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

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

Extended abstract

Here we present research work performed in the framework of the Greek national project DIANOEMA (GSRT, M3.3, id 35), focusing 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.

The design of the GSL corpus content has been driven by the demand to support sign language recognition as well as theoretical linguistic analysis. Content organisation makes a distinction between three parts: i) a list of lemmata representative of the use of handshapes as a primary sign formation component, and developed on the basis of measurements of handshape frequency of use in sign morpheme formation. This list provides a reliable test bed for initial recognition of single articulation units. Lemmata comprise (a) commands related to robot motion control and (b) simple and complex sign morphemes, representative of the basic vocabulary of GSL; ii) structured sets of elicited utterances, which form paradigms capable to expose the mechanisms GSL uses to express specific core grammar phenomena; iii) free narration sequences, intended to provide data of spontaneous language production for use in machine learning.
For morpheme level annotation the HamNoSys notation system was adopted. For sentence/phrase level annotation the ELAN annotation tool was used.

As regards 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:

1. Development of algorithms for automatic detection of the signer in a video frame
2. Selection of reliable and informative features, which facilitate and also define the gesture recognition process
3. Motion tracking and segmentation of the signer's movements (for instance hand, fingers, arm etc)

A system was developed for the detection and localization of the signer in the image. Also 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.

An example of segmentation and tracking of the image sequence from the GSL sign video using computer vision techniques is presented in Figure 1.

The module for automatic sign language recognition from multiple cues 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. The proposed recognition scheme can be decomposed in separate component models for sign trajectory and hand shape cues which are then fused. A novel approach is introduced by applying a combination of Self Organizing Maps (SOM) and Markov models for sign trajectory classification [Caridakis2008]. 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.

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 2, 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.

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 tele-operation 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.

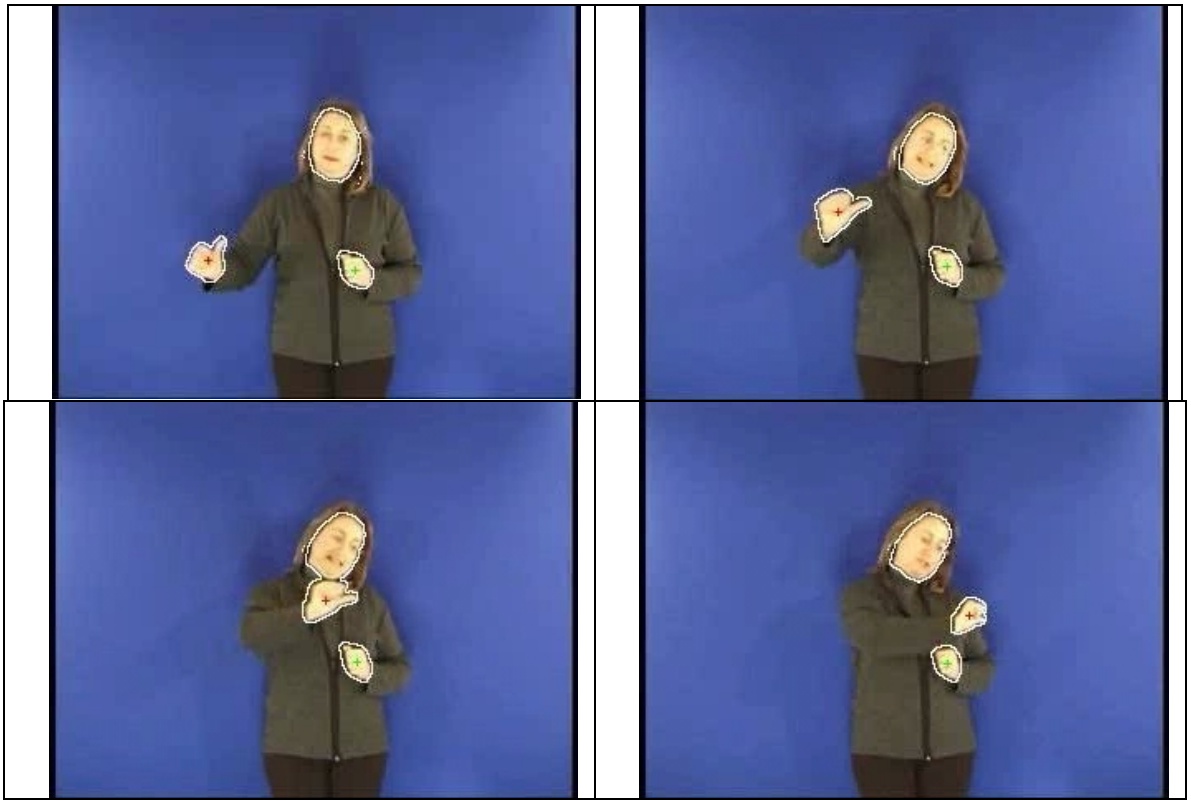

Figure 1. Segmentation and tracking of the image sequence from the sign video for the word “return” in GSL using computer vision techniques.

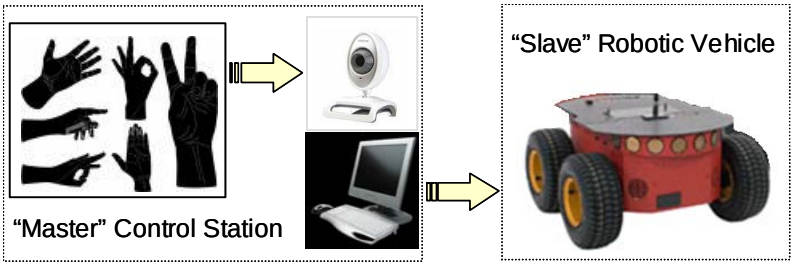

Figure 2. Gestural teleoperation of a mobile robotic vehicle: “Desktop” scenario.