Gesture Human-Computer Interface for Command and Control

José Miguel Sales Dias (*,**)
midias@microsoft.com

Pedro Nande (**)
Pedro.Nande@iscte.pt

Nuno Barata (**)
Nuno.Barata@iscte.pt

André Correia (**)
Andre.Correia@iscte.pt

(*) MLDC, Microsoft Language Development Center, MSFT LDA, Av. Prof. Aníbal Cavaco Silva, Edfício C1-C2, Tagus Park, Porto Salvo, Portugal, www.microsoft.com/portugal/mldc
(**) ADETTI/ISCTE, Associação para o Desenvolvimento das Telecomunicações e Técnicas de Informática, Edifício ISCTE, 1600-082 Lisboa, Portugal, www.adetti.pt

ABSTRACT

In this paper, we describe, test and validate through usability evaluation, a hand gesture recognition engine based on computer-vision. In our approach, we use a computer, with limited resources and a simple video camera, as a basic platform to enable the development of generic gesture-based Human-Computer Interfaces appropriate for Command and Control (C&C) tasks in computer applications and at interactive rates. With our technique, the system removes initially the background of captured images so that irrelevant pixel information is not considered in further analysis. The human hand motion is then detected, segmented and its contours localized, to allow a comparison with a pre-defined static hand poses library using various algorithmic approaches, some image-based other contour-based. Each hand pose, once recognised, can activate Human-Computer Interface command and control actions, organised in a hierarchy manner, which are specific for a given application. This work has been concluded in close collaboration with the Portuguese Foundation of the Communications, an institution owned by Portugal Telecom, the largest Portuguese operator. This cooperation has enabled the development and deployment of a demonstration application, using this platform, in the "House of the Future" test-bed, of the Communications Museum in Lisbon, Portugal. In the application, we enabled the command and control of some consumer devices linked within a home network, such as controlling the TV, the room lighting or the windows opacity characteristic. The paper also presents experimental results, regarding the precision and usability of the implemented system and discusses the best algorithmic approach for the purpose of identifying static hand poses at almost real-time rates and analyses future directions of our work.

Keywords

Gesture Recognition, Hand Poses, Computer Vision, Human-Computer Interaction, Command and Control, Portuguese Sign Language, Home Network

1. INTRODUCTION

The recent development of applications in the domain of computer vision has enabled the exploration of new approaches in the way humans interact with machines and with computers. The ideal form of interaction would be, for a given person, to have a familiar way of accessing specific computing application functionalities, without the need of using specific peripheral devices that could distract the user from the focus of his/her task. That could mean for the same person, to communicate with a machine much in the same way he/she would interact with other humans. There has been a considerable evolution in the issue of natural modalities of human-machine interaction, and new solutions have shown relevant improvements and have demonstrated that it is possible to use different modalities, such as speech, gesture or other natural media as multiple ways to communicate with the computer. With our work we are aiming at developing an open source gestures recognition engine that would offer the possibility to trigger user-specified command and control actions, activated by different hand gestures. The system requires a single video camera linked to a computer that captures the user's movements and recognizes hand gesturing, namely, a set of pre-defined static hand poses. A "static hand pose" is a gesture represented by a single hand pose with a spatial position that doesn't vary much in time. As an example, it could be a symbol of a sign language alphabet [Kadous95] [Niwa02]. This type of gestures, which are simple to

issue by humans, can trigger actions in the application level providing human-machine interaction. We have followed this approach in the "House of the Future" test-bed of the Communications Museum in Lisbon, Portugal, where we have developed a demonstration application that uses gesture-based HCI in a living room scenario. In our set-up the application has access to the home network of the house, which enables the user to interact and control some of the home networking linked devices like, a TV Set-Top-Box, the window glass transparency characteristics (opaque/transparent) and activating macro commands, such as opening/shutting the lights and opening/shutting down the blinds at nightfall. Our analysis of the Portuguese Sign Language signs [Hub98] has largely influenced our selection of the static hand poses, whose detection and identification as triggered our technique of gesture-based HCI. In synthesis, our system recognizes a set of static hand gestures based on computer vision that are used in an application that interprets those gestures and activates hierarchical actions, like controlling home devices and changing indoor's ambient conditions. With our technique we are able to achieve interactive rates (near real time performance, with an average of 6 frames per second on a Pentium IV with 3 GHz CPU, 512 MByte RAM, NVIDIA GeForce 4 with 64 MByte and 640x480 video input resolution).The paper is organized as follows: in section 2. RELATED WORK, we provide a short background in the issues of gesture data acquisition and recognition. In section 3. OPEN GESTURES RECOGNITION SYSTEM ARCHITECTURE, our system architecture, engine modules and dataflow, including the way we classify the type of recognised gestures and triggered actions, supported by our system. In section 4. PRECISION TEST RESULTS AND DISCUSSION, we detail our experimental performance testing results and discuss about the best studied methods for robust static hand shape recognition. In section 5. APPLICATION OF GESTURE HCI FOR COMMAND AND CONTROL, we describe the development, user interface, available functionalities and deployment of a demonstrator, which has been set in the "House of the Future" of the Communications Museum in Lisbon, Portugal. By using this test-bed we detail, in section 6. USABILITY EVALUATION, the usability evaluation results obtained with the mentioned application in the "House of the Future". Finally, in section 7. CONCLUSIONS AND FUTURE RESULTS, we derive conclusions of the research done and point out future directions for our work.

2. RELATED WORK

In this section we will present a short overview of related work in the domain of gesture recognition. There are two different approaches when acquiring data for gesture recognition: one based on sensory devices and the other a computer vision approach. There are also hybrid solutions that combine both fields, aiming towards a pragmatic solution.

2.1 Approach Based on Sensor Devices

This approach is focused in Mechanical, Electronics and Electromagnetic Engineering research and suggests the use of physical devices that can measure the variations of certain values that occur while performing a gesture, such as position changing, orientation, accelerations and forces made by the hand. The gesture analysis is essentially mechanical and electromagnetic, consisting in the evaluation of these physical parameters. Due to the specificity, complexity and cost associated to this type of electronic devices the systems based on this approach are usually less scalable and expensive. Nevertheless, there are academic developments of these solutions [Kadous05] [Geoffrey98] and commercial applications with some acceptance in the market.

2.2 Approach Based on Computer Vision

Gestures recognition systems based on computer vision are subject of widely diffused investigation works. Although theoretic approaches tend to perspective user gestures in the large and general problem of human motion, several results have focused on practicality and usability. Therefore, there are gesture-recognition specific systems of great interest. In [Chaudhurya00] [Niwa02] [Shirai02], we find gesture recognition systems applied to Sign Language interpretation. Other systems, like in [Aguilar03] are more directed for human-computer interface control. Recognition techniques, either image or vector based, all tend to acquire hand contour as a starting point for gesture perception and then its motion, have been somewhat influenced by human visual system processes. Background complexity is not frequently addressed, requiring gestures to be executed against a homogenous prepared scene. Related works from the computer vision domain, like advanced background subtraction [Davis99], can be useful for gesture recognition. Alternative approaches are based on mathematical descriptions of the whole captured image itself, as in [Maydt02], where Haar-like features are used to recognize trained objects, with an interesting potential for hand detection.

3. OPEN GESTURES RECOGNITION SYSTEM ARCHITECTURE

3.1 Architecture Overview

Our O.G.R.E, Open Gestures Recognition Engine (not to be confused with a popular 3D Engine [OGRE3D]), simplified modular system architecture and process flow, is depicted in Figure 1.

The system is conceptually divided in 3 abstract layers, from bottom (level 1) to top (level 3), which were defined in order to tackle the problem of defining a gesture recognition engine in a generic way:

- Level 1: Preparation, extraction and manipulation of visual information from a sequence of images.
- Level 2: Low level gesture recognition;
- Level 3: Interaction with the application layer.

In level 1, we have Video Capturing, responsible for image acquisition from the video camera, which feed the

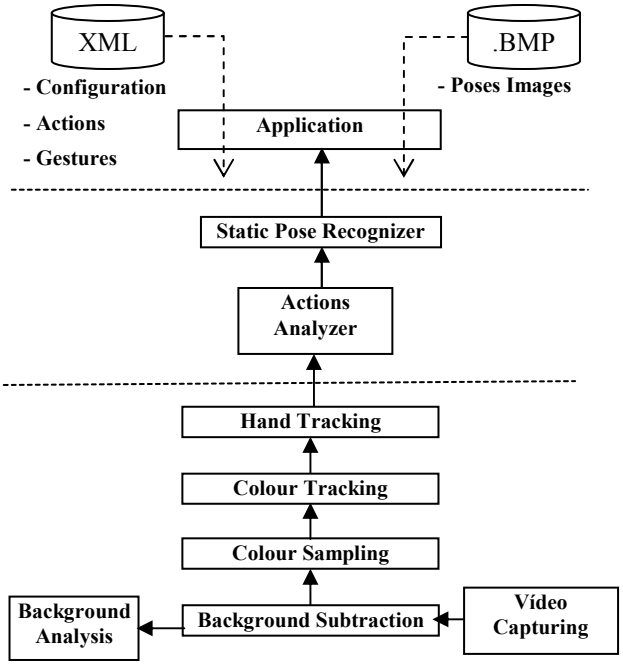

Figure 1. O.G.R.E. System Architecture.

Background Subtraction module. This last one provides the adaptive elimination of the static background, being aided in this task, by the Background Analysis Module. Once processed, images are sent to the Colour Sampling module, where the human skin tonality can be sampled or parameterized. With this information, the colour tracking module, scans subsequent images, looking for the skin Hue tone. With the found tonality area, the Hand Extraction module, will obtain the hand contour and pixmap mask, which will be optionally processed by the Contour Handling and B-Spline modules, to filter any contour noise, through mathematical approximation techniques. In Level 2, the parameters obtained in the previous Layer, are processed by the Actions Analyser Module, responsible for the XML configuration of the different gesture recognition algorithms and the invocation of the corresponding algorithm (only Static Pose Recognition is used in this paper, but the system also supports Simple Path Recognition, Staged Path Recognition and Free Path Recognition). Once a specific gesture is recognised and the corresponding Action is identified, by this Layer, the result is communicated to the Application Layer. Level 3 is composed by the Application module, which represents then user application that creates and invokes the OGRE engine. The XML Gesture-Action configuration file and the pixmaps with the known 2D static hand poses are supplied by the application to the engine.

3.2. Gestures and Actions Definitions

Static Poses

Static poses are rigid hand postures (a type of gesture according to our classification) which do not depend on the movement of the hand. They are trivially characterized only by their shape in the form of a 2D silhouette (a bitmap) stored in a user-specific database.

A static pose is defined in XML by the following notation:

```
<StaticPose name="A" filename="A.bmp">
</StaticPose>
```

Actions

Our engine introduces the abstract notion of actions, which are contextualized hierarchies of application dependent functionalities (requiring static hand gesture recognition for HCI), which feed the gestures engine and provide a guide for available recognition algorithms selection. Actions inform the application of the type of gesture to recognize at a given moment, thus minimising the need to activate all the different gesture recognition types simultaneously. Actions also simplify the communication between the engine and the application, since this last one does not need to know what gestures have been recognised, but only which actions have been activated. An Action corresponds to an application functionality. It is compose by an activation gesture and an optional termination gesture and, if required, by a list of optional child actions, which may be activated on a given application context.

There is an XML description of gestures and actions, which is useful for configuration purposes. A simple action is composed by a valid starting and an ending gesture (this last one is optional), which, respectively, trigger and terminate the action. In the context of this paper, a valid gesture is a StaticPose:

```
<Action name="mySimpleAction1">
   <Start gesture_name=”A”></Start>
   <End gesture_name=””B””></End>
</Action>
```

An action can also be composed of child actions previously defined, accessible in its context. These can be simple or composed actions:

```
<Action name="myComposedAction1">
   <Start gesture_name=”C”></Start>
   <End gesture_name=”D”></End>
   <Actions>
        <Action name=”mySimpleAction1”></Action>
        <Action name=”mySimpleAction2”></Action>
        <Action
name=”myComposedAction2”></Action>
   </Actions>
</Action>
```

There is a Root Action that determines the topmost hierarchical level and those identified as its children corresponds to the first hierarchical level of known actions:

```
<Action name="root">
   <Start gesture_name=""></Start>
   <End gesture_name=""></End>
   <Actions>
      <Action name="myComposedAction1"></Action>
      <Action name="mySimpleAction1"></Action>
      <Action name="mySimpleAction2"></Action>
   </Actions>
</Action>
```

For the purpose of the application installed in the “House of the Future”, we have identified one type of hand gestures, suitable for simple forward recognition: Static

Hand Poses. However, in addition to Static Hand Poses, the O.G.R.E engine also supports Simple Paths (primitive shapes such as circles, squares and triangles), Staged Paths (a mixture of Static Hand Poses and Simple Paths) and Free Tracking [Dias04], which are out of the scope of this paper.

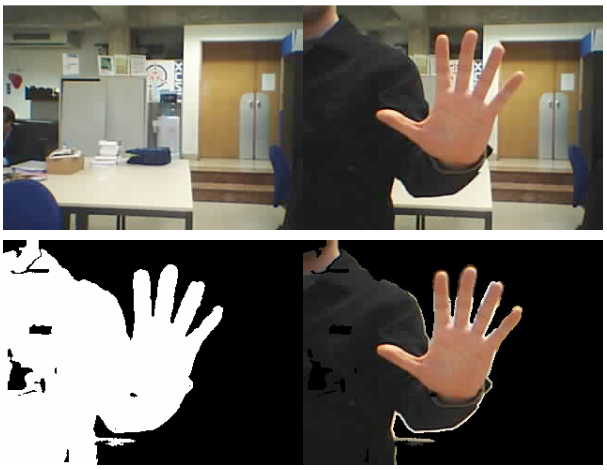

Figure 2. Subtracting the background. Top Left: the original background. Top Right: The moving user in foreground. Bottom Left: the mask with calibrated background removed and subsequent changes maintained. Bottom Right: Foreground user with subtracted background.

3.3. Engine Modules

Background Subtraction

Background subtraction [Davis99] is applied prior to any subsequent processing. It consists of a calibration period during which maximum and minimum per-pixel values in the YCrCb domain are stored and updated. After this initial period, foreground classification occurs, based on simple comparison between actual frame pixels YCrCb values and the stored background model, since its assumed that variations of actual frame pixels YCrCb values below the stored minimum or above the store maximum, classify them as foreground pixels (Figure 2).

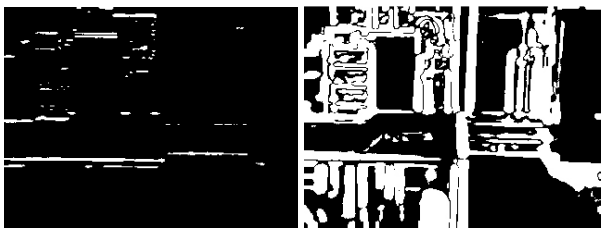

Figure 3. Background subtraction mask noise due to lighting changes (left) and camera positioning variation (right).

Background Analysis

This module is responsible for background deterioration detection (Figure 3). It has been observed that the background subtraction algorithm used is not resilient to environmental changes, such as light fading, scene decorative objects replacement and camera positioning instability. Therefore, for long functioning periods, it is necessary to robustly adapt to new scene conditions without user intervention. The algorithm used for detecting such changes is based on a timed observation of the background subtraction mask. In normal conditions, this mask is a binary image composed of both black and white pixels, representing background and foreground regions respectively. Foreground regions, as typically observed during user interaction, are a reduced number of large connected white areas (about one or two spots of a quarter image area). When deteriorated, this mask will contain several smaller regions spread throughout expected locations (the edges of contrasting background elements, such as closets, doors, tables, etc). These spots can therefore be classified as noise, as they are unwanted, disruptive elements for both background and colour segmentation. Their detection is done by finding small separated regions of 'white' value. For noise classification purposes, two measures are considered: 1) a maximum fraction of occupied area of all spots, relatively to the image (a configurable parameter), above which it is seen as noisy, and 2) the minimum noise coverage area, relative to the image's area, above which it is considered large enough to cause interference (also, a configurable parameter). A second type of noise is also observed when environmental conditions vary drastically (turning lights on or off, camera's field of view occlusion by large passing objects, etc). In all these cases there are no small spots visible, as the entire mask is constituted by a large white stain. This is also disruptive and therefore, considered noise. Its detection consists on simply establishing a maximum area for white value regions occupation (a configurable parameter), above which such elements are taken in consideration as noise. In both algorithms, noise detection triggers a positive alarm. During normal interaction a time window of such alarms is analyzed. If these positive alarms appear in a considerable number, recalibration will occur automatically (Figure 4). During calibration, no time window is analyzed, as a unique positive alarm is sufficient to destroy the entire statistical model

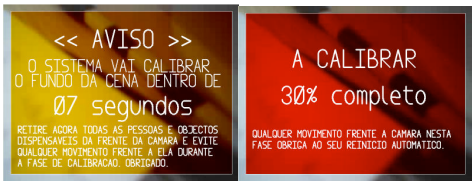

Figure 4. Automatic re-calibration of the scene: Left – Notification of the re-calibration; Right – re-calibration process.

Colour Sampling

In each frame, the hand is localized via colour information based on a Hue interval, either user specified by a specific initial interaction (the user approximates is hand to the video camera) or dynamically sampled, as in [Niwa02] or [Shirai02]. Our technique has proven to be experimentally robust to luminance variation and easily finds the skin tonality in a given environment.

Colour Tracking

Hue values obtained in the previous colour sampling phase, feed the CAMSHIFT algorithm (Continuously Adaptive Mean Shift [Bradski98]), which is then applied

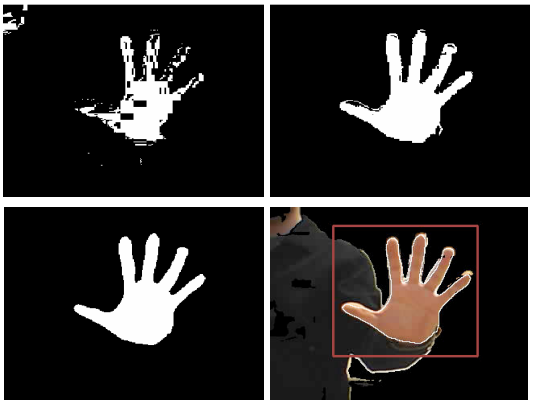

Figure 5. Extracting hand contour. Top left, the CAMSHIFT algorithm histogram back projection result. Top right: Result after YCrCb re-sampling. Bottom Left: result after morphological smoothing. Bottom Right: result after contour vectorization (white contour), with CAMSHIFT window limited to the green bounding box and mixture with foreground image.

to the current captured image. The CAMSHIFT algorithm computes a histogram back projection binary image (with reduced resolution, in respect to the initial image luminance, since the algorithm works in the sub-sampled chrominance domain), representing areas of a specified Hue tonality (the hand tonality, in our case). It also reduces the hand pose searching area to the largest connected component representing the user's hand Hue. Hand tracking is therefore guaranteed in the following frames (Figure 5).

Hand Extraction

The resulting hand contour image is still inadequate for static pose recognition. In order to extract a realistic hand silhouette, further processing is needed. In our algorithm, the histogram back projection binary image is used as a mask for YCrCb color space re-sampling, applied to the initial image with the background removed. With this, we intent to obtain richer information from image planes more suitable for edge extraction. Luminance and chrominances are sampled at different resolutions in order to achieve the best possible contour detail. With this process we are able to determine the average of YCrCb pixel values of the hand region, and perform a better classification of pixels, belonging or not, to the hand contour and interior. A smoothing morphological operation is then applied with an adequate structural filtering element for noise reduction. This element has dimensions which can be of 5x5, 7x7, 9x9 or 11x11 pixel, depending in the estimated silhouette dimension. The contour is then vectorized and, if necessary, a polygonal approximation sensitive to finger curvature is applied. This approximation is based on the best fit ellipse mathematical approach, as to obtain a measure of a given set of point's curvature, proportional to the ellipse eccentricity.

Actions Analyser

This module is the engine's core "intelligence". It analyses a specific action context and redirects gesture recognition into the adequate set of possible hand poses recognition algorithms (in the case of this paper, only static hand pose is considered).

Static Hand Pose Recogniser

In order to recognize static poses, several widely known algorithms for shape analysis were studied. We can divide these algorithms in two categories, Image Based Analysis and Contour Based Analysis. We have studied the following:

Image Based Analysis:

1. Template Matching [Bradski02]: Based in the convolution between two images at several scales in order to find a known template. It is scale invariant but variant to rotation and translation.
2. Discrete Cosine Transform Analysis: [Gianino84] A scale and rotation independent transformation in the frequency domain.

Contour Based Analysis:

3. Hu Moments [Bradski02]: These are a set of shape characteristics, invariant to rotation and scale metrics that can be useful for shape classification.
4. Pair-wise Geometrical Histogram (PGH) [Heuer84]: This method, computes the Histogram of distances and angles between the contour polygon's edges, which provides us with a unique contour signature. PGH is scale invariant and symmetry invariant as well (in relating to the *xx´* and *yy´* image axes). However, PGH is rotation variant.
5. Simple Shape Descriptors (SSD) [Lu02] [Fonseca00] [Eckhardt00] [Wirtht02], combine simple metrics which help describing shapes.
6. PGH-SSD Hybrid: This method corresponds to the authors efforts in combining the PGH and SSD advantages.
7. CALI [Fonseca00]: This is a software package, based in Fuzzy Logic, used normally to recognise sketched shapes in the context of calligraphic interfaces. The technique may bring advantage in the recognition of static hand poses, by introducing a probabilistic methodology in the recognition technique. It is invariant to rigid body transformations (scale, rotation and translation).

For static Hand Pose Recognition, the extracted hand silhouette is compared against a library of silhouettes templates or a library of silhouettes signatures (depending in the method), using one of the above algorithms. All algorithms were tested regarding its efficiency and precision, as described in the next section.

4. PRECISION TESTS RESULTS AND DISCUSSION

As a test case, static hand pose recognition algorithms were tested for the Portuguese Sign Language (PSL) 36 signs (more precisely, for 2D configurations of such signs, see Figure 6) and an Average Recognition Rate (%) was recorded. The selection of this test case, had to do with the fact that it constitutes a complex problem with a variety of 2D hand shapes, not yet addressed in the literature, for the particular PSL case (although considerable work does exist for other Sign Languages).

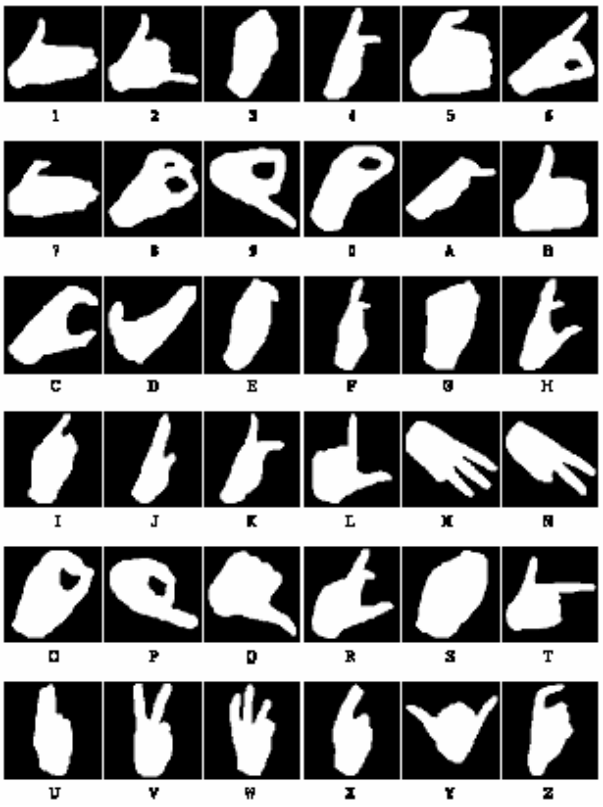

Figure 6. Portuguese Sign Language: 2D Sign configurations.

Processor	Clok Freq.	Graphics	Video Input Resolution[1]
Intel Pentium IV	3 GHz	NVIDIA GeForce 4	640 x 480

Table 1. Hardware configuration used for the precision tests.

The interest of the authors of pursuing, in the future, studies towards full PSL recognition, has influenced also the choice of this use case. These were the experimental conditions:

1. Each hand pose was tested ten times against a user library (that is, the subject performing the gestures was the same user whose hand poses had been stored in the library) and, against another person's library (the subject performing the tests was not the same one, whose hand poses images were in the library). Taking into account the number of PSL symbols, of libraries (user and generic, except for the CALI technique, which just use a generic library) and of hand pose recognition algorithms (7), this resulted in 4680 static hand pose user trials: $4680 = (36*10*2*6) + (36*10)$
2. The scene background consisted in a heterogeneous environment, with various objects with different colour hues (some approaching the skin tonality).
3. Given the existence of various light sources in the environment, and in order to minimise zones of super-exposition of light in the user hand, an area of shadow occupying the hand was created, to homogenize the luminous energy in the area of the hand.
4. The subject was wearing a long black shirt that effectively eliminated the harm skin, which interfered with the hand pose recognition algorithm.
5. The subject was knowledgeable of the PSL signs and the gesture technique to execute them, although he had no specific practice. With this option, the idea was to emulate an average user that has never practiced PSL and that could introduce human-originated errors in the process of generating the signs.
6. A given sign was considered to be recognised with success, if the system was able to identify it correctly almost immediately and, to stabilise the recognition during some seconds

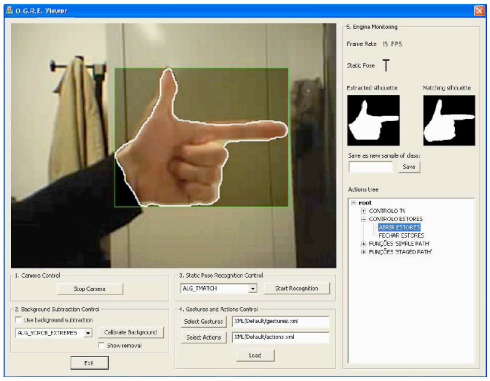

Figure 7. User Interface of O.G.R.E. Viewer, at the moment of identification of the static hand pose that matches the PSL sign "T.

Algorithms	Avg. Recognition Rate (%)	
	User Lib.	Generic Lib.
Template Matching	53,6	38,3
Discrete Cosine Transform	50,6	42,5
Hu Moments	25,8	17,8
PGH	58,1	30,3
SSD	5,8	2,8
PGH-SSD Hybrid	21,9	13,3
CALI	-	32, 5

Table 2. Testing Static Hand Poses for Static hand pose recognition algorithms for all the PSL symbols.

The hardware configuration used for the experiments, are described in Table 1, and an average of 6 frames per second were obtained during the trials. To aid the above described precision test process, an application was built (O.G.R.E. Viewer, Figure 7), which allows for the monitoring of the engine. The application enables:

- The visualization of the processed images, showing the detected hand contour, optionally with background subtraction.
- The selection of the pose recognition algorithm from a list of the 7 studied;

- The selection of the gesture and action XML configuration files;
- The identification of the recognised static pose and a tree to navigate in the context of triggered actions;
- The collection of static hand poses to add to generic hand pose library, use by the CALI algorithm.

The results presented in Table 2, show that Template Matching, Discrete Cosine Transform and PGH are, significantly, the best methods studied for robust shape recognition. Hu Moments are weak classifiers as well as Simple Shape Descriptors alone.

An attempt to improve PGH effectiveness was taken by creating a hybrid algorithm based in its integration with Simple Shape Descriptors. These can't robustly recognize individual symbols, but are useful when excluding some possibilities from the standard set of PSL. PGH would then be able to precisely pinpoint the performed pose from this shortened set. However, at the current stage, we were unable to effectively combine both algorithms own advantages, thus resulting so far on a poor hybrid method for shape recognition. The CALI technique, was able to obtain a success rate of 32,5% using a generic library (the technique supports a set of "generic" samples per pose). Although ranking below Template Matching and PGH, we believe that this technique has potential since it handles a training phase (which is inexistent in the other algorithms) and, therefore, in theory, could improve its performance by introducing in the library of the system, a larger number of better hand pose samples. However, the technique shows some limits, since not all the static pose shapes are adequate for recognition using this technique. Although the average recognition rate for the best algorithm (PGH) is less than 60%, when compared to similar state of the art systems having recognition rates up to 90%, we must consider that all possible symbols were included in the testing session. This means that highly correlated symbols have interfered in the process, thus causing false pose interpretations. When considered individually, we have identified a smaller set of Sign Language symbols which have a high standalone recognition rate (Table 2). When combined, and having selected an appropriate algorithm from the table above, the overall recognition rate can rise up to 90%. A carefully chosen subset (or a user defined set of creative poses), such as the ones depicted in Table 3, could be used in specific control functionalities. This was the approach taken in the application implemented in the "House of The Future", where the static hand poses shown in Figure 9, mostly related to Portuguese Sign Language Signs, were selected. After analysing the precision results, the Template Matching algorithm was the one used in the developed application we describe next, based on the algorithm good recognition precision results (Table 3).

Algorithms	Symbols
Template Matching	1,3,5,7,B,D,H,K,L,N,P,T,V,W,Y
DCT	1,2,5,7,9,A,B,D,L,N,P,T,Y
CALI	1, 5, 7, L
Hu Moments	5, H, L
PGH-SSD Hybrid	1,C,Z
SSD	T

Table 3. Subsets of PSL symbols with average recognition rates higher than 80%, for the Template Matching, DCT, CALI, Hu Moments, PGH-SSD Hybrid and SSD algorithms.

5. APPLICATION OF GESTURE HCI FOR COMMAND AND CONTROL

By developing a close collaboration with the Portuguese Foundation of the Communications, namely with the group that supports the citizens with special needs, a demonstration application has been set in the "House of the Future" test-bed of the Communications Museum in Lisbon, Portugal (Figure 8), enabling gesture HCI for command and control. Several demonstrators were already in place, such as, accessible Internet browsing for persons with vision impairments or speech activated commands to control house appliances. After identifying the opportunity to deploy gesture activated commands, especially oriented to citizens with hearing disabilities or motor impairments, other than the movements of the hands, we have collected the overall requirements of our application, namely:

Figure 8. Top: The application set-up as installed at the "House of the Future". Bottom: interacting via gestures with the house window characteristics.

- The universe of static poses should be reduced to a minimum, easily memorised;
- Since the application will not have any link with the interpretation of PSL (Portuguese Sign Language), the correspondence between the identified static poses, when these are used in such Sign Language, and the application symbols, should be changed in order not to confuse the user that is acquainted with PSL.

We have selected six static hand poses (Figure 9), which show small correlation and high recognition rate (above 80%) in the precision tests (Table 1), that will be used to

Figure 9. Six static hand poses of the left hand identified by our system, related to Portuguese Sign Language Signs. From top to bottom and left to right, we have the following signs: T, 1, Open Hand (not a sign), B, lying V and Y.

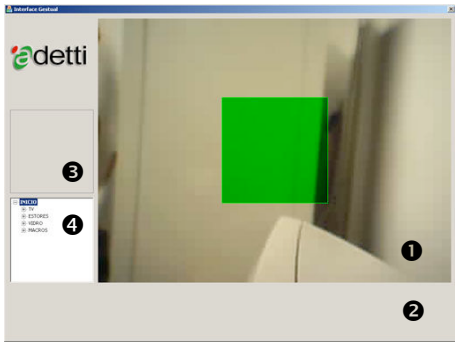

Figure 10. User's Interface. Field 1: Video Camera feedback; Field 2: Action Label; Field 3: Recognition feedback; Field 4: Actions Hierarchy Tree.

trigger user generated actions. The application commands (issued by gestures) were organised in a hierarchy manner, such that each hand pose would activate a menu option and, inside this menu, hand poses of the previous hierarchy level could be re-used.

Our O.G.R.E system architecture was integrated in a Home Network platform that controls home appliances, linked via ETHERNET, such as Set-Top-Box, TV Set, DVD player, Home lighting or Window Blinds developed by Portugal Telecom [Mar04]. A software module was developed to access a set of available Web Services that exposed the control of appliances of the house, for each interpreted and recognised gesture. By invoking these Web Services, it was possible to interface with the house equipment system command and control, therefore enforcing gesture commands for a set of selected equipments. A usage scenario in a living room, where a user interacts with home appliances (by means of selected static poses taken from the Portuguese Sign Language), was placed in operation. The application was installed in the child´s room, to exemplify the control of the surrounding home environment (Figure 10). The user's interface is very simple in order to promote straightforward demonstrations, during museum tours inside the "House of the Future" (Figures 8 and 10). It consists in a full-screen window where 80% of its area shows feedback from the users hand gesture action in front of the camera. There is an initial calibration process when the application is launched in order to perform background subtraction. After system calibration the user can perform hand poses to activate actions that allow the

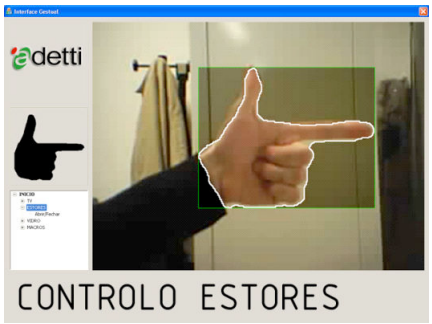

Figure 11. The application interface installed in the "House of the Future". The recognized static pose is presented to the user on the left, and the triggered action is indicated bellow (opening the window blinds).

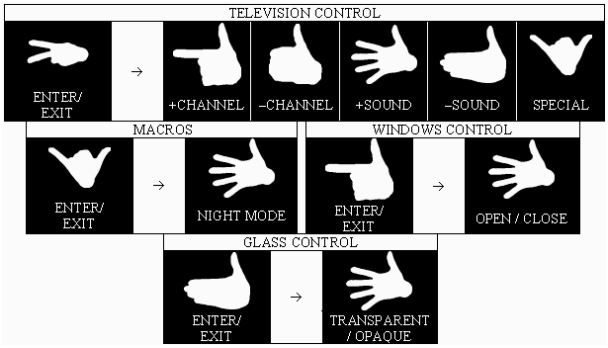

Figure 12. Available menus that can be browsed by performing specific hand poses.

command and control of house-hold devices. Figure 12 shows all available menu options. It is possible to control the TV set (select a channel up or down, turn the volume up or down or access pre-programmed channels), open or close the windows blinds, change the opacity of a polarised glass window and activate macros (set of simultaneous control actions inside the house, such as turn the light of or on or shutdown the windows blinds and night fall).When a drastic change of lightning or camera position occurs, the system launches an automatic recalibration process immediately (see section 4, Background Analysis).

6. USABILITY EVALUATION

The application was subjected to usability testing by 10 unpaid undergraduate students of both sexes, of the degrees of Engineering and Informatics at ISCTE in Lisbon, and five users of both sexes, which were support staff working in the "House of the Future. All the subjects had no apparent impairments. For this study, we have followed the model described in [Dias03]. The major aim of this testing phase was to understand how well adapted the system was to the needs and skills of the users. Three questionnaires were elaborated:

1. Questionnaire B.2, to assess the experience and opinion of subjects in relation to different user interface modalities, in relation to which our interaction technique could be compared.
2. Questionnaire B.3, filled after an audio-visual presentation, followed by a demo of the system in action, aimed at obtaining a first impression regarding the system functionality.
3. Questionnaire B.4, filled after the user trial, which followed a pre-established B.1 trial script.

The B.1 trial script guided the subject in a set of tasks that exercised his/her skills in executing all the gestures supported by the system. The subject was initially asked to browse the system menus, through static hand poses, that simulated the activation (in the lab), or activated (in the House), the remote control of the house devices, via gestures, such as the TV set (by increasing or decreasing the volume or snapping up and down the TV channels), the window binds (by opening or closing), or the window glass transparency (turning it transparent or opaque). The

subject was asked also to perform Simple Paths and Staged Paths [Dias04]. The answers to the Questionnaires were based in a quantitative scale of 7 levels, ranging from -3 (negative or a deficit assessment) to -3 (positive or even excessive assessment). Zero (0) was considered to be the neutral indifferent answer. The subjects were trained in the system concepts and in the tasks to be performed prior to the trials, which took a considerable amount of time. As mentioned, ten unpaid subjects of both sexes (but males in majority) were selected from the graduate and undergraduate courses in ISCTE, with ages ranging from 20 to 25 years of age, of the Engineering and Informatics courses. Five more exhibition guides in the "House of the Future" museum (mostly females of ages between 20 and 25 years of age), were also subjected to the same usability testing method. The results of questionnaire B.2, have showed that subjects were open to use "state-of-the-art" user interaction technologies, to assist in executing generic tasks. However, subjects were much more acquainted and, therefore, more biased towards mouse and keyboard interaction, than with other HCI modalities, such as gesture, although still favourable in trialling this last modality. The results of questionnaire B.3, have showed a high interest for the system, due to its high degree of novelty and unfamiliarity. When compared with other HCI modalities, subjects believed that it could be an alternative to data gloves. Static hand poses were considered to be easier than simple paths and these ones were believed to be more straightforward than staged paths. The global system impression was positive. Questionnaire B.4 results, have assessed the final opinion of the tested subjects. This could be compared with questionnaire B.3, were we could identify a decrease in the perceived ease of use of the system (producing physical gestures was more difficult than initially expected) and a small increase in the familiarity of use. In relation to B.3 results, the system was considered to be a viable alternative to remote commands of home equipment in addition to data gloves. Static hand poses were considered to be the simplest and easiest gesture to be executed, whereas the simple paths were more difficult and the staged paths were the ones that created the bigger difficulties in use. In general terms, the subject's reaction to the system was positive and they have considered that the system was easy to use, especially when good lightning conditions were made available.

7. CONCLUSIONS AND FUTURE WORK

In this paper, we have described the different architectural modules of a hand gesture recognition engine based on computer-vision. The system is configured with XML specifications that describe the type of gesture to be recognized in a given context with a number of possible static hand poses available. The system was evaluated, regarding its precision in recognizing with success certain hand poses. An experiment was set-up, where a subject was issuing static hand poses of Spelled Portuguese Sign Language, to assess the robustness of various algorithmic alternatives to handle the sub-problem of shape recognition, present in the hand pose understanding process. Our results have shown that Pair-wise Geometrical Histogram contour-based method, is the most effective in relation to the average symbol recognition rate metric, reaching the figure of 58.1% for the case of the own user library of symbols, followed by the Template Matching image-based method. If the test is only made with highly non-correlated symbols, the metric can rise up to 90%. In this case we have identified 15 symbols of PSL, whose recognition rate is higher than 80%. This means that these symbols van be used with success in gesture-based HCI tasks, where the basic gesture to issue is a static pose. This achievement is one of the original contributions of this paper, in addition to others, such as the integration of generic background removal with the gesture recognition technique, still achieving near real-time static hand pose recognition rates, (6 frames per second on a Pentium IV with 3 GHz CPU). We have also shown that our system can be generalized to application in general Human-Computer Interface tasks which require the static hand gesture recognition modality at interactive rates. A usage scenario in a living room, where a user interacts with home appliances (television, windows, lights, etc) by means of selected static hand poses, was developed and deployed in the "House of the Future" test-bed of the Communications Museum in Lisbon, Portugal, and it is still in operation at the time of writing of this paper. The application was subjected to usability testing by 10 users in our laboratory at ISCTE, and 5 more users inside the "House of the Future. In general terms, the subject's reaction to the system was positive and they perceived the system as being easy to use, especially when good lightning conditions were made available. Usability testing with people with special needs (hearing and motor impairments) would also be an enriching asset for this work and is planned in the mid-term. As a natural continuation of our work, we aim at bimanual gesture recognition, hand feature extraction for finger recognition and occlusion treatment and face motion detection, using vision based approaches, by possibly assessing other 2D contour detection techniques, such as Zernik moments [Prokop 02]. Approaches enabling more flexible hand pose recognition by compensating the position of a moving camera can also be addressed. By generating time-based vectors of characteristics related to hand, finger and face gestures, we plan to develop a statistical model based in Hidden Markov Models [Park 06], that can be trained with sufficient input gesture data of the mentioned type, in order to create robust and real-time gesture recognition engines, that can be applied to "hard" long-term problems, such as full Portuguese Sign Language interpretation.

8. ACKNOWLEDGEMENTS

The authors would like to thank Engº Gonçalo Areia, Engº Joel de Almeida, Drª. Isabel Manteigas, Sr. José Raposo (Fundação Portuguesa das Comunicações), Engº Pedro Santos, Engº Bruno Marques, Engª Clara Cidade, Dr. Tiago Alves (Portugal Telecom), Professor Joaquim Jorge, Engº Manuel da Fonseca (INESC-ID), Engº Pedro Santos (Microsoft), Engº Rafael Bastos (ADETTI), for

their valuable help in specifying and supporting our application.

9. REFERENCES

[Aguilar03] Aguilar, M., Joshua R., Hasanbelliu, E., "Facilitating user interaction with complex systems via hand gesture recognition". ACMSE'03.Knowledge Systems Laboratory, Jacksonville State University, 2003.

[Bradski98] Bradski, G., "Computer vision face tracking for use in a perceptual user interface". Intel Technology Journal. Microcomputer Research Lab, Santa Clara, CA, Intel Corporation, 1998.

[Bradski02] Bradski, G., "Intel Open Source Computer Vision library overview", Intel Labs, Intel Corporation, 2002.

[Chaudhurya00] Santanu Chaudhurya, Subhashis Banerjeeb, Aditya Ramamoorthya, Namrata Vaswania. "Recognition of dynamic hand gestures". Journal of The Pattern Recognition Society. Department of Electrical Engineering and Department of Computer Science Engineering, IIT Delhi, 2000.

[Davis99] Davis, L. S., Horprasert T., Harwood D. "A statistical approach for real-time robust background subtraction and shadow detection". Technical report, Computer Vision Laboratory University of Maryland, 1999.

[Dias03] Dias M, Jorge J, Carvalho J, Santos P, Luzio J, "Usability evaluation of tangible user interfaces for augmented reality", IEEE International Augmented Reality Toolkit Workshop: pp 54-61, 2003, Proceedings of ART03, 2nd IEEE International Augmented Reality Toolkit Workshop, Waseda Univ., Tokyo, Japan, Oct 07, 2003.

[Dias04] Dias, J. M. S., Nande P, Barata N, Correia A., "OGRE - Open Gestures Recognition Engine", in Araujo AA, Comba JLD, Navazo I, Souza AA, Eds., XVII Brazilian Symposium on Computer Graphics and Image Processing/II Ibero-American Symposium on Computer Graphics, SIBGRAPI/SIACG Curitiba, Brazil, Proceedings IEEE COMPUTER SOC: pp 33-40,, 17th-20th October 2004.

[Eckhardt00] Eckhardt, U., Latecki, L., Lakämper, R.. "Shape descriptors for non-rigid shapes with a single closed contour". IEEE Conf. on Computer Vision and Pattern Recognition (CVPR) 2000. Dept. of Applied Mathematics University of Hamburg, Germany, 2000.

[Fonseca00] Fonseca, M. J., Jorge, J. A. "Using Fuzzy Logic to Recognize Geometric Shapes Interactively", Proceedings of the 9th Int. Conference on Fuzzy Systems, (FUZZ-IEEE'00), pp 191-196, San Antonio, USA, 2000.

[Geoffrey98] Geoffrey E., Hinton S., Sidney Fels. "Glove-talk 2 - a neural-network interface which maps gestures to parallel formant speech synthesizer controls". IEEE Transactions on Neural Networks, Vol. 9, January 1998.

[Gianino84] P. D. Gianino J. L. Horner. "Phase-only matched filtering". Journal of Applied Optics, 1984.

[Heuer84] Heuer, J. Kaup, A., "Polygonal shape descriptors - an efficient solution for image retrieval and object localization", 34th Asilomar Conference on Signals, Systems and Computers. Siemens Corporate Technology, Information and Communications, Munich, Germany, 2000.

[Hub 98] Faria, I. H., Falé, I., Viana, M. C., Pereira, C., "A Língua Gestual Portuguesa Como Um Sistema Linguístico: análise de alguns verbos", Proceedings of XIV Encontro Nacional da Associação Portuguesa de Linguística, University of Aveiro, 1998.

[Kadous95] Kadous W.. "Grasp: Recognition of Australian Sign Language using Instrumented Gloves". Technical report, The University of New South Wales, Schools of Electrical Engineering and Computer Science and Engineering, October 1995.

[Lu02] Guojun Lu, Dengsheng Zhang. "A comparative study on shape retrieval using Fourier descriptors with different shape signatures". 2002. Fifth Asian Conference on Computer Vision (ACCV02). Gippsland School of Computing and Information Technology, Monash University, Australia, 2002.

[Maydt02] Maydt, J., Lienhart, R.. "An extended set of Haar-like features for rapid object detection", IEEE ICIP'2002, Intel Labs, Intel Corporation, 2002.

[Mar04] Bruno Marques. Web services da casa do futuro - Especificação Técnica. Technical report, Portugal Telecom - Sistemas de Informação, March 2004.

[Niwa02] Yoshinori Niwa, Kazuhiko Yamamoto, Terrillon, J., Pilpré, A. "Robust face detection and Japanese Sign Language hand posture recognition for human-computer interaction in an "intelligent" room". VI'2002. Office of Regional Intensive Research Project (HOIP), Softopia Japan Foundation, Faculty of Engineering, Gifu University, 2002.

[OGRE3D] www.ogre3D.org

[Park 06] Park, A., Lee, S., "Gesture Spotting in Continuous Whole Body Action Sequences Using Discrete Hidden Markov Models", in Sylvie Gibet, Nicolas Courty, Jean-François Kamp, Eds, Gesture in Human-Computer Interaction and Simulation, LNAI, Vol 3381: pp 100-112, 6th International Gesture Workshoop, GW 2005, Revised Selected Papers, 2006.

[Prokop 02] Prokop, R. J., Reeves, A. P. "A survey of moment-based techniques for unoccluded object representation and recognition". CVGIP Graphical models and Image Processing, 54(5): pp.438-460, 1992.

[Shirai02] Yoshiaki Shirai, Nobuhiko Tanibata, Nobutaka Shimada." Extraction of hand features for recognition of sign language words". VI'2002. Computer-Controlled Mechanical Systems, Graduate School of Engineering, Osaka University, 2002.

[Wirtht02] Michael A. Wirth. "Shape analysis and measurement". University of Guelph. CIS*6320 Image Processing Algorithms and Applications. Computing and Information Science Biocomputing Group, 2002.