Vers une reconnaissance d'état affectif à base de mouvements du haut du corps et du visage

Benjamin Allaert, Ioan Marius Bilasco, Adel Lablack

Laboratoire d'Informatique Fondamentale de Lille, Université de Lille 1, France

Résumé

L'émotion est une réaction complexe qui engage à la fois le corps et l'esprit. Elle peut être définie comme étant une réaction affective transitoire d'assez grande intensité provoquée par une stimulation venue de l'environnement. L'analyse des expressions corporelles a un rôle important dans le processus de reconnaissance de l'état affectif. Pour cela, nous proposons une approche de reconnaissance émotionnelle combinant deux canaux : le visage et le corps. Notre contribution s'appuie sur l'analyse du mouvement au sein du visage et du haut du corps qui sont synthétisés par des modèles de direction et de magnitude construits à partir des flux optiques. Ces modèles permettent de s'abstraire des bruits de détection à l'aide de l'extraction des caractéristiques principales des mouvements et constituent une base stable pour identifier les évolutions de l'état émotionnel et plus particulièrement de la valence et de l'arousal. Les modalités sont analysées individuellement et sont ensuite fusionnées dans un deuxième temps afin d'étudier l'apport informationnel issu de l'étude du mouvement de la personne dans sa globalité. L'approche proposée a enfin été validée avec succès sur un sous-ensemble de la base de données SEMAINE et permet de vérifier l'apport informationnel issu du mouvement dans la reconnaissance d'états affectifs.

The emotion is a complex reaction that involves both the body and the spirit. It can be defined as an affective reaction of high intensity usually caused by a stimulus that comes from the environment. The analysis of the body expression through movements is important in an affect recognition process. We propose an approach for affect recognition from two channels : face and body. Our contribution uses an analysis of facial and body movements through direction and magnitude models of motion constructed from optical flows. These models allow to remove the noise using the extraction of the main motion features and constitute a stable base to identify the evolutions of the emotional state and more specifically the valence and the arousal dimensions. Each modality is analyzed alone then combined to study the informative contribution of the user motion. The proposed approach has been validated successfully on a subset of SEMAINE database.

Mots clé : Reconnaissance d'émotions, analyse gestuelle, analyse du mouvement, analyse du visage

1. Introduction

Actuellement, il y a un vrai engouement autour des technologies permettant de reconnaître les émotions à partir de flux vidéo. La reconnaissance des expressions faciales et corporelles joue un rôle important dans une variété d'applications telles que l'automatisation de la recherche comportementale, le traitement audio-visuel de la parole, la téléconférence, l'e-learning, la sécurité aéroportuaire et le contrôle d'accès, etc.

La reconnaissance d'émotions à partir de flux vidéo s'est basée essentiellement sur les expressions faciales durant ces deux dernières décennies [HEF02] [KCY00] mais les systèmes proposés manquent de robustesse dans des environnements non contrôlés et en présence d'émotions spontanées, de variations de poses et d'éclairage, en présence de personnes âgées, etc. Bien qu'une étude fondamentale faite par Ambady et Rosenthal [AR92] ait suggérée que les expressions du visage et les gestes du corps semblent être les plus pertinents pour analyser le comportement humain, la reconnaissance d'émotions via les mouvements du corps ne commence que récemment à attirer l'attention des chercheurs [VDHdG11] [GP05] [CVC07].

Dans notre étude nous ciblons principalement les domaines de l'e-learning et de téléconférence. Dans ces domaines, il est important de garder l'attention de son auditoire pendant toute une présentation, ou du moins, d'être informé de la manière dont le public perçoit les messages transmis. C'est d'autant plus difficile si la conférence se déroule à dis-

tance, avec des participants présents par ordinateurs interposés. Pour y parvenir, il faut permettre à l'intervenant d'avoir un retour de l'impact de ses propos sur son audience.

Nous avons constaté que la problématique de la reconnaissance d'affect, ne peut être résolue convenablement sans recourir aux techniques de reconnaissance d'actions. Ce n'est pas le type d'actions qui nous intéresse ici, mais plutôt ses caractéristiques de mouvement. Ainsi, une étude du processus de reconnaissance est réalisée, couvrant les techniques de représentation, de classification et de fusion d'informations. A travers cet article, dont les contributions se déclinent principalement autour de ces trois axes, nous souhaitons étudier l'apport informationnel issu de l'étude du mouvement de la personne dans sa globalité (mouvement de la tête, mouvement des bras) par rapport à l'augmentation de la précision de la détection des émotions et à la quantification de leurs intensité.

L'article est organisé comme suit. Dans la Section 2, nous présentons brièvement l'état de l'art. La méthodologie proposée organisée en trois niveaux est décrite dans la Section 3. Nous présentons ensuite les résultats obtenus en utilisant la base SEMAINE [MVCP10] dans la Section 4. Afin de valider nos contributions, nous avons choisi un sous-ensemble de la base SEMAINE (utilisée dans les challenges AVEC) car elle présente des personnes assises face à la caméra en conversation avec un agent de manière très similaire à une vidéo-conférence. Les expressions faciales et corporelles sont spontanées car aucune instruction spécifique n'a été transmise aux participants. Enfin, nous concluons par discuter les pistes ouvertes par notre présente contribution dans la Section 5.

2. État de l'art

La définition de l'émotion est difficile à formaliser et dépend du contexte d'usage. Ainsi, là où un neurologue par exemple, sera attaché à des notions de facteurs somatiques ou d'activation cérébrale, un sociologue aura une vision bien plus globale et déterminera des valeurs liées à des paramètres sociaux. Selon les psychologues, certaines émotions de base sont universellement reconnues. Les descripteurs les plus couramment utilisés sont les six émotions de base : la colère, le dégout, la peur, la joie, la surprise, et la tristesse. La plupart des systèmes proposés tentent de reconnaitre un ensemble de prototype d'expressions émotionnelles sur le visage. Pour décrire les changements subtils du visage, le système d'action faciales (FACS) proposé par Ekman [HEF02] est largement utilisé. Une alternative à cette représentation catégorique est l'utilisation de trois dimensions : "agréable ou non agréable" (Valence), "réveil ou soumission" (Arousal) et "tension ou relaxation" (Stance). La Figure 1 illustre l'espace Valence/Arousal labellisé par Russel [Rus80].

L'essentiel de la littérature sur les émotions a été consacrée à l'étude d'une seule modalité qui est le visage. La plupart des travaux existants combinant différentes modalités sont orientés principalement sur les signaux vocaux et l'expression du visage [JSC*04]. L'émotion communiquée à travers les expressions corporelles a souvent été négligée. Le langage du corps est une forme de communication non ver-

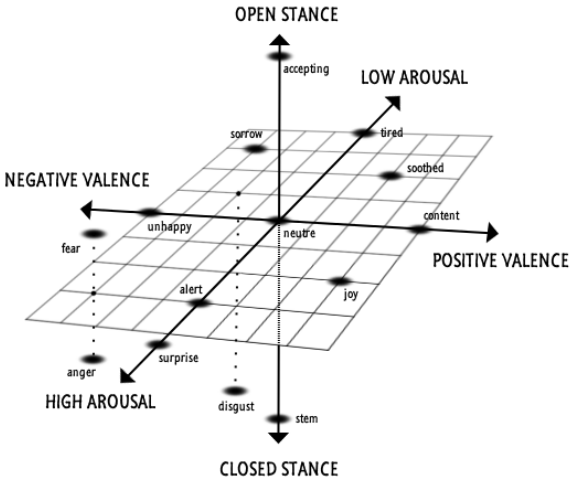

Figure 1: *Espace Valence/Arousal labellisé par Russel [Rus80].*

bale utilisé pour véhiculer certains messages qu'il est possible d'interpréter par l'observation attentive des gestes, expressions faciales et bien d'autres signaux et mouvements corporels. Van den Stock et al. [VDHdG11] ont constaté que la reconnaissance des expressions faciales est fortement influencée par l'expression corporelle. La plupart des signaux n'étant pas universels, ils doivent être interprétés en fonction du contexte, de l'émetteur, du récepteur, de la culture, etc. [JBS*09].

En analysant le mouvement global du corps (détendu ou contracté) et le mouvement du visage, des mains et des épaules, Gunes et Piccardi [GP05] arrivent à identifier les six états affectifs de base. Plus récemment, Pantic et al. [GNP11] analysent les mouvements des épaules et du visage en complément du canal audio afin d'identifier le niveau de valence et d'arousal pour en déduire l'état affectif.

En complément du geste, la posture exprime l'état psychologique : le degré d'assurance, de concentration, de maîtrise et de la situation en général. Les expériences présentées par Coulson [Cou04] suggèrent que l'interprétation de la posture du corps est comparable à la reconnaissance de la voix, et certaines postures traduisent les même émotions aussi bien que les expressions du visage. Castellano et al. [CVC07] ont présenté une approche pour la reconnaissance de l'émotion basée sur l'analyse des mouvements du corps et de l'expressivité du geste. Ils ont utilisé des données telles que l'amplitude, la vitesse et la fluidité des mouvements pour caractériser 4 émotions de base. La tristesse est représentée par une vitesse et une fluidité des mouvements lente alors que la joie est représentée par des valeurs importantes par exemple. Chen et al. [CTLM13] ont combiné MHI-HOG et Image-HOG à travers une méthode de normalisation temporelle pour décrire la dynamique du visage et des gestes du corps pour estimer l'état affectif. Hadjerci et al. [HLBD14] ont quant à eux utilisés l'information présente dans le mouvement pour estimer l'état affectif dans chacune des quatre dimensions.

On s'intéresse à l'étude des flux vidéo dans un système multi-canal, et principalement à l'étude du visage et du corps. En effet, le langage du corps est une forme de communication non verbale qui permet de véhiculer certains messages qu'il est possible d'interpréter par l'observation attentive de quelques zones bien précises. Il implique des gestes, des expressions faciales et bien d'autres signaux et mouvements corporels. Tous ces éléments font partie intégrante de la communication. De ces données, nous désirons identifier les états affectifs selon plusieurs dimensions, tout particulièrement la valence et l'arousal. Nous choisissons une représentation dimensionnelle, par rapport à une représentation discrète, car par rapport aux domaines ciblés où la durée de l'interaction est relativement longue, il est plus utile de connaître l'évolution de l'état de l'individu, que d'observer des réactions immédiates telles que la surprise, la joie, etc. Pour cela, nous explorons la caractérisation du mouvement à base de modèles de direction et de magnitude.

3. Méthodologie

Nous proposons une méthodologie organisée en trois niveaux pour la reconnaissance d'émotions : (i) Le bas niveau permet d'extraire certaines informations grâce à l'application des techniques de traitement d'images sur les flux vidéo pour en extraire les points caractéristiques, les zones en mouvement, etc. (ii) Le niveau intermédiaire englobe les descripteurs calculés à partir des caractéristiques de bas niveau tels que la trajectoire de déplacement, la vitesse moyenne, la direction moyenne du mouvement, etc. (iii) Le niveau sémantique dépend entièrement du domaine d'application. Son but est de reconstituer à partir des données du niveau intermédiaire des résultats sur l'analyse du comportement humain qui sont compréhensibles par les utilisateurs. La Figure 2 représente sous la forme d'une pyramide notre approche à trois niveaux.

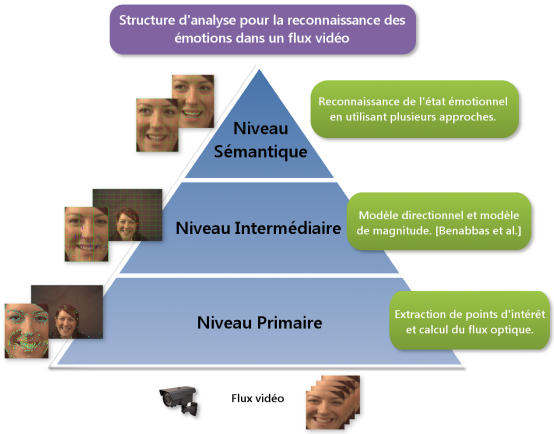

Figure 2: *Schéma en trois niveaux de l'approche de reconnaissance d'état affectif.*

3.1. Extraction des caractéristiques

Cette étape a pour but de quantifier le mouvement à partir des vecteurs de flux optique, afin d'estimer le modèle directionnel et le modèle de magnitude. Nous avons choisi d'utiliser le détecteur de coins de Harris [HS88] pour localiser les points d'intérêt. L'algorithme d'extraction des points d'intérêt de Harris est réputé pour son invariance à la rotation, au changement d'échelle, à la variation de luminosité et aux bruits dans les images. L'algorithme est rapide, ce qui convient aux applications temps réel. Il est aussi déterministe dans le sens où il retourne toujours les mêmes points d'intérêt pour une image donnée en gardant les mêmes paramètres pour l'algorithme. Ensuite, nous appliquons aux points d'intérêt la méthode d'estimation des vecteurs du flux optique KLT [LK81]. Cet algorithme nécessite comme paramètres les pixels de la première image dont nous souhaitons estimer le déplacement. Comme décrit par Baker et Matthews [BM04], l'algorithme recherche pour chaque point présent sur la première image, le point appartenant à la fenêtre de recherche qui lui correspond sur la deuxième image et qui minimise l'équation suivante :

$$\sum_{x,y}[T(x,y) - I(W(x,y;p))]^2 \quad (1)$$

où T est l'apparence du point dont on cherche la correspondance dans la deuxième image, I est la première image, (x,y) un point qui appartient à la fenêtre de recherche de correspondance, W est l'ensemble des transformations envisagées (dans ce cas, la translation) entre la première et la deuxième image et p représente l'ensemble des paramètres de la transformation.

Dans cette étude, nous nous intéressons à deux canal spécifiques : le visage et le corps. Pour chaque canal, nous employons des techniques spécifiques pour choisir les primitives (détection des points de Harris).

Concernant le visage, une étape de détection et de normalisation est nécessaire avant d'identifier les points d'intérêts. Nous appliquons l'algorithme de Viola et Jones [VJ04] pour détecter le visage dans les images. Il est ensuite nécessaire de normaliser les visages afin d'obtenir des modèles cohérents. En effet, le mouvement doit être calculé dans le même repère sur le visage, sinon les régions du visage ne sont plus agencées de la même manière entre deux images. L'algorithme de Danisman et al. [DBID10] s'appuie sur la détection des yeux afin de corriger l'orientation et la normalisation du visage, et permet de suivre le déplacement des yeux dans les images suivantes.

Il faut savoir que le visage n'est pas toujours en mouvement et que l'amplitude de ces mouvements n'est pas forcément constante. Nous cherchons à obtenir des points d'intérêt de qualité, c'est à dire sur des traits marquants du visage et avec un nombre adéquat en fonction de la région analysée.

L'étude du mouvement dans un visage est souvent liée aux contours. En effet, les régions dénuées de contours comme les joues ne sont pas porteuses d'informations. De ce fait, il est intéressant d'extraire les contours du visage pour analyser les mouvements. Pour cela, nous utilisons l'algorithme de Canny [Can86] pour extraire les contours. Parfois, des pré-traitements sont nécessaires pour augmenter la précision des contours (flou gaussien, égalisation d'histogramme, etc.).

Nous proposons une division de l'image en une grille de *M*x*N* blocs pour augmenter le niveau de précision. Ça nous permet également de couvrir autant que possible et de manière homogène le visage et de réduire fortement le temps de calcul de l'appariement des points. La taille de ces blocs et le nombre de points d'intérêt par bloc influent sur la précision du système. La sélection de ces paramètres sera étudiée dans la Section 4. La Figure 3 représente le processus d'extraction des points d'intérêt du visage.

Figure 3: *Processus d'extraction des points d'intérêts et des flux optique sur le visage.*

Afin de déterminer le nombre idéal de points d'intérêt sur le visage, nous calculons le nombre de pixels appartenant au contour dans chaque bloc et nous en déduisons le nombre de points caractérisant le mouvement. Nous appliquons ensuite le calcul du flux optique sur l'image filtrée par Canny et sur l'image suivante en niveau de gris. Cette technique permet de ne perdre aucun mouvement au niveau des contours.

Pour caractériser le mouvement du haut du corps, la méthode diffère quelque peu. En effet, la détection des contours est moins pertinente à cause de l'arrière plan. De ce fait, nous appliquons un algorithme d'extraction de silhouette (estimation des couleurs de l'arrière plan pour les enlever) afin d'obtenir uniquement les mouvements liés au corps. De la même manière, nous divisions l'image en *M*x*N* blocs afin d'augmenter la précision et de calculer le nombre de points d'intérêt par région. Ici, nous déterminons le nombre de points en fonction de la taille d'un bloc. Plus les bloc sont petits, plus le nombre de points d'intérêt par bloc diminue, ce qui a pour effet de réduire le bruit produit par un surnombre d'informations. Ensuite, nous appliquons l'algorithme du calcul des flux optique pour créer nos vecteurs de mouvement. La Figure 4 illustre le processus d'extraction des flux optique du corps.

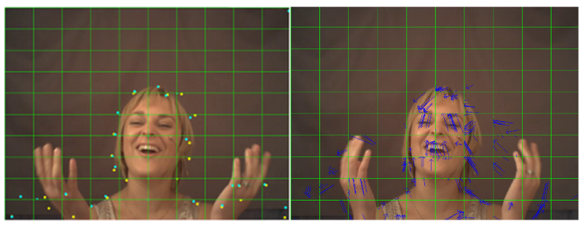

Figure 4: *Processus d'extraction des flux optique du corps.*

3.2. Modèle de direction et de magnitude

Cette étape estime le modèle directionnel et le modèle de magnitude pour l'intégralité de la séquence. Ces modèles serviront à constituer le modèle correspondant à l'évolution de l'état affectif contenu dans la séquence. Pour estimer nos modèles, nous adaptons les travaux de Benabbas et al. [BALD11], destinés aux actions, pour modéliser des changements d'états émotionnels.

La problématique de la reconnaissance d'état affectif, ne peut être résolue convenablement sans recourir aux techniques de reconnaissance d'action. Ce n'est pas le type d'action spécifique qui nous intéresse ici, mais plutôt ses caractéristiques de mouvement. Il est important d'étudier la différence entre l'étude des actions et l'étude des émotions. Pour l'un, nous recherchons à caractériser le mouvement alors que pour l'autre, nous nous intéressons plus particulièrement à l'étude du geste.

A la différence des actions, un geste est un mouvement du corps qui souligne une idée, révèle une pensée ou exprime une émotion. Les mouvements relatifs aux gestes sont plus difficiles à identifier car ils sont moins amples, plus rares et difficilement répétables. Il est donc nécessaire d'adapter les travaux liés à la détection d'actions pour parvenir à caractériser des gestes.

Après avoir calculé les vecteurs de mouvement dans chaque bloc, un algorithme de regroupement des données circulaires est appliqué aux orientations des vecteurs de flux optique. L'ensemble des *M*x*N* distributions circulaires associées est appelé "modèle directionnel". Par analogie, nous regroupons les magnitudes des vecteurs du flux optique dans chaque bloc grâce à des mélanges gaussiens. L'ensemble des mélanges gaussiens estimés représente le modèle de magnitude. Nous appliquons un seuil d'acceptation pour filtrer les données, ce qui permet d'enlever les mouvements, trop petits ou trop grands, qui ne caractérise pas un geste. La Figure 5 représente le modèle de direction et de magnitude construits à partir des flux optiques.

Figure 5: *Création du modèle de direction (b) et du modèle de magnitude (c) à partir des flux optiques (a).*

3.3. Reconnaissance

Cette étape a pour but de reconnaitre les changements d'états affectifs dans une vidéo en comparant les modèles de direction et de magnitude obtenus avec les séquences vidéos de référence. Afin d'analyser les évolutions des états affectifs, nous nous intéressons à l'étude du contexte. En effet, un état émotionnel se caractérise par une suite d'évènements

(gestes, expressions faciales et corporelles). En effectuant des regroupements sur les modèles de direction et de magnitude, nous voulons identifier une suite de mouvements afin de reconnaitre l'état affectif. Pour cela, différentes solutions sont envisageables, notamment la somme des modèles sur un intervalle prédéfini ou bien la sélection des plus grandes amplitudes dans chaque bloc de notre division. Il existe plusieurs techniques pour reconnaitre un état affectif à partir des modèles de référence. Nous en avons identifié trois.

La première solution consiste à comparer le modèle d'une séquence avec les modèles associés aux séquences de référence en utilisant une mesure de distance. L'état affectif associée au modèle ayant la distance la plus petite par rapport au modèle d'une séquence est retenue. Cette solution a l'avantage d'être rapide, et adaptée à une analyse locale (visage) et globale (corps).

La deuxième solution s'inspire des travaux de Gizatdinova et Surakka [GS07], où ils divisent le visage en 13 ROIs (Region of Interest) pour détecter les émotions. Une étude comparative [PRW*11] montre que cette solution permet d'améliorer le taux de reconnaissance par rapport à une division en *M*x*N* blocs. Cependant, cette technique s'applique uniquement au visage car à l'heure actuelle il n'existe pas de système équivalent aux AUs pour le corps.

Suite à une étude approfondie sur les unités d'action et l'évolution de l'état affectif, nous envisageons d'étendre les travaux de Gizatdinova et Surakka, en adaptant la division du visage en fonction des AUs identifiées comme pertinent. Nous voulons analyser des suites d'AUs et anticiper l'évolution de l'état affectif en fonction de l'ordre d'apparition des mouvements au sein du visage.

La troisième solution regroupe les deux solutions précédentes. L'idée est d'appliquer la solution globale sur le corps et la solution locale sur le visage. Grâce à cela, nous augmentons le taux de reconnaissance sur le visage tout en gardant les informations relatives au mouvement global.

Étant donné, que la solution deux et trois dépendent fortement de la normalisation du visage, qui fait l'objet de nos récents travaux, nous nous décidons d'étudier la première solution aussi bien sur le visage, que sur le haut du corps. Cette solution nous permet d'obtenir des premiers résultats pour analyser la pertinence de l'étude du mouvement dans la reconnaissance d'état affectif.

4. Résultats expérimentaux

Nous présentons dans cette section les résultat de l'expérimentation de notre approche sur un sous-ensemble de la base de données SEMAINE. Cette base permet d'étudier les signaux sociaux naturels qui se produisent dans des conversations entre humains et agents artificiels. Les vidéos ont été enregistrées à une fréquence de 50 images par secondes avec une résolution de 780x580 pixels. La base est composée de 31 vidéos d'apprentissage et de 32 vidéos utilisés pour les tests. La particularité de cette base réside dans le fait qu'elle est annotée en continue (traces) par au moins deux évaluateurs. L'analyse est effectuée sur l'évolution des dimensions affectives telles que l'activité (Arousal), l'anticipation, la puissance et la valence. Notre sous-ensemble d'étude est composé de 11 vidéos d'entrainement et de 8 vidéos de test provenant de deux sujets différents.

Nous nous intéressons tout particulièrement aux variations des émotions, c'est à dire aux changements de valeurs de la valence ou de l'arousal. Afin de catégoriser une émotion, nous nous inspirons du modèle de Russel, ce qui nous donne les quatre classes suivantes : Arousal +/ Valence +, Arousal +/ Valence -, Arousal -/ Valence +, Arousal - / Valence -. Chaque classe correspond à une augmentation ou une diminution de la valeur initiale, c'est à dire qu'un modèle se situant dans la classe Arousal +/Valence + n'a pas forcement une valence et une activité positives mais ses valeurs ont augmenté par rapport au modèle précédent.

Afin de valider nos résultats, nous optons pour la classification avec les SVM [CV95] puis nous utilisons LIBSVM [CL11] pour optimiser les paramètres du Kernel par rapport aux données dont nous disposons. Chaque modèle de direction et de magnitude est représenté par un vecteur. Pour cela, nous déterminons la classe du modèle en fonction de l'évolution des valeurs de l'arousal et de la valence, ce qui nous donne un label variant de 1 à 4. Concernant les index, ils correspondent aux blocs de l'image et les values, à leurs valeurs respectives. Nous décidons de réunir les données de direction et les données de magnitude dans le même classifieur afin d'obtenir une meilleure précision lors de la classification. De ce fait, tous les index impairs correspondent à la valeur de la magnitude et les index pairs à la direction. La construction de ce vecteur est représentée sur la Figure 6.

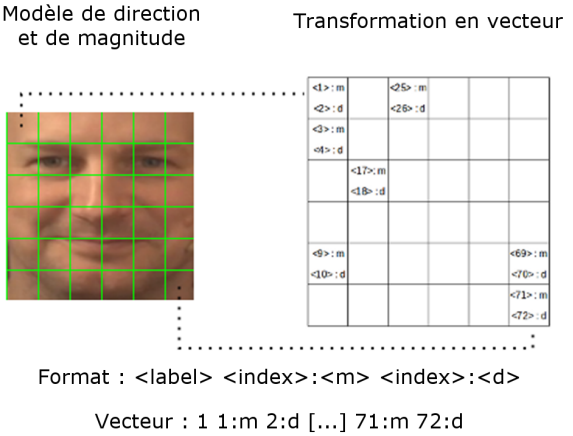

Figure 6: *Construction de vecteur à partir d'un modèle de direction et de magnitude*

La taille des vecteurs varie en fonction du nombre de blocs dans le modèle. De plus, il est possible d'ajouter d'autres données comme par exemple la puissance. Pour cela, il faut associer 3 index par bloc. Quant au nombre de vecteurs, il dépend fortement du nombre de divisions et du seuil d'acceptation appliqué aux magnitudes. En effet, plus le seuil est bas et le nombre de division élevé, plus l'algorithme construit de modèles. La Figure 7 présente le taux de reconnaissance avec une approche globale sur des visages non normalisés.

Sur la Figure 7, nous remarquons que le taux de reconnaissance varie en fonction de plusieurs paramètres. La précision du modèle est influencée par le nombre de blocs. En effet, plus la division est grande, plus il y a de points caractéristiques. Cependant, un nombre trop important de blocs peut réduire la précision du modèle car les données ne sont plus pertinentes pour calculer le mouvement. La taille des blocs dépend de la dimension des images. Lorsque nous appliquons la même division sur des images de tailles différentes, nous constatons une augmentation du taux de reconnaissance.

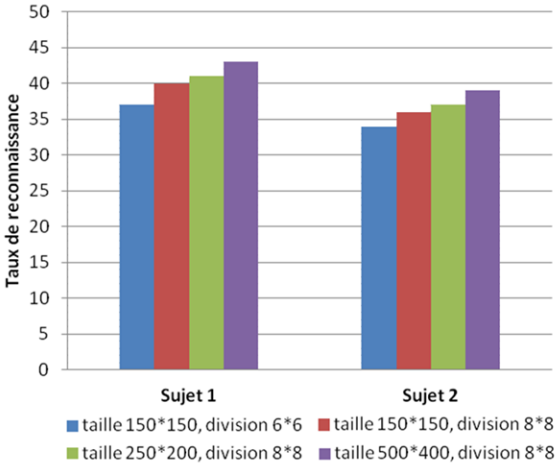

Figure 7: *Taux de reconnaissance sur le visage.*

En contrepartie, plus la taille de l'image est grande, plus le temps de calcul est long. De ce fait, il faut trouver le bon compromis entre précision et temps d'exécution par rapport à un contexte pratique spécifique. Etant donné que la normalisation n'est pas encore appliquée sur les visages, cela peut expliquer pourquoi le taux de reconnaissance n'est pas très élevé. La différence entre le Sujet 1 et le Sujet 2 est probablement liée au nombre de vidéos d'entrainement qui est plus conséquent chez le Sujet 1.

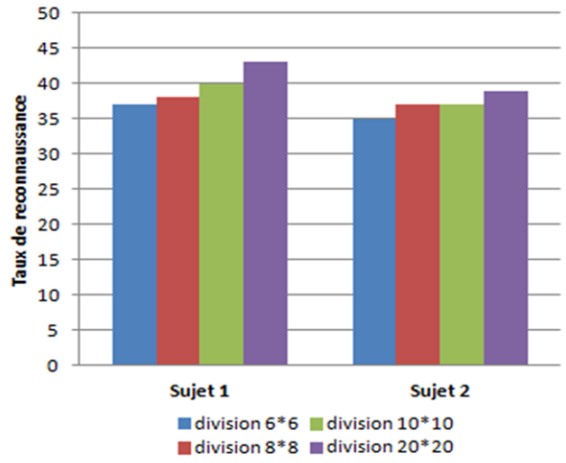

Figure 8: *Taux de reconnaissance sur le haut du corps.*

Nous faisons la même analyse sur les mouvements du corps et nous obtenons les résultats présentés sur la Figure 8. Nous remarquons que le taux de reconnaissance des mouvements du corps est plus ou moins équivalent à celui lié au visage. Au cours de ces conversations, les sujets ne bougent pas beaucoup ce qui implique que le classifieur n'arrive pas à dissocier les classes. De plus, les résultats du Sujet 2 sont inférieurs à ceux du Sujet 1 car ses vêtements se confondent avec l'arrière plan, de ce fait, nous obtenons uniquement les mouvements liés à sa tête.

4.1. Discussions

Il est important de préciser que la reconnaissance affective multicanaux ne vise pas à remplacer les expressions faciales ou les expressions corporelles, au lieu de cela, le but est d'explorer divers canaux communicatifs plus profonds et plus complets afin d'obtenir des corrélations relatives à l'état affectif.

Le langage du corps implique gestes, expressions faciales et bien d'autres signaux et mouvements corporels. Tous ces éléments font partie intégrante de la communication. Or, la complexité de chaque personnalité fait qu'il existe une multiplicité de gestes qui trahissent chacun d'entre nous. Cependant, le sens des gestes peut être interprété de plusieurs façons.

L'étude du mouvement permet donc d'identifier et de reconnaitre ces gestes. Cependant, l'interprétation des résultats n'est pas évidente. Ce problème réside dans le fait, qu'il n'existe pas une définition formelle du terme "émotion". Il est d'autant plus difficile de construire une base d'apprentissage lorsqu'on n'arrive pas à représenter l'objet de l'étude. Cette étude montre qu'il existe un nombre important de paramètres qui influencent le taux de reconnaissance. En effet, la dimension, la résolution de l'image et la division, sont des paramètres que l'on ne peut définir en amont car ils varient en fonction de la vidéo. De plus, la précision des modèles repose essentiellement sur la robustesse des algorithmes de normalisations et des descripteurs.

Au fil des résultats, nous obtenons un taux de reconnaissance meilleur sur le visage par rapport au corps. D'après ces données, nous considérons que le mouvement du corps est un complément d'information à celui du visage. Il permet de donner une intensité à l'état affectif, comme certaines AUs sur le visage.

Enfin, il existe plusieurs approches permettant d'identifier un état affectif. Il est possible de calculer un modèle à un instant précis de la vidéo, ou bien sur un intervalle. Durant l'étude nous avons constaté qu'il y a parfois des variations dans les courbes alors qu'aucun changement n'apparait dans le flux vidéo. En effet, l'état émotionnel dépend également du contexte, ce qui nous amène à construire des modèles de direction et de magnitude sur des intervalles et à identifier les mouvements déclencheurs en analysant les suites de mouvement. Bien entendu, les solutions sur les intervalles sont très variées, ce qui prouve une fois de plus la complexité et la richesse de cette étude.

5. Conclusion

Dans cet article, nous avons proposé une approche organisée en trois niveaux pour la reconnaissance d'états affectifs multi-canal. Notre système synthétise les mouvements du visage et du corps par des modèles de direction et de magnitude construits à partir des flux optiques. Les résultats expérimentaux permettent de vérifier l'apport informationnel issu du mouvement dans la reconnaissance d'états affectifs et montrent que cette solution est adaptable dans une approche globale ou locale. Dans nos futurs travaux, nous allons améliorer la normalisation des visages ce qui nous permettra d'analyser l'ensemble de la base SEMAINE. Nous analyserons également l'apport de la fusion multi-canal et de l'analyse du contexte dans l'amélioration du taux de reconnaissance.

Références

[AR92] AMBADY N., ROSENTHAL R. : Thin slices of expressive behavior as predictors of interpersonal consequences : A meta-analysis. *Psychological Bulletin. Vol. 111*, Num. 2 (1992), 256–274.

[BALD11] BENABBAS Y., AMIR S., LABLACK A., DJERABA C. : Human action recognition using direction and magnitude models of motion. In *International Conference on Computer Vision Theory and Applications (VISAPP)* (2011).

[BM04] BAKER S., MATTHEWS I. : Lucas-Kanade 20 years on : A unifying framework. *International Journal of Computer Vision (IJCV)* (2004).

[Can86] CANNY J. : A computational approach to edge detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)* (1986).

[CL11] CHANG C.-C., LIN C.-J. : LIBSVM : A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology. Vol. 2* (2011).

[Cou04] COULSON M. : Attributing emotion to static body postures : Recognition accuracy, confusions, and viewpoint dependence. *Journal of Nonverbal Behavior* (2004).

[CTLM13] CHEN S., TIAN Y., LIU Q., METAXAS D. : Recognizing expressions from face and body gesture by temporal normalized motion and appearance features. *Image and Vision Computing (IVC). Vol. 31*, Num. 2 (2013), 175–185.

[CV95] CORTES C., VAPNIK V. : Support-vector networks. *Machine Learning. Vol. 20* (1995), 273–297.

[CVC07] CASTELLANO G., VILLALBA S. D., CAMURRI A. : Recognising human emotions from body movement and gesture dynamics. *Affective Computing and Intelligent Interaction* (2007).

[DBID10] DANISMAN T., BILASCO I. M., IHADDADENE N., DJERABA C. : Automatic facial feature detection for facial expression recognition. In *5th International Conference on Computer Vision Theory and Applications (VISAPP)* (2010).

[GNP11] GUNES H., NICOLAOU M., PANTIC M. : Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. *IEEE Transactions on Affective Computing (TAC)* (2011).

[GP05] GUNES H., PICCARDI M. : Affect recognition from face and body : Early fusion vs. late fusion. *Systems, Man and Cybernetics* (2005).

[GS07] GIZATDINOVA Y., SURAKKA V. : Automatic detection of facial landmarks from au-coded expressive facial images. *International Conference on Image Analysis and Processing (ICIAP)* (2007).

[HEF02] HAGER J., EKMAN P., FRIESEN W. : The facial action coding system : A technique for the measurement of facial movement. *San Francisco : Consulting Psychologist* (2002).

[HLBD14] HADJERCI O., LABLACK A., BILASCO I. M., DJERABA C. : Affect recognition using magnitude models of motion. In *20th International Conference on MultiMedia Modeling (MMM)* (2014).

[HS88] HARRIS C., STEPHENS M. : A combined corner and edge detector. *Alvey Vision Conference* (1988).

[JBS*09] JACK R., BLAIS C., SCHEEPERS C., SCHYNS P., CALDARA R. : Cultural confusions show that facial expressions are not universal. *Current Biology* (2009).

[JSC*04] JING T. M., SCHMIDT X. K., COHN J., REED L., AMBADAR Z. : Multimodal coordination of facial action, head rotation, and eye motion during spontaneous smiles. *IEEE International Conference on Automatic Face and Gesture Recognition (FG)* (2004).

[KCY00] KANADE T., COHN J., YINGLI T. : Comprehensive database for facial expression analysis. *IEEE International Conference on Automatic Face and Gesture Recognition (FG)* (2000).

[LK81] LUCAS B., KANADE T. : An iterative image registration technique with an application to stereo vision. *International Joint Conference on Artificial Intelligence (IJCAI)* (1981).

[MVCP10] MCKEOWN G., VALSTAR M., COWIE R., PANTIC M. : The SEMAINE corpus of emotionally coloured character interactions. In *IEEE International Conference on Multimedia and Expo (ICME* (2010).

[PRW*11] POPA M., ROTHKRANTZ L., WIGGERS P., BRASPENNING R., SHAN C. : Facial action units recognition - a comparative study. *IEEE Transactions on Multimedia special issue on Multimodal Affective Iinteraction* (2011).

[Rus80] RUSSELL J. A. : A circumplex model of affect. In *Journal of Personality and Social Psychology, Vol 39(6)* (1980).

[VDHdG11] VAN DEN STOCK J., DE JONG S., HODIAMONT P., DE GELDER B. : Perceiving emotions from bodily expressions and multisensory integration of emotion cues in schizophrenia. *Social Neuroscience. Vol. 6*, Num. 5-6 (2011), 537–547.

[VJ04] VIOLA P., JONES M. J. : Robust real-time face detection. *International Journal of Computer Vision (IJCV). Vol. 57*, Num. 2 (2004), 137–154.