Apport d'un système d'analyse d'images pour l'étude de la Langue des Signes

Patrice DALLE, Boris LENSEIGNE, Céline HUDELOT

dalle@irit.fr, lenseign@irit.fr, hudelot@irit.fr
IRIT, Université Paul Sabatier,
118, route de Narbonne
31062 TOULOUSE cedex 4

Résumé

Les recherches concernant la LS sont actuellement en plein essort et de nombreux projets de recherche melent l'informatique à cette étude. Au-dela des systèmes de transcription/notation et des systèmes d'analyse des geste et de traduction, dont nous faisons l'état de l'art ; nous proposons un outil plus général d'analyse de la LS destiné à l'usage des linguistes. Après avoir étudié les besoins des chercheurs travaillant à partir de séquences d'images en LS, nous décrivons le mode de fonctionnement de notre système. Nous détaillons ensuite son architecture interne ainsi que la représentation des données que nous utilisons. Enfin nous faisons le bilan des outils à mettre en œuvre dan un tel système ainsi que de l'avancement de nos traveaux.

Mots-clés : *LSF, transcription de séquences d'images, système d'analyse de séquences d'images.*

La Langue des Signes Française (LSF) fait actuellement l'objet de recherches actives pour plusieurs raisons :

- Elle est de plus en plus utilisée dans la vie sociale et dans l'éducation des sourds
- les linguistes l'étudient pour elle même, mais aussi comme révélateur des problèmes liés à la langue et au langage
- enfin, pour les spécialistes de la vision par ordinateur elle constitue un cadre intéressant pour l'application des techniques de reconstruction 3D, d'analyse de mouvements, de suivi de gestes et d'interprétation d'images

cependant, les chercheurs manquent d'outils adaptés à cette étude. De plus, il est difficile pour le linguiste qui étudie la LSF, notamment à partir de supports vidéo, de déterminer quels paramètres sont pertinents pour cette

étude, et ce d'autant plus que le choix de ces paramètres sera biaisé par sa propre connaissance du sens du discours ainsi que par ses hypothèses sur la grammaire de la LSF. Pour notre part, nous nous intéresserons à l'apport de l'outil informatique, et de l'analyse d'images en particulier à cette problématique. Les recherches actuelles reliant l'informatique et la LS portent essentiellement sur deux domaines : la réalisation d'outils de transcription et de notation de la LS d'une part et l'analyse des gestes et des mouvements corporels dans des séquences d'images d'autre part. Nous proposons, ici, un outil d'analyse de la LS, c-à-d. un système qui vise à assister le linguiste dans l'analyse d'une séquence vidéo en LS. Ce dernier devra pouvoir, par exemple, lui demander de retrouver des configurations, faire des mesures ou visualiser des trajectoires et des évènement visuels. Cet outil doit permettre au linguiste d'étudier la LS ou d'illustrer son approche. Il ne doit donc pas être dépendant d'une théorie particulière de la LS.

1 État de l'art

Les applications reliant les langues des signes et l'informatique peuvent être regroupées en deux grands domaines : des outils de transcription et de la notation de la LS, des outils d'analyse des gestes et du mouvement dans des séquences d'images, associés à des outils de traduction automatique de la LS à partir de séquences vidéo. D'autres domaines d'application existes, notamment des dictionnaires interactifs pour une LS, mais nous n'en feront pas le détail dans cet état de l'art.

1.1 Les système de notation/transcription de la LS

Nous allons maintenant décrire les trois principaux systèmes des transcription/notation existants : la notation de Stokoe, HamNoSys et signWriting ; les deux derniers étant associés à une application logicielle. Une étude plus détaillée de ces systèmes de notation/transcription peut être trouvée dans [18].

le système de notation de Stokoe

Historiquement, le premier système de notation de la LS a été développé par Stokoe en 1960, dans le but de démontrer que l'ASL (American sign Langage) est une langue à part entière [28]. Ce système de notation décrit un signe selon trois paramètres :

- la configuration des mains
- l'emplacement où le signe est produit
- le mouvement qu'effectuent les mains

toutefois, il reste complexe à lire et à écrire dans la mesure où il a été developpé dans un but précis (prouver que l'ASL est une langue) et non pas pour être utilisé quotidiennement. De plus il est spécifique à cette langue.

HamNoSys

HamNoSys [1] (Hambourg Notation System) a été conçu par un groupe de sourds et entendants de l'université de Hambourg. Il s'agit d'un outil de recherche scientifique et non d'un système d'écriture. Il est en revanche prévu pour être applicable à toutes les langues des signes. Les signes sont décris par quatre paramètres : trois de ces paramétres sont communs à HamNoSys et Stokoe, le quatrième permet de décrire des composantes non manuelles. Ce système permet également de décrire l'expression du visage. Il reste néanmoins compliqué à utiliser et se rapproche plus d'un système de transcription que d'un système de notation.

Sign Writing

Il s'agit d'un système de notation [2] adapté du système de notation de la danse *DanceWriting*. Les deux systèmes ont été mis au point par Valérie Sutton. SignWriting décrit un signe selon cinq paramètres :

- le mouvement
- la configuration de la mian
- lieu où se fait le signe
- l'orientation de la main
- les signes grammaticaux non manuels

La principale particularité de SignWriting est sa représentation iconique des signes qui ne nécessite pas de connaissances particulières sur la LS pour être utilisée. De plus, il est facilement adaptable à différentes LS (il existe déja pour les langues des signes Américaines, Espagnoles et Portugaises). En revanche, il ne possède aucune information de prosodie.

1.2 L'analyse des gestes et du mouvement humain à partir de séquences d'images

L'analyse des gestes et des mouvements corporels est un domaine de recherche qui possède des applications variées et nombreuses: analyse de performances athlétiques, télésurveillance, Interface homme-machine, compression vidéo,... Nous ne présenterons dans cette partie que les travaux se rapportant à l'analyse des gestes de la Langue des Signes et ceux que nous

1. http://www.sign-lang.uni-hamburg.de/Projects/HamNoSys.html
2. http://www.signwriting.org

avons jugé intéressants pour un tel contexte. Des états de l'art relativement complets sur l'analyse du mouvement ou des gestes par un système de vision peuvent être trouvés dans [7], [4], [17] et [24].

Les difficultés du problème

Dans la réalisation d'un énoncé en langue des signes, les mains et les bras ne sont pas les seules parties du corps mises en jeu. Il faut aussi prendre en compte les mouvements du buste, des épaules, de la tête, la direction du regard et les expressions faciales. C'est pourquoi nous étudions le cas spécifique où le périphérique d'entrée est un système de vision, bien que certains travaux aient étéb mennés en utilisant notamment des gants numériques [1].

Le problème fondamental est donc d'analyser le mouvement de ces différentes parties du corps humain, sur des séquences longues d'images, de manière indépendante mais aussi relativement les unes aux autres. On se heurte donc aux trois problèmes fondamentaux en analyse de scènes dynamiques qui sont:

- la segmentation, c'est à dire l'extraction des zones d'intérêt (zones de l'image concernées par le mouvement) du reste de l'image
- le suivi du mouvement dans le but d'extraire l'information le caractérisant
- l'interprétation de celui-ci

Ces trois problèmes ne sont pas indépendants et ne peuvent pas être considérés comme des étapes successives dans l'analyse. D'autre part plus, nous avons une perception tridimensionnelle d'une scène en mouvement mais, nous ne disposons, via le système de vision que des projections 2D de celle-ci (perte de profondeur, problèmes d'occultations).

Dans le cadre de la reconnaissance de gestes, apparait le problème de la variabilité d'exécution de ces derniers. En effet le même geste peut être réalisé de manière différente, dans l'espace et dans le temps, selon le signeur. De plus, pour certains gestes de la LS, il existe une variabilité contextuelle (signes standards/signes variables) [5].

Enfin, dans le cadre de la reconnaissance continue des signes, il faut aussi tenir compte du phénomène de coarticulation selon lequel la réalisation d'un signe est influencée par le signe précedent et le signe suivant. Ce phenomène introduit des gestes supplémentaires.

Utilisation d'informations 3D

Vogler et Metaxas [29] [31] [30] de l'Université de Pennsylvanie travaillent sur un système de reconnaissance isolée et continue de phrases signées en ASL. Les données 3D sont acquises à partir de trois caméras. L'intérêt de ce

travail réside dans l'utilisation de phonèmes (plutôt que de signes entiers) comme unités de base du système de reconnaissance (MMC), en se basant sur une modèle phonologique séquentiel (description de Lidell et Johnson) de l'ASL. Sur une base de tests constituée de 400 phrases signées à partir de 22 signes de l'ASL, les auteurs affichent un taux de réussite de 84,85 % pour la reconnaissance de phrases et de 94,23 % pour la reconnaissance de signes. Ces travaux se rapprochent des travaux de Hienz [19] [21] [20] sur la langue des signes allemande dans lesquels la position de la main dans le repère tridimensionnel du signeur est retrouvée au moyen d'un modèle morphologique du bras. Une extension de cette méthode permet de retrouver l'axe du corps et la position des épaules. enfin, l'équipe Robovis de l'INRIA en collaboration avec l'équipe gestes et Images du LIMSI et l'équipe Intermédia de l'INT collaborent sur le projet Arc LSF dont l'objectif est d'aider à la conception de systèmes informatiques dédiés à la capture, la reconnaissance et à l'interprétation de la LS en se basant sur un système de suivi en 3D des gestes d'une personne à l'aide de la vision tridimentionnelle et dynamique [16].

L'utilisation de modèles d'apparence

Le principe est de représenter un objet ou une classe d'objets 3D par une collection de vue 2D de ces objets. C'est l'approche utilisée par Hamdam [25] dans son travail de thèse sur la détection, le suivi et la reconnaissance de la forme et du mouvement d'un objet à l'aide de modèles probabilistes gaussiens. Dans [8], Olivier Chomat et James Crowley de l'INRIA proposent un sytème de reconnaissance du geste et de l'activité humaine à partir d'un modèle d'apparence spatio-temporelle.
Cui et Weng [9] utilisent aussi cette approche dans le système SHOSLIF-M[3], système de reconnaissance des signes de l'ASL (American Sign LAnguage) à partir de séquence d'images en niveaux de gris. Ce système permet une reconnaissance simultanée du mouvement et de la configuration de la main. Ce système a été testé, avec un taux de réussite de 92,3 % sur un ensemble de 28 signes différents de la main présentant une grande variété de configurations et de mouvements.

Extraction, suivi et reconnaissance de la trajectoire de la main

Starner et Pentland [2], [27] ont proposé un système permettant la reconnaissance et l'interprétation continue d'un sous ensemble du vocabulaire de l'ASL. La segmentation de la main se fait en utilisant la signature de couleur caractéristique de la peau et la reconnaissance se fait à l'aide des Modèles de Markov Cachés . Sur un vocabulaire de 40 signes, leur taux de réussite est de 92%.

3. Self-organized Hierarchical Optimal Subspace Learning and Inference Framework for Movement

Dans [22], Imagawa présente un système de suivi en temps réel des deux mains pour la reconnaissance de la Langue des Signes japonaise. L'intérêt de ce système est de prendre en compte les deux mains du locuteur et de résoudre les problèmes d'occultations avec le visage (86 % de réussite). Un travail presque analogue utilisant des reseaux de croyance bayesienne est mené au sein du projet ISCANIT du groupe *Machine Vision* du département informatique de l'Université de Londres [26]. De même, dans le cadre du projet *GIST (Gesture Interpretation using Spatio-Temporal analysis)* du Beckman Institute, Yang et Ahuja ont proposé une méthode de reconnaissance non continue de gestes de l'ASL se basant sur l'extraction des trajectoires associées à une seul main en fusionnant plusieurs indices visuels (mouvement, couleur de la peau, indices géométriques de forme et de taille). La reconnaissance se fait en utilisant des réseaux de neurônes (TDNN : Time Delay Neural Network) avec 96 % de réussite [33].

Utilisation de blobs de mouvement

Cutler et Turk [10] de l'Université du Maryland ont proposé un système de reconnaissance de gestes basé sur l'estimation du flot optique et sa segmentation en blobs de mouvement. La reconnaissance se fait ensuite par l'application de règles sur les caractéristiques des blobs (taille, nombre, mouvement, mouvement relatif aux mouvements des autres blobs,...). C'est une méthode intéressante car elle n'impose pas de contraintes sur l'habillement, les conditions d'illumination ou le mouvement de la caméra mais elle est limitée à un certain type et nombre de gestes.

Bilan

Certains des systèmes présentés sont opérationnels mais ils sont souvent restreints à un nombre de gestes limités et trés peu de systèmes considèrent le mouvement du locuteur dans sa globalité, la plupart des systèmes ne considérent que les mouvements d'une seule main ce qui est insuffisant quand on sait que dans la réalisation d'un geste de la LS les mouvements des bras et des mains qui sont mis en jeu mais aussi ceux du buste, des épaules, de la tête et la direction du regard et les expressions faciales. De plus, peu de systèmes permettent la reconnaissance continue des gestes de la LS et même dans le cas contraire cette reconnaissance n'est possible qu'avec des phrases formées de signes standards. Aucun système n'a de dimension contextuelle. Ce travail reste encore du domaine de la Recherche.

2 Objectifs du système

notre travail, pour sa part, se situe dans un cadre différent. En effety, nous proposont unoutil d'analyse de la LS qui permettrait au chercheur l'utilisant de visualiser certains paramètre de la LS ou de vérifier leur pertinence.

2.1 Étude des besoins

Nous proposons un système destiné à assister le linguiste dans l'analyse de séquences vidéo d'un locuteur en LS.
Pour mener à bien cette tâche, le linguiste utilise d'ordinaire un magnétoscope à l'aide duquel il parcourt la séquence à la recherche d'évènements visuels qui vont lui permettre d'établir la structure du discours (figure : 1). Cependant, son travail d'analyse, ainsi que les résultats obtenus, sont influencés par la connaissance a priori qu'il a de la langue des signes et du contenu de la vidéo. Par exemple si nous prenons le cas de la segmentation de la séquence vidéo en signes, la question se pose de savoir si :

- le linguiste segmente la séquence de telle façon parce qu'il a reconnu le sens
- la segmentation se fait sur de critères purement syntaxiques et d'indices visuels que le linguiste reconnaît dans la séquence?

La détection d'indices visuels par le système de TI (Traitement d'Images) permettra au linguiste de se baser sur des critères objectifs. De plus l'analyse de la séquence sera guidée par l'objectif du linguiste. Il ne s'agit donc pas d'une tâche systématique. Notre système doit donc être suffisamment flexible pour permettre au linguiste de rechercher et de visualiser différents éléments dans une séquence vidéo (par exemple rechercher les clignements des yeux dans la séquence ou visualiser la trajectoire d'une main). Enfin, le système doit pouvoir dialoguer avec l'utilisateur pour lui permettre de définir des concepts de la langue des signes de façon interactive et de confronter la description de ces concepts avec les résultats obtenus.

2.2 Mode de fonctionnement proposé

Lors de l'analyse d'une séquence vidéo d'un locuteur en LS, la connaissance a priori qu'a le linguiste sur le sens du discours et sur la LS risque de l'amener à détecter des évènements visuels non parce qu'il les a vus, mais parce qu'il *sait* qu'il doit les trouver à ce moment. Ainsi, si nous faisons l'hypothèse que le clignement des yeux est un marqueur du transfert personnel (TP), on peut se demander si la détection de ces clignements est effectivement utilisée pour repérer les TP ou si d'autres indices (visuels ou issus de la compréhension du discours) ont permis de repérer ce transfert et donc de prédire un clignement des yeux qui est alors "vu" dans l'image même s'il n'est pas entièrement visible.

Notre système devra donc permettre de lever ce type d'ambiguïtés en demandant au linguiste de spécifier à l'avance les indices à prendre en compte. Il s'agit donc bien d'assister le linguiste dans sa tâche et non de l'automatiser.

FIG. 1 – *Exemple de transcription d'une séquence vidéo en LS (apport LS-Colin).*

Fragment :1	1a	1b	1c	1d	1f	1g	Fragment 2
Images **Direction des mouvements**							
Regard	Vers un point de lespace à gauche (TP)	Fermé (point de départ de la prise de décision (TP)	_____	Vers le locatif	Vers la camera	_____	Fermé (changement de scène)
Main dominante	.	.	Début de transfert de forme : " emplacement des jambes du personnage "	TS: début du déplac. De liactant	TS : déplac. De liactant vers le sol.	TS : Déplac. de liactant vers une cible (le mur) mouvement de " marcher " réalisé par les doigts	
Deux mains	TP : *" tenir les rennes "*	TP : *" tenir les rennes "*	.	.	.	.	TP : *" mettre les mains sur le mur "*
Main dominée	.	.	Début du locatif (cheval)	Locatif : cheval	Locatif : cheval	Locatif : cheval	.
Mimique faciale	Interrogation curieuse (yeux grands ouverts, sourcils soulevés, moue)	_____	résultative, prise de décision	résultative, décidé	résultative, décidé	résultative, décidé	résultative, décidé
Mouvement de la tête	.	.	Penchée complètement en arrière	Redressé	En face	En face	En face
Mouvement de la bouche	moue, lèvre supérieure avancée vers le haut.	_____	Lèvres serrées et contractées [mn]	_____	_____	_____	_____
Mouvement de la partie supérieure du corps	Droit, en face	_____	Légèrement penché en arrière	Droit, en face	Balancement des épaules vers la droite	Balancement des épaules vers la gauche	Droit, en face
Traduction approchée	*Assis sur son cheval, il se pose des questions ô*	*Ainsi, il prit la décision deô*		*descendre de son cheval etô*		*de marcher jusquau mur.*	*Il mit ses mains sur le murô*

Le linguiste pourra utiliser ce système pour justifier ou étayer sa propre technique de segmentation de la séquence (en fonction des indices visuels détectés par le système), vérifier la présence d'évènements visuels qui auraient pu lui échapper, effectuer des mesures (vitesse, accélération, ...) ou encore visualiser des évènements qui ne sont pas directement accessibles à l'observation de la séquence (occupation de l'espace, trajectoire, ...).

Le système doit comporter une interface à travers laquelle l'utilisateur va pouvoir dire de façon interactive ce qu'il recherche et/ou ce qu'il désire visualiser. Il dispose également d'une base d'opérateurs de traitement d'images pouvant être regroupés en opérateurs :

- d'analyse d'images
- de détection d'évènements
- de reconnaissance de configuration
- de mesures de paramètres
- et de visualisation (par exemple la mise en évidence d'une trajectoire)

ainsi que des opérateurs spécialisés dans l'analyse de primitives particulières à l'étude des gestes d'un personnage (détection de la peau, suivi d'une région en mouvement, ...). Le principe est de créer un dialogue entre le système et l'utilisateur au cours duquel ce dernier va être amené à décrire de façon incrémentale ce qu'il désire visualiser. Le système va alors chercher à configurer des chaînes d'opérateurs de traitement d'images (TI) correspondant à ces spécifications.

3 Conception du système

3.1 Architecture

La conception du système dont nous venons de présenter les fonctionnalités se place dans la lignée des travaux de l'équipe TCI dans le domaine de la formalisation du TI [14] [12] [13] [6] et de la conception incrémentale d'applications de TI [15] [23] [32] [3]. De façon générale, ces travaux visent à concevoir des systèmes permettant à un utilisateur, spécialiste de son domaine (ici la linguistique) mais pas du TI, de configurer des chaînes de TI en dialoguant avec la machine en termes de son domaine que le système va "traduire" en termes de chaînes d'opérateurs de TI.

Afin d'arriver à ce résultat, nous utilisons une architecture en trois modules (figure : 2):

- un système de base : il s'agit d'un système multi-agent où chaque agent encapsule un opérateur de TI. Ce système est capable d'évoluer selon

des critères internes d'utilité (il est doté d'une tendance interne à schématiser et structurer les données disponibles) et en fonction d'objectifs décrivant les entités de TI à extraire [32] [3]. son rôle ici, sera d'analyser une séquence d'images en la structurant à partir d'indices visuels spatio-temporels.

- un système intermédiaire doté de la même tendance à schématiser et à structurer les données dont il dispose que le système de base, il opère sur un langage qui décrit à la fois les données du TI produites et les chaînes d'opérateurs utilisées par le système de base. Il observe le comportement du système de base et cherche à détecter des régularités (il reconnaît et interprète les séquences d'opérateurs susceptibles de correspondre à la détection d'un évènement particulier). Il utilise ensuite cette interprétation pour définir les informations pertinentes à présenter à l'utilisateur, via une interface. En réponse aux réactions ou aux requettes de l'utilisateur il va faire évoluer l'élaboration du sens qu'il cherche à construire et va traduire cela en nouveaux objectifs pour le système de base.

- une interface homme-machine (IHM). Ce module présente les résultats issus du TI de façon explicite pour l'utilisateur. Elle lui fournit des outils d'interaction pour définir les nouveaux résultats attendus. La nature des outils proposés dépend de la nature des données présentées [23].

3.2 Une représentation multi-niveau

L'architecture que nous venons de décrire implique une représentation multi-niveau des informations contenues dans la séquence vidéo. Cette représentation se partage entre le système de base et le système intermédiaire. Il s'agit de passer de l'information contenue dans les images (des pixels) à l'information linguistique recherchée dans la séquence en passant par un certain nombre niveaux d'abstraction. Chaque niveau correspondant à la représentation de l'entité considérée dans un espace particulier. L'instanciation (éventuellement partielle) des ces niveaux à un instant donné de l'analyse est appelée *contexte*. En définissant, pour chaque entité, des mécanismes (de spécialisation/généralisation, structuration/déstructuration et nomination) permettant de passer d'un niveau d'abstraction à l'autre, il devient possible de combiner une analyse ascendante (de l'image au sémantique) avec des mécanismes de prédiction. De plus, l'instanciation d'un contexte à un instant t va permettre de prédire (du sémantique vers l'image) l'état de ce contexte pour l'instant $t+1$. On parle alors de *prédiction temporelle* (figure 3).

Cette représentation se retrouvant au niveau de l'interface, elle va permettre au linguiste d'exprimer sa requette à différents niveaux d'abstraction,

FIG. 2 – *Architecture générale d'un système d'analyse de séquences vidéo en LS mettant en évidence les interactions entre les différents modules.*

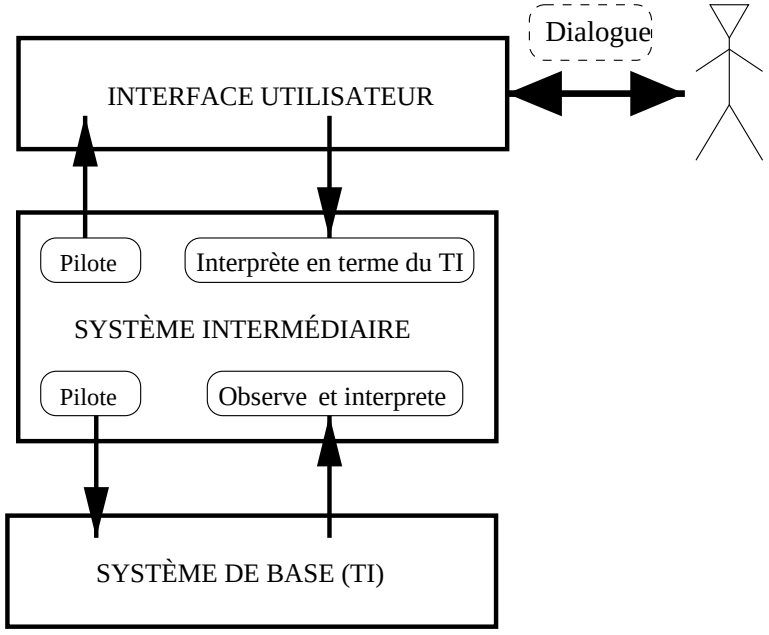
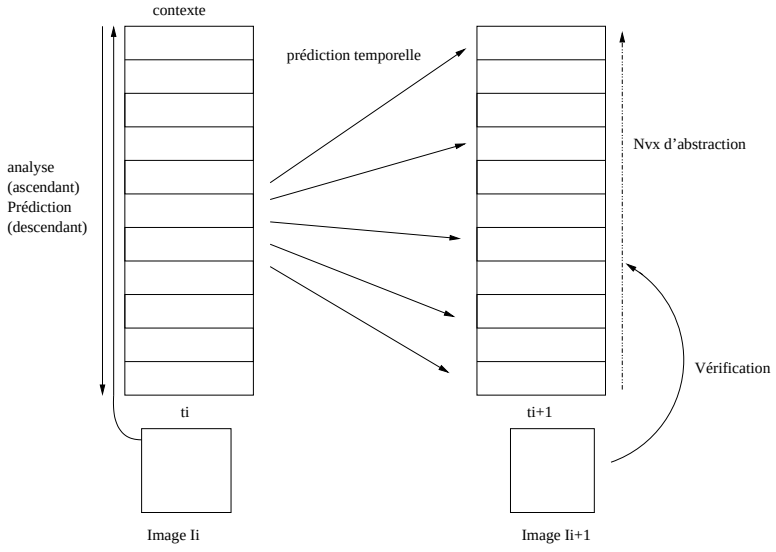


FIG. 3 – *Les mécanismes inter-niveau mis en œuvre durant l'analyse.*



c-à-d. de considérer un élément précis (un clignement d'œil par exemple) mais également de retrouver des configurations plus complexes (ex : les situations de transfert personnel). Dans ce cas, il sera alors possible de décrire un nouveau concept comme étant un ensemble de concepts déjà connus présentant des propriétés particulières.

4 Mise en œuvre

4.1 les outils nécessaires

Notre système s'appuie sur trois ensembles d'outils particuliers :

- des outils de caractérisation et de reconnaissance d'indices visuels dynamiques
- des outils de visualisation spécifiques aux gestes et à la LS
- des formalismes supportant le dialogue et les interactions entre le système intermédiaire et l'utilisateur d'une part, le système intermédiaire et le système de base d'autre part

Dans la première catégorie, nous retrouvons les opérateurs de TI permettant la détection des indices caractéristiques des éléments de la LS recherchés dans le cas d'une analyse ascendante (Cf. 3.2) ; ou, dans le cas d'une analyse descendante, des opérateurs permettant de vérifier la présence d'indices visuels correspondant à l'occurrence d'un évènement qui a été prédit. Ces outils constituent la bibliothèque d'opérateurs utilisés par le système de base (Cf. 3.1).

Les outils de visualisation, eux, sont spécifiques à l'interface. Ils vont permettre de présenter de façon intelligible à l'utilisateur les résultats obtenus à la suite d'une requête. Dans cette catégorie, nous retrouvons des outils permettant de visualiser une trajectoire, une direction de pointage où encore l'occupation de l'espace autour du signeur, etc...

Enfin, les formalismes mis en œuvre constituent un support pour l'implémentation des mécanismes permettant au système intermédiaire d'interagir avec les deux autres. Ils constituent en effet une représentation formelle des concepts manipulés par chaque module, représentation qui permet l'application de mécanismes de transformation sur ces concepts et donc l'évolution du système indépendamment ou non des requêtes de l'utilisateur.

4.2 État du projet

Pour l'heure actuelle, nous disposons de maquettes de l'interface et du système de base faisant chacune l'objet d'une thèse au sein de notre équipe

[32] [23]. De plus nous travaillons actuellement au développement et à l'évaluation d'outils de détection et de caractérisation d'indices visuels dynamiques(Cf. 4.1). Nous constituons une bibliothèque de tels opérateurs et testons leur validité dans le cadre de notre projet (figure 4).

A ce jour, nous disposons d'opérateurs permettant l'extraction de la silhouette du locuteur, la détection de zones de peau, la détection de zones en mouvement et d'un outil de visualisation de l'historique du mouvement (tMHI: time Motion History Image).

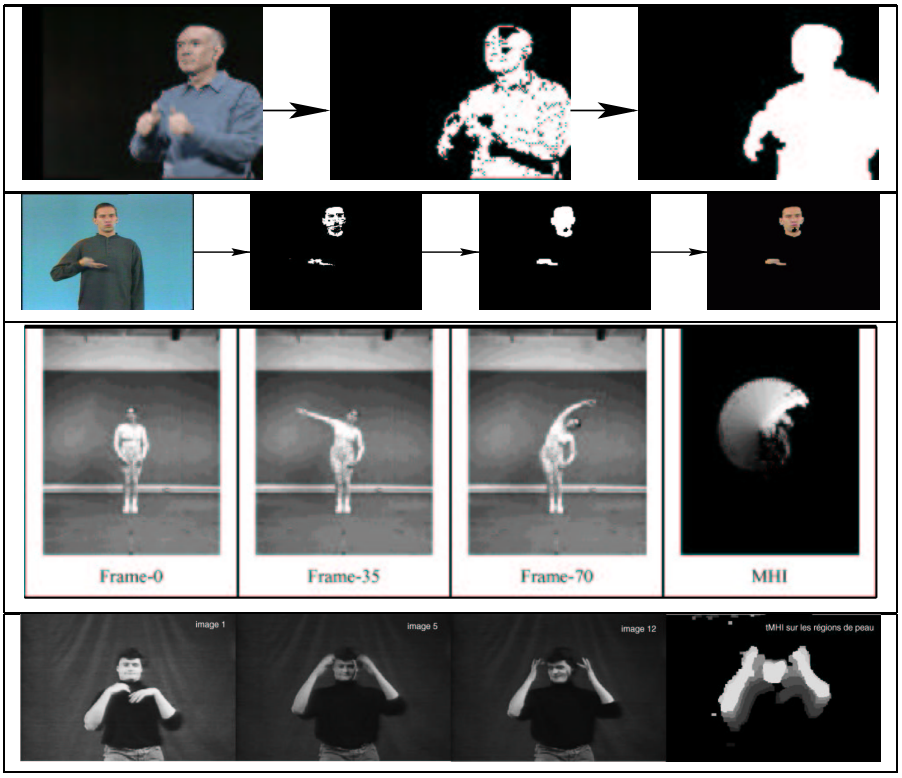


FIG. 4 – *De haut en bas: détection de la silhouette ; détection de la peau; exemple de tMHI (d'après [11]); tMHI appliqué aux zones de peau détectées au préalable.*

En revanche, la réalisation du système complet et l'étude des mécanismes à mettre en œuvre est un travail de recherche à plus long terme. Toutefois, les résultats obtenus avec les opérateur en cours de développement sont, eux exploitables indépendamment du système complet et laissent prévoir des retombées intermédiaires.

5 Conclusion et perspectives

Nous avons présenté un système d'analyse de séquences d'images d'un locuteur en LS. Ce système exploite des recherches menées dans le cadre du projet cognitique LS-COLIN sur la recherche et la visualisation d'indices visuels spatio-temporels. Il les intègre dans une architecture visant à construire une représentation caractéristique des signes du locuteur à différents niveaux et à les confronter aux analyses du linguiste. Il y ajoute la notions de séquence d'images et de mouvement. Enfin, en dotant les linguistes qui étudient la langue des signes d'un outil supplémentaire pour l'analyse de cette langue, nous souhaitons améliorer les connaissances que nous avons sur cette langue et ainsi participer à sa valorisation.

Références

[1] BRAFFORT A. Reconnaissance et compréhension de gestes; application à la langue des signes. *Thèse pour l'obtention du grade de Docteur de l'Université Paris XI-SUD, UFR Sciences, LIMSI*, juin 1996. Doctorat Informatique.

[2] STARNER T. PENTLAND A. Real-time american sign language recognition from video using hidden markov models. *M.I.T Media Laboratory Perceptual Computing Section, Technical Report TR-375*, 1995.

[3] MAGNIEN Y. ABCHICHE Y., DALLE P. Construction adaptative de concepts par structuration d'entités de traitement d'images. *RFIA*, 2002 à paraître.

[4] CAI Q. AGGARWAL J.K. Human motion analysis: A review. *Computer Vision and Image Understanding*, 73(3):428–440, mars 1999.

[5] CUXAC C. La langue des signes française. les voies de l'iconocité. *Faits de Langue, Ophrys*, 2000.

[6] DALLE P. CAPDEVIELLE O. Formulation d'objectifs de traitement d'images. *IHM'95, 7èmes journées sur l'Ingéniérie des Interfaces Homme-Machine, Toulouse*, octobre 1995.

[7] SHAH M. CEDRAS C. Motion-based recognition: A survey. *Image and Vision Computing*, 13(2):129–155, 1995.

[8] CROWLEY J.L. CHOMAT O. Utilisation de champs réceptifs spatio-temporels pour la reconnaissance de l'apparence locale d'activités. *Congrès ORASIS*, avril 1999.

[9] WENG J. CUI Y. Appearance-based hand sign recognition from intensity image sequences. *Computer Vision and Image Understanding*, 78:157–176, 2000.

[10] Turk M. CUTLER R. View-based interpretation of real-time optical flow for gesture recognition. *IEEE International Conference on Automatic Face and Gesture Recognition*, avril 1998.

[11] BOBICK A. DAVIS J. The representation and recognition of human movement using temporal templates. *Proceedings on Computet Vision and Pattern Recognition*, pages 928–934, juin 1997.

[12] DALLE P. DEJEAN P. Un langage de description de concepts pour la formulation d'objectifs d'analyse. In *5èmes journées ORASIS, Pôle Vision du GDR-PRC 'CHM'*, pages pp. 219–224, Clermont-Ferrand, 20 au 24 mai 1996.

[13] DALLE P. DEJEAN P. Image analysis operators as concept constructors. In *IEEE Southwest Symposium on Image Analysis and Interpretation*, pages pp. 66–70, San Antonio, Texas (USA), avril 1996.

[14] DALLE P. DEJEAN P. Modèle symbolique de la donnée de traitement d'images. In *10ème congrès RFIA*, pages pp. 746–75, Rennes, janvier 1996.

[15] DALLE P. DEJEAN P. Planification en traitement d'image : approche basée sur les données. In *11ème congrès RFIA*, pages pp. 75–84, Clermont-Ferrand, janvier 1998.

[16] FAUGERAS O. DELAMARRE Q. Suivi multi-caméras de personnes et modèles 3d articulés. *INRIA Projet RobotVis*, 1999.

[17] GRAVILA D.M. The visual analysis of human movement: A survey. *Computer Vision and Image Understanding*, 73(1), janvier 1999.

[18] MAVIANNE F. Vers un outils informatique d'édition de la langue des signes fraçaise assisté par le traitement d'images. Technical report, DEA Informatique de l'Image et du Langage. IRIT - UPS, Toulouse, France, 2001.

[19] OFFNER G. HIENZ H., GROBEL K. Real-time hand-arm motion analysis using a single video camera. *Proccedings of the Second Internatioonal Conference of Automatic Face and Gesture Recognition, Killington, USA*, pages 323–327, 1996.

[20] OFFNER G. HIENZ H., GROBEL K. Automatic estimation of body regions from video images. *International Gesture Worshop, Bielefeld, Allemagne*, pages 135–145, 1998.

[21] OFFNER G. HIENZ H., GROBEL K. An automatic video-based sign language recognition system as part of a sign printing system. *IEEE. International Conference on Intelligent Enginneering Systems, VIenne, Autriche*, pages 163–168, 1998.

[22] IGI S. IMAGAWA K., LU S. Color-based hands tracking system for sign language recognition. *Proc. 3rd IEEE International Conference on Automatic Face and Gesturerecognition*, pages 462–467, 1998.

[23] DALLE P. NOUVEL A. Description des entités du traitement d'images pour la conception interactive d'applications. In *ORASIS 2001*, pages pp. 349–358, Cahors, 5 au 8 juin 2001.

[24] HUANG T.S. PAVLOVIC V.I., SHARMA R. Visual interpretation of hand gestures for human-computer interaction : A review. *IEEE TRansactions On Pattern Analysis and Machine Intelligence*, 19(7), juillet 1997.

[25] HAMDAM R. Détection, suivi et reconnaissance des formes et du mouvement par modèles probabilistes d'apparence. *Thèse pour l'obtention du grade de Doctuer de l'Université Louis PASTEUR, Strasbourg1*, janvier 2001. Doctorat Electronique, Electrotechnique et Automatique, spécialité Traitement d'Images et Vision par Ordinateur.

[26] GONG S. SHERRAH J. Tracking discontinuous motion using bayesian inference. *Computer Vision and Image Understanding*, 73(3):428–440, mars 1999.

[27] PENTLAND A. STARNER T., WEAVER J. Real-time american sign language recognition using desk and wearable computer based video. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(12):1371–1375, decembre 1998.

[28] CRONEBERG C. STOKOE W., CASTERLINE D. *A dictionnary of american sign langage*. Gallaudet College Press, 1965.

[29] METAXAS D. VOGLER C. Asl recognition based on a coupling between hmms and 3d motion analysis. *Proceedings of the International Conference on Computer Vision, Mumbai, India*, pages 363–369, Janvier 1998.

[30] METAXAS D VOGLER C. Parallel hidden models for american sign language recognition. *Proceedings of the International Conference on Computer Vision, kerkyra, Greece*, septembre 1999.

[31] METAXAS D. VOGLER C. Toward scalability in asl recognition: Breaking down sign into phonemes. *Gesture Workshop'99, Gif-sur-Yvette, France*, mars 1999.

[32] ABCHICHE Y. Conception d'une plate-forme pour la configuration d'opérateurs de traitement d'images par système multi-agent. Technical report, DEA Informatique de l'Image et du Langage. IRIT - UPS, Toulouse, France, 1999.

[33] AHUJA N. YANG M.H. Recognizing hand gestures using motion trajectories. *Proceedings of the 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, juin 1999.