

Documentation of Pointing and command gestures under mixed illumination conditions: video sequence database

M. B. Holte and M. Störring

Introduction

Robust computer vision based gesture recognition is important for future human computer interfaces. In collaborative Augmented and Virtual Reality, e.g., for urban planning, where several people work together on a virtual model of a town, a set of hand gestures such as select, copy, paste, and move would be more convenient than using devices like the mouse. A typical place to use (future) gesture interfaces is an office environment with cluttered background and mixed illumination conditions. The combination of artificial indoor illumination and outdoor illumination (through windows) may cause very high intensity variations and large changes in the colour of the illumination ranging from rather bluish outdoor to yellowish-reddish indoor. A cluttered background and such illumination conditions make the low level segmentation of computer vision based gesture interfaces often fail. In particular skin colour like objects and illumination colour changes are difficult to cope with, whereas the problem of high intensity ranges will be solved by future camera technology – high dynamic range cameras are already available that can capture much higher ranges than the human eye.

The sequences of this database were recorded using a rather high quality (low noise) 3CCD computer vision camera. The scene contains objects that appear under certain illuminations skin colour like, and the illumination colour ranges from indoor to outdoor illumination. The range of intensities is, however, not that drastic than it would be when mixing indoor and outdoor light sources. All light sources used are artificial ones and their intensities were adjusted such that the number of under and overexposed pixels is small. Although this is not realistic it is not considered as a “constraint” since camera technology that can cope with high intensity ranges is on its way.

This document describes the scene setup and camera/software settings utilized for capturing the 16 video sequences of hand gestures.

The document includes issues as, the chosen gesture vocabulary, the scenario script performed by the test persons, scene setup, camera calibration/software settings, annotation and download of video sequences. These information can be found in the following five sections:

- Gesture Vocabulary
- Scenario Script
- Scene Setup
- Camera Calibration
- Annotation
- Gesture Video Sequence Database

Gesture Vocabulary

The gesture vocabulary consists of 13 gestures. Where 9 gestures are static and 4 are dynamic. All other hand movements and postures are included in an “unspecified gesture”. The 13 chosen gestures are shown in Figure 1 to Figure 5. Figure 1 illustrates the static gestures and Figure 2 to Figure 5 the dynamic gestures.

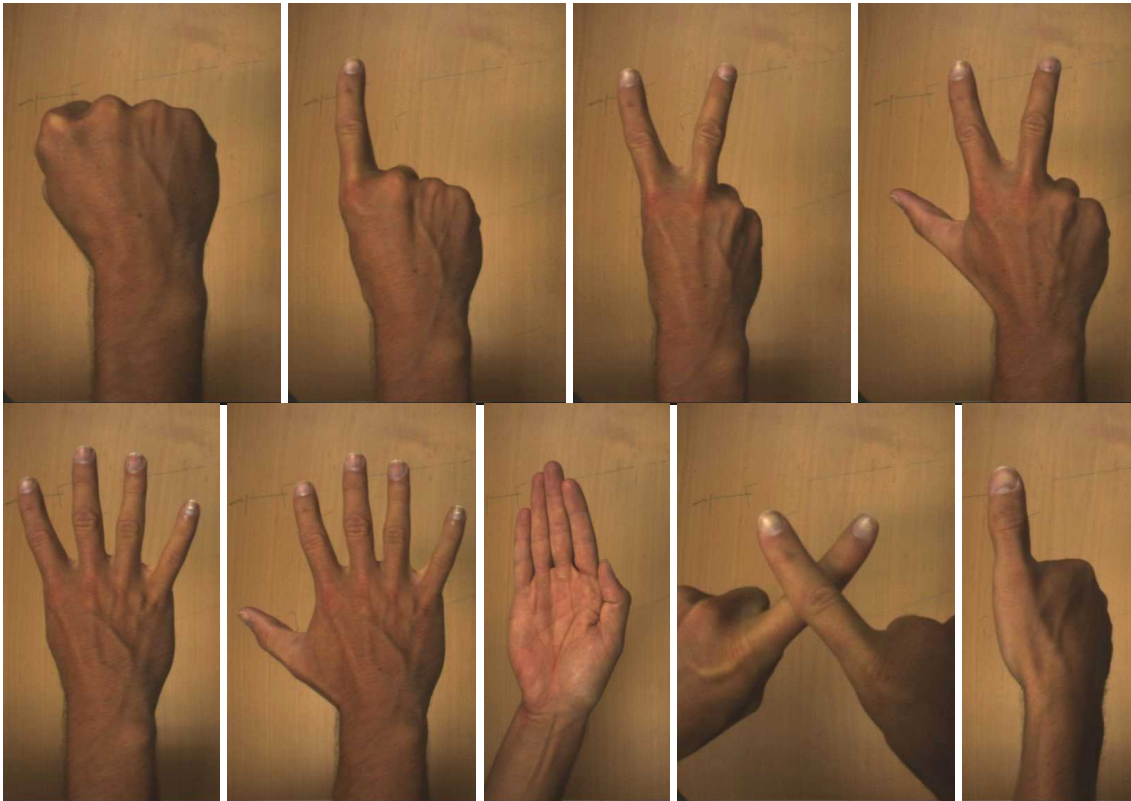


Figure 1 Static gestures: Residue, Point, Copy, Paste, Properties, Deselect, Menu, Delete, Yes/confirm.

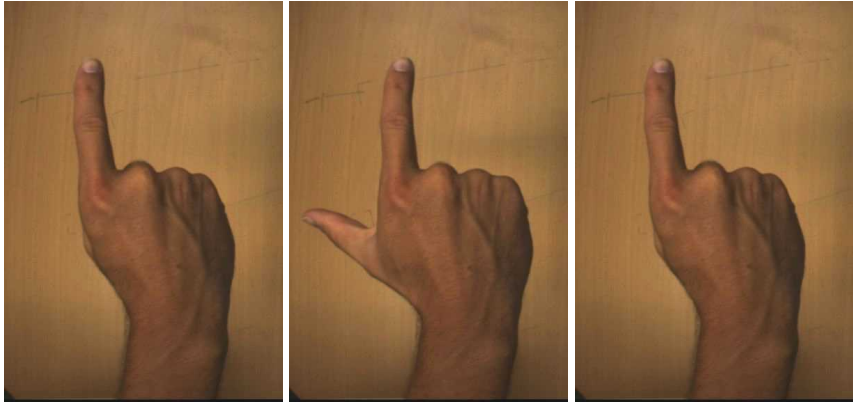


Figure 2 Dynamic gesture: Select

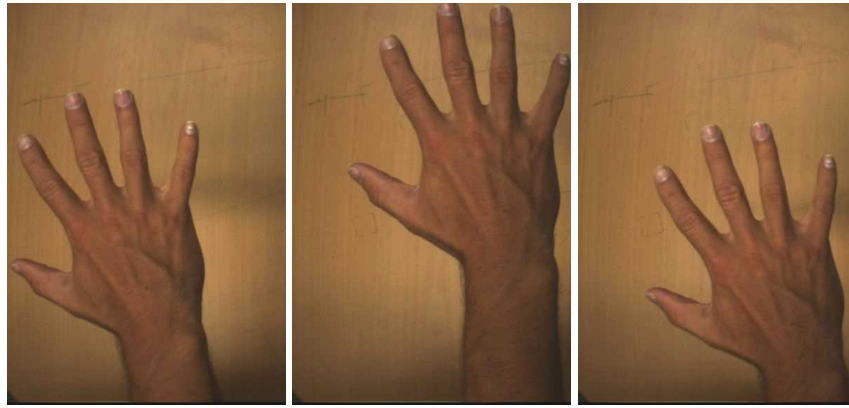


Figure 3 Dynamic gesture: Select all.

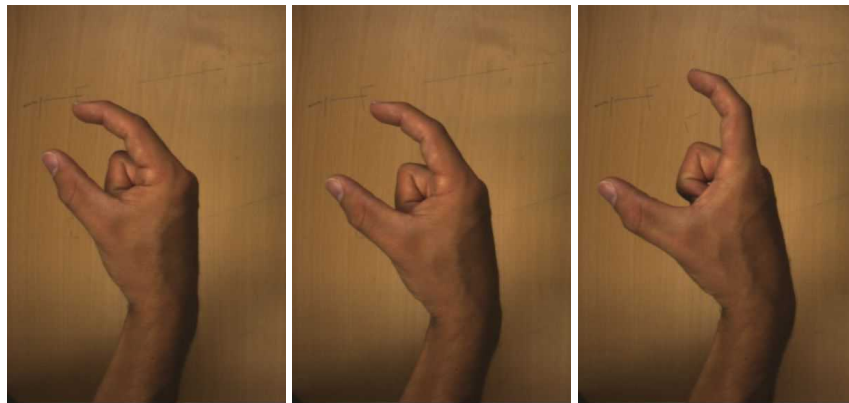


Figure 4 Dynamic gesture: Scale

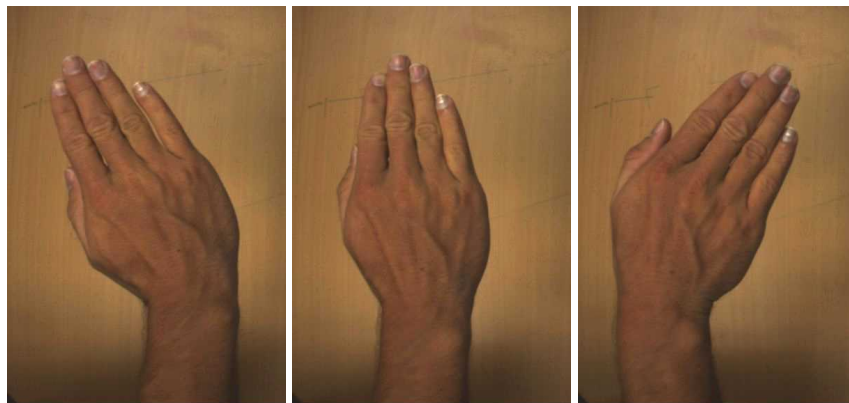


Figure 5 Dynamic gesture: No/undo.

Scenario Script

The scenario script is created to make the test persons imagining that they interact with the object placed on the table. Furthermore, some of the objects are moved during the scenario to introduce background changes in addition to hand movements. The full scenario script is listed in the following.

Scenario Script for Test Persons

1. Start with the fist in the middle of the camera view.

2. Point with the index finger and move the hand to the pen. Select (click) the pen and move to the paper. Move to the pencil case, close the hand, deselect and close the hand again.
3. Add the marker pen to the left of the compass, with the other hand. Remove the other hand from the field of view.
4. Point and move to the rubber, select (click) it, close the hand, copy and close the hand again. Point and move to the book, close the hand, paste and close the hand again.
5. Point and move to the paper, select (click) it, close the hand, delete and close the hand again. Confirm no and close the hand.
6. Shift the position of the pencil case and the pen, with both hands. Close the hands and remove the other hand from the field of view.
7. Point and move to the calculator, select (click) it, close the hand, choose properties and close the hand again. Confirm yes and close the hand.
8. Point and move to the compass, select (click) it, close the hand, scale up, down and up again and close the hand again.
9. Point and move to the middle of the view, close the hand, select all objects, close the hand, deselect them and close the hand again.
10. Point and move to the book, close the hand, display the menu and close the hand again. Confirm yes and close the hand.
11. Point and move to the middle of the field of view and make the fist gesture.

Additionally, the test person has to comply with the following restrictions:

- When moving the hand from one object to another, the hand should represent the point gesture, with only the index finger stretch.
- The residue gesture should be done before shifting to another gesture.
- Only slow movements are allowed during the recording process. Any fast movements will blur the sequences.
- The gestures should be performed perpendicular to the camera view angle and close to the surface of the table.
- When performing a gesture the hand/fingers should be in a strict position. For instance the fingers should be fully stretched.

Scene Setup

The captured scene includes a messy table environment with normal stuff for paper work. As these objects are used by the test persons to interact with, it is important that the scene setup do not get too complex, hence the test persons can get confused. The objects on the table are placed, so the distance to the edge of the field of view is large enough, to secure that the whole hand is captured during the recording process.

The light setup is arranged so that the table is split up in two parts with the same intensity (measured with luxmeter). One side of the table has a color temperature of 2600K and the other 4700K. The two light sources are of *type Photax 3200K head with Philips PF 308 E12 Argaphoto-B 240V / 500W* light bulbs. The scene setup is illustrated in Figure 6 and Figure 7.

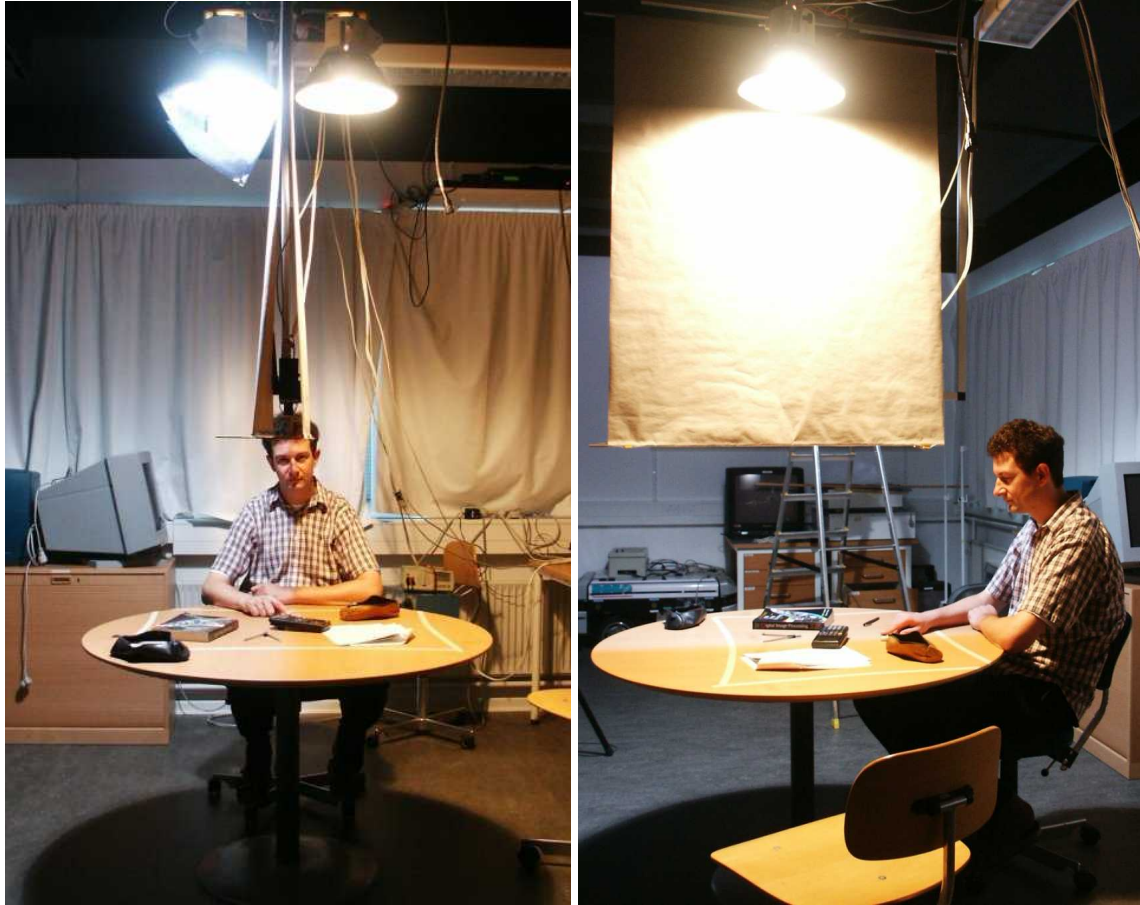


Figure 6 Scene setup including a test person observed from the front (left) and from the side (right).



Figure 7 Scene setup. Left: Hands illuminated by the split light configuration. Right: Table environment used during recording.

Camera Calibration

The utilized camera and frame grabber are of type *JAI CV-M90* and *Picasso PCI-3C* respectively. The aperture of the camera is set to 2.2, white balanced has been performed with a colour temperature of 3040K and the camera is calibrated to have an offset/black current close to zero. Further camera and frame grabber software settings concerning Iris control, RGB gain and offset are illustrated in Figure 8 and Figure 9.

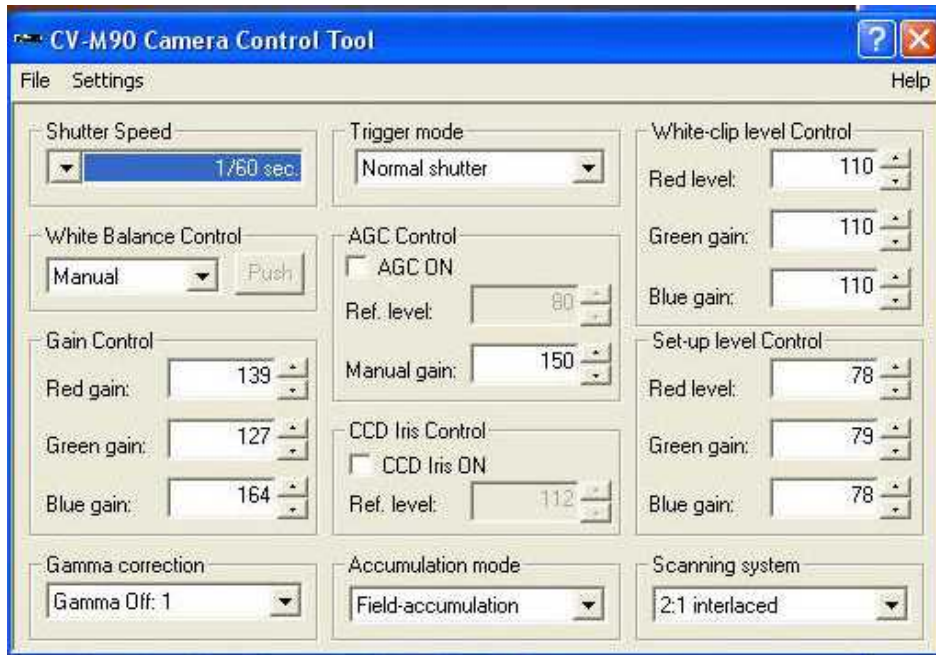


Figure 8 Camera settings.

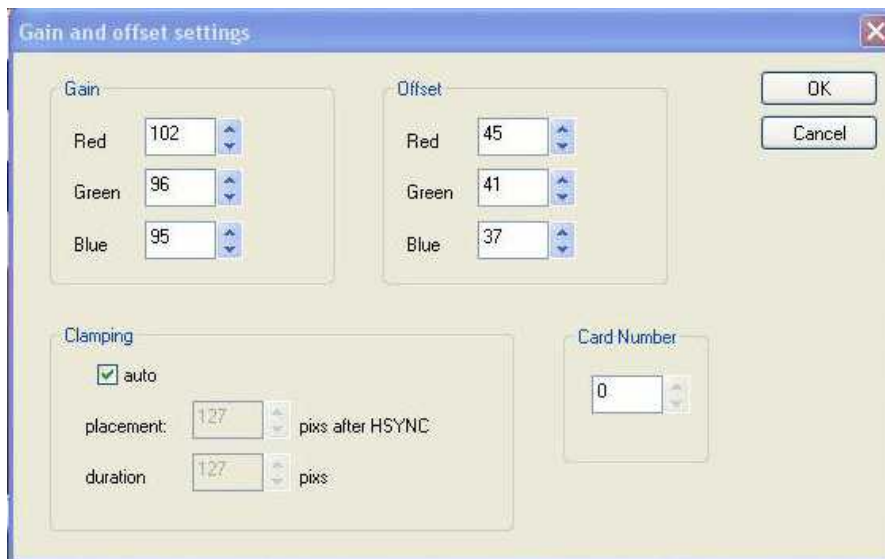


Figure 9 Frame grabber settings.

Black current??

To investigate the linearity of the camera response, a series of images is captured of a static scene while changing the exposure. Figure 10 shows the camera response function as the relation between the intensity and the exposure at a certain point in the image sequence. Figure 11 shows the result of plotting the intensity value for five of the neutral coloured squares of the Macbeth ColorChecker captured with the camera. As the figures illustrates, the camera response is nearly linear and no further gamma correction or nonlinearity compensation has to be done.

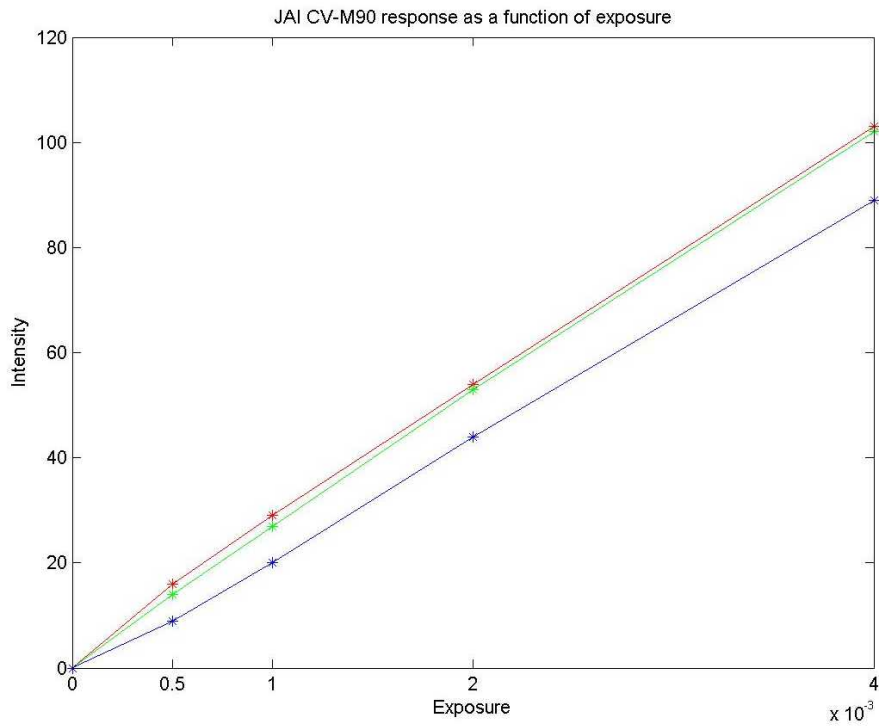


Figure 10 JAI CV-M90 camera response function based on images captured with different exposures.

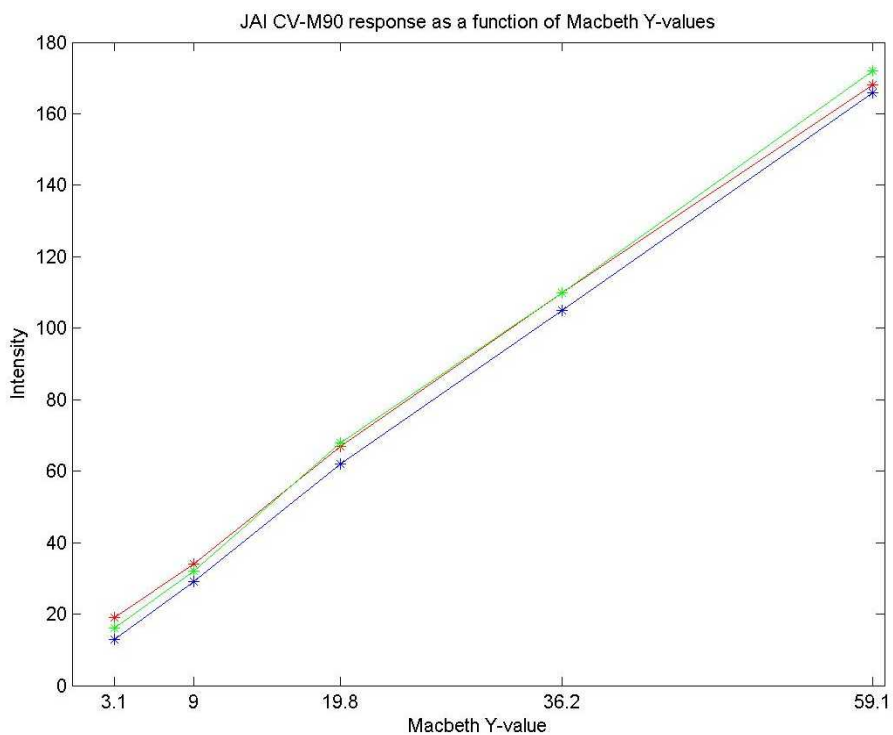


Figure 11 JAI CV-M90 camera response function based on Macbeth Y-values

A final check is carried out to investigate the RGB gains. This is done to avoid saturated intensity values or values close to zero in important parts of the sequences. Figure 12 shows some selections on a captured image, which are processes to calculate the minimum,

maximum and mean intensity values for each RGB color channel. Furthermore the r and g chromaticities are computed and plotted to check the influence of the split light configuration with two different color temperatures.

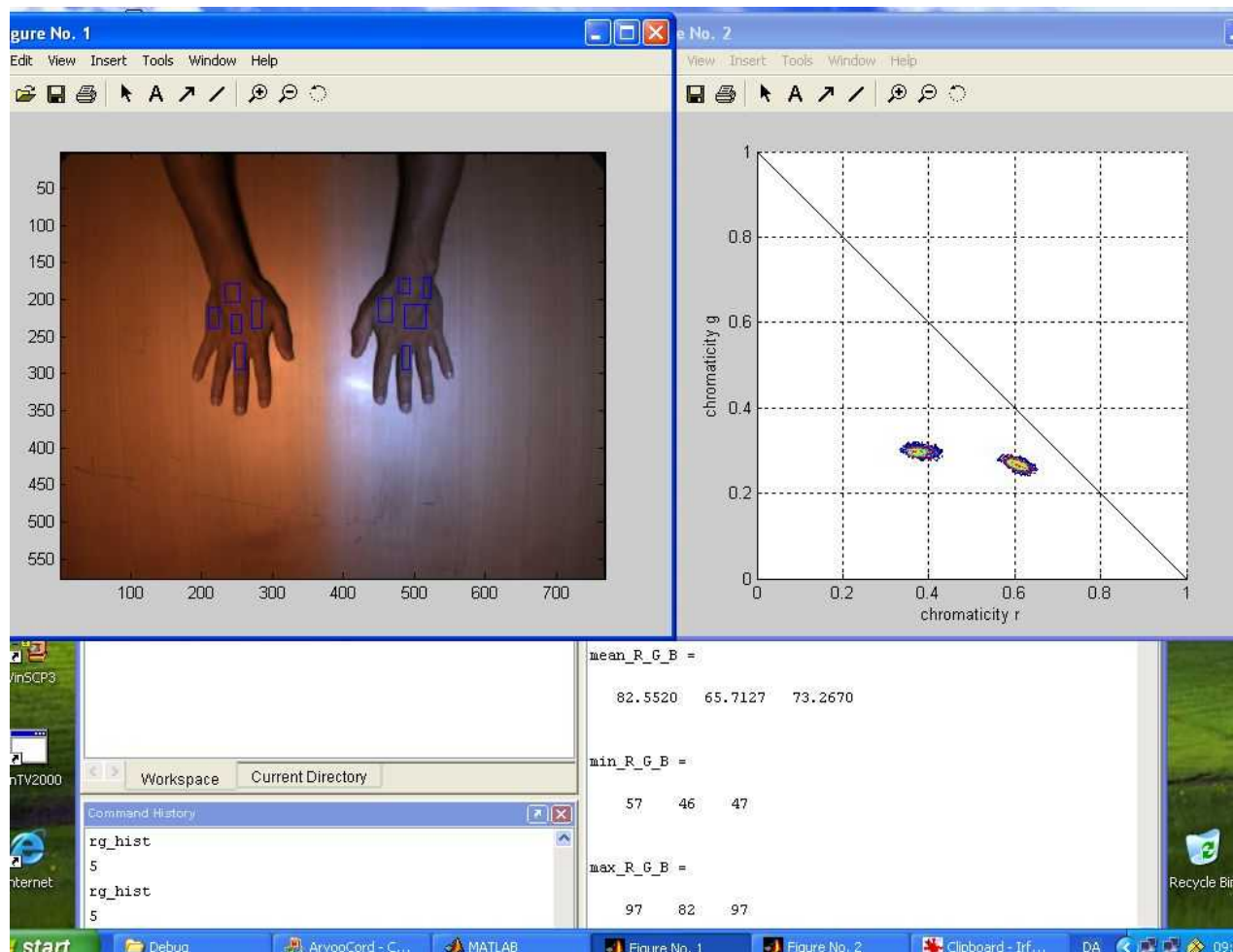


Figure 12 Investigation RGB gains to check intensity and chromaticity values.

Annotation

The recorded video sequences are annotated by the program, *FGAnno*, to provide the ground truth for all sequences. This program is specially developed to annotate video sequences with hand gestures. The user interface can be seen in Figure 13. The structure of the annotation files are also illustrated in Figure 14 and further specified in the following.

The annotation files includes the test person in the top of the file and the list of gestures afterwards. In the *EVENTLIST* all the events a test person has performed during a sequence is listed. Each event contains the elements:

- Event - The event number.
- Annotation Layer - The level of annotation. All video sequences has been annotated an Layer 3, which describe a hands posture and movement on sequence level.

- Person - The personID represented by a number. All person numbers are set to 00 because only one person is recorded in a video sequence.
- Gesture - The performed gesture represented by a number 00 - 13.
- Action_Start - The start of a gesture. A gesture starts when the hand of a test person begins a movement to impose a certain gesture.
- Stroke_Start - The start of a stroke. A stroke starts when the hand posture and movement is completely representing the gesture and in a strict position.
- Stroke_End - The end of a stroke. A stroke ends in the last frame where the hand posture is stict and the movement is completely representing the gesture.
- Gesture_End - The end of a gesture. A gesture ends in the last frame where the hand of a test person finish a movement to end a certain gesture.

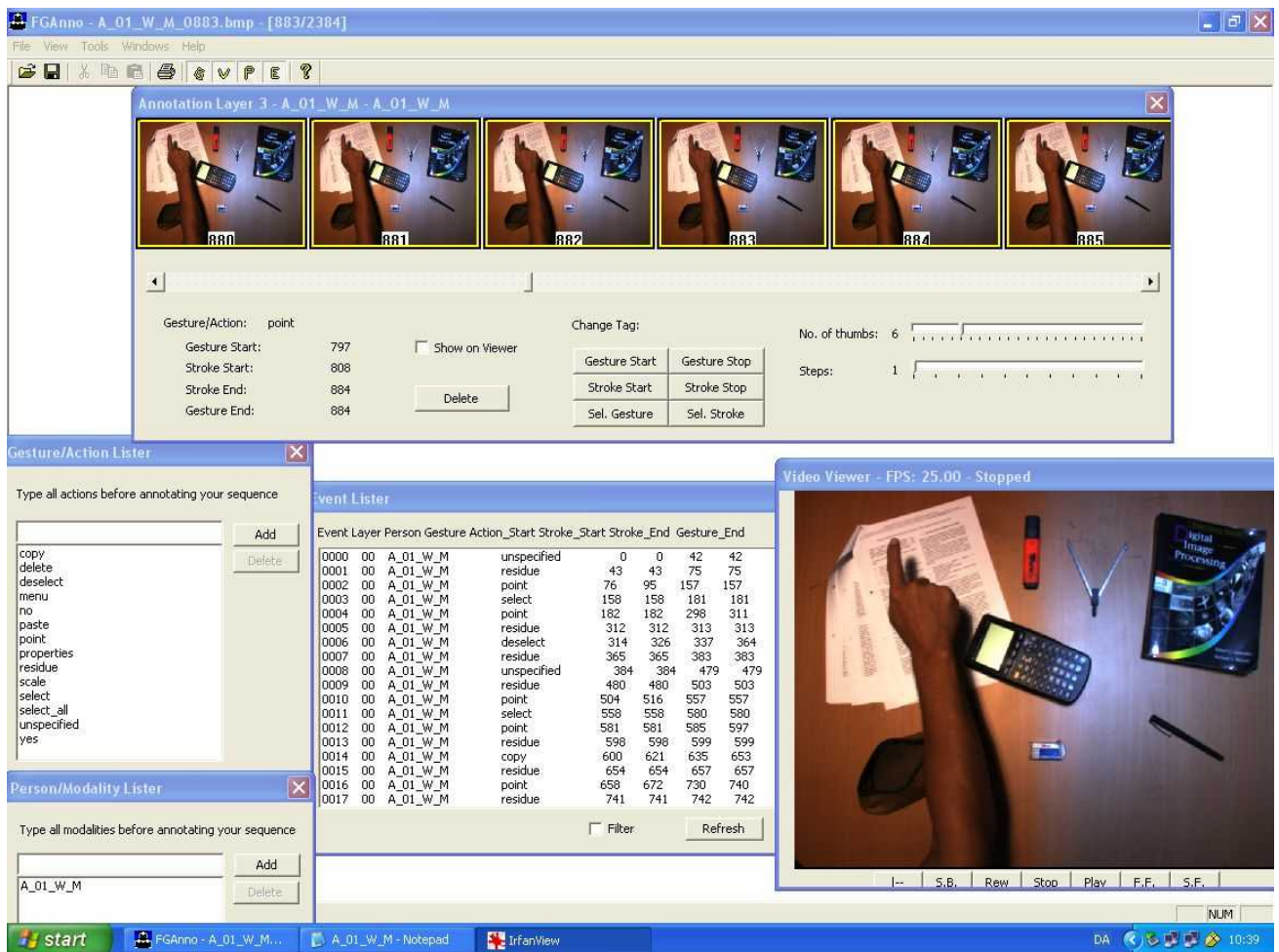


Figure 13 The utilized annotation program with Person Lister, Gesture Lister, Event Lister, Video Viewer and Annotation windows.

```

A_01_W_M - Notepad
File Edit Format View Help
#PATH ./
#TYPE bmp
#Person_Modality_List
00 A_01_W_M
#ENDLIST
#Action_Gesture_List
00 copy
01 delete
02 deselect
03 menu
04 no
05 paste
06 point
07 properties
08 residue
09 scale
10 select
11 select_all
12 unspecified
13 yes
#ENDLIST
Layer Name Type Modality Visible
#Annotation_Layers
00 A_01_W_M 3 0 1
#ENDLIST
Object format: object imagenumber layer no_of_contours
Contour format: contournumber inner no_of_points mp1 mp2
[x y]*no_points
#OBJECTLIST
#ENDLIST
#FINGERLIST
#ENDLIST
Event Layer Person Gesture Action_Start Stroke_Start Stroke_End Gesture_End
#EVENTLIST
0000 00 00 00 12 A_01_W_M_0000.bmp A_01_W_M_0000.bmp A_01_W_M_0042.bmp A_01_W_M_0042.bmp
0001 00 00 08 08 A_01_W_M_0043.bmp A_01_W_M_0043.bmp A_01_W_M_0075.bmp A_01_W_M_0075.bmp
0002 00 00 06 06 A_01_W_M_0076.bmp A_01_W_M_0095.bmp A_01_W_M_0157.bmp A_01_W_M_0157.bmp
0003 00 00 10 10 A_01_W_M_0158.bmp A_01_W_M_0158.bmp A_01_W_M_0181.bmp A_01_W_M_0181.bmp
0004 00 00 06 06 A_01_W_M_0182.bmp A_01_W_M_0182.bmp A_01_W_M_0298.bmp A_01_W_M_0311.bmp
0005 00 00 08 08 A_01_W_M_0312.bmp A_01_W_M_0312.bmp A_01_W_M_0313.bmp A_01_W_M_0313.bmp
0006 00 00 02 02 A_01_W_M_0314.bmp A_01_W_M_0326.bmp A_01_W_M_0337.bmp A_01_W_M_0364.bmp
0007 00 00 08 08 A_01_W_M_0365.bmp A_01_W_M_0365.bmp A_01_W_M_0383.bmp A_01_W_M_0383.bmp
0008 00 00 12 12 A_01_W_M_0384.bmp A_01_W_M_0384.bmp A_01_W_M_0479.bmp A_01_W_M_0479.bmp
0009 00 00 08 08 A_01_W_M_0480.bmp A_01_W_M_0480.bmp A_01_W_M_0503.bmp A_01_W_M_0503.bmp
0010 00 00 06 06 A_01_W_M_0504.bmp A_01_W_M_0516.bmp A_01_W_M_0557.bmp A_01_W_M_0557.bmp
0011 00 00 10 10 A_01_W_M_0558.bmp A_01_W_M_0558.bmp A_01_W_M_0580.bmp A_01_W_M_0580.bmp
0012 00 00 06 06 A_01_W_M_0581.bmp A_01_W_M_0581.bmp A_01_W_M_0585.bmp A_01_W_M_0597.bmp
0013 00 00 08 08 A_01_W_M_0598.bmp A_01_W_M_0598.bmp A_01_W_M_0599.bmp A_01_W_M_0599.bmp
0014 00 00 00 00 A_01_W_M_0600.bmp A_01_W_M_0621.bmp A_01_W_M_0635.bmp A_01_W_M_0635.bmp
0015 00 00 08 08 A_01_W_M_0654.bmp A_01_W_M_0654.bmp A_01_W_M_0657.bmp A_01_W_M_0657.bmp
0016 00 00 06 06 A_01_W_M_0658.bmp A_01_W_M_0672.bmp A_01_W_M_0730.bmp A_01_W_M_0740.bmp
0017 00 00 08 08 A_01_W_M_0741.bmp A_01_W_M_0741.bmp A_01_W_M_0742.bmp A_01_W_M_0742.bmp

```

Figure 14 Structure of annotation file with Person List, Gesture List , Annotation Layers and Event List.

Further Annotation Work

Annotation on layer 1 and 2 , which describe the hands posture and movement on pixel and image level. The annotation layer 1 includes information about the position in pixels of a contour of the hand and layer 2 concerns tracking of all visible fingers. The *FGAnno* user interface for annotation on layer 1 and 2 is shown in Figure 15.

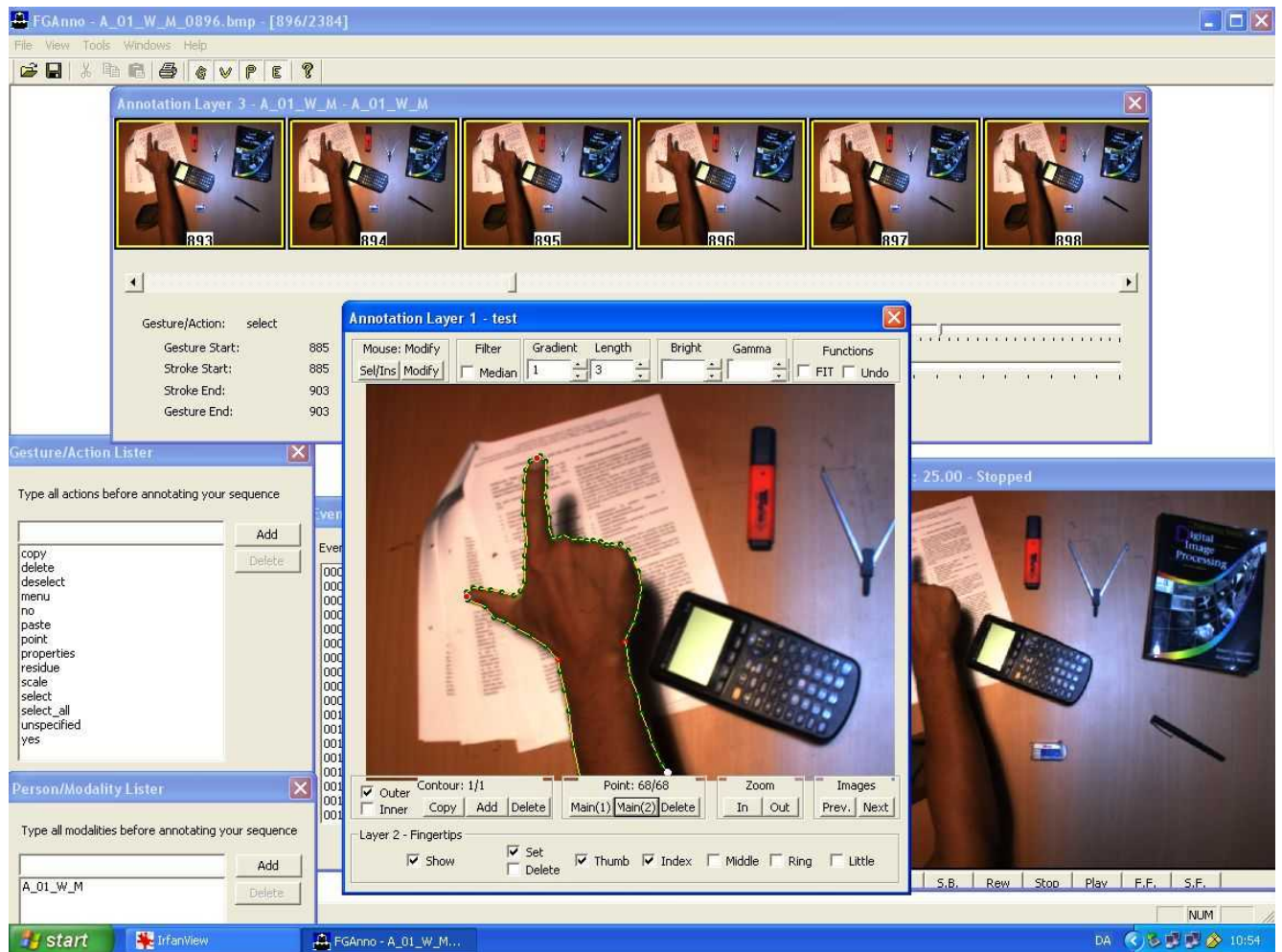


Figure 15 Annotation program setup for annotation on layer 1 and 2.

Gesture Video Sequence Database

The video sequences are recorded in PAL resolution, 768 x 576, and each image in the video sequences are named *sequenceType_personID_skinType_sex_frameNumber.fileName*. The sequenceType is set to A for all the available sequences, hence all sequences has been recorded as the same type of sequence. The personID is a number from 01 to 16, skinType is either (*White, Middle east, Yellow, Black*), the sex is either (*Male, Female*), and frameNumber is a four digit number (0000 >> XXXX). The Annotation files are nemed in the same way, just without the frameNumber.

A compressed video sequence can be found [here](#) and the full sequences will be downloadable soon.