DIANOEMA: Visual analysis and sign recognition for GSL modelling and robot teleoperation

Stavroula-Evita Fotinea[1], Eleni Efthimiou[1], George Caridakis[2], Olga Diamanti[2], Nikos Mitsou[2], Kostas Karpouzis[2], Costas Tzafestas[2] and Petros Maragos[2]

[1]ILSP – Institute for Language and Speech Processing / ATHENA R.C.
Artemidos 6 & Epidavrou, Maroussi GR-15125, Greece
evita@ilsp.gr, eleni_e@ilsp.gr
[2]ICCS - National Technical University of Athens, School of ECE, Athens 15773, Greece,
gcari@image.ece.ntua.gr, olga.diam@gmail.com, nmitsou@mail.ntua.gr,
kkarpou@image.ece.ntua.gr, ktzaf@softlab.ntua.gr, maragos@cs.ntua.gr

Abstract. This paper presents the scientific framework and achieved results of the DIANOEMA project, in the framework of which visual analysis and sign recognition techniques have been explored on Greek Sign Language (GSL) data aiming at GSL modelling and a pilot application for robot teleoperation. The project's accomplishments comprise the Greek Sign Language Corpus (GSLC) creation and annotation, video analysis algorithms for automatic visual detection and gesture tracking, a probabilistic recognition scheme for Automatic Sign Language Recognition from multiple cues, and the pilot application on teleoperation of a Mobile Robot Vehicle.

Keywords: Sign language, sign gesture recognition, Deaf communication, HCI.

1 Introduction

The scientific areas of visual analysis of video sequences, sign language gesture recognition, and sign language linguistic modelling, as well as their applications as interfaces for machine/robot teleoperation constitute a considerably broad spectrum of interdisciplinary scientific problems which is related to at least three different scientific fields. Specifically, in the field of image analysis and computer vision, an important research area is concerned with the analysis and recognition of video sequences containing gestures, such as the signs used in sign language. Their effective and useful recognition requires a good level of understanding of sign language, which can only be reached via a systematic collection and development of an annotated corpus for sign language. Thus, the analysis of static and dynamic signs is crucial to the scientific field of natural language processing as well. Also, in the fields of human-computer interaction (HCI) and robotic systems tele-operation, many additional technical applications for communication and teleoperation can emerge, that could be based on video gesture recognition. Such applications can support impaired people or result in the creation of more flexible and user friendly interfaces.

Research work in DIANOEMA focused on the following activities:

i) Development of innovative image analysis and computer vision algorithms for the effective visual analysis of video sequences, aiming at sign detection and tracking;

ii) Creation of a video-corpus of the Greek Sign Language (GSL) and annotation and modelling of an indicative subset of it;

iii) Automatic recognition of indicative categories of GSL gestures using automatic computer vision systems pre-trained on the GSL corpus, and combining AI techniques, machine learning and probabilistic analysis for the estimation of gesture instantiations.

iv) Integration of the above into a pilot application system of robot tele-operation, on the basis of a pre-defined vocabulary of simple signs for the tele-operation control.

2 The DIANOEMA scientific accomplishments

2.1 Greek Sign Language Corpus (GSLC) : creation and annotation

The design of GSLC content has been led by the demand to support sign language recognition as well as theoretical linguistic analysis. In this respect, content organisation makes a distinction between three parts:

The first part comprises a list of lemmata which are representative of the use of handshapes as a primary sign formation component. This part of the corpus is developed on the basis of measurements of handshape frequency of use in sign morpheme formation, but it has also taken into account the complete set of sign formation parameters. In this sense, in order to provide data for all sign articulation features of GSL, the corpus also includes characteristic lemmata with respect to all manual and non-manual features of the language.

The second part of GSLC is composed of sets of controlled utterances, which form paradigms capable to expose the mechanisms GSL uses to express specific core grammar phenomena. The grammar coverage that corresponds to this part of the corpus is representative enough to allow for a formal description of the main structural-semantic mechanisms of the language.

The third part of GSLC contains free narration sequences, which are intended to provide data of spontaneous language production that may support theoretical linguistic analysis of the language and can also be used for machine learning purposes as regards sign recognition.

The target of sign recognition imposed the demand for the collection of lists containing representative lemmata, capable to exhibit the articulation mechanisms of the language. These lists may provide a reliable test bed for initial recognition of single articulation units. Lemmata lists comprising the first part of the GSLC involve two categories, (i) commands related to robot motion control and (ii) simple and complex sign morphemes, representative of the basic vocabulary of GSL.

The definition of annotation features assigned to a given signing string, reflects the extent of the desired description of grammatical characteristics allotted to the 3-dimensional representation of the linguistic message.

For morpheme level annotation the HamNoSys notation system was adopted. For sentence/phrase level annotation the ELAN annotation tool was used (Figure 1).

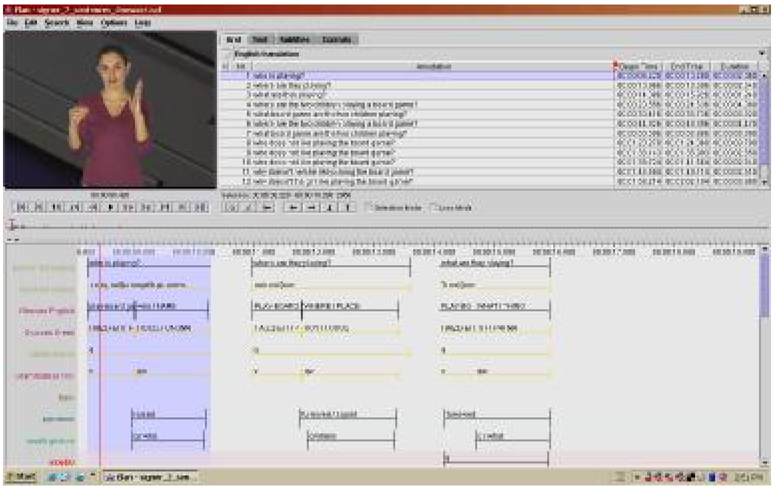

Fig. 1. Annotation markers and retrieval data for GSL in the ELAN environment.

2.2 Video Analysis Algorithms for Automatic Visual Detection and Gesture Tracking

The research objective here was to design and implement a front-end computer vision system to meet the application requirements. Work has been divided to the following tasks:

i) Development of algorithms for automatic detection of the signer in a video frame

ii) Selection of reliable and informative features, which facilitate and also define the gesture recognition process

iii) Motion tracking and segmentation of the signer's movements (for instance hand, fingers, arm etc)

Firstly a system was developed for the detection and localization of the signer in the image and secondly, methodologies were examined for the extraction of visual features, suitable for gesture recognition applications. The detection subsystem was used for the initialization of the tracking system and for the re-initialization of the system in case a tracking failure occurs.

One of the most important components of a reliable video analysis system for SL is the accurate tracking of the signer and the precise retrieval of the geometrical configuration, namely the segmentation of the arms, hands, or even the fingers of the signers, in a sequence of images. The informational content and meaning of the signs can be represented, to a large extent, in this sequence of geometrical features (Figure 2).

An example of segmentation and tracking of the image sequence from the GSL sign video using computer vision techniques is presented in Figure 3.

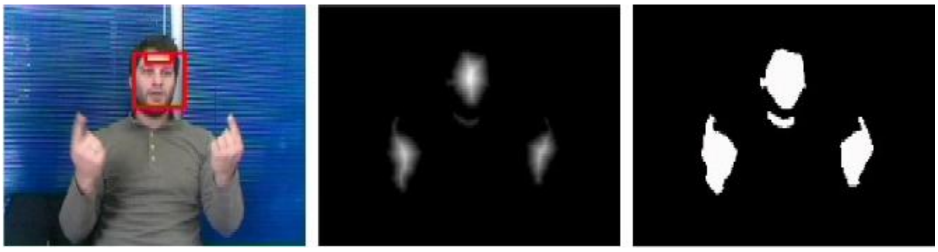

Fig.2. Example of sequences of geometrical features to represent the informational content and meaning of the signs.

Fig.3. Segmentation and tracking of the image sequence from the sign video for the word “return” in GSL using computer vision techniques.

2.3 Probabilistic recognition scheme for Automatic Sign Language Recognition from multiple cues

This module focuses on a novel classification scheme incorporating Self-organizing maps, Markov chains and Hidden Markov Models. Extracted features describing hand trajectory, region and shape are used as input to separate classifiers, forming a robust and adaptive architecture whose main contribution is the optimal utilization of the neighboring characteristic of the SOM during the decoding stage of the Markov chain, representing the sign class. Although an abundance of architectures have been proposed [4] for automatic sign language recognition achieving true signer independence and signer performance invariant recognition still remains a challenging issue.

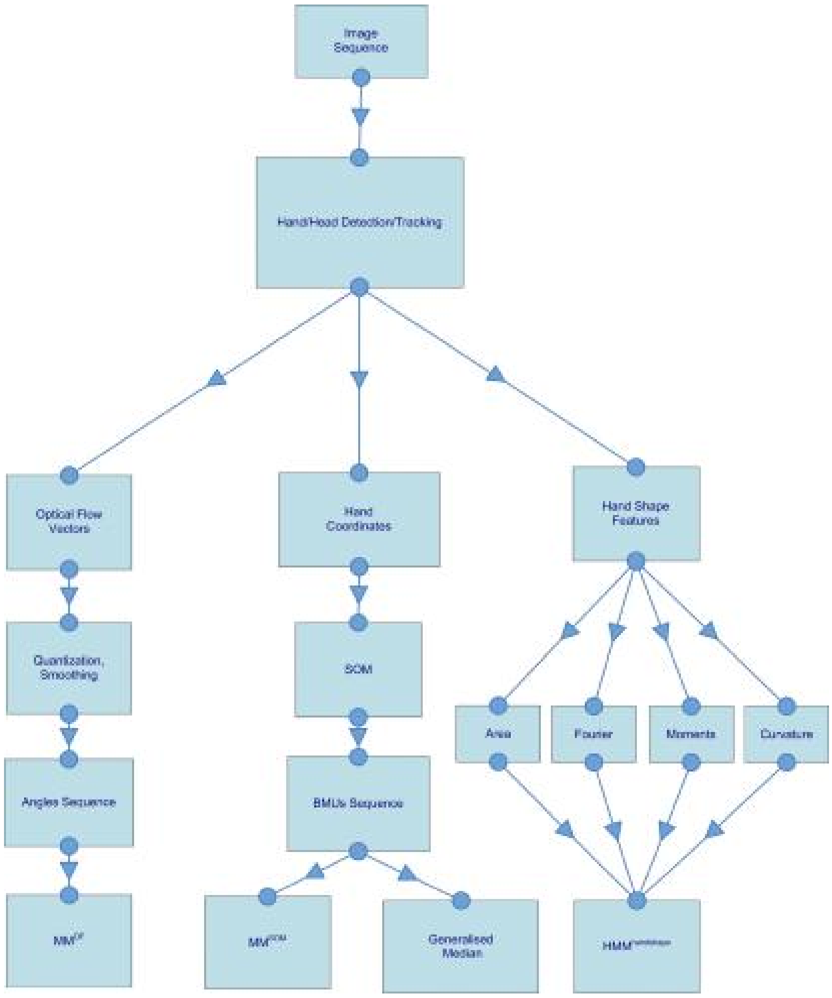

Fig.4. Recognition system architecture.

The proposed recognition scheme can be decomposed in separate component models for sign trajectory and hand shape cues which are then fused as can be seen on Figure 4. A novel approach is introduced by applying a combination of Self Organizing Maps (SOM) and Markov models for sign trajectory classification [3]. The extracted features used in the trajectory module include the trajectory of the hand and the direction of motion in the various stages of the gesture. This classification scheme is based on the transformation of a sign representation from a series of coordinates and movements to a symbolic form and on building probabilistic models using these transformed representations. The abstraction of the symbolic form enriches the classification scheme with adaptability, due to the incorporation of the neighboring function of the SOM during the decoding phase. Concerning hand shape, Hidden Markov Models are used to classify each sign instance into one of the models created by training a unique model per corresponding class per feature set. Our study indicates that, although each of the two sets of features (trajectory and hand shape information) can provide distinctive information in most cases, only an appropriate combination can result in robust and confident user independent sign language recognition.

2.4 Pilot Application: Teleoperation of a Mobile Robot Vehicle

In the frames of the DIANOEMA research project, described in this paper, and beyond, we are considering applications of multi-modal human-machine interfaces, incorporating vision-based human interaction modalities by means of natural and intuitive (hand, body or facial) gestures. In this context, a pilot application has been developed that concerns hand-gestural teleoperation of a mobile robotic vehicle. The first step was to design an appropriate “vocabulary”, which consists of a small set of hand signs (for the time being, static hand postures) that constitute a robot command language. A “desktop” teleoperation scenario was selected, as illustrated in Figure 5, where the gestural commands of the human operator are issued remotely, from a master control station that supports all the necessary computer vision gesture recognition operations.

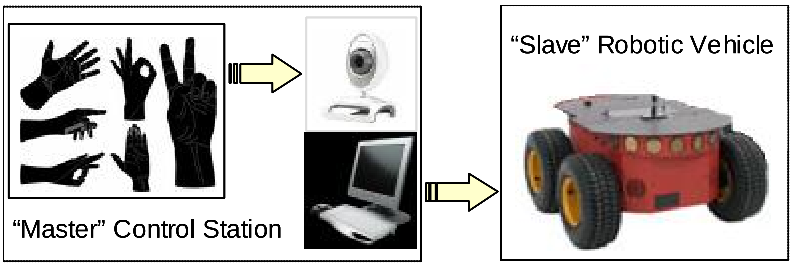

Fig.5. Gestural teleoperation of a mobile robotic vehicle: “Desktop” scenario.

A multi-level teleoperation architecture has been considered, inspired by related work in the field of telerobotics [5][6]. The system supports: (a) low-level, direct teleoperation commands (such as: <move-forward>, <rotate><right> [| <left>] etc.), (b) mid-level shared-autonomy commands (based on autonomous, sensor-based robot behaviours, such as: <follow-wall><left> [| <right>] etc.), and (c) high-level operations (e.g: <go-to-room><#value>), which are implemented and included in the command set. For the first pilot application scenario, a high-level teleoperation sequence was implemented, which consists of issuing a command of the above third type (autonomous mode of operation). Autonomous mobile-robot navigation algorithms have been implemented and tested experimentally, including: (a) path-planning, (b) collision avoidance, and (c) continuous localization and motion correction (based on static geometric landmarks). Experiments have been conducted at the premises of the NTUA-ECE Intelligent Robotics Laboratory, using a Pioneer 3-DX indoor mobile robot platform (manufactured by MobileRobots Inc., formerly ActivMedia Robotics). Further experiments are planned for the near future, in order to assess performance, both in terms of real-time static hand gesture recognition, as well as the robustness of the mobile robot navigation algorithms, within stochastic and dynamic environments.

3 Future Research

Since sign recognition involves interaction of a number of different technological and scientific domains, improvement of performance is highly dependant on the mutual feedback and effective interdisciplinary research. DIANOEMA team has entered a new period of research effort with emphasis on:

- Computer vision with experimentation on the development of technologies for automatic sign language recognition based on visual information
- HCI with development of more flexible and human “friendlier” interfaces
- Sign language modelling for use in Natural Language Processing environments
- Robotics, by researching intelligent robot systems in the service of humans, i.e. developing autonomous navigation of robot vehicles via tele-operation

Acknowledgments. This work has been partially funded by the European Union, in the framework of the Greek national project DIANOEMA (GSRT, M3.3, id 35).

References

1. Efthimiou, E. & Fotinea, S-E.: Annotation and Maintenance of the Greek Sign Language Corpus (GSLC). Proceedings of Sign Language Corpora Workshop, LREC 2008, Marrakech (2008)
2. Efthimiou, E. & Fotinea, S-E.: GSLC: Creation and annotation of a Greek Sign Language Corpus for HCI, Lecture Notes in Computer Science (LNCS), C. Stephanidis (Ed.): Universal Access in HCI, Part I, HCII 2007, LNCS 4554, pp. 657–666 (2007)
3. Caridakis, G., Diamanti, O., Karpouzis, K., Maragos, P.: Automatic Sign Language Recognition: vision based feature extraction and probabilistic recognition scheme from multiple cues, 1st ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA), July 15-19, Athens, Greece (2008)
4. Ong, S., Ranganath, S. Automatic sign language analysis: a survey and the future beyond lexical meaning, Transactions on Pattern Analysis and Machine Intelligence, 27, 873-891 (2005)
5. Sheridan, T. B. (1992). *Telerobotics, Automation and Human Supervisory Control,* The MIT Press, 1992.
6. Tzafestas, C. S., "Virtual and Mixed Reality in Telerobotics: A Survey", in: *Industrial Robotics: Programming, Simulation and Applications* (Edited by: Kin-Huat Low), Advanced Robotic Systems International 2007, Published by pIV pro literatur Verlag Robert Mayer-Scholz, Mammendorf, Germany (ISBN 3-86611-286-6), pp.437–470, Feb. 2007.