

## Une approche de segmentation de la Langue des Signes Française \*

François Lefebvre-Albaret, Patrice Dalle  
IRIT - Université Paul Sabatier, 118 Route de Narbonne  
31062 Toulouse cedex 9  
lefebvre@irit.fr, dalle@irit.fr

**Résumé.** L'article présente une nouvelle méthode de segmentation temporelle des vidéos en Langue des Signes Française (LSF). Nous commençons par un bref état de l'art sur les méthodes actuelles de traitement des vidéos en LSF. Après avoir expliqué pourquoi de telles méthodes de segmentation basées sur la reconnaissance des signes sont inadaptées aux applications que nous envisageons, nous listons différentes approches utilisées dans le cadre du traitement du signal et montrons comment elles peuvent être appliquées à la segmentation de la LSF. Nous présentons notre algorithme de segmentation basé sur le traitement des mouvements apparents des mains. Les résultats sont analysés pour mettre en évidence les différentes sources d'erreurs de segmentation. Nous abordons finalement la manière dont d'autres paramètres comme les configurations manuelles et les expressions du visage pourraient être utilisées pour améliorer leur précision. Nous mettons également en évidence plusieurs applications que pourrait avoir un tel travail dans les domaines de la pédagogie ou de la synthèse de la LSF par signeur virtuel.

**Abstract.** The article presents a new way of segmenting French Sign Language (LSF) video sequences. Starting from a short literature review on LSF video processing. We first explain why the recognition-based segmentation is not adapted to our application. Then, we list several approaches used for and show how those methods could be applied to sign language video segmentation. We describe the segmentation algorithm based on hand motion analysis during the sign realisation. The segmentation results are analysed to point out the main error sources. At the end, we show how the analysis of hand configurations and face expressions could be used to increase the segmentation accuracy. We also give rise to all application of our work for the french sign language pedagogy or for its use with a signing avatar.

**Mots-clés :** Langue des Signes Française, Segmentation, Analyse des mouvements, Programmation dynamique.

**Keywords:** French Sign Language, Segmentation, Motion analysis, Dynamic programming.

---

\* Cette étude est réalisée dans le cadre du projet SESCA (Systèmes pour l'Enseignement de la LSF et la Communication par Avatar : <http://www.irit.fr/sesca>) et est financée par la région Midi-Pyrénées et par la société Websourd (<http://www.websourd.org>).

## 1 Introduction

L'étude de vidéos en Langue des Signes Française (LSF) nécessite une première étape de segmentation temporelle. Ce découpage de l'énoncé signé peut déjà être effectué manuellement grâce à des outils d'annotation comme ANVIL (Kipp, 2003) ou ANCOLIN (Braffort *et al.*, 2001) mais il pose encore de nombreux problèmes. Outre le temps important nécessaire pour la segmentation, le résultat est extrêmement variable car chaque personne a ses propres critères pour évaluer le début et la fin de signe. Ceci nous conduit à nous questionner sur la problématique de segmentation et à proposer une approche originale permettant d'accélérer cette étape de traitement de la vidéo.

Le présent article dressera dans un premier temps un bref état de l'art de différentes méthodes qui ont déjà été utilisées dans le cadre d'analyse de vidéo en Langue des Signes (LS). Après avoir montré en quoi le problème de segmentation temporelle d'un énoncé en LS nécessite la mise au point d'une autre approche qui prendrait mieux en compte les observations sur la structure des unités gestuelles (Channon, 2002), nous décrirons les différentes méthodes utilisées en segmentation du signal et nous montrerons quelles adaptations sont nécessaires pour les utiliser dans le cadre de la segmentation d'une vidéo en LSF.

Nous exposerons ensuite une première mise en œuvre de notre méthode de segmentation basée sur l'analyse du mouvement effectué lors de la réalisation d'un signe.

Le problème de l'évaluation de la segmentation temporelle sera soulevé pour définir des critères permettant de qualifier l'exactitude du découpage temporel d'une vidéo en signes. Ces critères seront utilisés pour évaluer notre algorithme de segmentation. Les sources d'erreurs seront enfin examinées de manière à dégager les optimisations nécessaires du traitement proposé.

L'article s'achève sur un ensemble de perspectives d'utilisation de nos résultats à des buts pédagogiques et linguistiques. Nous évoquerons également les perspectives d'exploitation des informations issues de la segmentation pour l'animation de signeurs virtuels.

## 2 Approches existantes d'analyse de vidéo en LS

Actuellement, la plupart des équipes de recherches analysant les vidéos en LSF se focalisent sur le problème de la reconnaissance de signes isolés ou intégrés dans des phrases.

- Certaines représentations modélisent un signe comme une succession de mouvements et de tenues (Liddell & Johnson, 1990). Ces représentations peuvent être utilisées pour segmenter les énoncés en LS. C'est notamment ce qui est réalisé dans (Kim *et al.*, 2001). Le procédé de segmentation des vidéos en LS coréenne est basé sur des modèles de Markov à 5 états (repos, préparation, mouvement, rétraction/répétition, fin). Les résultats semblent très satisfaisants (95 % de réussite) mais on ne dispose malheureusement pas du protocole d'évaluation. Outre la segmentation, l'algorithme proposé permet également de caractériser les différentes primitives de mouvement : cercle, arc de cercle, ligne droite.
- Les méthodes les plus employées actuellement dans le domaine des traitements de vidéos en LS exploitent des Modèles de Markov Cachés. C'est notamment cette méthode qui est utilisée dans l'étude (Bauer & Hienz, 2000) où un taux de reconnaissance de 92% est atteint sur un corpus de 97 signes inclus dans différents énoncés. L'identification des mains dans

la vidéo est facilitée par l'utilisation de gants colorés. Notons que le taux de reconnaissance chute à 53% en utilisant uniquement les positions des mains.

- Un exemple plus récent est le travail de (Bowden *et al.*, 2004). Le processus proposé pour reconnaître les signes se déroule en plusieurs étapes. Après un suivi des mains, des épaules et une caractérisation des configurations, chaque instant de la réalisation du signe est caractérisé par une position relative des mains, l'emplacement des deux mains, le mouvement de chaque main et la configuration (S, G, B, H ou V). Un Modèle de Markov Caché combiné à une analyse en composantes principales de ces paramètres permet un taux de reconnaissance de près de 98 % sur un corpus de 43 signes à l'issue d'un apprentissage d'une occurrence seulement pour chaque signe. Même si ces résultats sont très encourageants, il est tout de même à noter qu'il s'agit d'une reconnaissance de signes isolés toujours exécutés de face. On trouve dans (Ong & Ranganath, 2005) des tableaux récapitulatifs des différentes approches basées sur la vision, au niveau du signe ou de ses composants.

### 3 Quelle approche pour une segmentation de la LSF ?

Les approches existantes que nous venons de citer permettent de reconnaître les signes et fournissent *in fine* une segmentation des vidéos analysées. Elles présentent l'inconvénient de ne pouvoir reconnaître que les signes déjà appris. De nombreux types d'unités de transfert (Cuxac, 2000) ne pourront de toute façon pas être appris car leur nombre est beaucoup trop important. Notre objectif est différent car nous souhaitons segmenter les vidéos de LSF en signes élémentaires de manière à permettre l'animation d'un signeur virtuel et à pouvoir, si besoin, effectuer des changements de signes dans les phrases capturées. Ceci implique d'être capable de traiter les signes standards aussi bien que les unités de transfert. Il faut également veiller à ce que les segments temporels obtenus contiennent l'intégralité du signe, mais n'incluent ni de configurations, ni de mouvements des signes voisins : si le segment comporte une partie du signe suivant, il ne pourra pas être utilisé dans un autre contexte.

Une nouvelle approche de segmentation est donc nécessaire. Nous abordons dans la partie qui suit différentes familles de méthodes utilisées dans le cadre de la segmentation. Nous montrons ensuite comment chacune de ces méthodes peut être adaptée au problème de la LSF. Un parallèle sera parfois dressé avec le problème de la segmentation d'image pour illustrer les problèmes à surmonter dans l'application des différentes méthodes.

#### 3.1 Approche "frontière"

La première famille de procédés vise à déterminer directement les frontières de la région à segmenter. Ceci est possible en utilisant des opérateurs locaux détectant des variations du signal qui caractérisent les bornes spatiales ou temporelles de la zone à segmenter. En traitement d'image, il s'agira par exemple des filtres détecteurs de contours.

En LSF également, les débuts et les fins de signes sont caractérisés par plusieurs éléments :

- Les changements de configuration réelle de la main (il faut cependant bien garder à l'esprit qu'à configuration identique, un changement d'orientation de la main au cours d'un signe changera la forme apparente de la main sur la vidéo),
- Les changements brusques de vitesse ou points de rebroussement,
- La mise en mouvement de la tête et des épaules du signeur,
- Les contacts ponctuels.

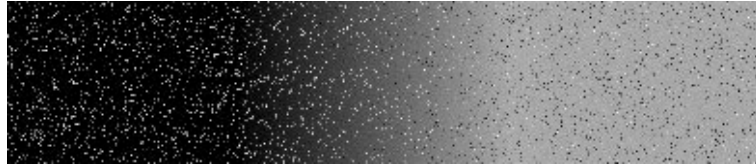


FIG. 1 – Exemple de domaine de transition progressif entre plusieurs régions

Il est important de souligner que ces événements ponctuels sont souvent présents en début et en fin de signe, mais peuvent également être observés au cours de la réalisation des signes.

A ceci s'ajoute le fait qu'il est souvent difficile de caractériser précisément le début et la fin de chaque signe en raison des problèmes de co-articulation. Ici, on peut faire une analogie claire avec la segmentation d'une image où les transitions entre les différentes régions seraient progressives et bruitées [Figure 1]. Comme en traitement d'image, ces opérateurs sont sensibles et précis mais facilement bruités.

### 3.2 Approche "région"

La deuxième famille de méthodes de segmentation se base directement sur les régions. Chaque région est caractérisée par un attribut commun à tous ses constituants. Dans le domaine du traitement de l'image, il peut s'agir par exemple de la luminosité, de la couleur, ou de la texture. Il existe aussi en LSF certaines propriétés qui peuvent être observables tout au long de la réalisation du signe :

- La vitesse constante des mains,
- La répétition d'un mouvement à l'identique (c'est le cas dans les signes à répétition),
- L'immobilité d'une main par rapport à une autre,
- La symétrie des signes ,
- Le contact prolongé entre les deux mains,
- La permanence d'une configuration (ici, l'invariance de la configuration apparente implique forcément l'invariance de la configuration réelle).

C'est sur cette méthode que nous basons notre processus de segmentation.

### 3.3 Approche "reconnaissance"

Les approches précédentes permettent d'effectuer une segmentation de l'image sans aucune connaissance *a priori* sur la structure de son contenu et procèdent par des comparaisons d'attributs à l'intérieur du signal. D'autres familles de traitement plus spécifiques identifient dans le signal une information particulière ; il s'agit donc ici de comparer le signal à un modèle de référence appris au préalable. Les processus de reconnaissance et de segmentation sont alors menés simultanément. En raison de l'aspect temporel de la LSF, les approches basées sur les Modèles de Markov Cachés (MMC) sont particulièrement bien adaptées à cette tâche. C'est d'ailleurs la raison pour laquelle les MMC sont largement utilisés par les algorithmes dédiés à la reconnaissance de signes (cf. § 2).

Si cette approche est actuellement utilisée pour la reconnaissance de l'intégralité du signe, nous pourrions également l'utiliser ultérieurement dans notre méthode de segmentation pour exploiter les informations de configurations manuelles.

### 3.4 Adaptation des critères de segmentation

Se pose alors la question de la mise en œuvre des différentes familles de méthodes citées précédemment. Quelle que soit l'approche retenue, elle nécessitera un certain paramétrage. Plusieurs cas de figures sont envisageables :

- Les valeurs des paramètres sont fixées une fois pour toutes à la création de l'application.
- Les valeurs des paramètres sont calculées une seule fois au début de l'exécution en fonction du contexte ou du contenu de l'énoncé à segmenter.
- Les valeurs des paramètres sont recalculées au cours de l'évolution et peuvent varier en fonction des traitements effectués.

Ce paramétrage de la chaîne de traitement peut faire appel à un apprentissage de réglages optimaux grâce à un corpus de référence ou être fixé par le concepteur ou l'utilisateur de l'algorithme de segmentation.

Pour illustrer l'importance de l'ajustement des paramètres de segmentation, prenons l'exemple d'un opérateur qui détecterait les mouvements particulièrement lents effectués au cours d'un signe (les mouvements très lents pourraient s'avérer intéressants dans la caractérisation des signes car nous avons observé qu'ils sont souvent associés à un transfert situationnel). Il est impossible de donner *a priori* un seuil de vitesse apparente de la main à partir duquel on pourra décréter qu'un signe est effectué lentement. Ce seuil de vitesse dépendra naturellement de plusieurs facteurs :

- La manière de signer du locuteur,
- Le type d'énoncé analysé.

Une comparaison d'une vitesse locale à l'histogramme des vitesses des mains dans la séquence vidéo permettra souvent de trouver les seuils les plus adaptés pour qualifier la vitesse.

## 4 Une approche alternative pour la segmentation

Contrairement aux approches de type "reconnaissance" (cf. § 3.3) que nous avons abordées lors de l'état de l'art, la méthode que nous avons développée pour la segmentation de signes s'inscrit plutôt dans une approche de type "région" (cf. § 3.2). Les segments temporels correspondant aux signes sont identifiés grâce à la permanence de caractéristiques comme la symétrie ou la répétition de mouvement durant toute la durée du signe. Le paragraphe suivant détaille les grandes lignes du fonctionnement de l'algorithme de segmentation.

### 4.1 Schéma global de la chaîne de traitement

Dans un premier temps, les seuls indices utilisés lors de la segmentation des signes proviennent de l'étude du mouvement des mains. Nous montrerons que sur cette seule information, des résultats intéressants peuvent être obtenus. D'autres caractéristiques comme l'exploitation des configurations ou de l'expression du visage pourront être ajoutés ultérieurement.

Le processus de segmentation que nous avons implémenté se déroule en plusieurs étapes : Le traitement initial de la vidéo consiste en un suivi des mains et de la tête du signeur dans la vidéo étudiée. Ceci peut notamment être réalisé grâce à un logiciel de suivi tel que celui proposé par (Gianni *et al.*, 2007).

Un opérateur humain est encore nécessaire pour pointer, lors du visionnage de la vidéo à vitesse réelle ou au ralenti, une frame de la vidéo - que nous appellerons amorce dans la suite de cet article - de manière à ce que chaque signe possède une et une seule amorce. Les signes de la séquence vidéo étant numérotés entre 0 et  $K-1$ , le numéro de l'amorce correspondant à la frame désignée pour le  $k^{ième}$  signe sera noté  $a_k$ .

Parallèlement, un algorithme attribue à chaque segment temporel de moins de  $N$  frames inclus dans la vidéo une mesure de confiance indiquant s'il s'agit plutôt d'un signe ou d'une transition. Cette mesure de confiance tient compte de plusieurs caractéristiques détectées automatiquement sur le segment temporel comme les symétries ou les répétitions.

Enfin, on détermine à l'aide d'un algorithme issu de la programmation dynamique l'enchaînement le plus probable de signes et de transitions pouvant caractériser la vidéo.

## 4.2 Opérateurs de traitement du mouvement

Pour l'instant, l'algorithme de segmentation que nous avons implémenté se base exclusivement sur le traitement du mouvement apparent (2D) des mains dans la vidéo. Les opérateurs caractérisent donc uniquement les trajectoires des deux mains dans les images. Le programme de segmentation se base sur trois grandes familles d'opérateurs actuellement implémentées :

Les opérateurs de relation entre les vitesses des deux mains :

- Symétrie sagittale
- Symétrie centrale
- Translation (les deux mains ont des vitesses à peu près égales à tout instant)
- Main droite statique (la main droite a une vitesse négligeable par rapport à la main gauche)
- Main gauche statique

Les opérateurs sur les répétitions de mouvement (pour la main droite, puis pour la main gauche) :

- Mouvement répété deux fois
- Mouvement répété trois fois
- Mouvement non répété

Les opérateurs sur la trajectoire d'une main (pour la main droite, puis pour la main gauche) :

- Mouvement rectiligne accéléré puis décéléré
- Mouvement rectiligne uniforme
- Autre primitives<sup>1</sup>

Il y a donc en tout  $N_a=17$  attributs qui qualifient chaque segment temporel (5 attributs de relations, 6 attributs de répétition et 6 attributs de trajectoire). A l'intérieur d'une même famille, ces attributs sont mutuellement exclusifs.

Ces différents opérateurs sont appliqués à chaque segment temporel de moins de  $N$  frames contenu dans la vidéo. Nous appellerons  $S_{ij}$  dans la suite de notre article ces segments temporels commençant à la frame  $i$  et finissant à la frame  $j$ .

Chaque segment temporel  $S_{ij}$  sera donc caractérisé par  $N_a$  attributs [Figure 2]. Ces attributs que nous noterons  $\sigma_{ij}^m$  ( $m \in [1..N_a]$ ) sont des réels compris entre 0 et 1 produits par les différents

<sup>1</sup>On notera l'absence de certains opérateurs de traitement des trajectoires comme la détection des arcs de cercle, ou de répétition comme la détection des allers-retours. Ces opérateurs seront certainement ajoutés dans une version ultérieure de notre logiciel s'ils améliorent le taux de segmentation correcte des signes.

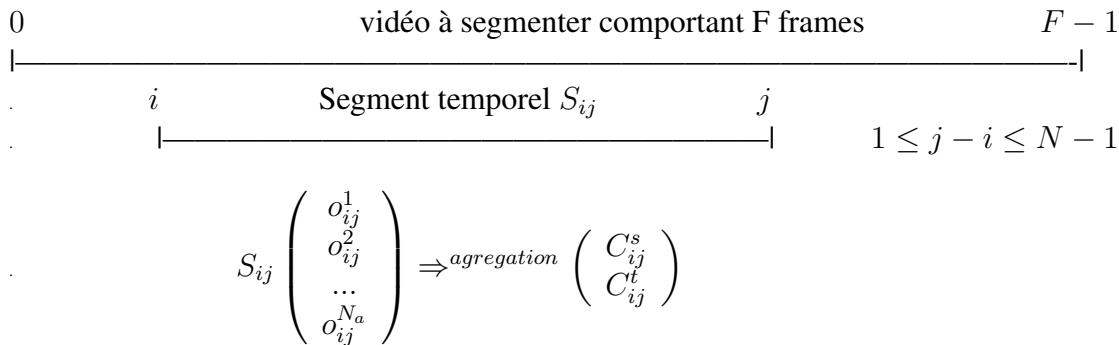


FIG. 2 – Représentation schématique d'un segment temporel inclus dans la séquence vidéo

opérateurs. Par exemple l'attribut mouvement symétrique vaudra 1 si le signe est parfaitement symétrique, 0,7 s'il se rapproche d'un signe symétrique, etc ...

Pour l'instant, la méthode d'agrégation des différents attributs consiste à sélectionner le meilleur critère dans chaque famille d'opérateurs et à combiner par pondération les critères sélectionnés ; d'autres méthodes d'agrégation seront étudiées par la suite. Cette méthode permet de caractériser chaque intervalle  $S_{ij}$  par deux mesures de confiance  $C_{ij}$  ( $0 < C_{ij} \leq 1$ ). La première notée  $C_{ij}^s$  indique si l'intervalle est un signe et la deuxième  $C_{ij}^t$  indique si l'intervalle est une transition.

### 4.3 Etape de programmation dynamique

Lors de la dernière étape de traitement de la vidéo, un algorithme de programmation dynamique permet de trouver le meilleur enchaînement de segments temporels  $S_{ij}$ . Dans les relations qui suivent,  $i(l)$  et  $j(l)$  représentent respectivement les instants de début et de fin du  $l^{ime}$  segment temporel. La recherche de la meilleure segmentation consiste à satisfaire simultanément l'ensemble des relations suivantes :

- Toute la vidéo est segmentée (les frames de la vidéo sont numérotées de 0 à  $F - 1$  et les segments temporels sont numérotés de 0 à  $2K$ ) :

$$i(0) = 0, j(2K) = F - 1$$

- Un segment ne peut pas avoir une durée nulle (il dure au moins deux frames) :

$$\forall l \in \mathbb{N} / l \leq 2K - 1 \quad 1 \leq j(l) - i(l) \leq N - 1$$

- Les segments temporels correspondant aux signes et aux transitions sont jointifs :

$$\forall l \in \mathbb{N} / l \leq 2K - 2 \quad j(l) = i(l + 1)$$

- La qualité de la segmentation est optimale :

$$\operatorname{argmax} \left( \sum_{k=0}^{k=K-1} \ln(C_{i(2k+1)j(2k+1)}^s) \cdot (j(2k+1) - i(2k+1)) + \sum_{k=0}^{k=K} \ln(C_{i(2k)j(2k)}^t) \cdot (j(2k) - i(2k)) \right)$$

Nous avons mis au point cette formule pour combiner les mesures de confiance en prenant en compte la longueur des intervalles temporels.

- Il y a une alternance de signes et de transitions et chaque segment temporel correspondant à un signe contient exactement une amorce :

$$\forall k \in \mathbb{N} / k < K \quad i(2k + 1) < a_k < j(2k + 1)$$

## 5 Évaluation du procédé de segmentation proposé

Le programme fournit une segmentation de la vidéo d'entrée sous la forme d'une alternance de signes et de transitions. Se pose alors la question de la manière de l'évaluer. Etant donné que le but de notre étude est à terme de fournir des données réutilisées pour l'animation de signeurs virtuels, nous avons souligné qu'il fallait éviter au maximum d'inclure à l'intérieur des segments temporels correspondant aux signes, tout mouvement ou configuration "parasite" appartenant aux signes précédents ou suivants.

Pour chaque signe appartenant à la vidéo segmentée, on définit donc trois classes de segmentation calculées en référence à une "vérité terrain" (segmentation manuelle par un expert en LSF). La première contient tous les signes correctement segmentés, la deuxième contient les signes partiellement segmentés et la troisième contient les signes ignorés par l'algorithme de segmentation.

Lorsqu'un signe est partiellement segmenté, on constate les erreurs suivantes :

- Le nombre de répétitions ou d'aller-retours dans le segment temporel ne correspond pas à celui du signe de la vidéo.
- Certains contacts ou points de rebroussement caractéristiques du signe sont absents du segment temporel.
- Une partie du signe précédent ou suivant est visible dans le segment temporel.

D'autre part, on observe que lorsque le signe comporte peu de frames, sa segmentation est souvent erronée.

### 5.1 Résultats

Notre algorithme de segmentation a déjà été évalué sur deux vidéos. La première vidéo de deux minutes provenant du corpus LS-COLIN (Cuxac, 2002) raconte en LSF les attentats du 11 septembre à New York. Nous obtenons 87 signes bien segmentés, 80 signes segmentés partiellement et 36 non segmentés.

Nous avons depuis segmenté une autre vidéo de 30 secondes tirée d'un bulletin d'information en LSF produit par la société Websourd<sup>2</sup>, partenaire de notre étude. Ce deuxième exemple nous a permis de valider notre approche sur un autre type de production et de quantifier les différentes sources d'erreurs de segmentation.

Dans la vidéo étudiée, le signeur filmé en plan américain est habillé en noir et porte des manches longues. La résolution de l'image utilisée lors de la segmentation est 320x240 et la fréquence d'échantillonnage est de 25 images par seconde. Aucun marqueur coloré additionnel n'est utilisé pour la détection de la tête et des mains.

A l'issue de la segmentation, 22 signes ont été correctement segmentés, 16 n'ont été que partiellement segmentés et 3 n'ont pas été segmentés. Des évaluations plus approfondies sont bien sûr nécessaires, elles sont en cours.

---

<sup>2</sup>La vidéo que nous avons segmenté est téléchargeable à l'adresse : [http://websourd.nnx.com/mediav0/public\\_html/lecteur-flv/lecteur-flv.php?urlFLV=IMG/flv/1D001-97.flv](http://websourd.nnx.com/mediav0/public_html/lecteur-flv/lecteur-flv.php?urlFLV=IMG/flv/1D001-97.flv)



## 5.2 Analyse des sources d'erreur

Pour aller plus loin dans l'exploitation des résultats, il n'est pas inutile d'identifier précisément la source des différentes erreurs de segmentation.

Parmi les signes segmentés partiellement :

- 1 segment n'a pas permis d'identifier le mouvement (trop peu de frames).
- 6 segments contenaient un mouvement incomplet du signe (manque de répétition ou de contact par exemple).
- 5 segments incluaient une configuration de transition vers un signe voisin.
- 4 segments incluaient un mouvement de transition vers un signe voisin.

Les signes non segmentés se situent au début et à la fin de la séquence vidéo. Ceci est dû à un défaut de notre algorithme de traitement.

Notre taux de segmentation correcte de 53 % est déjà très encourageant et pourrait être amélioré en prenant en compte d'autres paramètres comme les configurations et en retravaillant sur l'algorithme de segmentation pour éviter les effets de bords.

## 6 Conclusion

Il apparaît qu'une fois sur deux, la raison de la segmentation partielle des signes tient à une imprécision de quelques frames seulement sur l'estimation de la position de début ou de fin d'un signe. Il serait donc imaginable d'utiliser une approche de type "frontière" pour affiner la segmentation et gagner ainsi en précision. En outre, cela permettrait de caractériser les signes par d'autres attributs comme les points de contacts par exemple. Une autre piste serait l'adoption d'une approche multi-échelle qui permettrait la prise en compte d'indices de plus haut niveau comme les début et fin de phrase dans la segmentation.

L'utilisation d'une telle démarche de segmentation articulée sur la définition même du signe présente plusieurs intérêts :

- Du point de vue linguistique, elle pose clairement la question des critères de détection du début et de la fin du signe et nous interroge sur la définition même du signe et permettrait d'étayer les théories actuelles sur la phonologie des unités gestuelles.
- Pour l'exploitation des corpus, une telle approche de segmentation assistée permettrait de réduire par six environ le temps de segmentation manuelle (nous ne prenons ici en compte que le temps de saisie des amorces en faisant défiler le film au ralenti).
- Dans une perspective de synthèse de l'énoncé par un signeur virtuel, les caractéristiques de symétries, de primitives de mouvement et de répétitions que nous obtenons semblent compatibles avec les descriptions géométriques des signes proposées dans (Filhol *et al.*, 2007) et laissent entrevoir une application possible pour l'aide à l'animation de signeurs virtuels.

Il faut maintenant continuer à améliorer le processus de segmentation pour le rendre plus fiable et pouvoir caractériser les segments temporels des signes plus précisément. Il serait nécessaire pour cela de détecter également les positions des coudes et des épaules dans la vidéo (possibilité d'obtenir la posture 3D). D'autres informations comme l'orientation de la tête (Parashar, 2003) ou l'utilisation des configurations manuelles permettraient vraisemblablement d'améliorer les résultats de la segmentation.

## 7 Remerciements

Nous tenons à remercier J.Dalle qui a effectué l'ensemble des segmentations manuelles de référence qui ont servi à l'évaluation de notre algorithme.

## Références

- BAUER B. & HIENZ H. (2000). Relevant features for video-based continuous sign language recognition. In *FG '00 : Proceedings of the Fourth IEEE International Conference on Automatic Face and Gesture Recognition 2000*, p. 440, Washington, DC, USA : IEEE Computer Society.
- BOWDEN R., WINDRIDGE D., KADIR T., ZISSERMAN A. & BRADY M. (2004). A linguistic feature vector for the visual interpretation of sign language. In T. PAJDLA & J. MATAS, Eds., *ECCV (1)*, volume 3021 of *Lecture Notes in Computer Science*, p. 390–401 : Springer.
- BRAFFORT A., CHOISIER A., COLLET C., CUXAC C., DALLE P., FUSELLIER I., GHERBI R., JAUSIONS G., JIROU G., LEJEUNE F., LENSEIGNE B., MONTEILLARD N., RISLER A. & SALLANDRE M.-A. (2001). Projet LS-COLIN. Quel outil de notation pour quelle analyse de la LS ? . In *R-LSF'01, Toulouse, 23/11/2001-24/11/2001*, p. 71–86 : Université Toulouse 2 - Le Mirail.
- CHANNON R. (2002). *Signs are single segments : phonological representations and temporal sequencing in ASL and other sign languages*. PhD thesis, University of Maryland, USA.
- CUXAC C. (2000). *Les voies de l'iconicité*. Ophrys.
- CUXAC C. (2002). "langage et cognition" , rapport de fin de recherche du projet ls-colin. <http://www.irit.fr/LS-COLIN>.
- FILHOL M., BRAFFORT A. & BOLOT L. (2007). Signing avatar : Say hello to Elsi ! In *International Workshop on Gesture in Human-Computer Interaction and Simulation (GW), Lisbonne, Portugal, 23-25/05/07*.
- GIANNI F., COLLET C. & DALLE P. (2007). Robust tracking for processing of videos of communication's gestures. In *International Workshop on Gesture in Human-Computer Interaction and Simulation (GW), Lisbonne, Portugal, 23-25/05/07*.
- KIM J.-B., PARK K.-H., BANG W.-C., KIM J.-S. & BIEN Z. (2001). Continuous korean sign language recognition using automata based gesture segmentation and hidden markov model. In *ICCAS2001*, Jeju, Korea : Jeju National University.
- KIPP M. (2003). Annotation of video and spoken language, user manual. <http://www.dfki.de/kipp/anvil2/doc/anvil40-manual.pdf>.
- LIDDELL S. K. & JOHNSON R. E. (1990). American sign language : the phonological base. *Sign Language Studies*, **64**.
- ONG S. C. & RANGANATH S. (2005). Automatic sign language analysis : A survey and the future beyond lexical meaning. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **27**(6), 873–891.
- PARASHAR A. S. (2003). *Representation and interpretation of manual and non-manual information for automated American Sign Language recognition*. PhD thesis, University of South Florida, USA.