DOI: https://doi.org/10.3217/978-3-85125-609-3-31

Real Time Hand Tracking Method for an Immersive Application

Ouissem Ben Henia(✉)

King Khalid University Abha, Kingdom of Saudi Arabia
o.benhenia@gmail.com

Abstract. This paper presents a real time fingertip tracking method used for immersive application. The first step of the method consists of segmenting the input image captured by a video camera and detects the area representing the human hand. The latter one is identified by means of skin color detection method. The output of the first step is a black and white image where the hand area is coloured in black while the background is in white. In the second step, the hand shape is analysed and the fingertips are detected using convex-hull method. The barycentre of the fingertips is then calculated and to estimate the position of the hand. Using the estimated position in each frame, many applications belonging to human computer interaction and immersive domains could be proposed. In this paper, we propose an immersive application which consists of moving virtual objects according to the estimated hand position.

Key words: Augmented Reality · Image Processing

1 Introduction

Immersion into a virtual world could be defined as the feeling of being present physically in a non real world. To reach that perception, different parameters could be taken into account such as images, sounds, sensors and stimuli. Hardware devices and softwares are involved into the process of creating the virtual world. For the hardware part, different devices could be used such as video-cameras(2D or 3D), head mounted display, data gloves,etc. The software part corresponds to the program allowing the interaction between the human being and the virtual world and so produces the perception of immersion. An important domain allowing the immersion in a virtual world is called human-computer interaction (HCI). This domain is becoming more and more present in our daily life. Using video-cameras, researchers in computer vision brought on a wave of applications for human-computer interaction and specially for video games and 3D animation. Many of the applications are based on human body tracking. In this paper we present a hand tracking method to create an immersive application. In the next part, we present a related work on hand motion tracking. The section number 3 explains how the segmentation is achieved to detect the hand observed in the images. The section number 4 presents the method to detect

DOI: https://doi.org/10.3217/978-3-85125-609-3-31

the fingertips. Finally we present our application and the experimental results obtained using a 2D video camera.

2 Related work

Several methods on hand motion tracking using video-cameras were proposed during the last two decades [1][2][3][4]. Both 2D videos cameras and 3D cameras were used[5][6]. The advantage of using 2D cameras is the affordability comparing with the 3D ones but less data can be exploited using the first type of cameras. In fact, all vision based methods for hand tracking start by extracting features from the images. Only 2D features are used in the case of simple video-cameras while a third dimension is used for the 3D videos cameras.

The methods could then be divided in different categories. The first one deals with a 3D parametric model and the tracking is defined as a minimization problem where a cost function is minimised to obtained the best parameter values that make the model pose match with the hand one[7]. The cost function is also called a dissimilarity function where extracted features from hand images and model ones are used to compare the poses from real hand and 3D model. The dissimilarity function is minimized by means of minimsation algorithm such as the simplex approach proposed by Nelder and Mead [8] or using statistic methods such as particle filter [9]. The second category is mostly called data-driven method and uses a database of gestures calculated before the tracking process. It consists of matching the real hand pose with ones stored in the database through regression or classification techniques[10]. The author used coloured glove to improve the matching between input images and the ones stored in a database of images. The most important problem with this kind of methods is the huge number of hand gestures which makes impossible to create a database with all hand poses. In general a limited number of gestures is used to create a real time application.

Another category of methods simplifies the hand tracking problem to fingertips tracking. The method, then consists of detecting the fingertips for each input image and track them in a sequence of video. The advantage in this category is the fast tracking process which allows to create real time application easily.

Kumar and Shubham [11] proposed a segmentation algorithm which estimates hand position be calculating the centroid of orientation of the observed hand shape in a video sequence. Chen and al. [12] proposed a similar method which the identifies the finger areas for segmentation and recognition of the hand gestures. Finger detection algorithm was proposed by Zhou and al. to detect the finger using a 3D model represented by cylindrical surfaces and parallel features are then used to localize the hand fingers. Lee and al.[4] analyzed the hand shape using the curvature of the calculated contour in order to detect the fingertips. In a previous work, we proposed a fingertip tracking method using convex hull [1]. In this paper we propose a fast tracking method to apply for immersive application. Our method estimates the hand centroid position by calculating the

DOI: https://doi.org/10.3217/978-3-85125-609-3-31

barycentre of the fingertips. Convex hull is used to detect the fingertips. Our method is tested on real video sequence to move virtual object.

3 Segmentation

The first step of our method is called segmentation and it's considered as a fundamental part of the tracking process because it can greatly affect the final results. The segmentation consists of dividing the image into two areas : foreground and background. The foreground represents the hand in black colour while the rest of the image will be in white and represents the background. At the end of this step, a black and white image is obtained. In order to clarify this process, the following subsections will describe how images and colours are stored in the computer and then the process to achieve the segmentation.

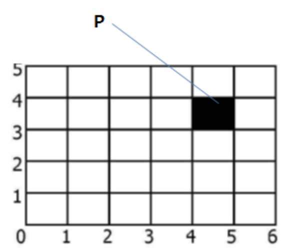

Fig. 1. Representation of an image as a matrix of pixels

In the computer, an image could be represented as a matrix of pixels as shown in figure 1 where each pixel will correspond to a color. Especially the pixel is a value encoding the color. We can differentiate three type of images: black and white images, grey scale images and colored ones. According to the type of the image, the pixels will take different values.

- Black and white image is also called digital, bi-level or two-level image. It means each pixel is stored as a single bit i.e a 0 or 1.
- Grey scale image is a matrix where a pixel represents the intensity of the grey color. In the computer the grey intensity is a value between 0 and 255.
- Colored image is also a matrix of pixels but each pixel will correspond to a vector of three values, each one between 0 and 255. The three values will represent one color.

There are many other types of images like the 3D ones or multispectral images but in this work, only the three kind of images presented before are considered. The next subsection gives an overview of some important color spaces that can be used to detect the skin color and explains the chosen(HLS) one in this work.

DOI: https://doi.org/10.3217/978-3-85125-609-3-31

3.1 Color spaces

It's important to notice there are many color spaces that can be used to represent the color. The following subsections present two of them.

RGB colour space This color space is the most used one and very easy to understand. It's based on using 3 values between 0 and 255 which encode the blue, green and red colors(Figure 2). Even if it's the easier one to understand, it doesn't deal with the main problem in computer vision which is the light variation. In fact, two images representing the same scene but taken in different conditions of lighting will be considered as completely dissimilar by the computer. In fact, the R,G,B values of the pixels will be completely different.

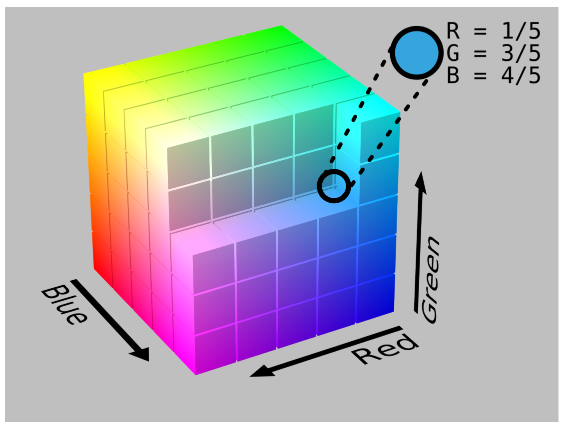

Fig. 2. RGB colour space [13]

HSL color space HSL stands for hue, saturation, and lightness is also often called HLS(Figure 3). This space colour deals better with the lighting variation because the colour is divided in pure colour information (with the H and S values) and the light one with the L value. For more clarifications : the pixels of two images representing the same scene but with different lighting will have the hue and saturation values very close. This colour space was used in this work and gave better results that the RGB one.

3.2 Skin Color detection

To achieve the segmentation it's important to distinguish the colours and select only the ones representing the hand. To detect the colour representing the hand, the proposed system starts by covering a square with the hand as shown in the figure 4.

DOI: https://doi.org/10.3217/978-3-85125-609-3-31

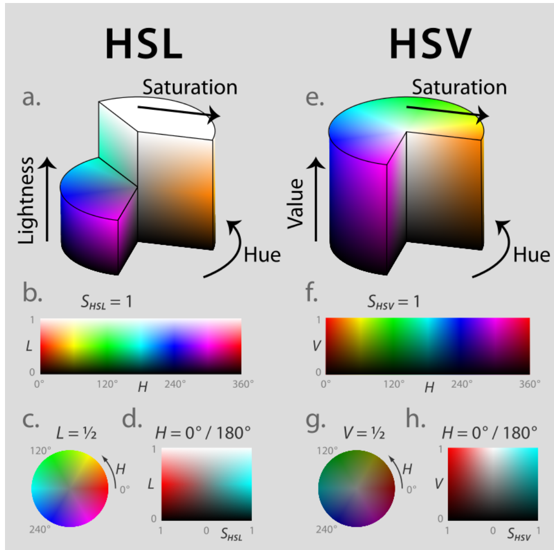

Fig. 3. HSL colour space [13]

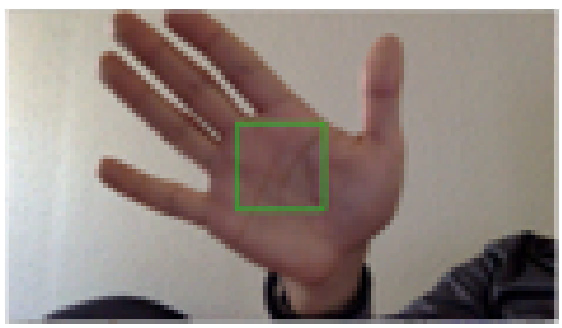

Fig. 4. initialisation of the process : covering a square with the hand and save the color hand

The median value of the pixels contained in the square is computed. According to the median value, two thresholds maxValue and minValue are defined using an empirical method and all the values between them are considered as the skin color. During the segmentation process, for each input image all the pixels encoding the hand will be colored with black color while the background will be colored with the white one.

4 Fingertips detection process

The first step of the fingertips detection process is to computer the hand contour from the black and white hand image. For that purpose, a method proposed by

DOI: https://doi.org/10.3217/978-3-85125-609-3-31

canny[14] is used. The second step. which is one of the most important part in that work. is to compute convex hull of the contour. The convex hull of a set of N points could be defined as the smallest perimeter fence enclosing the points. In our case the set of points consists of the contour points obtained by canny method. It could also be defined as the smallest convex polygon enclosing the points. The convex hull and convexity defects of the hand(Figure 5) are computed using the Sklansky's algorithm [15]. The points of the convex represent the points where the contour of the hand is convex, in other words the fingertips. The convexity defects represents the points where the hand contour is concave, in other words the points between the fingers (Figure 5).

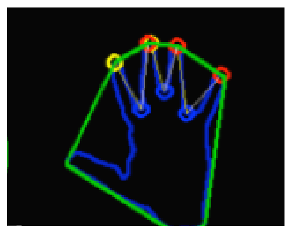

Fig. 5. hand contour (in blue colour) with detection of fingertips and the convexity defects (points between fingers)

5 Application and experimental results

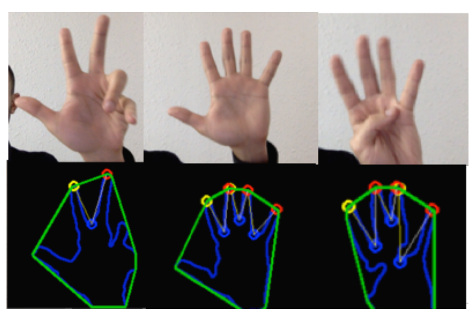

Fig. 6. Results : three poses of the hands with the results of the tracking. The first line represents the input images and the second one shows how the fingertips are detected

Once the fingertips detected, the barycentre is calculated. This latter represents the hand position in the image. The 2D coordinates of the barycentre will

DOI: https://doi.org/10.3217/978-3-85125-609-3-31

be used as the centre of a 3D virtual cube generated using the OpenGL library[1]. Our application is implemented using a machine with an Intel i5 processor 2,7 GHz and Intel Iris Graphics 6100 1536 MB. A video camera providing 2D images with a size of 480x620. The figure 6 shows the results of tracking the fingertips. The last figure (Fig7)shows how our application allows to move the 3D object according the position of the hand. In fact the first line represents the input images taken by a video camera and the second line shows the 3D object (Cube) following the hand position.

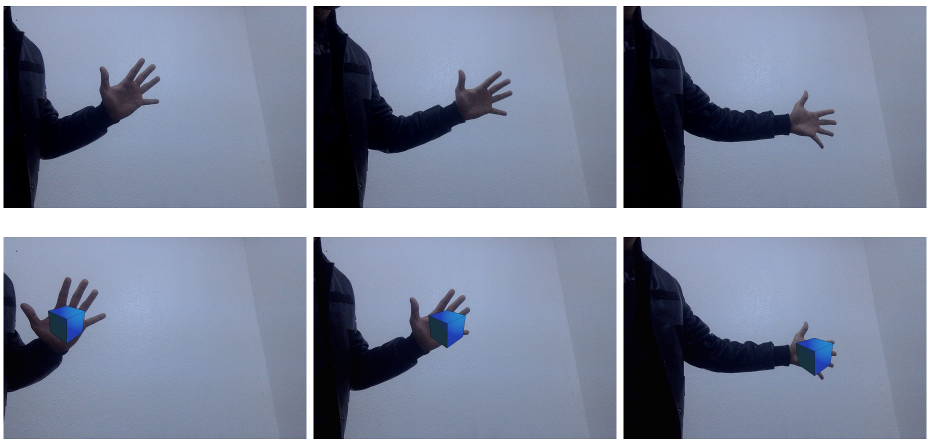

Fig. 7. Moving 3D cube according the hand position. First line : the input images. Second line: results.

6 Conclusion

In this paper, a real time fingertips tracking method was proposed and used to create an immersive application for moving virtual objects. The fingertips tracking method is based on HSL colour space segmentation algorithm and allows to detect the skin colour. The hand contour, convex-hull and convexity defect are then calculated in order to detect the fingertips. The barycentre of the latter represents the 2D position of the hand. Matching a virtual object position with the hand one allows to move the 3D object and create an interaction with a virtual world. Our proposed application uses only one hand. In future work, both hands can be involved in an immersive application to move, scale and rotate virtual objects which can provide a more complete human computer interaction system.

[1] www.opengl.org

DOI: https://doi.org/10.3217/978-3-85125-609-3-31

References

1. Ben Henia, O. Real Time Vision based method for finger counting through shape analysis with convex hull and PCA Techniques. *Internationa review on computers and software* **12.** doi:ouissemJournal2017 (2017).
2. Tkach, A., Pauly, M. & Tagliasacchi, A. Sphere-meshes for Real-time Hand Modeling and Tracking. *ACM Trans. Graph.* **35,** 222:1–222:11. ISSN: 0730-0301 (Nov. 2016).
3. Manresa, C., Varona, J., Mas, R. & Perales, F. J. Real ?Time Hand Tracking and Gesture Recognition for Human-Computer Interaction. *Electronic Letters on Computer Vision and Image Analysis* (2000).
4. Lee, D.-H. & Lee, S.-G. Vision-Based Finger Action Recognition by Angle Detection and Contour Analysis. *ETRI Journal* (2011).
5. Ben-Henia, O. & Bouakaz, S. *A NEW DEPTH-BASED FUNCTION FOR 3D HAND MOTION TRACKING* in *Proceedings of the International Conference on Computer Vision Theory and Applications (VISIGRAPP 2011)* (2011), 653–658. ISBN: 978-989-8425-47-8. doi:10.5220/0003378106530658.
6. Sridhar, S., Mueller, F., Oulasvirta, A. & Theobalt, C. *Fast and Robust Hand Tracking Using Detection-Guided Optimization* in *Proceedings of Computer Vision and Pattern Recognition (CVPR)* (2015). http://handtracker.mpi-inf.mpg.de/projects/FastHandTracker/.
7. Ben Henia, O., Hariti, M. & Bouakaz, S. *A two-step minimization algorithm for model-based hand tracking* en. in *18th International Conference on Computer Graphics, Visualization and Computer Vision (WSCG)* (Feb. 2010), 189–196. http://liris.cnrs.fr/publis/?id=4574.
8. Nelder, J. A. & Mead, R. A Simplex Method for Function Minimization. *The Computer Journal* **7,** 308–313 (Jan. 1965).
9. Isard, M. & Blake, A. CONDENSATION ffdfffdfffd Conditional Density Propagation for Visual Tracking. *International Journal of Computer Vision* **29,** 5–28 (1998).
10. Rosales, R., Athitsos, V., Sigal, L. & Sclaroff, S. *3D hand pose reconstruction using specialized mappings* in *ICCV* (2001), 378–385.
11. Dey, S. K. & Anand, S. Algorithm For Multi-Hand Finger Counting: An Easy Approach. *Advances in Vision Computing: An International Journal(AVC)* **abs/1404.2742.** http://arxiv.org/abs/1404.2742 (2014).
12. Chen, Z.-h., Kim, J.-T., Liang, J., Zhang, J. & Yuan, Y.-B. Real-Time Hand Gesture Recognition Using Finger Segmentation. *The Scientific World Journal* (2014).
13. wikipedia. *https://en.wikipedia.org/wiki/HSL_and_HSV*
14. Canny, J. A Computational Approach to Edge Detection. *IEEE Trans. Pattern Anal. Mach. Intell.* **8,** 679–698. ISSN: 0162-8828 (June 1986).
15. *Sklansky J Finding the Convex Hull of a Simple Polygon. PRL 1 number pp 79 83 (1982)*