Ibero-American Symposium on Computer Graphics - SIACG (2006)
P. Brunet, N. Correia, and G. Baranoski (Editors)

# Skin Detection for Hand Gesture Interfaces Using Assimilative Background

J. Santurde, D. Borro and L. Matey

CEIT and TECNUN (University of Navarra). Manuel Lardizabal 15, 20018 San Sebastian, Spain

**Abstract**

*This paper describes the study realized about hand movement capture systems oriented to gesture interfaces. Thanks to this study we have developed a skin colour based hand detection application. The state of the art already done provides a good starting point in the documentation stage of a hand detector creation for a gesture interface. This paper describes new strategies like assimilative background that combined with HSV, they result in a more robust skin detector.*

Categories and Subject Descriptors (according to ACM CCS): I.4.6 [Image Processing and Computer Vision]: Segmentation --- *Pixel classification,* I.4.8 [Image Processing and Computer Vision]: Scene Analysis --- *Color*, H.5.2 [Information Interfaces and Presentation]: User Interfaces --- *Natural language*

## 1. Introduction

In the last years, information processing capacity of computers has grown very fast but that is not the case of human-computer interaction systems. Now, we take advantage more efficiently of their computational power and we receive more information and faster but, essentially, the way to interact with computers has changed very few.

Common input devices like keyboard, mouse and joystick do not offer the natural and intuitive interaction that requires three-dimensional virtual environments, collaborative virtual environments (CVE) and ambient intelligence systems.

Gesture language is a powerful communication tool that complements the verbal communication and it can be used to interact with computers as people do in real life for talking each other. Gesture based interfaces are far more intuitive (essential in Human Computer Interaction), and they permit more immersive interaction with virtual environments. It would not be necessary to learn the use of intermediate devices because the interaction is pure human natural gestures.

To integrate these gesture interfaces in an ambient intelligence environment it is necessary to recognise the gesture in a non invasive way, i.e., the user must not wear any kind of marker or device.

In this article, firstly we present the state of the art of the hand capture methods, after that, in Section 3 three main colour spaces are compared to detect skin. In section 4 we describe the hand detection application based on colour segmentation that we have developed, finally in Section 5 conclusions and future work are drawn.

## 2. Previous work

Last years, many researches [AL04], [KTKN*98], [Que94] have deal with hand gesture interfaces. Different approaches can be broadly divided into methods based on devices that the user has to wear, methods based on computer vision and those that combine both methods in order to compensate the weaknesses of each one.

### 2.1. Hand tracking based on devices

### 2.1.1. Data gloves

Most device based approaches use data gloves. Data gloves are gloves equipped with sensors that register hand joints flexion, abduction of each finger, etc. General purpose commercial and non-commercial data gloves have been used in hand gesture based interfaces: LaViola [Lav99] did a practical guide about data gloves and Sturman [SZ94] writes a glove based input survey.

Although data gloves require lower computational cost this property does not compensate the discomfort imposed by devices that require to be connected by wires.

### 2.1.2. Other tracking devices

Also, it is possible to use magnetic or acoustic devices but they are too sensitive to interferences and they have limitations in distance, position, orientation and tracking precision (some of them only capture global motion of the hands). Mulder [Mul94] describes different tracking devices in his survey.

## 2.2 Computer vision.

### 2.2.1 Fundaments, advantages and disadvantages

In the beginning, glove based approaches became popular because of the difficulty of capturing hand movement, but in the last years computer vision has become more and more popular, mainly because data gloves restrict the expressiveness.

The freedom and naturalness offered by computer vision compensates the higher computational cost that it requires. Another advantage of computer vision is that cameras are becoming cheaper. LaViola's survey [Lav99] analyzes advantages and disadvantages of the two data collection systems: data gloves based and computer vision based.

Computer vision based hand detection objective is to recognize hand gestures in a sequence of images captured by one or more cameras using image processing algorithms, pattern matching, features, filters, etc. It is a non-invasive method so it allows more natural interaction without intermediate devices.

The greatest restrictions in computer vision based gesture recognition are cluttered backgrounds, changing lighting conditions and hand properties (colour, poses, finger occlusion, etc). To achieve robustness, people tend to use controlled environments with plain backgrounds and constant lighting sources.

Concerning to the hand colour segmentation, colour gloves have been used with unique colour [KTKN*98], [Sta95] or gloves with coloured rings to generate a special codification [Dom94]. John Underkoffler who is a researcher from Raytheon military company, is investigating an interface inspired in the movie "Minority Report". The user manipulates images projected on a panoramic screen with a pair of reflective gloves. A mounted camera keeps track of hand movements and a computer interprets gestures.

Anyway, the less invasive way is the hand detection without any glove or marker. In the following section, we describe different methods to achieve this.

### 2.2.2 Marker less vision based hand detection

Robust hand detection without markers could be achieved using colour segmentation, shapes and features or combining both methods.

Human hand has a lot of joints and its motion is very articulated which makes very difficult to detect, track and recognize, so the main tendency is the colour segmentation reinforced with pattern matching.

Shape analysis includes image feature extraction, statistics and models. Most techniques are posture recognition techniques like feature extraction, template matching, principal components analysis, active shape models, causal analysis, etc. instead of detection techniques, that is where this paper is focused on. LaViola [Lav99] discusses about each technique.

Detection is the first stage where the position of silhouettes or the features is calculated. Use of features in detection stage is not robust because fingers occlude each other and lighting conditions are changing. Also, computational cost increases with the number of features and the size of the image to search in.

Hand detection phase of a gesture based human computer interaction system needs a low computational cost technique in order to reserve free resources for more complex stages like gesture identification and classification. Colour segmentation, concretely skin colour segmentation, allows faster image processing because it is independent of geometric variations and it is robust against fingers occlusions and resolutions changes.

### 2.2.3 Skin detection methods

Although skin detection is a complex task because skin colour has strong dependence of lighting conditions and it may become similar to background colour, in the last years there have been a lot of studies in this area, mainly for facial recognition in biometry and also in face tracking for videoconferences.

The main two methods in image colour segmentation are those which classify each pixel as skin or non-skin independently from its neighbours and those which take into account the spatial arrangement of pixels called Region-Based methods [KBS02].

Region-based methods are robust in cluttered backgrounds but pixel based methods are faster, require less computational cost and it is possible to remove noise from background in an easy way.

Both methods require to choose a colour space where colour will be represented and a technique to model skin colour distribution (the way we will determine if a pixel or a region is skin colour or not).

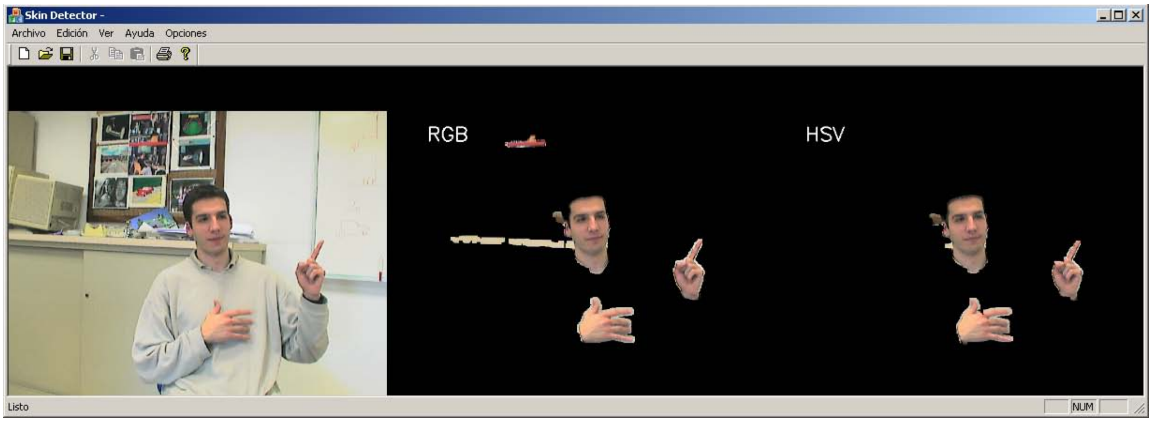


***Figure 1:*** Results using RGB and HSV colour spaces

#### 2.2.3.1 Colour spaces

There are a lot of colour spaces and most of them have been studied to get the best skin colour segmentation.

The general tendency is to consider that colour spaces that do an explicit discrimination between luminance and chrominance are better for skin colour segmentation.

For example, Zarit [ZSQ99] and [LY02] compared different combinations of colour spaces and skin colour modelling strategies. Zerit [ZSQ99] compared five colour spaces: HSV, HS, normalized RGB, YCrCb and CIELab using two skin modelling methods: LUT and Bayes classifier. Modelling skin colour with LUT they get better results in HSV and HS colour spaces and using Bayes classifier very few variations were detected between different colour spaces.

Albiol [ATD01] demonstrates that there is an optimum skin detection configuration for every colour space analyzed and this configuration has a structure independent from colour space.

Following, we describe the three most popular colour spaces for skin detection. In Vezhnevets survey [VSA03] a wider list of colour spaces for skin detection is analyzed.

**RGB:** It describes colours as a combination of three coloured rays (red, green and blue). Although is one of the most popular colour-spaces in image processing, mixing of luminance and chrominance makes it too sensitive to lighting changes so it is not adequate for skin detection.

Our main conclusions after applying RGB colour space in our application are the following:

- Detection of a smaller skin colour area in video sequence than using other colour spaces like HSV.
- Low robustness to lighting changes.
- Higher rate of false positives, especially with red objects (clothes, wooden doors, etc.)

**Normalized RGB:** Each component is normalized and after normalization one of the components can be removed in order to save memory. This method reduces the dependence to brightness. Vezhnevets [VSA03] used it for matt surfaces. Ignoring ambient light, normalized RGB is invariant to changes of surface orientation relatively to the light source.

**HSV:** The colour is defined as the combination of dominant colour (hue), colourfulness of an area in proportion to its brightness (saturation) and lightness. It differentiates luminance and chrominance properties so it seems to be one of the best colour space for skin detection.

This colour space is used in our skin detection application because we get best results (an example it is shown in Figure 1):

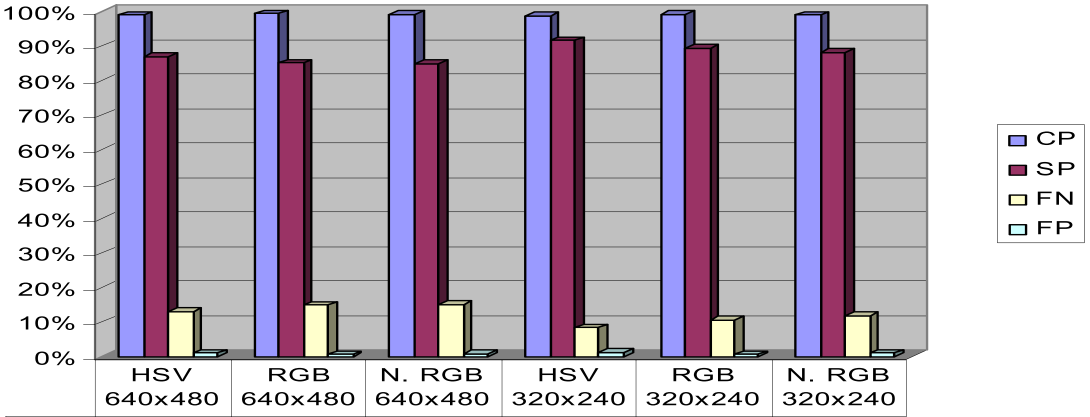
- Lowest rate of false positives and lowest rate of false negatives.
- Highest robustness in changing lighting conditions.
- Larger skin colour areas.

#### 2.2.3.2 Colour modelling

Skin modelling methods can be broadly divided in three groups: those who define explicitly a skin colour region, those based on parametric techniques and those based on non-parametric techniques.

a) Explicitly defined skin modelling

This method consists in defining skin region in a colour space using rules. These rules are based on practical studies where skin colour distributions are analyzed in a previous training process under different lighting conditions [SF96].



| | HSV 640x480 | RGB 640x480 | N. RGB 640x480 | HSV 320x240 | RGB 320x240 | N. RGB 320x240 |
|---|---|---|---|---|---|---|
| CP | 98,93546 | 99,31420 | 99,13240 | 98,72558 | 99,23860 | 98,98828 |
| SP | 86,77435 | 85,10394 | 84,81297 | 91,61910 | 89,30685 | 88,11478 |
| FN | 13,22564 | 14,89606 | 15,18702 | 8,380892 | 10,69314 | 11,88521 |
| FP | 1,064535 | 0,685791 | 0,867594 | 1,274414 | 0,649740 | 1,011719 |

***Figure 2:*** Results of pixel discrimination over 43 images captured in different indoors enviroments. CP (correct pixels): percent of all captured pixels that has been correctly classified (skin pixels and non skin pixels). FN (false negatives): percent of skin pixels classified as non skin. FP (false positives): percent of image pixels of non-skin classified as skin. SP (skin pixels): percent of skin pixels classified correctly.

The greatest advantage of this method is that thanks to its simplicity, is possible to build a very fast skin colour classifier even in high resolutions. The most important issue is the right election of the colour system and the definition of not too much restrictive rule group robust against lighting changes.

In Section 4 our skin classifier is described. It uses rules to define skin region but they are rules with parameters that allow more robustness when environments change.

b) Non-parametric skin modelling

This method consists in estimating a skin colour distribution from previous training without creating an explicit skin colour model. This distribution is usually represented non-analytically using colour probability histograms.

Different probability map creation techniques are: Normalized Lookup Tables (LUT), Bayes classifiers, artificial neural networks, etc. These methods and their advantages are described in [Lav99].

Using this method, the classification process is faster and the training stage is simpler but it requires a lot of space for skin colour probabilities storage, classification results are linked on the previous training and it is not enough robustness when environment changes.

c) Parametric Skin modelling

Analytically represented skin colour model is generated using representative information from training process. With this technique the skin is defined in a generic mode that is interpolable to different environments and conditions.

To model the probability distribution of skin colour, it can be used an unimodal Gaussian probability density function or multimodal Gaussian mixtures.

## 3. Colour spaces comparative for skin detection: HSV, RGB and Normalized RGB

A comparative evaluation of pixel classification performance of three colour spaces (RGB, Normalized RGB and HSV) have been done using bounds based detection method to define what constitutes skin colour.

### 3.1 Performance metrics

Following, used images and performance metrics are described:

The testing was done with 43 images. The images were taken from real time video sequences captured in different indoors environments with different lighting conditions. Two different resolution images have been employed 640x480 and 320x240 pixels.

The skin regions of the images were marked by hand using this mask as guide for false positives and false negatives detection.

Our algorithm is focused to skin detection in image sequences taking into account the results of the assimilative background so each capture was analysed with its background mask to discriminate the background areas of the picture. Noise filters had been not employed in this test, only pixel discrimination for every pixel of the image taking into account the background mask.

Four different metrics have been used to evaluate the pixel classification results. CP (correct pixels) is the percent of all captured pixels that has been correctly classified (skin pixels and non skin pixels). FN (false negatives) is the percent of skin pixels classified as non skin. FP (false positives) is the percent of image pixels of non-skin classified as skin. SP (skin pixels) is the percent of skin pixels classified correctly. The results of our experiments can be consulted in Figure 2.

### 3.2 The results

In RGB colour space test different video sequences were analyzed to create a skin colour model describable with rules in RGB coordinates.

In HSV colour space our skin classification algorithm was used to classify the pixels. This algorithm is based on parametrized rules (see Section 4 for hand detection application details).

In all colour spaces best results are obtained in 320x240 resolution because there are less isolated pixels similar to the skin colour (false positives and noise) and number of false negatives decreases in the contour of the hand. Nevertheless 640x480 resolution allows better results in the gesture identification process of a future hand gesture interface.

In both resolutions skin pixels classification is better with HSV but total number of correct classified pixel is lower than using RGB because the high quantity of false positives.

In 640x480 normalized RGB results are situated between HSV and RGB, most times closer to RGB. On the other hand, skin colour pixels results (SP) in 320x240 with normalized RGB are even worst than RGB.

Is important to take into account that this data only represents the number of pixels classified correctly independently from they position. The position of a skin coloured pixel in the image is very important in the hand detection process because an isolated false positive can be corrected with a noise filter, but a false negative in the centre of the hand area may alter the result of the detection. This lower rate of false negatives is the reason because, although the number of false positives with HSV is higher, it is a much better colour space for hand detection.

## 4. Skin colour based hand detection application

Following, we provide the description of the proposed skin colour based hand tracker. This tracker is oriented to integrate in a hand gesture interface or body pose and body action estimation applications.

Skin colour detection method is based in each pixel colour independently from its neighbours. Each pixel is classified as skin or non-skin using a colour region explicitly defined by parametrized rule set in HSV colour space. Peer et al. 2003 [PKS03] used a similar method defining a rule set to classify skin colour pixels in RGB.

Skin colour pixels are grouped forming blobs according to their proximity. These blobs will represent each hand allowing their tracking and access to their properties.

The method used has various advantages compared with existing approaches:

- It permits a fast and a computationally efficient detection because it does not use complex geometric models releasing more resources for more complex tasks that will be necessary in the classification stage.
- It does not require previous off-line colour learning processes.
- It detects skin colour objects robustly, even under changing lighting conditions thanks to its calibration system and the parametrized rule set that adapts the classifier to each environment.
- It uses an assimilative background detector so it is not necessary a uniform or static background.
- Other approaches use too low resolutions (320x240 or lower) to maintain detection speed in real time but our proposal can work even with 640x480 pixels resolution so, the gesture identification process is better.
- To identify the hands between other skin colour areas and get properties like position, area, etc., high level features are extracted from silhouettes: image moments.

A prototype implementation of the proposed skin detector operates with live video in real time at 20fps on a Pentium IV 2 GHz without special hardware. The camera used is a Sony 1394 DFW-V500 with 6mm lens.

Following, we describe the different phases that our hand tracker has.

### 4.1. Phases

The proposed method consists in four steps as it can be checked in Figure 3:

1- After calibration process (see calibration Section 4.2.1), all pixels of image that are not part of the assimilative background are classified as skin or non-skin using parametrized rules with values obtained from calibration process. At the end of this step, skin colour pixel position mask is obtained.

2- The mask is filtered to remove possible flickering due to artificial lighting, in which each pixel looks at its neighbours (making erode and dilate operations).

3- Using a connected components algorithm, pixel clusters grouped in silhouettes are identified in the mask. From these silhouettes is possible to extract high level parameters.

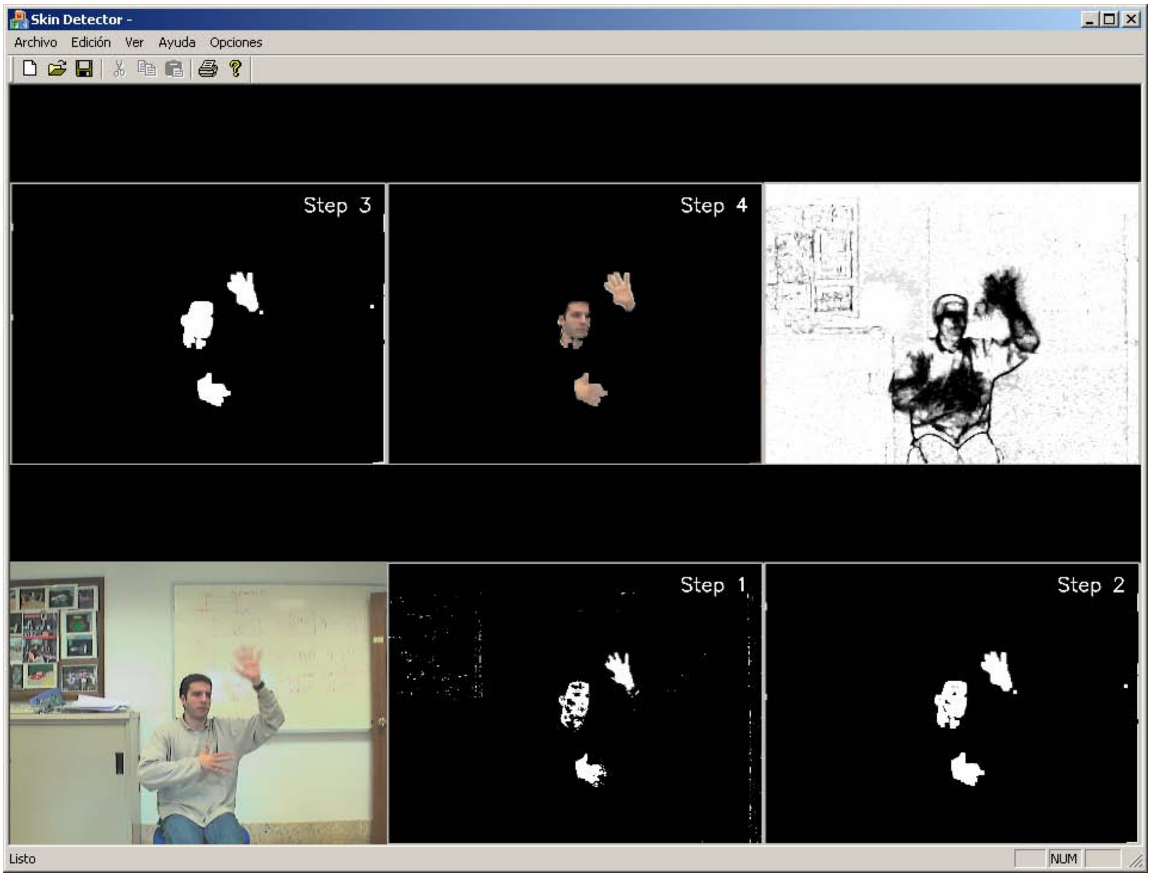


***Figure 3:*** Proposed method steps

4- Finally, the detected silhouettes are filtered according to their area to discard too small and too large shapes. After this step only hand (and face if it was necessary) silhouettes will appear in the image.

### 4.2. Detection of skin colour areas

Main problem of colour based tracking is the inconsistence of the colour when the lighting conditions change. Illumination is not constant so, a classifier trained for a specific skin colour can not work fine because changing the light also it changes the skin colour that camera captures.

To confront this issue, most approaches assume controlled lighting conditions. Our proposal combines initial friendly and robust calibration system (see Section 4.2.1) with HSV colour space to get better precision in the skin colour definition.

HSV differentiates luminance and chrominance properties reducing the effect of lighting changes in skin classifier.

Other important issue of skin colour segmentation are cluttered backgrounds. This challenge is solved in our application using an assimilative background. The assimilative background classifier does not take into account areas of the image that are static too long time.

### 4.2.1. Calibration

The hand tracker uses a friendly calibration system to adapt better to the environment lighting conditions and user skin tone (from very light skin tone to dark tone). Calibration stage is friendly, fast and it does not require extra effort by the user because hand poses and positions that are required are very natural and intuitive.

Calibration stage has 2 phases:

- Firstly the background is learned using the assimilative background detector. User has to avoid any movement in the image areas bounded by red rectangles.

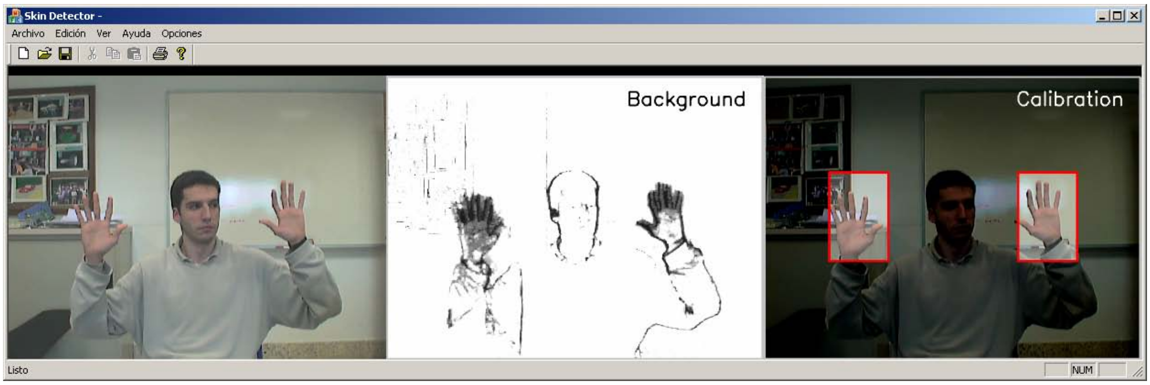


***Figure 4:*** Friendly calibration system

- When background is detected the user is asked to put the hand in two concrete areas delimited by red squares in different poses (one example is shown in Figure 4): palms to the front, closed, opened, etc. In each posture only hand pixels will be used to calculate skin colour reference values that will be used as parameter for the discrimination rules of the core of the classification algorithm adapting it to the environment lighting condition and user skin tone.

### 4.2.2. Assimilative background

The application has its own interpretation of the image background every time. The application background will be composed of the position of all pixels that has spent a specific time period without colour changes. Skin detection algorithms do not take into account the pixels that are considered as background in the classification process.

Thanks to this idea, it is not necessary to define a known initial background or constant background because changes in background (for example, somebody left the door opened when he left the room) will be learned progressively in real time. The approach registers number of frames that each pixel continues without significant colour changes and when it takes too long, these pixels are added to the background.

## 5. Conclusions and future work

### 5.1. Conclusions

We have realized a study of different hand movement capture systems that can be useful for those that are looking for an approach or device to create a hand gesture based human computer interface.

Different skin modelling methods and colour spaces have been reported and a comparative of three colour spaces (HSV, RGB and normalized RGB) for skin colour detection has been done.

A skin colour based hand detection application has been developed. It detects the hands using a pixel based skin colour classifier with parametrized rules. Results with this tracker are robust, fast and reliable hand detection against different skin tone or lighting conditions changes.

### 5.2 Future work

This hand tracker will be integrated in hand gesture interface applications, especially our interest is focus on the 3D visualization area in which the user could interact with virtual objects using his hands.

Other interesting applications could be human body pose estimation systems and person tracking systems for behaviour studies, character animation, TV and virtual decorates, etc.

## References

[AL04] ARGYROS, A. A. and LOURAKIS, M. I. A.: *Tracking skin-colored Objects in Real-time*. Cutting Edge Robotics Book, 2004.

[ATD01] ALBIOL, A., TORRES, L. and DELP, E.: Optimum Color Spaces for Skin Detection. In *proceedings of the International Conference on Image Processing* (2001), pp 122-124.

[Dom94] DOMER, B.: *Chasing the Colour Glove: Visual Hand Tracking*. Master´s thesis, Simon fraser University, 1994.

[KBS02] KRUPPA, H., BAUER, M. A., and SCHIELE, B.: Skin patch detection in real-world images. In *proceedings of the DAGM 2002*, (2002), pp. 109-117.

[KTKN*98]KUMO, YOSHINORI, TOMOYUKI, I., KANG-HYUN, J., NOBUTAKA, S., and YOSHIAKI, S.: Vision-Based Human Interface System: Selectively Recognizing Intentional Hand Gestures. In *proceedings of the IASTED International Conference on Computer Graphics and Imaging*, (1998), pp. 219-223.

[Lav99] LAVIOLA, J. J., *A Survey of Hand Posture and Gesture Recognition Techniques and Technology*.

Technical Report, Brown University: Providence, Rhode Island, 1999.

[LY02] LEE, JAE Y. and YOO, SUK I.: An elliptical boundary model for skin color detection. In *proceedings of the International Conference on Imaging Science, Systems, and Technology*, (2002).

[Mul94] MULDER, A., *Human movement tracking technology*. Technical Report 94-1, School of Kinesiology, Simon Fraser University, 1994.

[PKS03] PEER, P., KOVAC, J., and SOLINA, F.: Human Skin Colour Clustering for Face Detection. In *proceedings of the Eurocon 2003*, (2003).

[Que94] QUEK, F.: Toward a vision-based hand gesture interface. In *proceedings of the Virtual Reality System Technology Conference*, (1994), pp. 17-29.

[RN95] REKIMOTO, J. and NAGAO, K.: The World through the Computer: Computer Augmented Interaction with Real World Environments. In *proceedings of the ACM Symposium on User Interface Software and Technology*, (1995), pp. 29-36.

[Sta95] STARNER, T. *Visual recognition of american sign language using Hidden Markov Models*. Massachusetts Institute of Technology. Master's Thesis, MIT, 1995.

[SF96] SAXE, D. and FOULDS, R.: Toward robust skin identification in video images. In *proceedings of the FG'96*, (1996), pp379-384.

[SZ94] STURMAN, D. J. and ZELTZER, D.: A survey of glove-based input*.* *IEEE Computer Graphics and Applications*, 14, (1994), 30-39.

[VSA03] VEZHNEVETS, V., SAZONOV, V., and ANDREEVA, A.: A survey on pixel-based skin color detection techniques. In *proceedings of the Graphicon*, (2003).

[WH01] WU, Y. and HUANG, T. S.: Hand modeling, analysis and recognition*.* *Signal Processing Magazine, IEEE*, 18, 3 (2001), 51-60.

[ZSQ99] ZARIT, B. D., SUPER, B. J., and QUEK, F. K. H.: Comparison of Five color models in skin pixel Classification. In *proceedings of the ICCV'99 Int'l Workshop*, (1999), pp. 58-63.