Skin detection in video under changing illumination conditions

Maricor Soriano, Birgitta Martinkauppi, Sami Huovinen
Machine Vision and Media Processing Unit
Infotech Oulu and Dept. Of Elec. Eng.
P.O.Box 4500 FIN-90014 University of Oulu, FINLAND
{msoriano,jbm,shuovine}@ee.oulu.fi

Mika Laaksonen
NOKIA Mobile Phones
P.O.Box 50 (Elektroniikkatie 10)
FIN-90571 Oulu, FINLAND
mika.h.laaksonen@nokia.com

Abstract

Techniques for color-based tracking of faces or hands often assume a static skin color model. However, skin color perceived by a camera can change when lighting changes. In common real environment multiple light sources impinge on the skin. Therefore, for robust skin pixel detection, a dynamic skin color model that can cope with the changes must be employed. We show that skin detection in video can be enhanced by exploiting the knowledge of the range of possible skin colors for the camera used. In normalized color coordinates this range has a distinct shape we call the skin locus. We developed an adaptive histogram backprojection technique where the skin color model is updated by pixels in the search region which fall in the skin locus. We demonstrate increased detection capability with webcam videos of faces taken successively under daylight, incandescent lamp, fluorescent light and a combination of these light sources.

1. Introduction

Skin color may not be enough for tracking faces and hands in video but it is often the first cue sought. Skin color is distinct and since color is a low level attribute of the image, any operation done with color will be fast. Afterwards, skin pixel candidates can be further processed to get shape, texture or motion cues. However color changes as the current illumination differs from the one used in camera calibration and failures are certain if a fixed skin model is used.

One solution may be to correct for colors before processing - a body of work in color science known as color constancy [2], [3]. Certain color constancy algorithms effect a global change on the whole scene whereas in many real-life cases, there can be localized illumination variations. Objects like skin can appear to cameras to have multiple colors as a result for example, when a person is indoors but near the window, part of the face can be illuminated by daylight and the other part by fluorescent light. The retinex algorithm [6], [1] can work under multiple illuminants but like most color constancy algorithms, it can only work well when prior assumptions are satisfied, for example, that the scene is colorful enough (gray world assumption), and that the illumination is slowly and smoothly varying. Violation of these requirements may lead to unstable or grayed out colors.

Raja, McGenna and Gong [9] demonstrated that color model adaptation is another solution to the problem of varying illumination in tracking. They modelled their object colors as a Gaussian mixture and recursively adapted the mean, covariance and prior probabilities of each Gaussian cluster using color from a subimage within the bounding box.

In this work we try to make skin pixel detection cope in unconstrained illumination conditions instead of recovering colors. We take the approach of adapting the skin color model but unlike in [4] and [9] we make no assumptions of the form of the distribution. Instead we use the actual skin histogram as a nonparametric skin color model and employ histogram backprojection [11] to label skin pixel candidates. A key feature in our approach is that we use the knowledge of possible colors that skin can take under different illumination conditions. It has been observed that the set of all possible skin chromaticities in normalized color coordinates follow a shell-shaped structure which hence we will call the skin locus [10]. For the camera used we measured and modeled this cluster. To adapt, we update the histogram with pixels belonging to the skin locus from the bounding box found by a face tracking algorithm. Unlike in [12] the technique is robust to scaling and does not require a frontal face.

Mobile video communications is one target area for this technique. Since laptops can now have built-in webcams and third generation mobile phones may have video cameras, the mobility of these tools mandates that face processing techniques be illumination invariant. By using this technique for labelling skin pixel candidates, existing tracking techniques can also be made more robust. The skin locus

0-7695-0750-6/00 $10.00 © 2000 IEEE

and histogram adaptation is explained in Section 2. Section 3 gives the experiments and results and Section 4 concludes the paper.

2. Method

2.1 Shape of the skin locus

Measurements of skin reflectance, light spectral power distribution and camera channel sensitivities allow the computation of ideal RGB values for different skin types [7]. Conversion to normalized color space *(r, g)* chromaticity where $r = R/(R+G+B)$, $g = G/(R+G+B)$ reduces brightness dependence. Actual measurements have shown that dark, yellowish and pale skin have almost the same chromaticity [5], [10]. It has also been shown in [10] that for a camera white balanced for one illuminant, the chromaticity of skin follows a curve similar to the Planckian locus when imaged under light sources of different correlated color temperatures. In real situations, more than one light source may be shining on the face at a time. If on one side of the face there is shining one light source and on the other side another light source with different spectral content then in the image of the face the chromaticity vary between chromaticities caused by these two sources.

2.2. Adaptive histogram backprojection

An initial estimate of the skin model is the 2-D histogram $S(r,g)$ obtained from cut-out skin regions in the face. The frame to be segmented is transformed into rg-space and each pixel p_i with chromaticity (r_i, g_i) is assigned the value of the histogram at $(r_i, g_i), S(r_i, g_i)$. A variation is to use a ratio histogram $R(r,g)$, which is $S(r,g)$ divided by the whole image histogram $I(r,g)$ to penalize colors which are not part of the model or are present also in the background, thus increasing the contrast between skin and background pixels. We use ratio histograms for skin model in our experiments. A parametric model for skin color (e.g. Gaussian fitting) tends to smoothen the actual distribution and requires a distance metric (e.g. Mahalanobis) to assign skin probability. With ratio histogram and histogram backprojection, no fitting is necessary because the histogram itself is used as the model, and probabilities are assigned by simple table lookup, thus leading to a faster labeling.

Further processing may then be done on the backprojected image. Let a tracking algorithm take as input the backprojected image and output the bounding box around the face. If illumination conditions cause the apparent skin color to change then the current model will only produce few lighted pixels in the face region unless the skin model is updated. We propose to do adaptation by determining pixels in the bounding box which belong to the skin locus and to use these pixels to compute the current ratio histogram. This way the skin pixels are chosen automatically.

2.3. Tracking

A simple tracking scheme was used to indirectly test the efficacy of histogram adaptation. The backprojected image is thresholded and a majority 3x3 filter is applied (if there are 5 or more ones in the window then middle pixel is set to one). The largest connected component is then assumed to be face and a bounding box is drawn around it. Skin pixels are then extracted from the bounding box using the skin locus constraint and ratio histogram is updated using the histogram of these pixels and the histogram of the whole image.

The bounding box is then expanded proportionally to define a search region where the face will likely be. Histogram backprojection is perfomed on this region and the tracking and adapting process repeats. Tracking is suspended if no clusters are found (for example, when no reliable color information is available such as when the image is too dark or too bright). In this case the old bounding box is used until clusters are found again.

3. Results

3.1 Skin locus

Still images of faces under different illuminants (CIE D65 (daylight), incandescent lamp, TL84 fluorescent, and horizon sunlight) and camera white balancing conditions where captured with a 1CCD QuickVideo webcam, following the procedures in [7]. Sections of skin were cut out and their normalized chromaticities where plotted in rg-space. Skin color ranged from red to orange to blue. The set of possible skin colors seen by the camera is shown in Figure 1. A simple model for this locus of points is a pair of intersecting quadratic functions defining the upper bound and the lower bound of the cluster.

For each r, the maximum and minimum g was used to estimate the upper and lower quadratic functions respectively. Note however that the white point ($r = g = 0.33$) is completely within the locus model. To avoid selecting whitish pixels, a circle with a radius of 0.02 about the white point is drawn in the locus. A pixel is labeled as skin-candidate if it falls within the locus but outside the circle. For the upper bound, the quadratic coefficients found are A_{up}=-1.8423,b_{up}=1.5294 c_{up}=0.0422 while the lower bound coefficients are A_{down} = -0.7279 b_{down}=0.6066 c_{down}=0.1766. Skin pixels are determined by

$$Skin(r,g) = \begin{cases} 1 & (g < g_{up}) \cdot (g > g_{down}) \cdot (W_r > 0.004) \\ 0 & otherwise \end{cases} \quad (1)$$

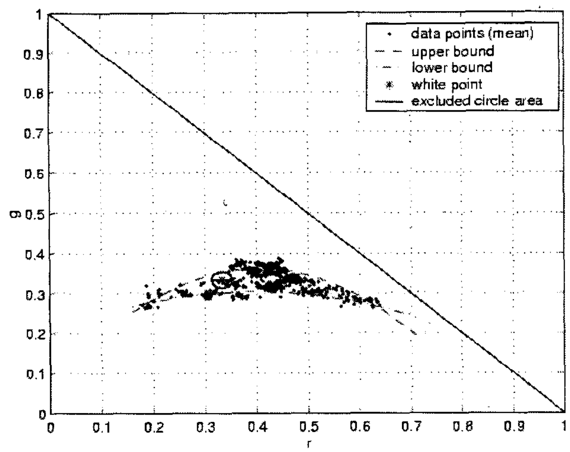

Figure 1. Skin locus of QuickVideo camera.

where $W_r = (r - 0.33)^2 + (g - 0.33)^2$, $g_{up} = A_{up}r^2 + b_{up}r + c_{up}$, and $g_{down} = A_{down}r^2 + b_{down}r + c_{down}$.

The skin locus occupies 3.2% of the total area of the chromaticity space and defines the range of possible skin color. The ratio histogram on the other hand, defines skin color probability and at any one time, occupies a subset of the locus.

3.2. Skin detection and tracking

Figure 2 shows some frames from a tracking movie made to show different skin color. The camera was mounted on a laptop and was allowed to white balance for an incandescent source (color temperature 2856 K) in a dark room before automatic white balancing was turned off. The subject carried the laptop from the dark room out into the office room where there was fluorescent light. The subject moved around the room while the light in the dark room was switched to D65 (bluish light) and the subject returned. The three leftmost frames show the skin color changing from normal flesh tones to yellowish orange to bluish. We assume that a perfect bounding box is found by using manually selected ground truth bounding boxes as the adaptation window. The second column shows the skin labelling results without adaptation and the third column shows with adaptation. Figure 3 displays the adapted ratio histogram. With adaptation even bluish skin pixels are labelled correctly.

Figure 4 and Figure 5 shows some tracking results for the same movie in Figure 2. Around frame 300 the face became too dark and the tracker froze. Tracking using a fixed skin model (Figure 4) was not able to recover when sufficient light was again present because the skin color has significantly changed. On the other hand, tracking with adaptation

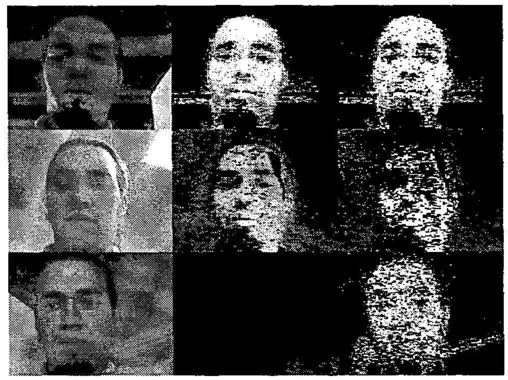

Figure 2. Skin labeling result. 1st column: original frames, (top to bottom, frames 100, 200, 400), 2nd: detected pixels with fixed ratio histogram, 3rd: detected pixels with adaptive ratio histogram.

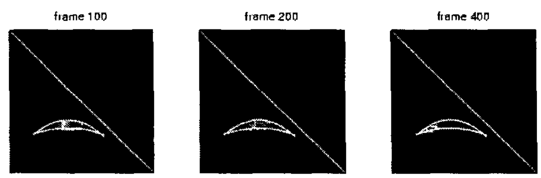

Figure 3. Skin ratio histograms from frames 100, 200 and 400.

Figure 4. Face tracking result no adaptation. Inner box is bounding box found by tracking algorithm. Outer box is search region. Frame numbes left to right, top to bottom are 110, 185, 285, 385, 435, and 460. The last three images show that the tracker has frozen.

using the skin locus recovered (Figure 5). Tracking results may be found at http://www.ee.oulu.fi/research/imag/color/.

4. Discussion and conclusions

A method to select training pixels for adapting skin model histogram for skin pixel labelling by backprojection is discussed. A knowledge of the range of possible skin colors observed with the camera used under different illumination conditions is used as a criteria to select pixels in the tracking bounding box for updating the histogram. In [8] the same idea is presented: possible changes in a color patch are limited in a region in chromaticity space. In our case, we calibrated the camera under four different simulated illuminants and have three illumination changes for each calibration. This allows us a wider operation range where it is possible to adapt to the color change. In [8], a clustering algorithm is used to find cluster representatives in the chromaticity space for both object and background, our method is simpler because we only have to define the cluster boundaries and the knowledge of background is not needed. Finally, our work is different in that our approach is a low level operation, filtering, while in [8] is a high level operation, recognition.

Adaptive skin color modelling using the skin locus is shape and scale invariant and is shown to improve tracking even under mixed illumination. As with other updating schemes, if unchecked the histogram can adapt to other objects with skin color. We have observed this failure in our experiments too, but noticed that it has dependence on the quality of the bounding box.

The locus is of course camera-specific. It may be found experimentally or, in principle, may be calculated. A database of illuminants, skin spectral reflectance and a knowledge of the camera sensitivities (for example, supplied by the manufacturer) will allow the user to compute the camera skin locus.

(M. Soriano's permanent address is with the National Institute of Physics, University of the Philippines, 1101 Diliman, Quezon City).

Figure 5. Face tracking with adaptation for frames 110, 185, 285, 385, 435, and 460. Last three images show that the tracker was able to recover.

References

[1] D. Jobson, Z. Rahman, and G. Woodell, "A multi-scale retinex for bridging the gap between color images and the human observation of scenes", *IEEE Trans. Image Processing*, 6(7):965-976, 1997.

[2] G. Finlayson, B. Funt, and K. Barnard, "Color constancy with shadows", *Perception*, 23:720-725, 1995.

[3] D. Forsyth, "A novel algorithm for color constancy", *Intl. J. Computer Vision*, 5:5-36, 1990.

[4] W.-C. Huang, and C.-H. Wu, "Adaptive color image processing and recognition for varying backgrounds and illumination conditions", *IEEE Trans. Industrial Electronics*, 45(2):351-357, 1998.

[5] M. Hunke, and A. Waibel, "Face locating and tracking for human-computer interaction", *Proc. 28th Asilomar Conf. Signals, Systems and Computers* 2:1277-1281, 1994.

[6] E. H. Land, "Recent advances in retinex theory", *Vision Research*, 26:7-21, 1986.

[7] E. Marszalec, B. Martinkauppi, M. Soriano, and M. Pietikäinen, "Physics-based face database for color research", *J. Electronic Imaging*, 9(1):32-38, 2000.

[8] J. Matas, R. Marik, and J. Kittler, "Illumination invariant colour recognition", in E. Hancock(ed.) *Proc. British Machine Vision Conference*, 469-479 1994.

[9] Y. Raja, S. McKenna, S. Gong "Tracking colour objects using adaptive mixture models", *Image and Vision Computing*, 17(3-4):225-232, 1999.

[10] M. Störring, H. Andersen, E. Granum, "Skin colour detection under changing lighting condition", H. Araujo and J. Dias (ed.) *7th Symposium on Intelligent Robotics Systems*, 187-195, 1999.

[11] M. Swain, D. Ballard "Color indexing", *Intl. J. Computer Vision* 7(1):11-32, 1991.

[12] T.-W. Yoo, and I.-S. Oh, "A fast algorithm for tracking human faces based on chromatic histograms", *Pattern Recognition Letters* 20:967-978, 1999.