

2D Versus 3D Colour Space Face Detection

Jure Kovač, Peter Peer, Franc Solina

`jure.kovac@link.si`, `{peter.peer, franc.solina}@fri.uni-lj.si`

University of Ljubljana,

Faculty of Computer and Information Science

Tržaška 25, SI-1000 Ljubljana, Slovenia

19th January 2003

Abstract

Computer vision is one out of many areas that want to understand the process of human functionality and copy that process with intention to complement human life with intelligent machines. For better human-computer interaction it is necessary for the machine to see people. This can be achieved by employing face detection algorithms, like the one used in the installation “15 Seconds of Fame” [7]. Mentioned installation unites the areas of modern art and technology. Its algorithm is based on skin colour detection. One of the problems this and similar algorithms have to deal with is sensitivity to the illumination conditions under which the input image is captured. Hence illumination sensitivity influences face detection results. One of the aspects from which we can observe illumination influence is the choosing of the proper colour space. Since some colour spaces are designed to eliminate the influence of illumination when describing colour of an object, an idea of using such a colour space for skin-colour detection was taken under consideration and some of the methods were researched and tested.

KEYWORDS

computer vision, automatic detection, human face, face candidates search, skin-colour determination, 2D colour space, 3D colour space, illumination independence

1 Introduction

1.1 Installation “15 Seconds of Fame”

The installation “15 Seconds of Fame” [7] is an interactive art installation, which intends to make instant celebrities out of common people by putting their portraits on the museum wall. The idea was inspired by the quotation of the famous artist Andy Warhol: “In the future everybody will be world famous for fifteen minutes” and by the pop-art style of his work. The installation looks like a valuable framed picture (Fig. 1). LCD monitor and digital camera are built into the picture. Camera is connected to a computer, which controls the camera and processes captured images. Special software contains algorithm for face detection, which looks for faces in captured images. Among them it chooses one for further processing. In the next step a randomly chosen portrait is processed with randomly chosen filter and random colouring is applied afterwards. In such a way the portrait in a pop-art fashion arises, which is afterwards shown on screen for 15 seconds.



Figure 1: LCD computer monitor dressed up like a precious painting. The round opening above the picture is for the digital camera lens.

The face detection algorithm that is used by the installation “15 Seconds of Fame” [7] uses 3D colour space (RGB) for detecting skin colour pixels. With the help of heuristic rules it is determined, whether a certain pixel of input image corresponds to the skin colour. Note that the original face detection algorithm [6] was developed to work best under standard daylight illumination (CIE illuminant D65) [4]. After the skin colour classification is done for every pixel of the image, the skin region segmentation takes place. Unsuitable regions are then eliminated on the basis of geometric properties of the face. Remaining regions represent faces. Fig. 2 illustrates the described process. The skin colour for example for fair complexion is determined with

the following rules [7, 6], which describe the skin cluster in the RGB colour space:



Figure 2: Stages in the process of finding faces in an image in the installation “15 Seconds of Fame”: 1) downsize the resolution 2048×1536 of the original image to 160×120 pixels, 2) eliminate all pixels that can not represent a face, 3) segment all the regions containing facelike pixels using region growing, 4) eliminate regions which can not represent a face using heuristic rules.

```
% The skin colour at uniform daylight illumination =>
```

```
 $R > 95$  AND  $G > 40$  AND  $B > 20$  AND
```

```
 $\max\{R, G, B\} - \min\{R, G, B\} > 15$  AND
```

```
    % RGB components must not be close together -
```

```
    % greyness elimination
```

```
 $|R - G| > 15$  AND
```

```
    % also R and G components must not be close together -
```

```
    % otherwise we are not dealing with the fair complexion
```

```
 $R > G$  AND  $R > B$     % R component must be the greatest component
```

```
OR
```

```
% The skin colour under flashlight or (light) daylight lateral illumination
```

```
=>
```

```
 $R > 220$  AND  $G > 210$  AND  $B > 170$  AND
```

```
 $|R - G| \leq 15$  AND    % R and G components must be close together
```

$R > B$ AND $G > B$ % B component must be the smallest component

This is how we get the basic information for face candidates. Therefore skin colour determination is essential for further algorithm execution and the quality of results, in great extent, depends on described step.

1.2 Motivation

Since skin detection is performed in the RGB colour space, previously described rules are very much subject to the influence of illumination. Some colour spaces have a property of separating the luminance component from chromaticity component and with that at least partial independence of chromaticity and luminance is achieved. Such colour spaces are for example YUV, rg, HSI, TSL, etc. Mentioned property implies a possibility of searching skin colour elements in the 2D chromaticity space, where colour is described by only two dimensions and the third dimension is ignored. By doing that we achieve at least partial illumination independence when searching for skin colour elements and in such a way we try to improve the final results of face detection. This hypothesis is going to be checked in the following sections of this paper.

Also, images reconstructed under standard illumination [5], when we try to eliminate the influence of non-standard illumination from images, show that the amount of noise affects the quality of reconstruction. The amount of noise is increased in images with mild illumination. Which means that dark images contain more noise than images taken under bright illumination conditions. The amount of noise affects the quality of reconstruction results because noisy colours are transformed to wrong colours. If in certain case a noisy skin area is transformed in described fashion, this will have an influence on the compactness of the reconstructed area. Hence skin region segmentation performed by face detection algorithm can be seriously harmed. The 2D skin colour determination might diminish the influence of described type of noise.

1.3 Structure of the paper

In the next section we describe tested 2D colour space methods for finding skin-like pixels. The comparison of effectiveness of different methods for finding skin colour is given in Section 3, using the presented face detection scheme. We conclude the paper in Section 4.

2 2D colour space skin clustering methods

The skin cluster model is realized in YUV ($Y C_r C_b$) colour space [3]. This realization is done in two different ways.

In the first case we describe skin colours with cluster determined by two central curves $\overline{C}_b(Y)$ in $\overline{C}_r(Y)$ and by deviation curves $W_{C_b}(Y)$ in $W_{C_r}(Y)$. Here we can notice that all curves are dependant on Y component, which represents luminance. Equations of mentioned curves are the following:

$$\begin{aligned}
 W_{C_i}(Y) &= \begin{cases} W_{L_{C_i}} + \frac{(Y-Y_{min})(W_{C_i}-W_{L_{C_i}})}{K_l-Y_{min}} & ; Y < K_l \\ W_{H_{C_i}} + \frac{(Y_{max}-Y)(W_{C_i}-W_{H_{C_i}})}{Y_{max}-K_h} & ; K_h < Y \\ W_{C_i} & ; \text{else} \end{cases} \\
 \overline{C}_b(Y) &= \begin{cases} 108 + \frac{10(K_l-Y)}{K_l-Y_{min}} & ; Y < K_l \\ 108 + \frac{10(Y-K_h)}{Y_{max}-K_h} & ; K_h < Y \\ 108 & ; \text{else,} \end{cases} \\
 \overline{C}_r(Y) &= \begin{cases} 154 - \frac{10(K_l-Y)}{K_l-Y_{min}} & ; Y < K_l \\ 154 + \frac{22(Y-K_h)}{Y_{max}-K_h} & ; K_h < Y \\ 108 & ; \text{else,} \end{cases} \quad (1)
 \end{aligned}$$

where i in $W_{C_i}(Y)$ is b or r , $W_{C_b} = 46.97$, $W_{L_{C_b}} = 23$, $W_{H_{C_b}} = 14$, $W_{C_r} = 38.76$, $W_{L_{C_r}} = 20$, $W_{H_{C_r}} = 10$, $K_l = 125$, $K_h = 188$, $Y_{min} = 16$ and $Y_{max} = 235$. These parameters are evaluated on the basis of skin cluster constructed from tests described in [3]. For every pixel of input image it is determined whether it corresponds to the colour of the skin or not. If it does, it corresponds to the following conditions:

$$\begin{aligned}
 \overline{C}_r(Y) - \alpha W_{C_r}(Y) &< C_r(Y) < \overline{C}_r(Y) + \alpha W_{C_r}(Y) \\
 \overline{C}_b(Y) - \alpha W_{C_b}(Y) &< C_b(Y) < \overline{C}_b(Y) + \alpha W_{C_b}(Y), \quad (2)
 \end{aligned}$$

where α represents a constant. Selected value α in [3] equals 1. This choice is quite unpleasant since it causes that too many pixels are recognized as skin colour. By the use of small test set we manage to tune the value of α . It was shown that the best results are provided by the value of $\alpha = 0,56$. This value is also used in the described tests. The graph of skin colour cluster is shown on Fig. 3.

In the second case the skin colours can be represented with an ellipse. In this case we adopt the following constraint. In the colour space YUV the chromaticity is independent of luminance, which is not entirely true. In reality the chromaticity is nonlinearly dependent on luminance Y. The

equation of an area that has a form of an ellipse (see Fig. 4), with which we can describe the skin colour cluster, can be written as:

$$\frac{(x - e_{c_x})^2}{a^2} + \frac{(y - e_{c_y})^2}{b^2} \leq 1$$

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} C_b - c_x \\ C_r - c_y \end{bmatrix}, \quad (3)$$

where $c_x = 109.38$, $c_y = 152.02$, $\theta = 2.53$ radians, $e_{c_x} = 1.6$, $e_{c_y} = 2.41$, $a = 25.39$ and $b = 14.03$. This parameters also reflects the properties of skin colour cluster described in [3]. When chromaticity (C_r, C_b) leads to the pair (x, y) , that is inside the ellipse, then this colour corresponds to the skin colour.

In the continuation we refer to the described methods as 3D (our 3D method), 2D_a (for the first 2D method) and 2D_b (for the second 2D method).

3 Comparison of methods

Described methods for skin colour determination were tested on two testing sets [5]. Experiments were first performed on the set of images gathered at the first public showing of the installation. The testing set is composed of 40 images taken under standard (CIE illuminant D65) [4] illumination conditions with one or more faces on each image.

The results of this test are shown in Tab. 1 and are in comparison to the 3D method for skin detection much worse. The reason lies in the fact that beside true skin colour pixels many other pixels that do not represent

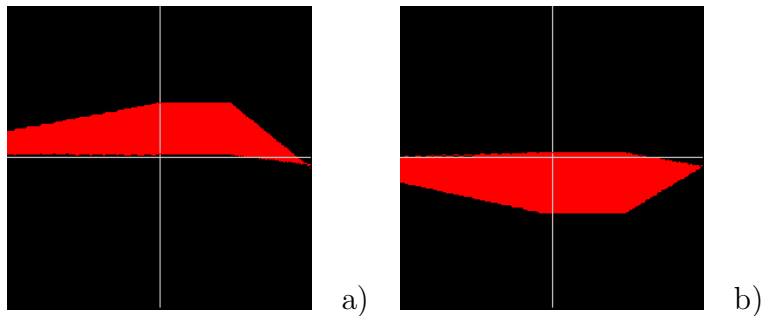


Figure 3: The graph of skin colour cluster of the first method for the 2D skin colour determination shows: a) function $C_r(Y)$ (first line in Eq. (2)), b) function $C_b(Y)$ (second line in Eq. (2)).

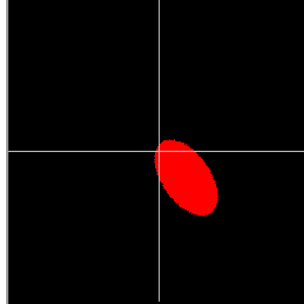


Figure 4: The skin cluster of the second method for the 2D skin colour determination: x axis of coordinate system corresponds to component C_r and y axis corresponds to component C_b (see Eq. (3)).

skin colour were recognized as skin. These pixels obstructed face detection algorithm when it performed skin colour segmentation. As a result face regions were connected with the neighbouring regions and as such could not satisfy the face shape criterion. Hence they were eliminated from further processing. An example of this problem is shown on Fig. 5. In Tab. 1 we can notice that the algorithm Grey World [2, 5] diminishes the influence of this effect. Gray World algorithm is based on the presumption that the average surface colour on the image is acromatic. It presents simple and fast method for colour compensation, i.e. for elimination of influence of non-standard illumination from images.

The second testing set contained approximately $3 \cdot 20 = 60$ images with one or more faces on each image. They were taken under 3 different non-standard illumination conditions. The subsets *red*, *blue* and *green* were generated by illuminating the scene with the corresponding colour light. First, illumination detection on images was performed with correlation technique [1, 5], then the images were reconstructed under approximation of standard illumination and finally face detection algorithm was applied on images. Note that the results of the face detection on the second set of images are in all cases TP/Det=0% and FN/All=100% if no method for eliminating the influence of non-standard illumination is applied previously to face detection algorithm. This can occur if face detection algorithm finds no skin colour in an image, which can often be the case in extreme illumination conditions.

Correlation method (Colour by Correlation [1, 5]) for eliminating the influence of non-standard illumination consist of two distinct steps: scene illumination detection and standard illumination reconstruction. In the first step, the algorithm determines with the help of preliminary knowledge which illumination out of the set of known illuminations is present in the image. In

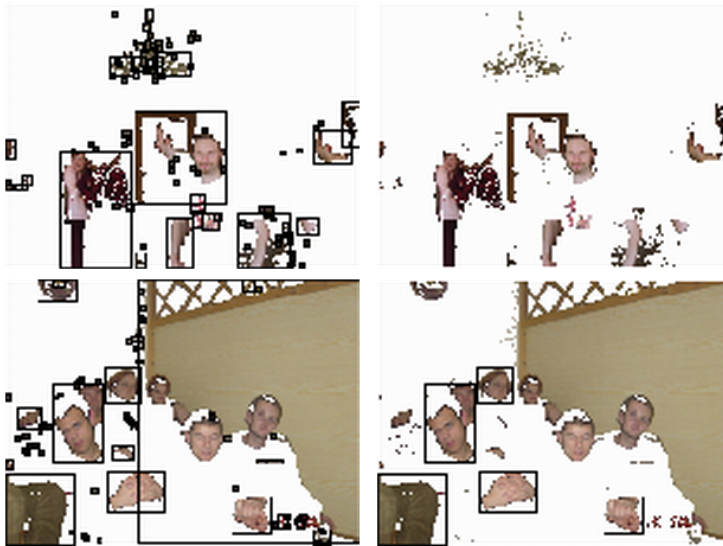


Figure 5: Two examples of bad results of face detection performed in the 2D colour space: All found regions are shown on the left side of the figure and all kept regions are shown on the right side of the figure. The reason for bad performance is explained in the text.

the second step, it applies the necessary transformations to reconstruct the standard (or other wanted) illumination.

Results of the test performed on this set showed that 2D methods for skin colour determination have in some cases positive influence on the results of face detection. From Tab. 2 it is obvious that described methods show a certain amount of improvement, first of all on FN/All percentage. This is obvious on subsets with darker illumination conditions (*red* and *green*). We noticed that a bigger noise level is present in the image after performing reconstruction of standard illumination on images from all three subsets. The 2D skin detection methods diminish the influence of noise and cause better compactness of skin regions, hence better segmentation is possible.

4 Conclusion

From the presented results we can conclude that for face detection algorithms that detect faces based only on colour information it is better to use skin detection in the 3D colour space. This fact also dictates the usage of specific constraints on the type of input image illumination. The methods that use 2D colour space for skin colour determination are more appropriate for face

standard				standard + GW			
Method	3D	2D_a	2D_b	Method	3D	2D_a	2D_b
All	109			All	109		
Detected	75	64	65	Detected	70	85	91
TP	68	37	35	TP	65	65	64
FP	7	27	21	FP	5	20	27
FN	40	71	73	FN	44	44	45
TP/Det	90,66	57,81	53,85	TP/Det	92,85	76,47	70,33
FN/All	36,70	65,14	66,97	FN/All	40,37	40,37	41,28

Table 1: The results of face detection for standard illumination conditions (**standard**) and the same set previously processed with Grey World algorithm [2, 5] (**standard + GW**). The results are given for our 3D method (**3D**) and for the presented 2D methods (**2D_a**, **2D_b**) for the skin colour determination. The table shows the number of all detections (Detected), the number of detected faces as true positives (TP), the number of false detections as false positives (FP) and the number of faces missed as false negatives (FN). TP/Det shows the percentage of true positives out of all detections and FN/All shows the percentage of false negatives out of all faces in the subset. For our installation the first percentage is extremely important, while the second one is merely informative, since we have consciously eliminated faces that were too small for further processing but were included in the number of all faces!

detection algorithms that beside colour based approach use also some other approach, e.g. feature based approach. The 2D methods define skin colour over greater colour specter, hence more exact process for face candidate region conformation is necessary. All arguments presented in this paper call our attention to the fact that the proper method selection with regard to the type of input image illumination is essential.

The hypothesis that 2D colour space face detection could improve our results, as written in Section 1.2, was checked. 2D methods indeed diminish the influence of noise in dark images, but in images that are captured under standard daylight illumination they label too many pixels as skin-like pixels. So the performance of our face detector is decreased. This confirms that we made the right decision to use the 3D method already at the beginning of the installation development.

The installation works in real time with surprisingly good results. To overcome the problem with non-standard illumination or with other words to achieve illumination independance, we instead of 2D colour space skin labeling approach use the algorithms for eliminating the influence of non-standard illumination from images, e.g. Gray World [2] and Colour by Correlation [1]. Now the installation can be exhibited almost anywhere.

blue + C				green + C			
Method	3D	2D_a	2D_b	Method	3D	2D_a	2D_b
All	41			All	23		
Detected	80	68	65	Detected	18	29	19
TP	29	22	26	TP	13	18	16
FP	51	46	39	FP	5	11	3
FN	13	20	16	FN	10	5	7
TP/Det	36,25	32,35	40,00	TP/Det	72,22	62,07	84,21
FN/All	30,95	48,78	39,02	FN/All	43,48	12,20	17,07

red + C			
Method	3D	2D_a	2D_b
All	34		
Detected	25	31	30
TP	23	27	27
FP	2	4	3
FN	11	7	7
TP/Det	92,00	87,10	90,00
FN/All	32,35	20,58	20,58

Table 2: The results of face detection for three non-standard illumination conditions previously processed with Colour by Correlation algorithm [1, 5] (**blue + C**, **green + C** and **red + C**). The results are given for our 3D method (**3D**) and for the presented 2D methods (**2D_a**, **2D_b**) for the skin colour determination. The table shows the number of all detections (Detected), the number of detected faces as true positives (TP), the number of false detections as false positives (FP) and the number of faces missed as false negatives (FN). TP/Det shows the percentage of true positives out of all detections and FN/All shows the percentage of false negatives out of all faces in the subset. The TP/Det is for our installation extremely important, while FN/All is merely informative for the performance of our face detector. Note that small faces are deliberately eliminated from further processing already by the face detection algorithm.

References

- [1] G. D. Finlayson, S. D. Hordley, P. M. Hubel, “Color by Correlation: A Simple, Unifying Framework for Colour Constancy,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 11, pp. 1209–1221, 2001.
- [2] B. Funt, K. Bernard, L. Martin, “Is Machine Colour Constancy Good Enough,” *Proceedings of the 5th European Conference on Computer Vision (ECCV’98)*, pp. 445–459, Freiburg, Germany, 1998.
- [3] R. L. Hsu, M. Abdel-Mottaleb, A. K. Jain, “Face Detection in Colour Images,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*,

vol. 24, no. 5, pp. 696–706, 2002.

- [4] Joint ISO/CIE Standard ISO 10526:1999/CIE S005/E-1998 CIE Standard Illuminants for Colourimetry, 1999, <http://www.cie.co.at/cie/> .
- [5] J. Kovač, “Eliminating the Influence of Non-Standard Illumination from Images,” *Diploma Thesis*, University of Ljubljana, Faculty of Computer and Information Science, 2002.
- [6] P. Peer, F. Solina, “An Automatic Human Face Detection Method,” *Proceedings of the 4th Computer Vision Winter Workshop (CVWW’99)*, pp. 122-130, Rastendorf, Austria, 1999.
- [7] Franc Solina, Peter Peer, Borut Batagelj, Samo Juvan, “15 seconds of fame - an interactive, computer-vision based art installation,” *Proceedings of the 7th International Conference on Control, Automation, Robotics and Vision (ICARCV 2002)*, pp. 198–204, Singapore, 2002.