

Requête vidéo dans une vidéo en langue des signes : Modélisation et comparaison de signes*

F. Lefebvre-Albaret

P. Dalle

Institut de Recherche Informatique de Toulouse
Université Paul Sabatier, 118 Route de Narbonne
F-31062 TOULOUSE CEDEX 9
lefebvre@irit.fr dalle@irit.fr

Résumé

Cet article présente une méthode originale de recherche d'une occurrence d'un signe dans une vidéo en Langue des Signes (LS). Bien que la LS soit une langue "parlée", la difficulté d'appliquer les méthodes traditionnelles du traitement automatique de la parole à ce domaine implique la mise au point de nouvelles approches. La méthode que nous présentons est basée sur une modélisation paramétrique du mouvement des signes. La requête est formulée en signant un exemple du signe à retrouver dans la vidéo. Le système utilise une caractérisation des signes basée sur le mouvement projeté des mains et de la tête. Les signes isolés sont regroupés en catégories en extrayant plusieurs paramètres de la position, des trajectoires et de la dynamique des mains et de la tête. Nous présentons les différents filtres permettant d'extraire les valeurs de ces paramètres et de calculer un score de similarité entre deux signes. Nous présentons enfin l'architecture du système de requête basé sur ces outils et évaluons le gain qu'il apporte à l'utilisateur pour l'assister dans sa recherche. Le problème traité peut être perçu plus largement comme la recherche d'une trajectoire en trois dimensions et pourrait trouver d'autres applications dans l'indexation automatique de vidéos ou la communication homme machine.

Mots Clef

Langue des Signes Française, LSF, Requête vidéo, Patron de signe.

Abstract

This article presents a new method to find a sign occurrence in a Sign Language (SL) video. Although the sign language is an oral language, it is hard to apply traditional methods dedicated to automatic speech processing. It is then necessary to develop new approaches. Our method is based on a parametric model of sign movements. The query consists in an example of the sign to look for in the target video. Our algorithm is only based on the projected trajectory of the head and the two hands in the videos. We present several

filters to extract the parameter values and compute a similarity score between two signs. We finally describe how these tools can be embedded in a system to assist the user in a sign query. We also quantify the amount of time saved in using our tool. The system could also be applied to look for gesture occurrences in the framework of coverbal gesture or human computer interaction.

Keywords

French Sign Language, LSF, Video query, Sign Pattern.

1 Introduction

Dans le domaine du traitement informatique du signal, les LS occupent une place à part à l'interface du traitement automatique du geste et de celui de la parole. Outre le fait que des paramètres très différents les uns des autres (configuration, emplacement, mouvement et orientation des mains, expression du visage, regard ...) sont impliqués dans la production de signes, la difficulté provient de la grande variabilité de réalisation d'un même signe.

Nous traitons dans cet article le problème de recherche d'une occurrence d'un signe dans une vidéo en Langue des Signes Française (LSF). Les vidéos cibles (dans lesquelles on souhaite effectuer la recherche) sont des narrations libres et des traductions de brèves d'actualité dont le lexique et les structures des énoncés n'ont pas été contraints. Ces vidéos ne comportent aucun marqueur additionnel pour faciliter le suivi des différentes parties du corps du signeur.

Nous abordons dans un premier temps le problème de modélisation du signe en LSF puis nous montrons ensuite comment la prise en compte de la variabilité des signes nous a amenés à modifier des méthodes dédiées au traitement automatique des LS. Nous exposons ensuite une méthode de caractérisation paramétrique des signes que nous utilisons dans le cas de requêtes vidéos. Nous évaluons enfin l'approche et quantifions le gain de temps obtenu lors du traitement de la requête.

2 Modélisation linguistique

Dès lors qu'on s'intéresse à la recherche d'un signe, il est nécessaire de définir une modélisation des signes.

* Cette étude est financée par la société Websourd et la région Midi-Pyrénées

Plusieurs approches ont été proposées par les linguistes et les informaticiens pour sélectionner les caractéristiques pertinentes pour l'identification parmi le flot d'informations disponibles.

Stokoe, dans [17] est l'un des premiers linguistes proposant une description des signes. Il dégage trois paramètres impliqués dans la caractérisation des signes : la configuration (posture de la main), le mouvement et l'emplacement des mains.

Par la suite, Liddell et Johnson [13] proposeront un modèle de signe sous forme d'un enchaînement de phases de mouvement et de tenues (modèle multi-segmental). Ce modèle permet de rendre compte de la synchronisation des paramètres lors de la production de signes. Durant chaque phase, les paramètres du signe sont décrits séparément. Cette représentation permet de mettre en évidence de nombreuses régularités dans l'organisation temporelle des signes (répétitions, dynamique des signes) et dans les relations entre les mains droites et gauches (symétries). N'importe quel type de signes peut être modélisé, même le plus improbable ayant une structure incompatible avec les structures des LS.

A cette modélisation multi-segmentale s'oppose une modélisation mono-segmentale proposée par [4]. L'auteur se base entre autre sur la constatation que les structures de répétitions à l'identique dans les signes de l'ASL sont beaucoup plus importantes que dans les mots anglais (respectivement 50% et 1%). Le signe est caractérisé par un ensemble de traits tels que la répétition, les relations main gauche/main droite, la direction du mouvement, l'emplacement. Ce modèle est beaucoup plus synthétique que le précédent car il ne modélise que des signes plausibles, incluant cependant la majorité des signes de l'ASL.

Actuellement, la majorité des algorithmes de traitement automatique des LS basés sur les modèles de Markov et le Dynamic Time Warping mettent implicitement en œuvre un modèle de type multi-segmental. Au contraire, les modèles de génération incluent de plus en plus de paramètres comme la symétrie ou les répétitions qui font plus référence au modèle mono-segmental [7] [14].

Nous nous plaçons dans le cadre de reconnaissance d'un signe isolé dont on ne dispose que d'une réalisation sous forme vidéo. Les données issues du suivi 2D dans l'image des mains dans cette vidéo sont parfois très bruitées. Pour cette raison, nous avons choisi d'exploiter le modèle mono-segmental de description des signes qui contraint davantage la forme du signe à reconnaître et apporte ainsi plus de robustesse.

Nous nous sommes toutefois assurés que les hypothèses à la base de ce formalisme permettaient bien de représenter le mouvement des signes de la LSF à partir du dictionnaire [15] comportant 4027 signes (hors signes composés). Comme R. Channon [4], nous observons une fréquence importante des répétitions (35% des signes) et des symétries entre mains gauche et droite (32%).

3 Méthodes de traitement existantes

De nombreux travaux ont déjà été effectués dans le champ de la reconnaissance automatique de signes. Les meilleurs résultats ont été obtenus à partir de signaux obtenus par des techniques de capture de mouvement [8][19] où jusqu'à 5000 signes peuvent être correctement segmentés avec près de 90% d'identification correcte. Cependant, dans le cas de données obtenues par des systèmes mono-vision dont on ne possède pas les caractéristiques techniques (vidéo obtenues sur le web par exemple), ce qui est le cas de nos travaux, des taux de reconnaissance similaires (90%) ne parviennent à être obtenus que sur des effectifs ne dépassant pas quelque centaines de signes. La majorité des études tentent d'adapter les techniques dédiées au traitement automatique de la parole à la reconnaissance de signes à l'instar de [21]. Les principales familles de méthodes utilisées sont les Modèles de Markov Cachés (MMC) [21][16][2][18][3][6] et le Dynamic Time Warping (DTW) [9] [1] qui sont basées sur la programmation dynamique.

Il est intéressant de noter que l'adaptation de ces méthodes de reconnaissance traditionnelles s'oriente de plus en plus vers une prise en compte de la nature paramétrique des signes :

- Une première approche consiste à dissocier les différentes composantes du signe (mouvement et configuration de chaque main) qui sont traitées séparément mais synchronisées. C'est le cas dans [18][3][6]. Ces méthodes réduisent la taille des modèles utilisés dans le cadre de la reconnaissance.
- Une autre approche tend à grouper les signes de structures similaires. Ainsi [10] utilise des MMC pour modéliser des primitives de mouvement récurrentes en Langue des Signes Coréenne. La modification de la structure des chaînes de Markov permet également de modéliser les répétitions. C'est également la méthode que nous avons retenue dans [11] pour mettre au point un outil de segmentation en signes où nous prenons également en compte la notion de relation de symétrie entre les deux mains.
- Une dernière approche paramétrise les modèles utilisés par les DTW ou les MMC. C'est le cas dans [20] où les états prennent en compte une variation d'orientation et d'amplitude du signe. [1] prend aussi en compte la translation d'un signe en différents points de l'espace de signation.

4 Méthode de traitement retenue

Ces évolutions récentes sont fondées sur le plan linguistique (cf §2) et montrent des performances meilleures que les techniques traditionnelles. Il nous a paru intéressant de combiner la prise en compte de la synchronisation durant la réalisation du signe, la catégorisation des signes et la prise en compte de leur variabilité, dans un même système de caractérisation de signe.

La méthode que nous proposons est basée exclusivement sur le mouvement projeté des centroïdes des mains et de la tête. Nous nous plaçons donc dans un espace de dimension 3 (2D+t). Ces points ont été suivis au préalable grâce à la méthode décrite dans [12]. Nous n'intégrons pas encore les configurations et orientations manuelles ainsi que les autres paramètres non manuels pour la caractérisation du signe bien que leur intérêt soit évident. Le problème de reconnaissance de ces paramètres est d'ailleurs actuellement un problème ouvert. Nous souhaitons donc évaluer les performances que l'on peut atteindre en utilisant uniquement le mouvement des mains et de la tête. Comme de nombreux signes ne diffèrent que par les configurations et orientation manuelles, notre but n'est pas de reconnaître les signes, mais plutôt de définir une similarité entre plusieurs signes basée sur le mouvement, pour permettre ainsi de réduire l'espace de recherche d'une occurrence de signe dans une vidéo à l'instar de [1].

Dans cet esprit, nous utilisons un modèle paramétrique pour caractériser des signes isolés en utilisant les paramètres suivants : emplacements relatifs des mains, emplacement des mains par rapport à la tête, contact, géométrie du signe, dynamique du signe, type de relation main gauche/main droite, orientation et amplitude. Les filtres de détection de paramètres retenus sont appliqués à chaque intervalle temporel $[t_1, t_2]$ de la vidéo mais n'autorisent pas de déformation non linéaire du temps contrairement aux méthodes HMM et DTW. Ceci est intentionnel car nous pensons qu'une simple dilatation temporelle des signes rend mieux compte de la structure régulière observée dans les signes à répétitions, très fréquents en LSF (cf. 2).

Nous exposerons dans un premier temps les filtres utilisés dans le cadre de la classification. Nous montrerons ensuite comment les résultats des différents filtres peuvent être combinés pour obtenir des scores d'appartenance d'un signe à une catégorie de mouvement ou pour évaluer la similarité de deux signes.

5 Caractérisation des signes

Notre système utilisé reçoit en entrée les coordonnées 2D des mains droite $(X_r(t), Y_r(t))$, gauche $(X_l(t), Y_l(t))$ et de la tête $(X_h(t), Y_h(t))$ mesurées durant l'étape de suivi de la vidéo. La caractérisation du type de mouvement est réalisée grâce aux vitesses instantanées des mains droite et gauche. Ces vitesses sont ré-échantillonnées de manière à ce que chaque signe soit caractérisé par N vitesses instantanées pour chaque main (nous avons choisi $N = 16$ dans notre implémentation). Ces vitesses seront notées $(V_{x_r}(i), V_{y_r}(i))$ et $(V_{x_l}(i), V_{y_l}(i))$ par la suite. Les traitements employés dans l'algorithme sont de deux types :

- Les **filtres de catégorisation** (découpage discret) permettent de donner une mesure d'appartenance d'un segment temporel de vidéo à une catégorie de signe.

Pour le moment, nous ne prenons en compte que 6 cas : les mouvements balistique, aller/retour, balistique répété, circulaire, circulaire répété et angulaire, car ils sont parmi les plus représentés en LSF. Nous distinguons également les signes où une seule main est en mouvement de ceux où les trajectoires des deux mains sont symétriques. Nous prenons donc en compte en tout $6 \times 2 = 12$ catégories.

- Les **opérateurs de comparaison** indiquent la similarité de deux signes à partir d'un de leurs paramètres. Ces opérateurs portent sur des paramètres dont la valeur évolue de manière continue. Les différents opérateurs de comparaison que nous utilisons portent sur l'amplitude, l'orientation, la position relative des mains, l'emplacement du signe.

5.1 Filtres de catégorisation

Les filtres servant à la catégorisation d'un signe prennent en compte la géométrie, la dynamique du signe et les relations entre les mouvements des mains droite et gauche. Chacun de ces filtres prend en entrée les vitesses ré-échantillonnées des deux mains et retourne une mesure de confiance comprise entre 0 et 1.

Les filtres ont pour l'instant été créés manuellement et correspondent à la dynamique de signe modélisée par [14]. Nous envisageons d'affiner ces modélisations à l'aide d'un apprentissage basé sur des exemples de signes appartenant aux différentes catégories.

Le **filtre de géométrie** permet de caractériser la forme de la trajectoire manuelle. Il est constitué d'un tableau de N angles $[\alpha_1, \alpha_2 \dots \alpha_N]$ qui représentent les orientations des vitesses discrétisées du modèle de signe. La mesure d'appartenance du signe à la catégorie géométrique peut être calculée de la manière suivante dans un repère (\vec{u}, \vec{v}) :

$$\vec{Lin} = \sum_{i=1}^N (V_x(i)\cos(\alpha_i) + V_y(i)\sin(\alpha_i)) \vec{u} + \sum_{i=1}^N (V_y(i)\cos(\alpha_i) - V_x(i)\sin(\alpha_i)) \vec{v}$$

Ceci revient à "dérouler" la trajectoire du geste en appliquant des changements de repère aux vitesses instantanées.

$$GS = \frac{\|\vec{Lin}\|}{\sum_{i=1}^N \sqrt{V_x^2(i) + V_y^2(i)}}$$

On notera que ce score est indépendant des rotations du signe dans le plan 2D de la vidéo.

Il est alors possible de déterminer l'orientation du mouvement de la main dominante¹ grâce à la mesure $\theta = (\vec{u}, \vec{Lin})$

Une fois l'orientation du signe déterminée, il est possible de projeter les vitesses instantanées sur la trajectoire du modèle de signes de cette catégorie.

$$V'(i) = V_x(i).\cos(\alpha_i - \theta) + V_y(i).\sin(\alpha_i - \theta)$$

¹main droite pour un droitier

On peut alors appliquer un **filtre de dynamique** qui permet de caractériser le profil des vitesses à l'intérieur du signe. Le filtre est constitué d'un profil type de vitesses discrétisé en N points $[V_p(1), V_p(2) \dots V_p(N)]$. La mesure d'appartenance du signe à la catégorie dynamique peut être calculée avec le produit scalaire normé :

$$DS = \frac{\sum_{i=1}^N V'(i) \cdot V_p(i)}{\sqrt{\sum_{i=1}^N V'^2(i)} \sqrt{\sum_{i=1}^N V_p^2(i)}}$$

D'autres mesures permettent de caractériser la dépendance des deux mains. Comme l'ont déjà mis en évidence plusieurs études comme [7], la grande majorité des signes est soit effectuée avec une seule main mobile, soit avec deux mains symétriques. Il convient toutefois d'être prudent, car du fait de la projection du mouvement 3D en 2D, les types de relations de symétrie entre les deux mains sont susceptibles de varier en fonction de l'orientation du signeur (cf fig. 1).



Figure 1: Biases de projection

Le filtre suivant MS permet de détecter une **main statique** durant l'exécution du signe.

$$MS = \frac{\sum_{i=1}^N (\|\vec{V}_r(i)\| - \|\vec{V}_l(i)\|)}{\sum_{i=1}^N (\|\vec{V}_r(i)\| + \|\vec{V}_l(i)\|)}$$

Le filtre SYM permet de détecter une éventuelle **relation de symétrie entre les deux mains** ($V_{r_x}(i) \approx \pm V_{l_x}(i)$ et $V_{r_y}(i) \approx \pm V_{l_y}(i)$).

$$C_x = \frac{sgn(\sum_{i=1}^N V_{r_x}(i) \cdot V_{l_x}(i)) \sqrt{\sum_{i=1}^N V_{r_x}(i) \cdot V_{l_x}(i)}}{\sqrt{\sum_{i=1}^N \max(V_r^2(i), V_l^2(i))}}$$

La mesure C_y est calculée de la même manière.

$$SYM = \sqrt{C_x^2 + C_y^2}$$

Comme on peut le voir sur le graphique 2, il est possible de déduire à partir des signes de C_x et de C_y le type de symétrie impliqué dans le signe. Toutefois, nous préférons utiliser une mesure d'angle continue σ , moins affectée par le biais de projection.

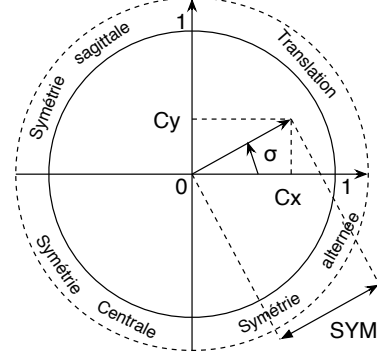


Figure 2: Types de symétries

A partir des filtres ci-dessus, nous pouvons obtenir une mesure de confiance qui indique si un signe appartient à une catégorie particulière. Les résultats des différents filtres sont fusionnés de manière probabiliste. On note respectivement (DS_r, GS_r) et (DS_l, GS_l) les mesures de confiance obtenues pour les mains droite et gauche. Les scores d'appartenance d'un signe à une catégorie sont :

Pour un signe effectué avec la main droite :

$$SCORE = GS_r \cdot DS_r \cdot MS$$

Pour un signe effectué avec la main gauche :

$$SCORE = -GS_l \cdot DS_l \cdot MS$$

Pour un signe effectué avec les deux mains :

$$SCORE = SYM \cdot \sqrt{GS_r \cdot DS_r \cdot GS_l \cdot DS_l}$$

5.2 Opérateurs de comparaison

D'autres mesures que nous ne détaillerons pas sont introduites pour évaluer la similarité entre deux signes. Elles permettent chacune une comparaison d'un paramètre du signe requête avec le signe cible et retournent une mesure de ressemblance comprise entre 0 et 1 qui indique le degré de conservation du paramètre entre le signe requête et le signe recherché:

AMP : Amplitude du mouvement,

CONT : Contact entre les deux mains,

TR : Placement relatif des deux mains,

PL : Placement des mains par rapport à la tête.

OR : Orientation du signe (mesurée par l'angle θ).

TS : Type de symétrie (mesurée par l'angle σ).

5.3 Mesure de similarité entre deux signes

Le score de similarité du signe requête S_s avec un segment temporel de la vidéo cible sera calculé en utilisant une fusion des résultats des filtres de catégorisation et des opérateurs de comparaison. La manière dont sont fusionnés les résultats dépend de la catégorie du signe à rechercher. Nous obtenons en tout 4 formes de fusions différentes. L'exemple ci-dessous correspond au cas d'un signe où les deux mains effectuent un mouvement balistique et sont liées par une translation. Dans le calcul du score de similarité suivant, les mesures de confiances

$GS_r, DS_r, GS_l, DS_l, SYM$ concernent le segment à caractériser avec la catégorie cat_s :

$$SCORESIM = (SYM * GS_r * DS_r * GS_l * DS_l) * (CONT * AMP * TR * PL * OS * TS)$$

La mesure de ressemblance finale entre le signe requête S_s et le segment temporel de la vidéo cible est donc calculé par un produit des résultats des filtres de catégorisation et des opérateurs de comparaison .

6 Système de requête vidéo

Le problème que nous cherchons à résoudre consiste à trouver dans une vidéo V_t toutes les occurrences $S_t(j)$ d'un signe S_s présent dans une vidéo source V_s . Plutôt que de résoudre directement ce problème comme dans [1], nous choisissons de le scinder en plusieurs étapes :

A - Caractérisation du signe requête S_s , détermination de sa catégorie cat_s

1. La vidéo requête comportant le signe requête est enregistrée et on effectue le suivi des deux mains du signeur et de sa tête par la méthode décrite dans [12].
2. Des scores sont calculés pour chaque intervalle temporel $[t_1, t_2]$ et chaque catégorie cat de signe.
3. Les différentes propositions classées par score sont présentées sous forme d'images de signe requête (à la manière de figure 1). L'utilisateur peut alors choisir d'après la forme des flèches la catégorie cat_s du signe et la relation entre les deux mains qui correspondent le mieux au signe recherché.

B - Recherche dans la vidéo

1. Un suivi de la tête et des mains est effectué dans la vidéo-cible V_t .
2. Les filtres de la catégorie cat_s et les opérateurs de comparaison sont appliqués sur tous les intervalles temporels de V_t de moins de T_{max} (2 s dans notre implémentation).
3. Les segments de vidéo sont présentés à l'utilisateur en les classant par score de similarité avec le signe source.

7 Evaluation

Le but de notre recherche étant de faciliter la requête d'un signe dans une vidéo signée, il est donc naturel de vérifier dans quelle mesure notre chaîne de traitements permet à l'utilisateur de gagner du temps en ne regardant que les parties de la vidéo cible susceptibles de contenir une occurrence du signe recherché. Pour avoir un panel représentatif des différentes catégories de productions en LSF, les vidéos seront à la fois tirées de récits libres et de traductions de brèves fournies par la société Websourd. L'évaluation a consisté à émettre 103 requêtes différentes. Certains signes étant présentés plusieurs fois, il y avait 178 bonnes

réponses possibles. Nous avons volontairement écarté les signes de moins de 3 frames (soit 0.1s) qui sont difficilement identifiables en l'absence de contexte, même par un signeur. Par contre, bien que notre méthode soit basée sur le mouvement, nous n'excluons pas les signes dont la dynamique apparente est faible car ils représentent une partie significative du lexique et car ils peuvent être caractérisés par des critères comme la position relative des mains ou l'emplacement du signe. Pour chaque requête, notre algorithme extrait chaque segment temporel de la vidéo cible et les classe suivant leur taux de similarité avec la requête vidéo. On note à chaque fois le rang relatif de la proposition correcte² correspondant à la requête. Le diagramme qui suit présente les effectifs de rangs relatifs obtenus³.

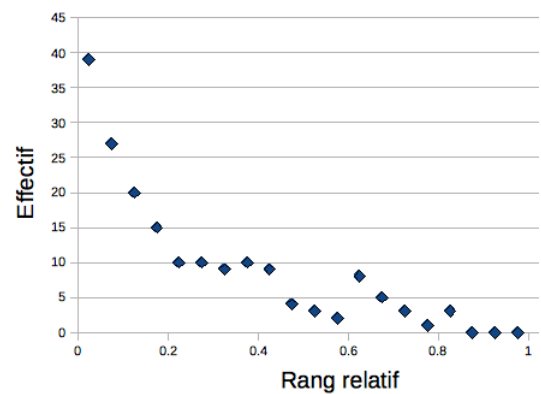


Figure 3: Effectif par classe de rang relatif de la bonne réponse à la requête

D'après les résultats, près de 50% des réponses aux requêtes sont présentes dans les premiers segments temporels présentés à l'utilisateur représentant seulement 15% du nombre de segment total. Il est possible de déduire de l'allure de la courbe que la recherche sera environ deux fois plus rapide en visualisant les segments proposés par notre algorithme dans l'ordre de classement. Cette moyenne couvre toutefois d'énormes disparités en fonction de la catégorie de signe. Les signes amples comme le signe [Immeuble] présent dix fois dans notre corpus de test peut être localisé en moyenne 15 fois plus vite grâce à notre méthode. Les signes impliquant un mouvement moins ample sont beaucoup plus difficiles à localiser dans la vidéo. On peut émettre l'hypothèse que d'autres paramètres comme l'orientation de la main et la configuration sont dans ce cas nécessaires pour obtenir une caractérisation satisfaisante.

Lors de notre évaluation, nous avons observé des modifications systématiques de paramètres entre les signes filmés dans la requête et les signes correspondant dans la vidéo de

²Le segment extrait de la vidéo est considéré comme correct lorsqu'il contient plus de la moitié d'un signe correspondant à la requête vidéo.

³Le rang relatif correspond au rang de la réponse correcte divisé par le nombre de segments de vidéos présentés à l'utilisateur.

corpus test. En particulier, il est fréquent que les signes à répétition perdent une de leur répétition lors de leur utilisation dans un énoncé signé. Malgré tout, le fait que de multiples critères soient utilisés pour caractériser le signe permet à notre algorithme de donner des résultats satisfaisants même dans ce cas.

8 Conclusions et perspectives

Les résultats obtenus permettent déjà d'envisager des applications à très court terme pour naviguer plus rapidement à l'intérieur de longs documents en LSF. L'architecture envisagée peut également permettre par la suite d'intégrer d'autres paramètres que le mouvement comme la forme de la main ou l'expression du visage dans le processus de recherche. Il serait cependant intéressant d'affiner les filtres utilisés par des mécanismes d'apprentissage.

Comme l'ont montré des études comme [5], les contraintes de symétrie ou de répétition se trouvent également dans d'autres types de gestes impliqués dans la communication comme les gestes coverbaux. On peut donc également utiliser nos algorithmes dans ce cadre. De même, la modélisation paramétrique de geste peut également apporter un gain significatif dans le domaine des interfaces homme machine [20]. Une adaptation de notre algorithme à ces deux domaines permettrait de mieux prendre en compte la variabilité dans la production d'unités gestuelles.

References

- [1] J. Alon, *Spatiotemporal gesture segmentation*, Mémoire de thèse, université de Boston, 2006.
- [2] B. Bauer et K. Kraiss, *Video-Based Sign Recognition Using Self-Organizing Subunits*, ICPVR, Quebec, 2002.
- [3] M. Brand, N. Oliver, et A. Pentland, *Coupled hidden markov models for complex action recognition*, IC-CVPR, Porto Rico, 1997.
- [4] R.E. Channon, *Signs are single segments: phonological representations and temporal sequencing in ASL and other sign languages*, mémoire de thèse, université du Maryland, 2002.
- [5] R.E. Channon, *The Symmetry and Dominance Conditions Reconsidered*, Journal Proceedings from the Annual Meeting of the Chicago Linguistic Society, Chicago Linguistic Society, vol. 40, pp 45–47, 2008.
- [6] J. Deng et H.T. Tsui, *A Two-step Approach based on PaHMM for the Recognition of ASL*, ACCV, Melbourne, 2002.
- [7] M. Filhol *Modèle descriptif des signes pour un traitement automatique des langues des signes*, mémoire de thèse, Université Paris-1, 2008.
- [8] W. Gao, G. Fang, D. Zhao, et Y. Chen. *Transition movement models for large vocabulary continuous sign language recognition*, Face and Gesture, Séoul, pp. 553-558, 2004.
- [9] U. Han, G. Awad et A. Sutherland, *Subunit Boundary Detection for Sign Language Recognition Using Spatio-temporal Modelling*, ICVS, Bielefeld, 2007.
- [10] J.B. Kim, K.H. Park, W.C. Bang, J.S. Kim et Z. Bien, *Continuous korean sign language recognition using automata based gesture segmentation and hidden markov model*, ICCAS, Jeju, Corée, 2001.
- [11] F. Lefebvre-Albaret, P. Dalle, *Une approche de segmentation de la Langue des Signes Française*, TALN, Avignon, 2008.
- [12] F. Lefebvre-Albaret, P. Dalle, *Analyse de vidéos en langue des signes : méthodes et stratégies*, ORASIS, Trégastel, 2009.
- [13] S.K. Liddell and R.E. Johnson, R.E. *American Sign Language: the phonological base*, Sign Language Studies, vol. 64, 1990.
- [14] O. Losson *Synthèse du geste communicatif. Application à la Langue des Signes Française*, mémoire de thèse, université de Lille, 2000.
- [15] B. Moody, *La Langue des Signes*, IVT, tome 2 et 3, 1997.
- [16] Y. Pei, T. Starner, H. Hamilton, I. Essa, J.M. Rehg, *Learning basic units in American Sign Language using discriminative segmental feature selection*, ICASSP, Taipei, 2009.
- [17] W.C. Stokoe, D. Casterline, C. Croneberg, *A dictionary of American Sign Language on Linguistic principles*, Linstok Press, 1978.
- [18] C. Vogler et D. Metaxas, *Parallel hidden Markov models for American sign language recognition*, ICCV, Kerkyra, Grèce, 1999.
- [19] C. Vogler and D. N. Metaxas, *Handshapes and movements: Multiple-channel American Sign Language recognition*, Gesture Workshop, Genève, pp. 247-258, 2003.
- [20] A.D. Wilson et A. Bobick, *Hidden Markov models for modeling and recognizing gesture under variation*, International Journal on Pattern Recognition and Artificial Intelligence, Vol. 15, pp. 123-160, 2001.
- [21] M. Zahedi, P. Dreuw, D. Rybach, T. Deselaers, J. Bungeroth et Hermann Ney, *Continuous Sign Language Recognition – Approaches from Speech Recognition and Available Data Resources*, LREC Workshop on the Representation and Processing of Sign Languages, Genève, pp. 21-24, 2006.