COLOUR SPACE COMPARISON FOR SKIN DETECTION IN FINGER GESTURE RECOGNITION

Apostolos Tsagaris[1], Sotiris Manitsaris[2]
[1]Multimedia Technologies and Computer Graphics Lab, Department of Applied Informatics, University of Macedonia, Thessaloniki, Greece
[2]Robotics Lab, Department of Mathématiques et Systèmes, MINES ParisTech, Paris, France

ABSTRACT

The aim of this research is to compare the RGB, RG normalized, HSV and YCrCb colour spaces, which are used for skin detection in finger gesture recognition. Two comparison methods have been used: a) the first method calculates the deviation of the pixels identified as finger gesture and it is based on statistics by calculating how much the distribution of the coordinates of the fingertips match well with the normal distribution on a video (fingertip detection) and b) the second method uses the Precision and Recall metrics the provides the recognition accuracy of the system (finger gesture recognition). These methods have been tested with the PianOrasis system that provides 2D real-time recognition of fingerings and finger gestures, using different spaces based on a low-cost optical camera. The results of the research highlight that the most appropriate colour space for finger gesture recognition and especially for the recognition of finger strokes is the RG normalized. It has the minimum pixel deviation on the recognition and the best for fingertip detection. The precision rate is up to 93,45% and the recall rate up to 89,86%.

KEYWORDS: *mathematical morphology, skin model, gesture recognition, computer vision, real-time, interaction, finger tracking, colour space.*

I. INTRODUCTION

In Human-Computer Interaction (HCI), traditional input devices such as keyboard, mouse and joystick become less suitable in the interaction between human and computer since it is more natural for Human-Human Communication (HHC) to communicate with his/her body. Especially, hands and fingers have a a significant role in HHC. They used to accompany or even replace spoken communication in some cases, such as in sign languages, where hands completely replace spoken communication. Nowadays, computer vision gives a lot of potential for motion capturing without to put limits to the user and this offers the possibility to create more natural gestural interfaces. With regards to vision-based interfaces that uses gestures to control real or virtual processes, the colour of the skin consists a key factor. The skin is a full-fledged sense organ that plays the key role in the sense of touch. Its palmar surface embeds most of the sensory nerve endings which enable the sharpness of tactile perception [1].

The skin that covers the dorsum of the hand and digits is greatly different from the skin that covers the palm and the volar surface of the digits. Contrary to the later, the skin of the dorsal aspect is thin, pliable, and attached very loosely to the underlying tissues. This property enables it to move and stretch as much as necessary during the motion of the digits. Characteristic folds in the skin are even present on top of the interphalangeal knuckles, to account for skin extension during flexion of the digits. The dorsal skin has no particular sensory role; in this respect it is similar to the skin covering most of the rest of the body [2].

This paper presents the most important colour spaces that are used to detect skin regions on the image. The goal of this research is to find the most suitable for finger tracking and gesture recognition by comparing the detection of fingertips using different colour spaces. For this comparison, PianOrasis (*associating the terms Piano and Orasis - meaning «vision» in greek*) has been used. PianOrasis is a computer vision system for the recognition of finger gestures performed on a surface or in space. It recognises simultaneously the gestures of all the five fingers of a hand. Scale and rotation invariance techniques are also integrated to the system, increasing this way the recognition accuracy for finger gestures performed in space [3].

The paper is organized in the way that in Section 2 an overview of the suitable colour spaces used for skin detection is presented and in Section 3 a survey of relevant work. Section 4 presents an overview of the methodology and Section 5 the system architecture. Section 6 shows the evaluation method, Section 7 the experimental results and in Section 8 some conclusions and future work are presented.

II. COLOUR SPACES USED FOR SKIN DETECTION

A colour space is a method by which we can specify, create and isualize colour. As humans, we may define a colour by its attributes of brightness, hue and colourfulness. A computer may describe a colour using the amounts of red, green and blue phosphor emission required to match a colour. A printing press may produce a specific colour in terms of the reflectance and absorbance of cyan, magenta, yellow and black inks on the printing paper.

A colour is thus usually specified using three co-ordinates, or parameters. These parameters describe the position of the colour within the colour space being used. They do not tell us what the colour is, that depends on what colour space is being used.

2.1. RGB - RG

Red, Green, Blue (RGB) colour space is the most common colour space used for digital images. The main purpose of the RGB colour model is for the sensing, representation, and display of images in electronic systems, such as televisions and computers, though it has also been used in conventional photography. RGB is developed as an additive colour space and it has high correlation, non-uniformity and mixing of chrominance and luminance data. Therefore RGB is not suitable for colour analysis and colour based recognition [6] but a conversion of the RGB space to normalized RG space wich renders system less dependent on variations in brightness and allows the identification of the skin model as a set of values. Normalized RG space is based on the following mathematical formulas:

$$r = R/(R+G+B) \quad (1)$$

$$g = G/(R+G+B) \quad (2)$$

This procedure depicts every pixel of RoIRGB into RoIrg via the function N. It is about the depiction of pixels from the 3D space RGB in the 2D space RG:

$$\begin{aligned} &\forall p_i \in RoI^{RGB} \\ &N : RoI^{RGB} \to RoI^{rg}, \\ &N(R, G, B) = (r, g) \end{aligned} \quad (3)$$

The result is exported to a binary image containing either the value 1 for skin pixels (foreground) or 0 for the rest of the colours (background). Sometimes, given that the skin model is not perfect, small areas of background are considered by the system as belonging to the foreground and vice versa. This problem can be solved by applying methods of mathematical morphology for noise reduction.

2.2. HSV – Hue Saturation Value

Image data can be stored using various parameters. The well-known RGB colour space uses red, green, and blue colour components. The components of the HSV (Hue, Saturation, Value) colour space in contrast uses quantities that are more related to the every day colour description. Hue is the parameter we would define as “the colour”, saturation means the concentration, and value is the measure of brightness. As stated in [16] this colour space is less sensitive to lighting changes than RGB, and the hue value is invariant for the various skin "colours". In fact, it is only the saturation that increases for dark skinned people

$$H = \arccos \frac{\frac{1}{2}((R-G)+(R-B))}{\sqrt{((R-G)^2+(R-B)(G-B)}} \quad (4)$$

$$S = 1 - 3\frac{\min((R,G,B)}{R+G+B} \quad (5)$$

$$V = \frac{1}{3}(R+G+B) \quad (6)$$

The HSV colour spaces were developed to be more "intuitive" in manipulating colour and were designed to approximate the way humans perceive and interpret colour. They were developed when colours had to be specified manually, and are rarely used now that users can select colours visually or specify Pantone colours. The HSV colour space is preferred for manipulation of hue and saturation (to shift colours or adjust the amount of colour) since it yields a greater dynamic range of saturation. One disandvantage is that the costs for conversion for HSV colour spaces are expensive and the pixels with large and small intensities are discarded as hue and saturation becomes unstable.

2.3. YCrCb

YCrCb is an encoded nonlinear RGB signal and the transformation simplicity and explicit separation of luminance and chrominance components makes this colourspace attractive for skin colour modeling [17], [18]. As for skin colour segmentation, researchers have proposed the used of YCbCr colour space which contains luminance (Y) and chromaticity (CbCr) information. The separation of brightness information from the chrominance and chromaticity in YCbCr colour space can reduces the effect of uneven illumination in an image [19]. Therefore, YCbCr are typically used in video tracking and surveillance.

The equation for RGB conversion to YCbCr can be seen in (7)

$$\begin{aligned} Y &= 0.299R + 0.587G + 0.114B \\ C_r &= R - Y \\ C_b &= B - Y \end{aligned} \quad (7)$$

III. STATE OF THE ART

Skin detection can be defined as the process of selecting which pixels of a given image correspond to human skin. This procedure starts with the collection of skin images by several people as a sample, an image library as we call it. The criteria are to find a sample as diverse as possible so the results would be much more effective and accurate. Also, there needs to be a large number of pictures so the quality of the model remains in high levels. This implies that these images need to be distinct from the aspect that lighting and capturing angles are different in each one. Each sample consists of pixels that are pictured with a three-component vector (R, G, B) which is a point in the colour space.

Skin colour is affected by ambient light which is unknown in many situations; different cameras produce different colours, even from the same person, under the same illumination conditions; and finally, skin colours change from person to person.

Several colour spaces have been proposed in the literature for skin detection applications [4,24,25]. Many researchers have proposed robust and accurate solutions for detecting hand gestures using pattern recognition. But from reviews, detection using skin colour cue is fast to robust and can minimize the processing time [5]. Manresa, et al. [6] proposed a real-time control of a videogame based on hand gesture recognition in which they used colour spaces for hand segmentation. The application acquired promising result but using only average of the depths of the convexity defects to give the result not so reliable.

The skin colour segmentation can be done by many methods. Stergiopoulou and Papamarkos [7] proposed YCbCr segmentation. As YCbCr is also used in video coding and then no transcoding is needed, this colour space has been used in skin detection applications where the video sequence is compressed [8,9].

Bretzner et al. [10] use a hand detector by including skin colour information in the form of a probabilistic prior. Meanwhile hand is the only object in the plan, [11] propose to use the uniform YCbCr colour space in order to dissociate the luminance from the colour information to extract hand for gesture recognition. However, it is not reliable to model a skin colour for people of high variations of skin colours and under different lighting conditions. Limitations arise from the fact that human skin has common properties and that it can be defined in various colour spaces after applying colour normalization. So the model has to accept a wide range of colours, making it more susceptible to noise.

Peer et al [12] proposed RGB segmentation. Also In [13] two components of the normalized RGB colour space (rg) have been proposed to minimize luminance dependence. Ribeiro and Gonzaga [14] proposed different approaches of real time GMM (Gaussian Mixture Method) background subtraction algorithm using video sequences for image segmentation.

The HSV colour space has been proposed in [15] because it is more related to human colour perception. J. Yang et al [21] proposed a design and implementation of a command system based on static hand gesture recognition with the use of HSV colour model. It takes advantage of Fourier descriptors and K nearest neighbor classification to distinguish four different gestures, for the control of a Microsoft Office PowerPoint application. Experiment results show that the system can meet the requirements of practical applications.

In many cases researches uses a combination of colour models. Aznaveh et al. [22] are using mixed vector space of RGB and HSV and also neighbourhood pixels (vectors). There are two phases, one is based on skin colour and the other considers the neighbourhood pixels. A mixed skin model combined YCrCb Gauss model and normalized rgb has been proposed by Du, Y. and Yang, N [23]. The algorithm use the colour information which is simple and fast, so the detection speed is faster than other algorithm, and it can be used in some real-time system. D. Chai and K. N. Ngan [26] are using the YCrCb colour space because this format is typically used in video coding and an effective use of the chrominance information for modeling human skin colour can be achieved in this colour space.

In our application, the requirement is of fast computation. So, here we are simply making use of R/G ratio skin colour segmentation. This method gives the advantage of the speed of processing, more precision of the crop on the image and is easier to implemented. Also the users can move their hands freely in time and space without keeping their hand in a specific position with regards to the camera position.

IV. METHODOLOGY'S OVERVIEW

Finger motions are captured in multiple frames with a low-cost optical camera. The camera is placed several centimeters over the input surface, with a well-defined angle facing the working area. To make the system able to detect skin regions on a video, we developed a Skin Model (SM) (Figure 1). Based on the SM, the skin detection aims at the segmentation between dermal (*foreground*) and non-dermal regions (*background*) on the image and outputs a binary image. We determined the Region of Interest (ROI) by obtaining samples of colour pixels of the skin and nails.

The main feature of skin colour segmentation is to choose the suitable colour space. We are using four colour spaces to compare the results of the hand recognition. Red, Green, Blue (RGB) colour space is the most common colour space used on images. Normalised RG and HSV spaces are very common too.

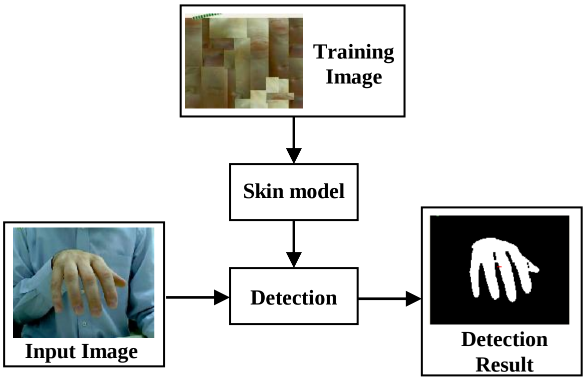


Figure 1. Skin Colour Detection

After the skin detection, filters that use mathematical morphology (Erode, Dilate and Gauss etc) are applied on the image in order to extract the hand from the background and the noise. It is very important to apply these filters with the appropriate parameters to obtain the best result on the image. Afterwards the fingers will be identified based on the geometric properties of the hand's posture.
The identification of fingers on the image becomes extremely difficult, especially when the distance between the tips of the fingers is very small. To address this problem, the binary image is imported into the hand segmentation algorithm and a set of image processing methods are applied: (a) the simplification of the binary image by reducing the noise and extracting the silhouette of the hand and (b) the decomposition of the image by extracting the contour of the hand and fingertips.
The candidate finger points are calculated by computing the Euclidean distances between the centroid and the coordinates of the pixels belonging to the contour of the fingers. The calculation of the local maxima of the Euclidean distances contributes to the identification of the fingers. To increase the quality of the finger identification scale and rotation techniques are applied on the captured frames of the camera [20].
The centroid of the hand silhouette is calculated based on:

$$X = \frac{\sum_{i=1}^{s} x_i}{s} \quad \text{and} \quad Y = \frac{\sum_{i=1}^{s} y_i}{s} \qquad (8)$$

where, x_i and y_i are the x and y coordinates of the i-th pixel in the hand region and s is the total number of pixels in the hand region. Finally we calculate the Δ x and Δ y distances of the centroid and the fingertips (Figure 2). We define a threshold of 20 pixels for the fingering.

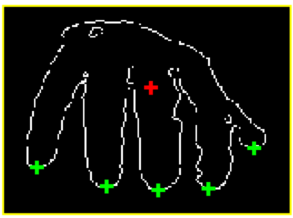


Figure 2. Distances Calculation

V. THE SYSTEM ARCHITECTURE

The methodology is evaluated via its implementation using a programming environment for real-time applications called Max/MSP and along with several toolboxes, such as Jitter and FTM (Figure 3). The implementation follows the concept of a low-cost computer vision system that will be able to recognize in real-time finger gestures performed in space, on a surface or on a keyboard.

The system works either in real-time by recording and processing the image directly from the camera, either with pre-recorded videos. The resolution of the videos is 320x240.

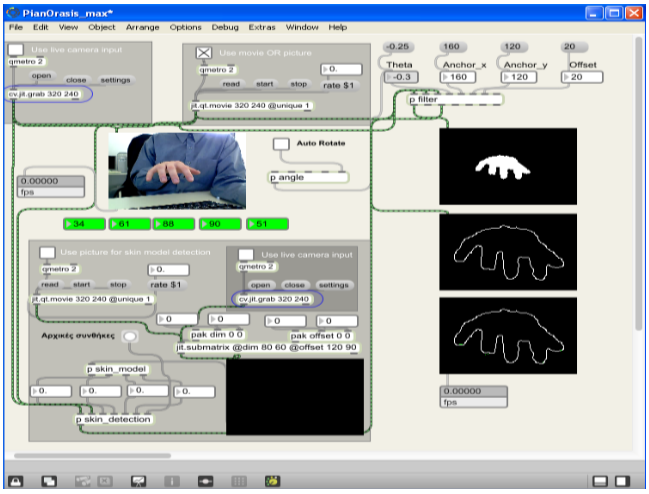


Figure 3. PianOrasis evaluation Software

The main patch gives the user the opportunity to use an already captured video or to capture straight from the camera. The object “cv.jit.resize” transforms the captured video from the original size to 320x240, which is the default for our program. This object uses bicubic interpolation to resize a matrix. In this case, we could use the “jit.matrix” object to resize the video, but “cv.jit.resize” generates more accurate results. Then, this matrix is imported in the patch “SkinDetection”. In the skin detection patch we detect the skin pixels on the images based on the 4 described colour spaces.

After the skin detection there are some pixels that do not really correspond to dermal pixels. We used a mathematical morphology approach in order to remove these pixels from the binary. At first, we applied an Open filter (object “cv.jit.open”). The morphological Open consists of applying erode and dilate operations sequentially. The output of this stage is the centroid of the hand silhouette that will be used later on to calculate the positions of the fingertips.

As a second stage, the hand segmentation will take place. At first, Gauss and canny filters are applied to get the contour of the hand silhouette. Afterwards, pixels on the hand contour are used to identify fingers and calculate position of the fingertips using Euclidean distances between maxima of the hand contour and the centroid of the hand silhouette. The output of this stage is a matrix with the coordinates of the five fingertips.

VI. EVALUATION METHOD

The evaluation is done with two approaches: a) the statistical approach that aims at the evaluation of the fingertip detection using the 4 colour spaces and b) the Precision and Recall metrics that aims at the evaluation of the gesture recognition accuracy.

The statistical values are the sum of squares of pixels deviations from the mean (DEVSQ), the value of how skewed the distribution of pixels is (SKEW) and the standard deviation (STDEVP). These statistical values are implement on the area in which the fingers are not moving. Thus can be measured the deviation of the recognized as dermal pixel.

DEVSQ calculates the mean of all the numbers and sums the squared deviation of each number from that mean. N are the number of values. The calculation formula is:

$$\sum_{i=1}^{N}(x_i-\bar{x})^2 \qquad (9)$$

SKEW returns a measure of how skewed a distribution is, relative to a normal distribution - that is, how asymmetric it is. Positive values indicate a distribution with a tail inclining to the positive side, and negative values a distribution with a tail inclining to the negative side. **SKEW** calculates:

$$\frac{n}{(n-1)(n-2)}\sum_{i=1}^{n}\left(\frac{x_i-\bar{x}}{s}\right)^3 \quad (10)$$

STDEVP returns the standard deviation from a number of values, calculated on the entire population. For a population of *N* values, the calculation formula is:

$$\sqrt{\frac{1}{N}\sum_{i=1}^{N}(x_i-\bar{x})^2} \quad (11)$$

The performance of the system is measured with the use of Precision and Recall metrics. Precision and Recall method is a parameterized method that balanced between accuracy and noise. Recall is the number of correctly detected strokes by the system divided by the total number of strokes (i.e. detected and not detected strokes). Precision is the ratio of correctly detected strokes to the total number of strokes detected by the system (i.e. correctly detected and wrongly detected strokes). Recall measures how sensitive the strokes detection module is and precision determines the accuracy of the system's module.

In terms, Precision is the probability that the detector' s event is valid, and recall is the probability that the ground truth data was detected.

Equations (12) and (13) provide mathematical definitions of precision (p) and recall (r) for convenience.

Precision is the fraction of retrieved strokes that are relevant to the search.

$$\Pr ecision = \frac{|\mathrm{Re}\,levant\_strokes \cap \mathrm{Re}\,trieved\_strokes|}{\mathrm{Re}\,trieved\_strokes} \quad (12)$$

Recall in information retrieval is the fraction of the strokes that are relevant to the query that are successfully retrieved.

$$\mathrm{Re}\,call = \frac{|\mathrm{Re}\,levant\_strokes \cap \mathrm{Re}\,trieved\_strokes|}{\mathrm{Re}\,levant\_strokes} \quad (13)$$

VII. EXPERIMENTAL RESULTS

The aim of our research is to find the most appropriate colour space for skin detection using an optical camera to capture finger movements. PianOrasis has been used as finger gesture recognition system to evaluate the 4 colour spaces. In order to evaluate the performance of PianOrasis for the static recognition of finger gestures performed in space (without keyboard) several scenarios have been used.

The statistical values are calculated only on the region where fingers are not moving, that can be measured the sum of squares of pixels deviations from the mean (DEVSQ), the value of how skewed the distribution of pixels are (SKEW) and the standard deviation (STDEVP). Table 1 shows the results of the statistical evaluation including square deviation, the skew distribution and the standard deviation.

28 videos with total duration of 548 sec have been used to compare the colour spaces. With 15 fps frame rate, 8220 frames were processed, in which 286 strokes were made. The normalized RG space gives the best results about the deviation of the correct recognized pixel from the gesture. It is obvious that the standard deviation for Ring finger is 2,06 pixels, for the middle finger is 2,58 pixels and for the index finger 2,81 pixels, quite smaller than the respective fingers of the other spaces.

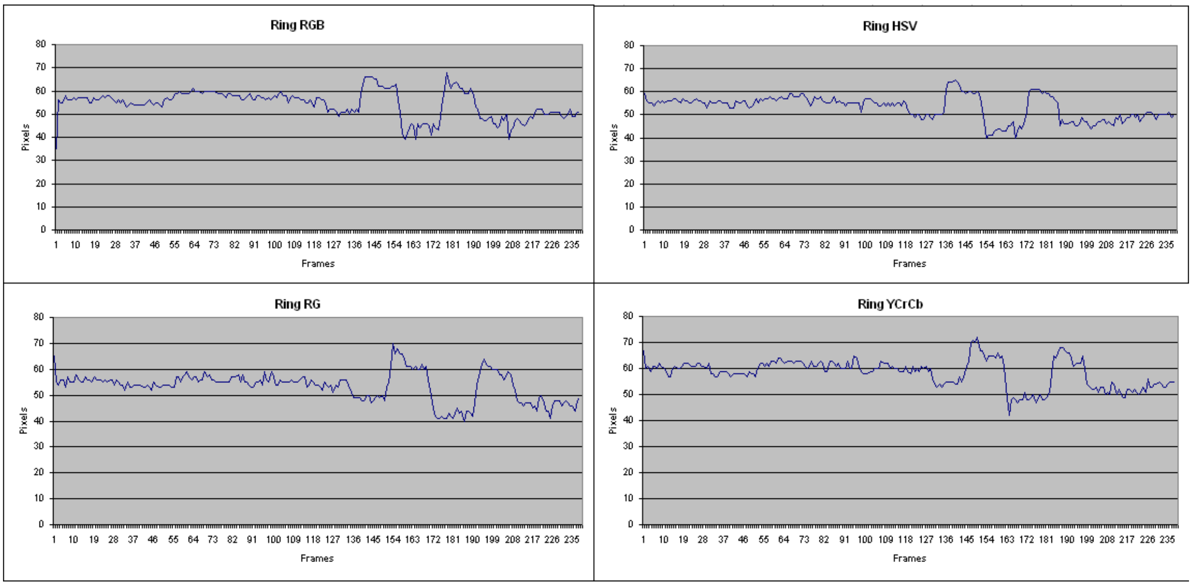
Table 1. Statistical evaluation of Colour Spaces

Ring Finger				
	RGB	**HSV**	**YCrCb**	**RG**
DEVSQ	1315,67	1079,53	828,42	591,97
SKEW	-2,57	-0,86	-0,56	0,14
STDEVP	3,08	2,79	2,44	2,06

Middle Finger				
	RGB	**HSV**	**YCrCb**	**RG**
DEVSQ	3060,9	2966,54	2491,85	1563,89
SKEW	-0,41	-0,42	-0,39	-0,24
STDEVP	3,67	3,52	3,47	2,58

Index Finger				
	RGB	**HSV**	**YCrCb**	**RG**
DEVSQ	2174,1	2156,44	2002,95	1648,64
SKEW	-0,81	-0,73	-0,65	-0,21
STDEVP	3,93	3,91	3,22	2,81

Figure 4 presents an excerpt of the video on which the fingertip detection has been applied.



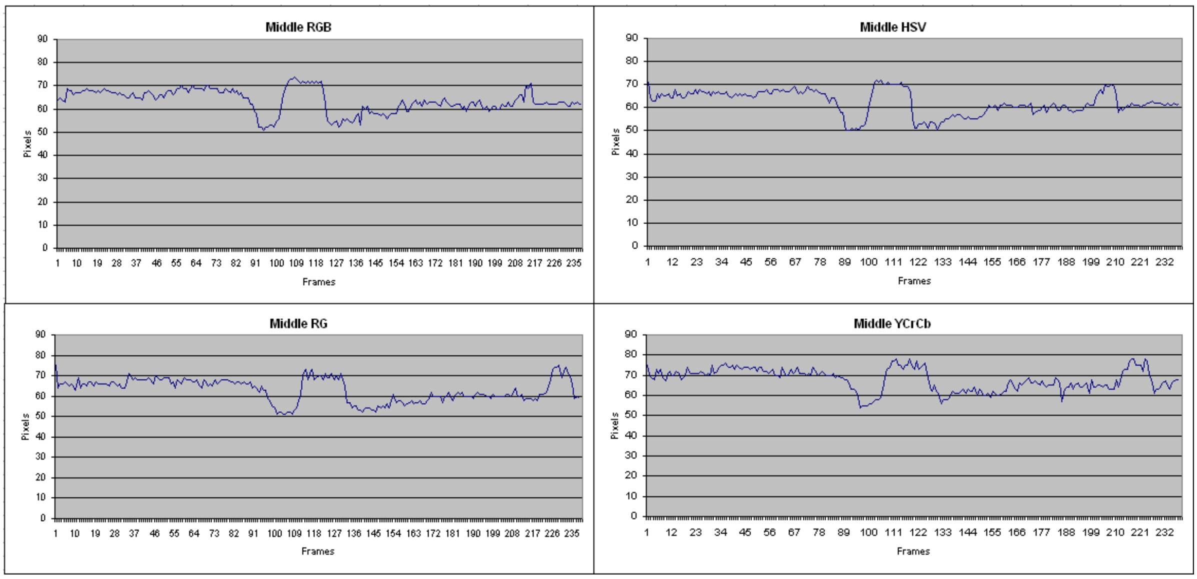


Figure 4. Fingertips positions

Precision and Recall metrics are presented in table 2. We have the following results from 286 strokes in total.

Table 2. Precission and Recall evaluation of strokes on different Colour Spaces

	Retrieved	Relevant	Not Relevant	Precis.	Recall
RGB	242	208	34	85,95 %	72,73%
HSV	231	205	26	88,74 %	71,68%
YCrCb	255	228	27	89,41 %	79,72%
RG	275	257	18	93,45 %	89,86%

Precision is the percentage of accurate stokes detected according to the total of strokes, while Recall is te percentage of strokes related to the total actual strokes. Thus, precision in RG colour space is 93,45% - better compared to other spaces (RGB – 85,95%, HSV – 88,74%, YCrCb – 89,41%). Recall from RG is 89,86% also better from other spaces (RGB – 72,73%, HSV – 71,68%, YCrCb – 79,72%).

VIII. CONCLUSION

We presented a comparison between the RGB, RG normalized, HSV and YCrCb colour spaces in terms of detection of the fingertips' position and gesture recognition accuracy. The aim of our research is to find the most appropriate colour space for finger gesture recognition based on skin detection. The comparison has been done using the PianOrasis system which is implemented in the max/MSP programming language. The hands of the users were in a semi-closed pose with a distance of 15 cm from the optical camera and with a maximum rotation of 10°. It has been asked to the users to perform music-like finger gestures performed in space. The system correctly identified the fingertips for all the frames. Both statistical evaluation and Precision and Recall metrics showed that the most appropriate space for this purpose is the normalised RG space. The results can be justified by the fact that there is no luminance information in the normalized RG colour space. That makes the skin detection less dependent on lightning variations than the other colour spaces. Nevertheless, a fully controlled room with predefined lightning conditions has been used for these experiments. The optimization of the algorithm in order to render the system completely independent from lightning variations can be a very interesting future work..

REFERENCES

[1] J. Randall Flanagan and Roland S. Johansson, (2002) “Hand Movements”. In: *Encyclopedia of the Human Brain*. Edited by Vilanayur S. Ramachandran. Volume 2. 4 volumes. New York, United States: Academic Press (Elsevier Science Ltd.), pages 399–414.

[2] R. Michalec, (2011) “Modeling and Control of Multifingered Dextrous Manipulation for Humanoid Robot Hands”, *Doctoral thesis*, France, 2011

[3] A. Tsagaris, S. Manitsaris, K. Dimitropoulos, A. Manitsaris (2011), “Intelligent invariance techniques for music gesture recognition based on skin modeling”, *12th IEEE International Symposium on Computational Intelligence and Informatics (CINTI 2011)*, 21-22/11/2011, Budapest, Hungary

[4] A. Albiolt, L. Torres E. Delpf, (2011) “Optimum colour spaces for skin detection”, *Proceedings. International Conference on Image Processing (2001),* Volume 1, Pages: 122- 124.

[5] E. Lin, A. Cassidy, D. Hook, A. Baliga, Tsuhan Chen, (2002) "Hand tracking using spatial gesture modelling and visual feedback for a virtual DJ system", *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces (ICMI'02),* Volume , Issue , Pages: 197- 202.

[6] C. Manresa, J. Varona, R. Mas, F.J. Perales, (2005) "Real-time hand tracking and gesture recognition for human-computer interaction", *ELCVIA(5),* No. 3, Pages: 96-104.

[7] E. Stergiopoulou and N. Papamarkos, (2006) "A new technique for hand gesture recognition", *Proceedings of IEEE International conferenceon Image Processing, Atlanta*, pp. 2657-2660, Oct. 8-11.

[8] A. Albiol, L. Torres, C.A. Bouman, and E. J. Delp, (2000) “A simple and efficient face detection algorithm for video database applications,” in *Proceedings of the IEEE International Conference on Image Processing,* Vacouver,Canada, vol. 2, pp. 239- 242.

[9] H. Wang and S-F. Chang, (1997) “A highly efficient system for automatic face region detection in mpeg video,” *IEEE Transactions on circuits and system for video technology,* vol. 7, no. 4, pp. 615-628.

[10] L. Bretzner, I. Laptev, and T. Lindeberg. (2002) “Hand Gesture Recognition Using Multi-scale Colour Features, hierarchical models and particle filtering*”. In Proc. Intl. Conf. Face and Gesture Recognition*, pages 423–428.

[11] A. Ben Jmaa1, W. Mahdi1, Y. Ben Jemaa2, and A. Ben Hamadou1, (2009) “ Hand Localization and Fingers Features Extraction: Application to Digit Recognition in Sign Language*”, The 10th International Conference on Intelligent Data Engineering and Automated Learning* “ IDEAL”. pp.151-159.

[12] P. Peer, J. Kovac, J. and F. Solina, (2003) ”Human skin colour clustering for face detection”, In: submitted to *EUROCON –International Conference* on Computer as a Tool .

[13] J. G. Wang and E. Sung, (1999) “Frontal-view face detection and facial feature extraction using colour adn morphological operators,” *Pattern recognition letters*, vol. 20, no. 10, pp. 1053-1068.

[14] H. Luchetti Ribeiro, A. Gonzaga, (2006) "Hand Image Segmentation in Video Sequence by GMM: a comparative analysis", *Symposium on Computer Graphics and Image Processing (SIBGRAPI'06),* 357 - 364, 8-11.

[15] K. Sobottka and I. Pitas, (1996) “Face localization and facial feature extracion based on shape and colour information,” *in Proceedings of the IEEE International Conference on Image Processing*, Lausanne, Switzerland, vol. 3, pp. 236241.

[16] G. R. Bradski, (1998) “Computer Vision Face Tracking For Use in a Perceptual User Interface”, *Intel Technology Journal*, Q2.

[17] S. L Phung, A. Bouzerdoum and D. Chai, (2002) "A novel skin colour model in YCbCr colour space and its application to human face Detection", *IEEE International Conference on Image Processing (ICIP'2002)*, vol. 1,289-292.

[19] R.-L. Hsu, M. Abdel-Motalleb and A.K. Jain, (2002) "Face detection in colour images", *IEEE Trans. Pattern Analysis and Machine Intelligence 24,*5, 696-706.

[19] P. Sebastian, V. Yap Vooi and R. Comley. (2008) “The effect of colour space on tracking robustness*”.* In *Industrial Electronics and Applications,* ICIEA 2008.

[20] A. Tsagkaris, S. Manitsaris, K. Dimitropoulos and A. Manitsaris (2011), “Scale and Rotation invariance for the recognition of finger musical gestures performed in space”, *European Workoshop on Visual Information Processing (EUVIP 2011)*, 4-6/06/2011, Paris – France

[21] J. Yang, J. Xu, M. Li, D. Zhang and C. Wang (2011), “A Real-time Command System based on Hand Gesture Recognition”, *Seventh International Conference on Natural Computation*, 26-28/07/2011, Shanghai-China

[22] M.M., Aznaveh, H. Mirzaei, E. Roshan and M. Saraee, (2009) "A New and Improved Skin Detection Method Using Mixed Colour Space", *Human System Interactions Conference*, Vol. 60, 471-480.
[23] Y. Du and N.Yang, (2011) "Research of Face Detection in Colour Image Based on Skin Colour". *Energy Procedia* 13, 9395–940.
[24] S.L. Phung,, A. Bouzerdoum, D. Chai (2005), " Skin segmentation using colour pixel classification: analysis and comparison.", *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 27(1), 148–154.
[25] K. Kakumanu, S. Makrogiannis, N. Bourbakis, N. (2007), "A survey of skin colour modeling and detection methods.", *Pattern Recognition*, 40(3), 1106–1122
[26] D. Chai and K. N. Ngan, (1999) "Face Segmentation Using Skin-Colour Map in Videophone Applications", *IEEE Transactions on circuits and systems for video technology*, vol 9, no. 4.

AUTHORS

Apostolos Tsagaris was born in Germany in 1970. He received the Bachelor degree in Automation Engineer from Alexander Technological Educational Institute of Thessaloniki – Greece in 1994, MSc in Design of interactive and industrial products and systems from Aegean University - Greece, in 2005, Msc in Mechatronics from University of polytechnic from Catalonia – Barcelona – Spain in corporation with T.E.I of West Macedonia - Greece, in 2007 and Med in Adult Education from Open University of Patras – Greece, in 2010. He is currently a Lecturer in Robotics and CAD/CAM/CAE in Automation Department in ATEI (Greece). His main research interests include Automation, Mechatronic and Robotic systems, computer vision, CAD/CAM/CAE and System Integration . His involvement with those research areas has led to the co-authoring of 30 articles in refereed journals and international conferences. He has participated in several research projects.

Sotiris Manitsaris is a researcher at the Robotics Laboratory of MINES ParisTech. He has a university degree in applied mathematics in computer science from the Aristotle University of Thessaloniki, a diploma in music studies and a doctoral thesis in computer vision for the finger gesture analysis in music interaction recognition. He has a number of projects for the medical rehabilitation of hand functions using gesture recognition from the University of Macedonia in Greece. He has done a postdoctoral research at the SIGMA Laboratory (ESPCI ParisTech) for the visual-speech recognition and synthesis and at the industrial chair between MINES ParisTech and PSA Peugeot-Citroën for the technical gesture modelling and recognition. Dr. Manitsaris is also the coordinator of art technology research projects, his research interests include computer vision and music, finger gesture control of sound and stochastic modeling.