COLOUR NORMALISATION BASED ON BACKGROUND INFORMATION

Maria Vanrell, Felipe Lumbreras, Albert Pujol, Ramon Baldrich, Josep Lladós, J.J. Villanueva

Computer Vision Center
Dept. d'Informàtica, Universitat Autònoma de Barcelona
Edifici O, Campus UAB - 08193 Bellaterra, Barcelona)

ABSTRACT

This paper proposes an improvement on a well-known colour normalisation by the introduction of some knowledge on background. Comprehensive normalisation gives an invariant representation of the image colour. This invariant representation can be considered a canonical representation whenever image content is preserved and changes are only due to illuminant conditions. One of the steps of the normalisation is based on the grey-world normalisation that removes colour changes on each channel. Because a diagonal model is assumed, the independence of chromatic variations is also achieved if the channel normalisation is applied only with background mean in spite of image mean. This will allow to remove illuminant effects meanwhile no influence from the foreground is introduced on the normalised coordinates. It will provide an almost canonical colour space without an explicit estimation of the scene illuminant.

1. INTRODUCTION

In this paper we present a result towards a solution to deal with variations on colour image representations. Any computer vision method that intends to use colour as a visual cue has to consider its variability. Changes on scene conditions and non-controlled factors on image acquisition devices are some causes for this problem.

Human visual system (HVS) shows an excellent behavior dealing with colour changes, colour statements given by the HVS seem to be independent to those factors that can change colour perception, this ability has been called colour constancy.

To simulate the colour constancy ability of the HVS some colour constancy methods have been developed [1, 2, 3]. Their goal is to estimate the scene illuminant in order to be able to remove his effect and to construct a canonical representation of the image colour [4]. Other approaches try to deal with this colour constancy problem without estimating illuminant but computing a normalised representation of the image [5]. In this paper we bring together both approaches.

Work partially supported by projects FEDER 2FD97-1800 and CICYT TIC2000-0382

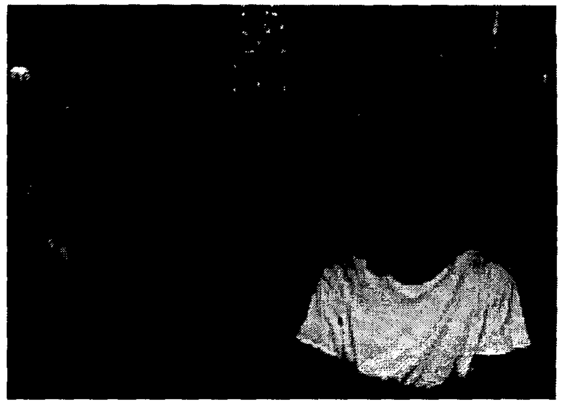

Fig. 1. A typical scene taken by the surveillance system. It shows a static background and some people coming to the counter to be registered.

Our proposal is made in the framework provided by a real application. It is a surveillance system devoted to store on a database the people appearance description. The final goal is to get all the people data given their appearance description.When people is registering to get into a building all their data is entered jointly with an image taken by the surveillance system. We can see an example of these images at figure 1. The computer vision module of this surveillance system is in charge of storing the appearance description built from the image as natural language annotations, that is, hair colour, clothes configurations, glasses, etc. Afterwards, the security staff will be able to access to people data by indexing from appearance.

This application has two crucial steps, the background subtraction to separate people (foreground) from the rest of the scene, and the segmentation of homogeneous colour regions. Colour regions perform a crucial role in constructing descriptions, since they are used to describe skin, hair and clothes. The colour segmentation step is held on the outputs of a colour naming module [6].

Both background subtraction and colour segmentation held on a common assumption on the images, that is, there is a constant region on the images, the background, for which only small differences on structure and illumination conditions occur along time. Background subtraction method models the background structure independently of illumination changes, and it allows to get an improved colour normalisation of the image that behaves as an illuminant-independent representation of colour without estimation of scene illuminant.

In the next section we will give a brief description of the background extraction method. Afterwards, we introduce the normalisation applied to colour representation and in section 4 we present our proposal on how to improve the normalisation based on the background knowledge. Finally, we illustrate the results of this proposal and give some conclusions of this work.

2. BACKGROUND SUBTRACTION

Given a sequence of images acquired at the same scenario, $\mathbf{I} = \{I_1, I_2, \ldots, I_n\}$, the goal of any background subtraction methods is to label their image pixels into : **i)** regions where the underlying scenario is visible (background), and **ii)** image regions where the scenario is being occluded (foreground), that in our application it is related to arriving people. The proposed background subtraction method [7] assumes that all the background points will present a chromatic stability all over the background image regions, which is always represented by a diagonal model (i.e, this model is able to represent small illuminant variations that can occur along time).

Some previous works (Pfinder [8], W4 [9]) broach this problem in a statistical way. These methods usually take a set of background (where the whole scenario is visible) reference images $\mathbf{B} = \{B_1, B_2, \ldots, B_m\}$. These images are used to estimate the probability density function (PDF) that an image pixel belongs to background. The PDF applied in these methods encompass in the same model the variations of the background pixels due to illumination changes and acquisition noise.

The application of these methods to our problem produces an over-relaxed background model, where slight changes on illumination produce the undesirable effect of increasing the number of possible foreground values being classified as background. To avoid this problem, the method applied in this work [7], splits the above methodology into two steps: **i)** An statistically robust estimation of the global linear transformation that adjusts the image to a noise free reference background image. **ii)** a statistical noise model that allows us to decide if the differences between the pixel values of a corrected image and those of the reference background model are expected for a background pixel.

Assuming zero mean gaussian noise, the reference background image $M(x,y)$ is then constructed averaging a set of background images $\{B_1(x,y), \ldots, B_l(x,y)\}$. Note that both the reference image, as well as the background images used to compute it, are automatically updated each time the elements of the scenario are rearranged. Then, assuming the above introduced model, the pixels of the background region of an image $I(x,y)$, can be expressed as:

$$I(x,y) = M(x,y) \begin{pmatrix} \alpha & 0 & 0 \\ 0 & \beta & 0 \\ 0 & 0 & \gamma \end{pmatrix} + \epsilon(x,y) \qquad (1)$$

where $\epsilon(x,y) = (\epsilon_R(x,y), \epsilon_G(x,y), \epsilon_B(x,y))$ is a zero mean gaussian term, and (α, β, γ) is the parameters whose existence is assumed by the diagonal model of the chromatic properties of the image background in our application. This makes that these parameters can be computed by a voting process on the solution of the equation systems solved for each pixel [7].

Once the linear adjustment has been determined, It is applied to the image $I(x,y)$, obtaining a new image $\tilde{I}(x,y)$. Ideally, the RGB pixels values of the background image region $\tilde{I}_{bck}(x,y)$ of $\tilde{I}(x,y)$ and the ones of the reference background image $M(x,y)$, now only differ in the zero mean gaussian noise term $\epsilon(x,y)$ of Eq.1.

This adjustment can also be applied to the set of background images, $\{B_1(x,y), \ldots, B_l(x,y)\}$, used. The resultant $\{\tilde{B}_1(x,y), \ldots, \tilde{B}_l(x,y)\}$ images, are then used to fit a gaussian model of the probability of obtaining a certain noise error vector $P_B(\epsilon(x,y))$ given that the (x,y) pixel belong to the background class. In this way, the background region for an image $I(x,y)$, can be determined as those image pixels for which the probability of its estimated $\epsilon(x,y)$ vector, is bigger than a certain threshold.

$$I(x,y) \in \tilde{I}_{bck} \iff P_B(\epsilon(x,y)) > thr \qquad (2)$$

3. COLOUR NORMALISATION

Once the background has been substracted from the image, next step is based on a colour naming module that is the basis for all the subsequent high level interpretations.

A robust colour naming system has to preserve independence from illuminant variations. Colour normalisation is a common approach to deal with this invariance property [5]. We have based this work on the comprehensive colour normalisation [10]. This normalisation is based on the iterative application of two normalisations on the image I, whose pixels are denoted by the tuples (r, g, b):

1. A lighting geometry normalisation, denoted by R

$$R(r,g,b) = \frac{1}{r+g+b}(r,g,b) \qquad (3)$$

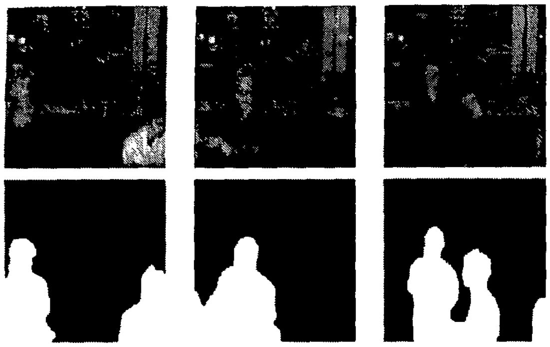

Fig. 2. Some results of the foreground segmentation. Upper line shows the original images, lower shows in white the foreground and in black the background.

this pixel normalisation allows to remove changes due to differences on the magnitude of the colour vector given by a scalar, s, that is

$$R(r,g,b) = R(sr,sg,sb) \quad (4)$$

2. A colour illuminant normalisation, denoted by C

$$C(r,g,b) = \left(\frac{rN/3}{\sum_{k=1}^{N} r_k}, \frac{gN/3}{\sum_{k=1}^{N} g_k}, \frac{bN/3}{\sum_{k=1}^{N} b_k}\right) \quad (5)$$

where r_k, g_k and b_k represent the k-th pixel value of an image I of N pixels. This is a channel normalisation that removes chromatic changes given by three scalar values, α, β and γ, hence

$$C(r,g,b) = C(\alpha r, \beta g, \gamma b) \quad (6)$$

where a linear diagonal model is assumed for colour changes [11].

these two normalisations are computed until idempotence is achieved. The final result is independent of the order in which these two steps are applied.

The obtained normalised coordinates, (nr, ng, nb), satisfy the following expressions

$$nr + ng + nb = 1 \quad (7)$$

$$\sum_{k=1}^{N} nr_k = \frac{N}{3}, \sum_{k=1}^{N} ng_k = \frac{N}{3}, \sum_{k=1}^{N} nb_k = \frac{N}{3} \quad (8)$$

4. NORMALISATION ON BACKGROUND POINTS

Comprehensive normalisation removes from the image all the illuminant effects. Normalised coordinates of an image pixel depends on its initial RGB values and on the image content. The influence of the image content is introduced by the channel normalisation (5), that divides each image pixel by the channel mean.

In our application, when someone appears in front of the registration desk, that is, a part of the background is occluded by this new foreground, the channel means vary depending on the colour values of the new foreground. Until now we have assumed that the ratio between the foreground and background areas favours in much to the background, this let us assume that we have small differences on the channel means, and normalised coordinates will not vary a lot from one image to another. If it is, then we can use normalised coordinates as in a canonical colour space. But, still there are small variations.

These small variations are not very important when we try to give name to colours with large areas on the chromatic diagram, but they become more significant in giving names to colours on small chromatic areas, as is the case of the skin colour.

In section 2 we have seen how to compute what pixels of an image belong to the background, we will denote this image region as I_{bck}. Then normalised coordinates can be computed avoiding undesirable effects of the foreground areas. This imply to introduce background information on the channel normalisation. Like in equation (6) the coefficients of the illuminant changes will be removed. Therefore, we will redefine the channel normalisation, C as C', it is shown in the proposed algorithm:

1. For a given image I composed by two regions: I_{bck} and I_{frg}, given by the background subtraction method, $N(I)$ is initialized by I.
2. Repeat
   (a) For every pixel (r,g,b) of $N(I)$, the modified channel normalisation, C', is computed:
   $$N(I)(r,g,b) := C'(r,g,b)$$
   where
   $$C'(r,g,b) = \left(\frac{N'/3r}{\sum_{k=1}^{N'} r'_k}, \frac{N'/3g}{\sum_{k=1}^{N'} g'_k}, \frac{N'/3b}{\sum_{k=1}^{N'} b'_k}\right)$$
   and (r'_k, g'_k, b'_k) are points of $N(I_{bck})$ and N' is the number of pixels belonging to I_{bck}.
   (b) For every pixel (r,g,b) of $N(I)$, the lighting normalisation, R, is computed:
   $$N(I)(r,g,b) := R(r,g,b)$$
3. Until Idempotence of $N(I_{bck})$ is achieved or $|N(N(I)) - N(I)|$ is small enough.

Convergence and uniqueness is only accomplished for the background points as it is in the comprehensive normalisation, also constraint 7 is assured for the whole image. On the other hand, the goal of removing the illuminant change coefficients is attained and in practice the modified algorithm always converge very quickly.

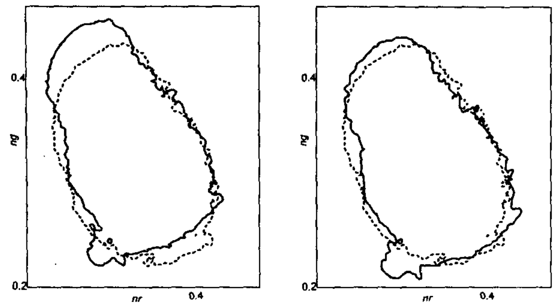

Fig. 3. Example of the stability introduced by the background consideration on colour normalisation. Dashed lines represent a gamut projection of a normalised background image. Normal lines represent the gamut of a normalised image with foreground, on the left the result with the comprehensive normalisation, on the right the result with the our proposed algorithm.

5. RESULTS AND CONCLUSIONS

In figure 3 we show an example of the stability introduced by the proposed method (BCCN) in comparison with the results obtained using the comprenhensive color normalisation (CCN). Left image of this figure shows the displacement between the projected gamuts of two images, one is a normalised image of a background scene, B (in dashed lines) with nobody in front of the register desk, that is, without foreground, and the second one corresponds to a normalised image with foreground, I (in solid line). Right image show the same images when the BCCN normalisation has been applied on the second image, I. The BCCN normalisation has the same effect than CCN when the image is B, a simple background.

From this figure, we can see a noteworthy improvement in the stability of the gamut computed using the proposed modification to the NCC.

In order to do an assessment of the improvement a measure of the displacement between both gamuts distributions has been computed using normalised correlation. This measure has been repeated for a set of 56 different images. The averaged correlation obtained using the CNN, and the proposed method are respectively 0.47 and 0.78, which is an improvement of 30% . These results suppose that the color-naming confusion is reduced.

The work presented in this paper shows how knowledge on image content can be considered on a colour normalisation in order to give colour constancy ability to an intelligent system that works in real conditions.

6. REFERENCES

[1] L.T. Maloney and B.A. Wandell, "Colour constancy: a method for recovering surface spectral reflectance," *Journal of the Optical Society of America*, vol. 3, no. 1, pp. 29–33, 1986.

[2] D.A. Forsyth, "A novel algorithm for color constancy," *International Journal on Computer Vision*, vol. 5, no. 1, pp. 5–36, 1990.

[3] Graham D. Finlayson, S.D. Hordley, and P.M. Hubel, "Colour by correlation: A simple, unifying approach to colour constancy," in *Proceedings of the ICCV*, 1999, vol. 2, pp. 835–842, New York.

[4] B.V. Funt, K. Barnard, and L. Martin, "Is machine colour constancy good enough?," in *Proceedings of the 5th ECCV '98*, 1998, pp. 445–459.

[5] Graham D. Finlayson and Gui Y. Tian, "Perfect colour constancy vs colour normalisation for object recognition," in *Proceedings of the Europto, SPIE Series*, 1999, vol. 3826, pp. 195–203, Munich, Germany.

[6] R. Benavente, G. Sànchez, R. Baldrich, M. Vanrell, and J. Lladós, "Normalised colour segmentation for human appearance description," in *Proceedings of 15th ICPR*, 2000, vol. 3, pp. 637–641, Barcelona.

[7] A. Pujol, F. Lumbreras, X. Varona, and J.J. Villanueva, "Locating people in indoor scenes for real applications," in *Proceedings of 15th. ICPR*, 2000, Barcelona.

[8] C. R. Wren, A. Azarbayejani, T. Darrell, and A. P. Pentland, "Pfinder: Real-time tracking of the human body," *IEEE Trans. Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 780–785, 1997.

[9] I. Haritaoglu, D. Harwood, and L. S. Davis, "W4: Who? when? ehere? what? a real time system for detecting and tracking people," in *Proceedings of Third International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 222–227, Nara, Japan.

[10] G.D. Finlayson, B. Schiele, and J.L. Crowley, "Comprehensive colour image normalization," in *Proceedings of the 5th ECCV '98*, 1998, pp. 475–490.

[11] G.D. Finlayson, M.S. Drew, and B.V. Funt, "Diagonal transforms suffice for color constancy," in *Proceedings of the 4th ICCV '93*, 1993, pp. 164–171.