Detecting and tracking the hand to create an augmented reality system

Khawla Benabderrahim[1] & Mohamed Slim Bouhlel[2]

[1](Sciences and Technologies of image and Telecommunications (SETIT) Research Unit, ISBS Sfax, (ISBS),
Sfax University, Tunisia)
[2]((SETIT Research Unit, ISBS Sfax, (ISBS), Sfax University, Tunisia)

Abstract: *To facilitate the access to computer systems, computer scientists choose to design interactive mechanisms which are more adapted than traditionally used devices such as keyboard or mouse.*
The use of gestures modalities has been extensively studied in recent years by the human machine interaction community. Augmented Reality is one of the mechanisms of interaction that allow implementing the human-machine communication. Indeed, its principle is to mix the real world and the virtual world.
Many applications already use Augmented Reality, particularly in the field of video games, film, business and also in the medical field which really interests us: our goal is to analyze medical images using augmented reality, it comes to insert 3D virtual objects, allowing to analysis the medical image and helping the doctor to make the right decision, on the doctor's hand.
In this paper, we describe our approach of detecting and tracking of hand in a video stream, we hybridize two methods to draw more benefits from each one. The first is to detect the hand by subtraction of the background and the second is to detect the hand by skin color.
Keywords: *Hand detection, subtraction of the background, skin color, acquisition, tracking of hand*

I. INTRODUCTION

Augmented reality is a computer assisted computer tool that allows you to insert virtual objects in the real world. It is also based on the superposition of a real image, which is returned by a camera, and a virtual image, modeled by a computer.
To increase the actual user's environment with virtual objects, two methods are proposed:
• Methods with marker: which are the most used for augmented reality applications indoors (prepared environment). This is to put, in the actual scene, markers that are used to calculate the 3D coordinates from three specific points recognized by the system.
• Methods without marker: based on image processing algorithms more advanced, making it possible to monitor a component of the real scene as the hand or face. For example, for monitoring the hand, the augmented reality system without marker, detects the position and orientation of the hand of the user, and superimposes a 3D model, positioned and oriented in the same way.
Our goal is to create an augmented reality system without markers that inserts virtual objects on the hand already tracked. The Figure 1 following describes different phases of our application.
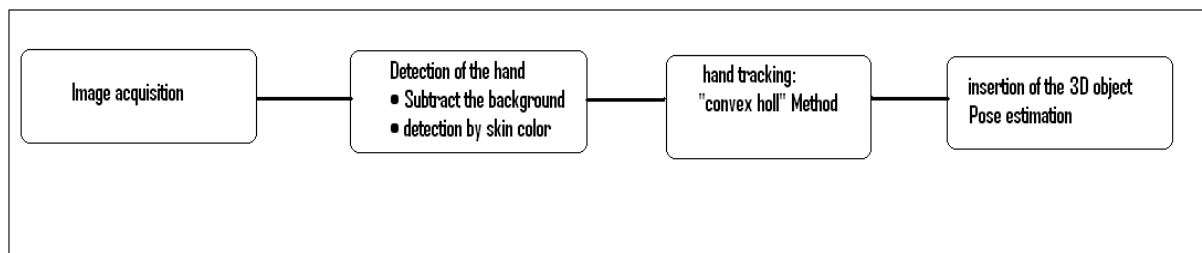

Figure 1: phases of our augmented reality system without markers

So, the process of augmented reality system begins by acquiring images from a camera.

II. IMAGE ACQUIRING

As part of our work, the implementation will be performed using webcams such as those built into our laptops.
The image produced by the webcam is in RGB format and the resolution will be used is 640x480 pixels.

The YCbCr color space, with it we work in our algorithms, is a way of representing the color space in video, it is divided into a luminance component (Y) and two chrominance components (Cb and Cr) . The chrominance information is the color. It is divided into three components are the primary colors red, green and blue. The luminance is the intensity of light. It can be evaluated using the following formula, which involves colors Red (R), Green (G) and Blue (B):

$$0.299 * R + 0.587 * G + 0.114 * B \quad 1$$

After acquisition of the image, the next step of our application is the detection of the hand.

III. DETECTION OF THE HAND

The purpose of this phase is to detect on an image, the area of interest. That is to say, in this case, distinguish the hand from the rest of image.

Several detection methods exist, all with advantages and disadvantages. None of them are perfect. Our choice fell on not one method, but two. We will perform the hybridization of two detection methods: the first involves a subtraction of the background and the second to detect the color of the skin.

The following Figure 2 shows the advantages and disadvantages of each method of detection chosen.
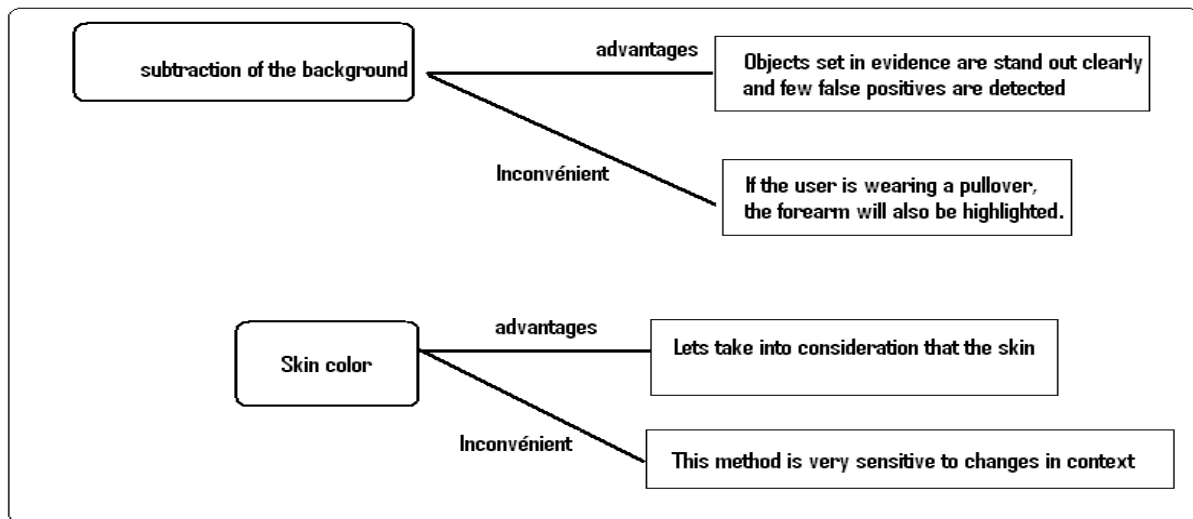

Figure 2: advantages and disadvantages of tow detection approaches

Starting by specify the first method: Subtraction of background.

1. Subtraction of background method:

This method is to learn the model of background to the computer. Subsequently, this model is compared to the current image to subtract the background previously known.

  Define now, what is a background. This is a generic concept that varies according to the application and the context. Indeed, the background may be light movement, such movement of a forest in the wind or the movement of the camera due to the respiration of the user. The change in light can also enter into consideration.

This technique operates satisfactorily with simple scenes. Unlike, there is the disadvantage that all pixels are independent and are not influenced by their neighbors. That is makes this method very sensitive and included noise in an image.

To solve this problem, it is possible to learn, to the computer, more complex model in which pixels around are taken on consideration. Of course, this implies an extension of memory and computation time allocated to the algorithm. For this reason complicated models are avoided. In our case, the fluidity of the stream video is very important to make it's interactions as smooth as possible.

  There is another way to get a model more slightly complex, whilst kipping an optimal fluidity. This is the "codebook". The principle of this technique is to create a dictionary representing significant statements of the background.To do this; a dictionary is created for each pixel constituting the image. Then, this dictionary is updated on the basis of several images.

2. Application of the background subtraction

 According to the background model, learned in the previous step, it is possible to extract from the image any object in the scope of the camera. The result of this processing is a binary mask: All white pixels are corresponds to the foreground, and every black pixel to the background (Figure 2).

To implement our approach of detection of the hand, we used the method of «background subtraction ", which is defined in the OpenCV library.

 OpenCV offers different styles of background subtraction, we choose codebook model. This model calibrates for some time to be exact for some frames, and can acquire all images, and calculates the mean and deviation of each pixel.

Therefore, this method gives us the ability to remove the background from the actual scene and we have detected the hand in the foreground. It comes to a black image with only the hand in white as is shown in Figure 3.
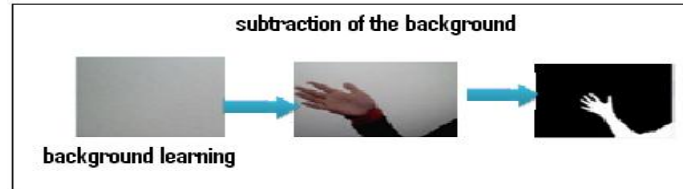

Figure 3: hand detection by background subtraction

IV. DETECTING THE REGION OF THE SKIN

This method consists to learn, to program, the skin color of the user. Once the program has learned that color, it is possible to highlight all the pixels corresponding to the color learned.

To detect the color of the skin, it must first choose the color space with it we should work.

1. Choice of the color space:

The choice of color space is important when processing the image. Indeed, each has its own characteristics.

To determine the convenient color space, we conducted a comparative study containing the results of tests with different color spaces (Figure 4).

We took an image and we analyzed each time the proportion of good results obtained with the following color spaces: RGB, HLS, Luv and YCbCr.

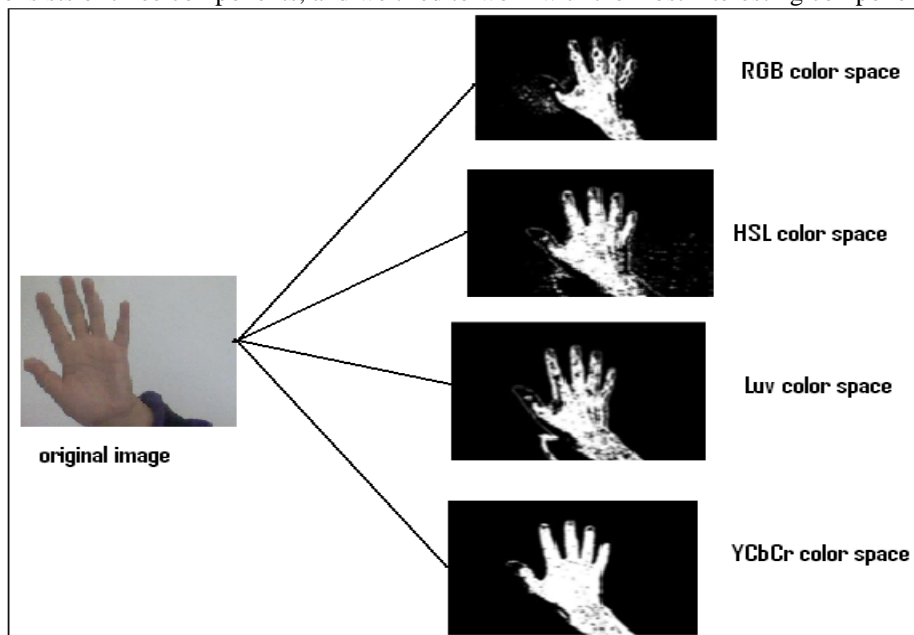Each area consists of three components, and we tried to work with the most interesting components.

Figure 4: Comparison of the results of recognition of the color of the skin in different color spaces

After analyzing the found results, we decided to work with the YCbCr color space because it is the space that provides us the good results that can operate in following our application.

2. Learning of the color of the skin:

Of course, since the image processing is done pixel by pixel, the color of these varies considerably, even if they are located in a contiguous area (Figure 5). To improve efficiency, it is necessary to consider a sufficiently large number of pixels. This allows us to establish a range of reference colors. The interval must be calculated and recorded for each channel used in the color space.

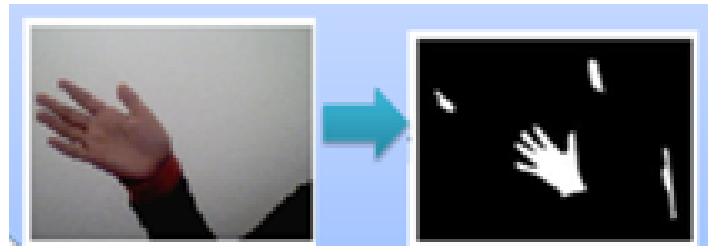Our detection results of the hand skin color are given by the following figure:

Figure 5: detecting the hand by skin color

The method of detecting the hand by skin color gives the results, so that the method of detection by subtracting the background. In our work, we want to find the best detection result and we seek to minimize noise in this found result using tow methods. That why, we propose to combine both methods together.
The results of the hybridization of two detection methods are presented by the figure below:
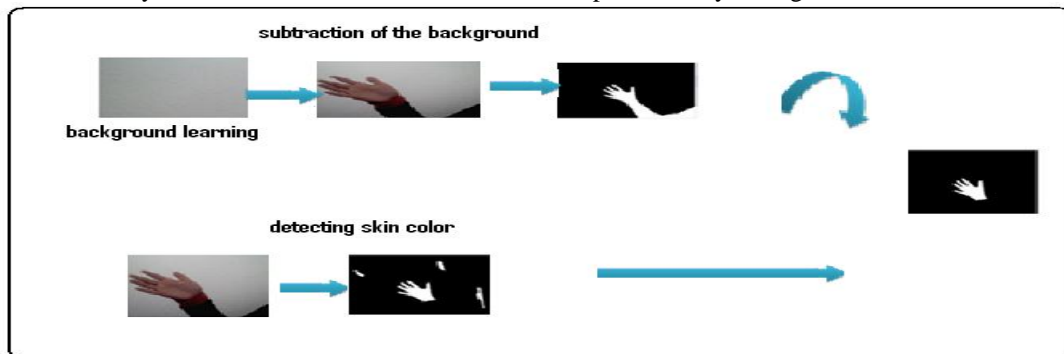
Figure 6: hand detection by hybridization of two methods: background subtraction and skin color

V. TRACKING THE HAND BY CONVEX HULL METHOD

In this section, we present the method, using successive images produced by a single camera, to find the position and orientation of an object in the scene relative to the camera. We are particularly interested in a interactive method that can be implemented in real time, that is to say usable for Augmented Reality applications .

The problem of estimating the pose of a 3D object from a flow of images is a difficult problem, especially in the context of interactive applications where the camera is manipulated by the user (held in the hand or attached to his head). Under these conditions objects, in the scene, undergoing any movements, sometimes very fast and it is difficult to predict. This type of interaction system is based essentially on the knowledge of the pose at each instant; the estimation mechanism must be the most reliable and able to detect estimation failures to reset. These difficulties are, in addition, to very strong constraints execution times especially when the target platform is an embedded system. This explains the rarity of commercial solutions of Augmented Reality based on computer vision.

The method we have chosen to model the hand and extract its data skillfully uses the properties of convex hulls. We use the method «convex hull «defines in the library OpenCV to find the fingertips. This method is used mainly to define the area of the hand (figure 7) .

Figure 7 Detection of the hand by "convex hull" method
The red line, delimiting hand, presents the "convex hull" of the hand, it is a convex set. This means that is connecting point of the two successive fingers by a red line and so on to form the "convex hull" of the hand.

The Figure 8, following, shows the detection and tracking of hand in our application
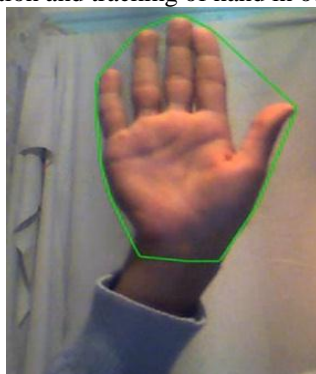

Figure 8: tracking of hand in our application

Detection begins when the user moves his hand and fingers in a position corresponding to an interaction. Then, the user must stay in this position for some time of validation. At the end of the detection time, the processing system by the interaction occurs. Tolerance time post- interaction is not always used according to the interaction, but is useful for the system to avoid unnecessary treatment because the user will anyway position too long for detection. The Feedback, displayed early detection of interaction, can help the user to know what he is doing to understand the system.

VI. Conclusion

This phase of analysis allows us to have a good view of the technologies used in the field of detection and tracking of the hand. We know, now, that we will use OpenCV. We also have a good overview of the overall operation and analysis of the algorithm for detecting hand gives us a guideline for the rest of our work.
The present Interactions and that imagined, also, give us a good idea of the work required for the creation of new interactions. This specification provides us a good outline which we will implement in our application to creating a medical image analysis system by tracking the doctor's hand and insert virtual objects which facilitates the interpretation of the medical image and helps the doctor to take the best decision.

REFERENCES

[1] Alexander Toshev, Ameesh Makadia, Kostas Daniilidis *"Shape-based Object Recognition in Videos Using 3D Synthetic Object Models"*, IEEE, pp.288-295, 2009.
[2] Pengwei LIU, Huiyuan WANG et al. *"Motion Compensation Based Detecting and Tracking Targets in Dynamic Scene", IEEE, pp.703-706, 2010.*
[3] Ritika1, Gianetan Singh Sekhon *"Path Estimation and Motion Detection of Moving Object in Videos", IOSR Journal of Computer Engineering (IOSRJCE) ISSN: 2278-0661 Volume 2, Issue 4,* PP 01-04, July-Aug. 2012.
[4] Mohammed Sayed and Wael Badawy, "*A novel motion estimation method for mesh- based video motion tracking"*, *IEEE, 2004, 337-340.*
[5] Li-Qun Xu, *"Simultaneous Tracking And Segmentation Of Two Free Moving Hands In A Video Conferencing Scenario", IEEE,* 2003, 49-52.
[6] Beiji Zou, Xiaoning Peng and Liqin Han, *"Particle Filter With Multiple Motion Models For Object Tracking In Diving Video Sequences", IEEE, 2008, 224-228.*
[7] Ihsan Rabbi1, 2, Sehat Ullah1, Siffat Ullah Khan*, "Augmented Reality Tracking Techniques: A Systematic Literature Review Protocol », IOSR Journal of Computer Engineering (IOSRJCE) ISSN: 2278-0661 Volume 2, Issue 2, July-Aug. 2012, PP 23-29*
[8] Milgram P, Drascic D, Julius J, et al. *« Merging Real and Virtual Worlds. Proceedings of IMAGINA",* 1995, PP 218-230.
[9] Milgram P, Takemura H, Utsumi A, Kishino F. *"Augmented Reality: A class of displays on the reality-virtuality continuum. In, SPIE Proceedings: Telemanipulator and Telepresence Technologies" ,* (Boston, MA); 1994, PP 282-292
[10] Fuhrmann AL, Splechtna R, Přikryl J, *"Comprehensive Calibration and Registration Procedures for Augmented Reality", In Proceedings Eurographics Workshop on Virtual Environments;* 2001.
[11] Wang X, Dunston PS, *"Design, Strategies, and Issues towards an Augmented Reality-Based Construction Training Platform. ITcon ", Vol 12, pg 363-380, http://wwwitconorg/2007/25* 2007:363-380
[12] Krevelen DWFV, Poelman R, *" A Survey of Augmented Reality Technologies, Applications and Limitations", The International Journal of Virtual Reality* 2010;9:1-20.