caractériser un objet et vérifier sa présence à une position donnée dans une image. Différentes méthodes ont été proposées dans la littérature, la plus simple étant d'utiliser une région d'intérêt centrée sur l'objet dans la trame initiale pour en extraire des descripteurs qui le caractérisent, tels que des histogrammes de couleur [1]. Une telle méthode n'est en général pas robuste aux variations de pose qui induisent

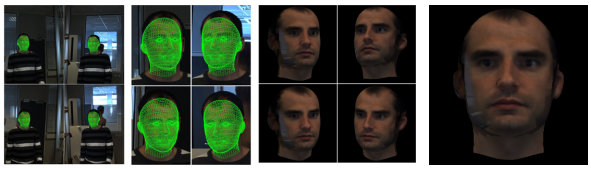
des changements d'apparence de l'objet dans les images, comme c'est le cas pour le visage du fait de sa géométrie tridimensionnelle. Basu *et al.* [2] résolvent ce problème en utilisant un modèle de visage 3D et en estimant conjointement sa position et son orientation. Cependant, leur algorithme repose sur une méthode de flot optique, qui n'est pas robuste en cas de changement d'illumination au cours de la séquence. Pour être robuste aux variations d'orientation et de luminosité, Babenko *et al.* [3] proposent de modifier le modèle d'apparence en ligne. Cette méthode possède des limites en cas de changement rapide de pose et présente un risque de mise à jour incorrecte si l'estimation courante de la position est mauvaise. Pour éviter de perdre des données consistantes, un apprentissage incrémental du modèle en ligne est proposé dans [4], permettant de conserver la connaissance antérieure de l'objet tout en enrichissant sa description avec l'ajout de vues le représentant sous de nouvelles poses. Comme dans la méthode précédente, dès lors que le modèle est mis à jour en ligne, le risque d'y ajouter des données erronées persiste. Nous proposons donc une méthode de suivi du visage dont le modèle est entièrement déterminé dès la première trame afin de garantir sa validité. La robustesse à des variations de pose est acquise grâce à la définition tridimensionnelle du modèle, qui induit une connaissance multi-pose de l'apparence de la tête. En outre, l'utilisation d'un modèle personnalisé permet de générer des descripteurs propres à chaque individu, ce qui renforce la qualité du suivi. Ainsi, il n'est pas nécessaire d'apprendre des descripteurs génériques qui requièrent de larges bases d'apprentissage pour couvrir la variabilité de la classe visages.

Pour effectuer le suivi d'un objet, une des méthodes les plus utilisées est le filtre de Kalman, qui impose cependant des conditions restrictives de linéarité sur le système et nécessite une adaptation particulière pour un suivi tridimensionnel à partir de vues 2D multiples [5]. Nous privilégions une méthode de suivi par filtre particulaire qui permet d'exprimer aisément le suivi de pose dans l'espace 3D en fusionnant les informations 2D des différentes vues. Dans cet article, différents algorithmes de filtre particulaire sont comparés, dont les méthodes multi-passes telles que le recuit simulé [6] et une variante prenant en compte les spécifités du visage (présentée dans la partie 4.3). Ces méthodes optimisent la distribution des particules en augmentant le nombre de rééchantillonnages pour chaque trame.

Nous présentons tout d'abord le système d'acquisition et le processus global d'authentification dans la partie 2. La partie 3 décrit la modélisation de la tête employée, ainsi que les paramètres à estimer au cours du suivi. Quelques rappels sur le filtrage particulaire sont donnés dans la partie 4, ainsi qu'une présentation de l'adaptation de cet algorithme à notre contexte et le détail des méthodes multi-passes. Les résultats obtenus avec les variantes de suivi sont présentés dans la partie 5. Nous terminons cet article avec les axes de recherche qui font suite à ces résultats sur le suivi de pose avec un modèle 3D personnalisé.



(a) Chaîne de traitement global



(b) Estimation des paramètres du modèle (c) Vue frontale

FIGURE 2 – Processus global de suivi et d'authentification

2 Processus global d'authentification

2.1 Authentification à la volée

La plupart des systèmes d'acquisition faciale nécessitent un comportement spécifique de la part de l'utilisateur, tel que l'immobilisation devant une ou plusieurs caméras. Cette contrainte d'arrêt ralentit considérablement l'étape de contrôle d'identité. L'objectif du sas présenté ici est d'acquérir le visage « à la volée », sans interaction spécifique de l'utilisateur, afin d'accélérer l'ensemble du processus d'authentification. Celui-ci s'effectue en plusieurs étapes. L'utilisateur scanne tout d'abord son passeport à l'entrée du système avant de pénétrer dans le sas. Au cours de son avancée vers la sortie, sa tête est suivie dans le repère 3D du sas et les paramètres du modèle de la tête sont estimés à partir des différentes vues (figure 2(b)) afin de correspondre au mieux au visage de la personne suivie. Bénéficiant d'une modélisation de la tête adaptée à l'individu, de nouvelles vues peuvent être générées, et en particulier la vue frontale (figure 2(c)) pour être comparée à la photo d'identité. L'ensemble du processus est résumé dans la figure 2(a). Les résultats exposés dans la suite de l'article concernent spécifiquement l'étape de suivi et d'estimation de la pose de la tête à partir du modèle adapté à la trame initiale.

2.2 Dispositif d'acquisition

L'acquisition des images est effectuée par le biais de quatre caméras situées à gauche et à droite de la sortie du sas, dans la configuration illustrée à la figure 3. Chaque caméra est

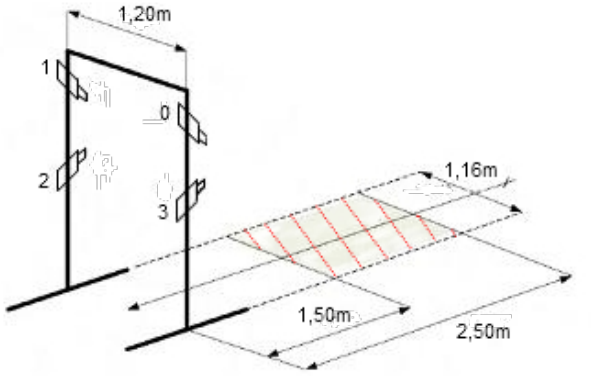


FIGURE 3 – Géométrie globale du système.
La partie hachurée correspond à la zone utile du système d'acquisition.

FIGURE 4 – Vues des 4 caméras en début de séquence

orientée vers un point central du sas, à hauteur moyenne de tête. Ces caméras sont calibrées par la méthode proposée dans [7] et acquièrent les images de manière synchronisée à une fréquence de 8 trames par seconde. Un exemple d'images acquises en début de séquence est donné à la figure 4.

3 Modèle de visage individualisé

3.1 Modélisation du visage et de la pose

Différentes modélisations du visage ont été proposées ces dernières années. Parmi les plus employées, on peut notamment citer l'*Active Appearance Model (AAM)* [8], qui est une modélisation statistique du visage, s'appuyant sur la position de certains points caractéristiques associée à une texture globale du visage. Blanz et Vetter [9] vont plus loin dans la caractérisation du visage en proposant un modèle 3D de forme associé à une description statistique de la texture. Une analyse par composantes principales effectuée sur un ensemble d'acquisitions 3D de visages permet alors d'obtenir une base de vecteurs propres définissant l'espace des visages. La forme d'un visage H peut donc être exprimée comme une combinaison de vecteurs propres :

$$H = \mu + \sum_{i=1}^{M} \lambda_i F_i \qquad (1)$$

où μ est le vecteur moyen, M le nombre de vecteurs propres retenus, et λ_i le coefficient associé à chaque vecteur propre F_i. La texture associée s'exprime de façon similaire.

La pose à estimer au cours de la séquence est décrite par :
- la position du centre de la tête : $X^c = (x_c, y_c, z_c)$ exprimée dans le repère associé au sas, [1]
- l'orientation de la tête, décrite par trois angles : le lacet, le tangage et le roulis, qui caractérisent respectivement la rotation autour des axes (Oy), (Ox) et (Oz).

3.2 Spécification des paramètres du modèle à l'initialisation

Comme nous l'avons rappelé dans l'introduction, les résultats de l'algorithme de suivi dépendent de la qualité des descripteurs utilisés. Le modèle de visage présenté précédemment est donc adapté à chaque personne à la première trame de la séquence afin de procéder au suivi avec des descripteurs spécifiques à chaque individu. Pour évaluer l'ensemble des paramètres du modèle, la position de la tête est tout d'abord estimée dans chaque image à l'aide d'un détecteur de visages [10]. Les poses en début de sas étant quasiment frontales dans chacune des vues, ce détecteur présente de bons résultats. Les informations issues des zones d'intérêt des quatre vues sont ensuite utilisées dans un processus d'ajustement similaire à [11] pour évaluer conjointement la pose et les coefficients λ_i optimaux (équation 1). L'ensemble de l'initialisation est illustré dans la figure 2(b). Les paramètres du modèle (forme et texture) sont ensuite fixés pour le processus de suivi dans le reste de la séquence.

4 Suivi de pose par filtre particulaire avec modèle 3D

4.1 Théorie du filtre particulaire

Nous nous plaçons dans un cadre bayésien pour effectuer le suivi de pose par filtre particulaire. La problématique du filtrage bayésien est la suivante : à partir d'un ensemble d'observations acquises dans le temps $(y_0, y_1, ..., y_t)$, l'objectif est d'estimer la densité *a posteriori* d'un état caché associé x_t.

1. Sauf mention explicite du contraire, toutes les coordonnées seront données dans ce repère.

A chaque instant t, l'état caché x_t est lié à l'état précédent x_{t-1} par la loi d'évolution temporelle suivante :

$$x_t = g(x_{t-1}) + \alpha_t \tag{2}$$

où g est une fonction (éventuellement non linéaire) caractérisant la dynamique du système, et α_t est le bruit associé à cette évolution. De cet état caché x_t peuvent être déduites les observations y_t correspondantes, suivant l'équation :

$$y_t = h(x_t) + \gamma_t \tag{3}$$

où γ_t est un bruit associé aux mesures.

Dans le cas où les fonctions g et h sont linéaires et les bruits α_t et γ_t gaussiens, le filtre de Kalman offre une réponse optimale à cette problématique [12]. D'autres méthodes ont été développées pour traiter le problème du filtrage bayésien dans des conditions moins restrictives, comme le filtre de Kalman étendu ou le filtre particulaire [13] que nous adoptons ici. En effet, l'adaptation du filtre de Kalman pour effectuer le suivi du visage en 3D nécessite par exemple des informations issues de détecteurs de points caractéristiques [5], que nous n'utilisons pas dans notre cas. Les seules informations disponibles sont donc l'intensité des pixels de l'image, et la fonction h générant ces observations est la fonction de projection d'un modèle de visage 3D sur une image. Du fait de la non-linéarité de cette fonction à cause des occultations, la méthode par filtre particulaire est privilégiée pour effectuer le suivi.

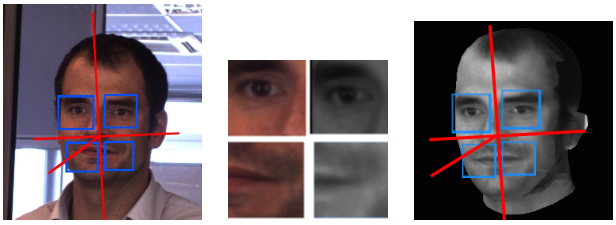
La particularité du filtre particulaire est d'approcher la densité *a posteriori* de l'état caché x_t par un ensemble de N éléments $x_t^{(i)}$ appelés particules en utilisant une méthode séquentielle de Monte-Carlo. A chaque particule $x_t^{(i)}$ est associé un poids $w_t^{(i)}$ qui traduit sa vraisemblance avec les observations. La densité de l'état x_t connaissant les mesures précédentes et courantes $(y_0, ..., y_t)$ est donnée par :

$$p(x_t|y_0, y_1, ..., y_t) \sim \sum_{i=0}^{N} w_t^{(i)} \delta(x_t^{(i)} - x_t) \tag{4}$$

où δ est la fonction dirac.

L'algorithme d'échantillonnage avec rééchantillonnage par importance (*Sampling Importance Resampling* ou *SIR* [14]), qui est l'une des déclinaisons les plus courantes du filtre particulaire, comprend trois étapes :

1. la prédiction, caractérisée par une distribution de transition $p(x_t|x_{t-1})$, rattachée à l'équation 2,
2. la correction, caractérisée par une distribution de vraisemblance $p(y_t|x_t)$, rattachée à l'équation 3,
3. le rééchantillonnage (optionnel), qui est une redistribution des particules afin de mieux caractériser la densité *a posteriori*. Celui-ci est effectué si l'ensemble des particules est trop disparate, ce qui est évalué par le quotient $\frac{1}{\sum_{i=1}^{N}(w^{(i)})^2}$, appelé nombre effectif de particules.



(a) Observations (b) Patchs (c) Données du modèle

FIGURE 5 – Patchs extraits du visage pour le calcul de vraisemblance

4.2 Adaptation au suivi de pose de la tête

Nous décrivons à présent l'adaptation du filtre particulaire à notre contexte pour effectuer le suivi de pose. Du fait de la variation d'apparence d'un visage en cas de changement d'orientation, nous estimons simultanément les trois angles de rotation et la position de la tête. L'état caché x_t recherché à chaque instant est donc la pose du visage, les observations y_t sont les images acquises à cet instant et $x_t^{(i)}$ représente la pose de la $i^{ème}$ particule.

L'étape de prédiction modifie l'état de chaque particule entre l'instant $t-1$ et l'instant t. Elle caractérise donc les changements de position et d'orientation de la tête entre deux acquisitions consécutives (la période étant de 135 millisecondes). La fréquence d'acquisition étant assez faible, nous choisissons une modélisation minimaliste du changement de pose au cours de l'avancée. La fonction g de l'équation 2 traduit simplement l'avancée selon l'axe principal du sas. Le reste des variations de pose entre $t-1$ et t est exprimé par l'ajout d'un bruit gaussien α_t.

Une fois l'état des particules mis à jour (équation 2), la deuxième étape consiste à réévaluer leur poids en fonction des observations faites à l'instant t. La procédure de mise à jour est la suivante :

1. Sélection des patchs : étant donné l'état $x_t^{(i)}$ d'une particule, des points 3D du modèle sont tout d'abord projetés sur chaque image. Des patchs sont ensuite extraits autour de ces projections (figure 5(a)). Par ailleurs, on calcule la pose de la tête dans le repère de chaque caméra pour générer la vue synthétique associée et on en extrait des patchs autour des mêmes points 3D projetés (figure 5(c)). La corrélation entre les patchs issus du modèle et ceux extraits des observations peut alors être calculée comme suit.
2. Similarité par critère de texture : le but étant de valider simultanément la position et l'orientation du visage, on compare les patchs pixel à pixel. La similarité est calculée par corrélation croisée normalisée de moyenne nulle (ZNCC), afin d'être invariante aux changements d'illumination du visage au cours de la séquence.
3. Fusion : soient N_v le nombre de vues, s_j^k le score de similarité pour le $j^{ème}$ patch dans la $k^{ème}$ vue, et N_p^k

le nombre de patchs extraits dans cette vue (ce nombre est dépendant de la vue, car la visibilité des points varie en fonction de la pose dans le repère de chaque caméra). Le score de similarité globale de la particule est alors donné par :

$$S_{global} = \prod_{k=1}^{N_v} \frac{\sum_{j=1}^{N_p^k} s_j^k}{N_p^k} \quad (5)$$

Les différents scores de similarité pour les patchs d'une vue sont sommés, pour empêcher d'invalider une particule à cause de modifications locales (clignement des yeux, ouverture de bouche). Pour valider la pose 3D de la tête dans le repère du sas, les projections dans chacune des vues doivent être valides, c'est pourquoi les scores de chaque vue sont multipliés. La similarité globale peut être reliée à la vraisemblance des observations compte tenu de l'état de la particule :

$$p(y_t|x_t^{(i)}) \propto e^{-S_{global}}. \quad (6)$$

4.3 Filtre particulaire multi-passes

A chaque instant t, nous estimons les six paramètres qui déterminent la pose. La dimension de l'espace de recherche étant relativement élevée, l'algorithme classique du filtre particulaire nécessite un grand nombre de particules pour que l'évaluation soit valide. Différentes méthodes d'optimisation du filtre particulaire ont été proposées dans la littérature pour réduire le nombre de particules nécessaires. Nous évaluons tout d'abord l'algorithme du recuit simulé (*Annealed Particle Filter* ou APF) proposé par Deutscher *et al.* dans [6]. Cette méthode consiste à augmenter le nombre de rééchantillonnages pour une trame donnée et à adapter récursivement la fonction de vraisemblance utilisée pour mettre à jour le poids des particules. A la $k^{ème}$ passe, la vraisemblance est donnée par :

$$p_k(y_t|x_t^{(i,k)}) \propto (e^{-S_{global}})^{\beta_k}, \quad (7)$$

où β_k est calculé de telle sorte que le nombre effectif de particules soit proche de $N/2$.

Dans le même ordre d'idée, nous proposons un algorithme de filtre particulaire à deux passes, en tenant compte des propriétés du visage pour expliciter les deux fonctions de vraisemblance utilisées. Comme l'illustre la figure 6(b), la fonction de vraisemblance décrite dans la partie 4.2 est piquée. De ce fait, si aucune des particules n'est assez proche de la pose réelle, les poids mis à jour ne caractérisent pas la distribution de l'état caché. Compte tenu de cette observation, nous définissons une première fonction de vraisemblance mieux adaptée à l'étape de correction. Comme précédemment, elle calcule la vraisemblance entre la vue fournie par chaque caméra et des images synthétisées du modèle, mais s'appuie sur un seul patch global décrivant l'ensemble du visage plutôt que sur un ensemble de patchs locaux.

Le changement de support pour le calcul de la corrélation influence la réponse des fonctions autour de la position réelle comme l'illustre la figure 6. L'utilisation d'un support global permet d'augmenter la robustesse à une erreur sur la position ou l'orientation du visage. En effet, les particules à proximité de la pose réelle seront valorisées après la première passe et légèrement bruitées pour affiner l'estimation avec une fonction de vraisemblance plus précise lors de la deuxième passe. L'algorithme 1 résume le filtre particulaire à deux passes que nous venons de présenter.

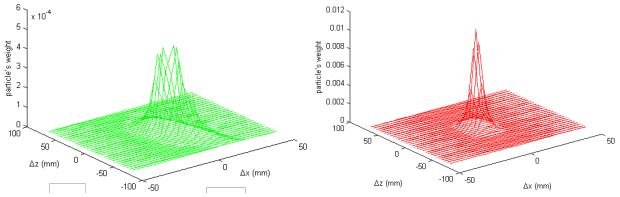
Algorithme 1 Filtrage particulaire à deux passes (2PF)

1: Trame 1 :
- Estimation de la pose (x_0) et des paramètres du modèle (forme et texture)
- Initialisation des N particules :
$(x_0^{(i)} = x_0, w_0^{(i)} = \frac{1}{N} \forall i)$
2: **Pour** trame $t = 2$ à nb_{trames} **Faire**
3: **Pour** $i = 1$ à N **Faire**
4: **Etape de prédiction :** $x_t^{(i)} = g(x_{t-1}^{(i)}) + \alpha_t^{(i)}$
5: $1^{ère}$ **passe :**
- $\forall$ caméra k : extraction d'un patch global (visage entier) et calcul de la corrélation s_k
- Fusion des scores s_k pour les N_v vues : $\prod_{k=1}^{N_v} s_k$
- Mise à jour des poids : $\tilde{w}_t^{(i)} = w_{t-1}^{(i)} p(y_t|x^{(i)})$
6: **Fin pour**
7: Normalisation ($\tilde{w}_t^{(i)} \leftarrow \frac{\tilde{w}_t^{(i)}}{\sum_{i=1}^{N} \tilde{w}_t^{(i)}}$), rééchantillonnage et ajout d'un bruit gaussien : $x_t^{(i)} \leftarrow x_t^{(i)} + \tilde{\alpha}_t^{(i)}$
8: **Pour** $i = 1$ à N **Faire**
9: $2^{ème}$ **passe :**
- $\forall$ caméra k : extraction de patchs locaux (figure 5) et calcul de la corrélation s_k^j.
- Fusion des scores s_k^j pour les N_v vues (équation 5).
- Mise à jour des poids : $w_t^{(i)} = \tilde{w}_t^{(i)} p(y_t|x_t^{(i)})$
10: **Fin pour**
11: Renormalisation
12: $x_i = \sum_{i=1}^{N} w_t^{(i)} x_t^{(i)}$
13: Rééchantillonnage (optionnel)
14: **Fin pour**
15: **Renvoyer** $x_1, x_2, ..., x_{nb_{trames}}$

5 Résultats

5.1 Génération de la vérité terrain

Les séquences de données utilisées pour l'évaluation de nos algorithmes sont issues du système d'acquisition présenté précédemment. La vérité terrain est générée manuellement en annotant neuf points caractéristiques du visage sur deux images pour chaque instant. Leurs positions 3D (X_i^V) sont ensuite calculées connaissant les contraintes épipolaires, et sont utilisées pour estimer la pose. Celle-ci est obtenue par une méthode de minimisation des moindres



(a) Vraisemblance globale : visage entier (b) Vraisemblance locale : sous-régions du visage

FIGURE 6 – Réponse de la fonction de vraisemblance autour de la pose réelle (variations en x et en z)

carrés itérative présentée en [15]. R_t étant la matrice de rotation caractérisant l'orientation de la tête et T_t la translation appliquée à son centre à l'instant t, le terme à minimiser est le suivant :

$$\sum_{i=1}^{9} \|X_i^V - (R_t X_i^M + T_t)\|^2 \quad (8)$$

où X_i^V est la position 3D issue des annotations du $i^{ème}$ point caractéristique et X_i^M est sa position 3D dans le repère du modèle générique.

5.2 Evaluation des algorithmes

Nous employons deux types de mesure pour quantifier l'erreur sur l'estimation de la pose. La première est la distance euclidienne entre la position du centre de la tête X_t^c issue de l'algorithme de suivi et sa position réelle T_t :

$$e_c = ||X_t^c - T_t|| \quad (9)$$

Nous évaluons également la qualité du suivi par une moyenne des erreurs sur la position 3D de N_{pc} points caractéristiques annotés manuellement :

$$e_{pc} = \frac{1}{N_{pc}} \sum_{i=1}^{N_{pc}} ||X_i^{FP} - X_i^V||, \quad (10)$$

où X_i^{FP} est la position du $i^{ème}$ point caractéristique estimée par l'algorithme de suivi et X_i^V sa position réelle. Les points utilisés sont les centres des yeux ainsi que les coins de bouche qui constituent les annotations les plus précises.

Apport d'un modèle tridimensionnel. La figure 7 illustre le gain en précision acquis grâce à l'utilisation d'un modèle 3D pour générer les données du modèle (figure 5(c)). L'erreur e_c est croissante avec la méthode s'appuyant sur les patchs 2D extraits de la trame initiale pour évaluer la vraisemblance des particules. Cela s'explique par l'écart croissant de pose entre la trame courante et la trame initiale lors de l'avancée de la personne. La figure 8 montre la région d'intérêt extraite autour de la projection du centre de la tête à trois instants de la séquence et illustre ce changement d'apparence. Ayant un modèle 3D, il est possible de générer des vues sous de nouvelles poses non frontales telles qu'elles apparaissent à partir du milieu de la

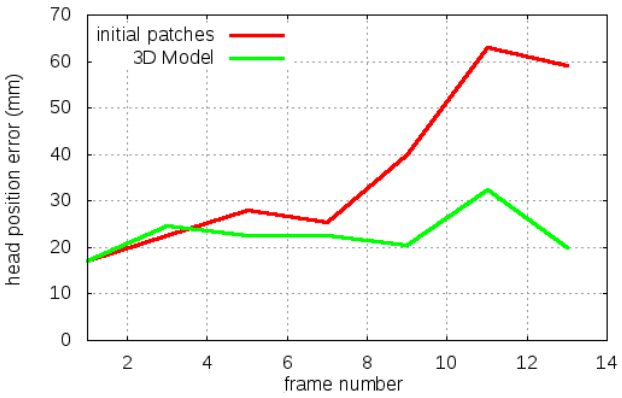


FIGURE 7 – Erreur (e_c) avec et sans utilisation de templates synthétisées à partir d'un modèle 3D

FIGURE 8 – Régions d'intérêt extraites autour du centre du visage projeté

séquence. En utilisant les descripteurs modèles adaptés au cours de l'avancée, les scores de corrélation restent valides pour l'évaluation des particules malgré les changements de pose.

Comparaison des méthodes multi-passes. Nous comparons également les résultats de suivi avec les différentes variantes du filtre particulaire présentées dans les parties 4.2 et 4.3 : la méthode SIR (Sampling - Importance Resampling), le recuit simulé (APF) et le filtre à deux passes (2PF). Dans chacun des cas, le modèle 3D instancié au début de la séquence est utilisé pour générer des vues 2D synthétiques et évaluer la vraisemblance des particules.
La figure 9 représente l'erreur e_{pc} au cours du temps. Les résultats ont été obtenus en moyennant l'erreur sur 14 séquences rejouées cinq fois, et utilisant 500 particules pour le SIR, 250 particules pour le 2PF et 100 (respectivement 250) particules et 5 (resp. 2) passes pour l'APF.
Notons tout d'abord que les méthodes multi-passes améliorent globalement les performances de suivi de la pose au cours de la séquence. Ce gain est dû à l'optimisation de

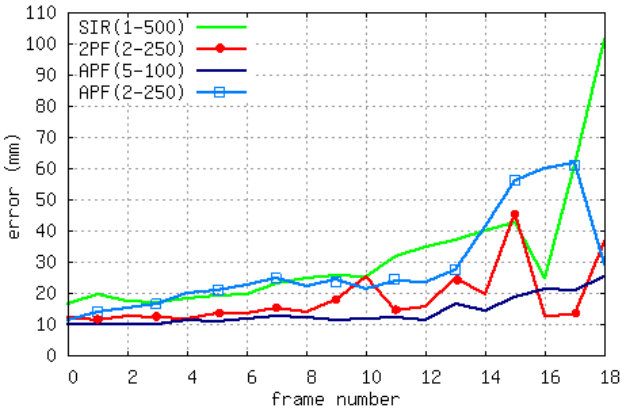


FIGURE 9 – Erreur (e_{pc}) avec des méthodes multi-passes : PF : SIR - 2PF : Algorithme 1 - APF : recuit simulé. Le nombre de passes et de particules est donné entre parenthèses

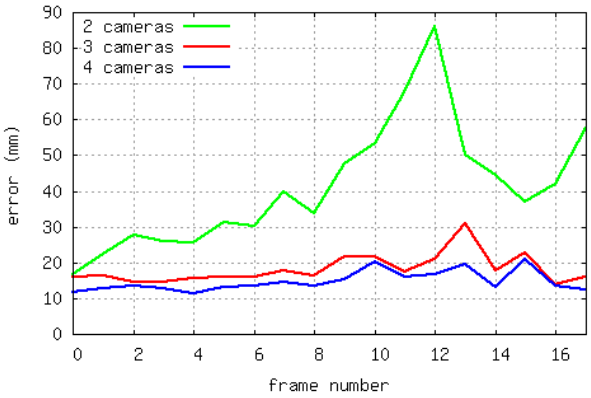


FIGURE 10 – Erreur (e_{pc}) en fonction du nombre de caméras

l'exploration de l'espace des poses, par le biais de rééchantillonnages successifs à chaque trame. Pour un nombre de passes fixé à deux, l'algorithme 2PF que nous proposons offre de meilleurs résultats que l'APF, et ce dès le début de la séquence. Le recuit simulé utilisant cinq passes améliore légèrement ces performances, au prix d'un coût de calcul plus élevé à cause des similarités calculées sur les patchs locaux à chaque passe. De plus, les deux fonctions utilisées dans la méthode 2PF sont entièrement déterminées, contrairement au recuit simulé où le paramètre β_k doit être optimisé à chaque étape. La figure 11 présente les résultats du suivi pour la méthode 2PF au début, au milieu et à la fin d'une séquence.

Influence du nombre de vues utilisées. Le temps de calcul étant lié au nombre de vues utilisées pour calculer la vraisemblance des particules, nous avons également évalué l'influence de ce paramètre sur la qualité du suivi. Les résultats donnés à la figure 10 ont été générés avec l'algorithme 2PF et 250 particules, pour 2, 3 et 4 caméras. Avec deux vues seulement, on note une erreur en augmentation tout au long de la séquence. L'ajout d'une troisième caméra permet de réduire considérablement l'erreur, qui reste alors inférieure à 2 cm sur les douze premières trames. Cela s'explique par le fait qu'en utilisant uniquement les deux caméras supérieures (paire employée pour les performances à deux caméras), il n'y a pas de disparité verticale, ce qui laisse des incertitudes sur l'estimation de la pose. L'ajout d'une troisième caméra située en-dessous des deux premières permet de lever cette incertitude. La prise en compte de la quatrième caméra n'améliore que légèrement les performances, car l'information qu'elle porte est en grande partie redondante avec les autres vues.

Influence de l'estimation du modèle individualisé. Lors de l'initialisation, le modèle est adapté au visage de la séquence en forme et en texture. Afin de vérifier la robustesse de l'algorithme aux variations individuelles de la forme, nos algorithmes ont également été executés avec un modèle de forme moyen (et une texture adaptée à l'individu comme précédemment). Les erreurs obtenues restent du même ordre car la précision de l'estimation des paramètres déformables du modèle n'est pas suffisante pour améliorer substantiellement l'attache aux données.
D'autre part, comme la texture est estimée à partir de la trame initiale où les visages sont globalement de face, la texture extraite au niveau de certaines parties du visage (joues, oreilles) est imprécise. De ce fait, les vues synthétiques générées sous des angles très différents de la pose initiale présentent des erreurs et limitent donc la plage d'angles sous laquelle le suivi est robuste. Pour augmenter cette plage, il faudrait bénéficier d'une connaissance plus complète du modèle 3D afin de générer des vues de synthèse valides sous des angles plus variés.

6 Conclusion et perspectives

Cet article présente une nouvelle méthode de suivi de pose s'appuyant sur un modèle 3D qui s'adapte aux spécificités de chaque visage. La connaissance tridimensionnelle permet de synthétiser des vues de la tête sous n'importe quelle pose, ce qui rend le suivi plus robuste malgré les variations d'orientation de la tête (et donc d'apparence dans les images). Le gain en précision est significatif comparativement à un filtre particulaire utilisant des patchs 2D extraits de la trame initiale. Par ailleurs, l'utilisation d'un filtre particulaire à plusieurs passes permet d'améliorer la qualité du suivi avec des temps de calcul du même ordre. En particulier, la prise en compte des caractéristiques du visage pour l'adaptation manuelle des fonctions de vraisemblance (2PF) permet d'obtenir des résultats meilleurs que ceux du recuit simulé à deux passes.

La finalité de notre application est d'authentifier les personnes par le biais d'une reconstruction tridimensionnelle de leur visage, qui peut être raffinée au cours du temps par l'ajout de nouvelles informations. Les axes de recherche à venir consistent à intégrer les paramètres λ_i du modèle (équation 1) dans l'état caché x_t et à filtrer leurs valeurs

FIGURE 11 – Pose estimée avec le 2PF - 500 particules. Les axes caractérisent la pose estimée par l'algorithme de suivi et les points blancs correspondent aux projections des points caractéristiques associées à cette pose.

au cours du temps. Il y aura donc conjointement des paramètres statiques et dynamiques à évaluer, comme cela est fait dans [16] pour des formes géométriques simples. Du fait de l'augmentation du nombre de variables à évaluer, de nouvelles méthodes d'optimisation du filtre particulaire seront nécessaires pour limiter le nombre de particules à utiliser.

Références

[1] P. Pérez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in *European Conference on Computer Vision*, pp. 661–675, 2002.

[2] S. Basu, I. Essa, and A. Pentland, "Motion regularization for model-based head tracking," in *Proceedings of the International Conference on Pattern Recognition*, pp. 611–616, IEEE Computer Society, 1996.

[3] B. Babenko, M.-H. Yang, and S. Belongie, "Visual Tracking with Online Multiple Instance Learning," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2009.

[4] Z. Yin and R. Collins, "On-the-fly Object Modeling while Tracking," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, 2007.

[5] C. Canton-Ferrer, J. R. Casas, M. Tekalp, and M. Pardàs, "Projective Kalman Filter : Multiocular Tracking of 3D Locations Towards Scene Understanding," in *Machine Learning for Multimodal Interaction*, vol. 3869, pp. 250–261, Springer, 2005.

[6] J. Deutscher, A. Blake, and I. Reid, "Articulated Body Motion Capture by Annealed Particle Filtering," *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, pp. 126–133, August 2002.

[7] Z. Zhang, "A Flexible New Technique for Camera Calibration," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, pp. 1330–1334, 2000.

[8] G. J. Edwards, C. J. Taylor, and T. F. Cootes, "Interpreting Face Images Using Active Appearance Models," in *International Conference on Face & Gesture Recognition*, pp. 300–305, 1998.

[9] V. Blanz and T. Vetter, "A Morphable Model for the Synthesis of 3D Faces," in *SIGGRAPH*, pp. 187–194, 1999.

[10] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 1, pp. 511–518, 2001.

[11] S. Romdhani, *Face Image Analysis using a Multiple Features Fitting Strategy*. PhD thesis, Universität Basel, January 2005.

[12] Y. Bar-Shalom and T. E. Fortmann, *Tracking and Data Association*, vol. 179 of *Mathematics in Science and Engineering*. 1987.

[13] M. Isard and A. Blake, "Condensation – Conditional Density Propagation for Visual Tracking," *International Journal of Computer Vision*, vol. 29, pp. 5–28, August 1998.

[14] A. Doucet, S. Godsill, and C. Andrieu, "On Sequential Monte Carlo Sampling Methods for Bayesian Filtering," *Statistics and Computing*, vol. 10, no. 3, pp. 197–208, 2000.

[15] S. Umeyama, "Least-Squares Estimation of Transformation Parameters Between Two Point Patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 13, no. 4, pp. 376–380, 1991.

[16] P. Minvielle, A. Doucet, A. Marrs, and S. Maskell, "A Bayesian Approach to Joint Tracking and Identification of Geometric Shapes in Video Sequences," *Image and Vision Computing*, vol. 28, pp. 111–123, January 2010.