CAIP'2001, LNCS 2124, pp. 341-348, Springer. Warsaw, Poland, September 2001.

Face tracking using the Dynamic Grey World Algorithm*

José M. Buenaposada, David Sopeña, Luis Baumela

Departamento de Inteligencia Artificial, Universidad Politécnica de Madrid
Campus de Montegancedo s/n, 28660, Madrid, SPAIN
{jmbuena,dsgalindo}@dia.fi.upm.es, lbaumela@fi.upm.es

Abstract. In this paper we present a colour constancy algorithm for real-time face tracking. It is based on a modification of the well known Grey World algorithm in order to use the redundant information available in an image sequence. In the experiments conducted it is clearly more robust to sudden illuminant colour changes than popular the rg-normalised algorithm.

Keywords: Face Tracking, Grey World, Colour Constancy

1 Introduction

In this paper we will study the problem of face tracking using colour. Trackers based on this feature, which is the most frequently used feature for face tracking [14], are used as an initial estimate or follow-up verification of face location in the image plane.

The primary problem in automatic skin detection is colour constancy. The [RGB] colour of an image pixel depends not only on the imaged object colour, but also on the lighting geometry, illuminant colour and camera response. For example [6], if the scene light intensity is scaled by a factor s, each perceived pixel colour becomes $[sR, sG, sB]$. The rg-normalisation algorithm provides a colour constancy solution which is independent of the illuminant intensity by doing: $[sR, sG, sB] \mapsto [sR/s(R+G+B), sG/s(R+G+B)]$. On the other hand, a change in illuminant colour can be modelled as a scaling α, β and γ in the R, G and B image colour channels. In this case the previous normalisation fails. The Grey World (GW) algorithm [6] provides a constancy solution independent of the illuminant colour by dividing each colour channel by its average value.

In this paper we introduce a colour constancy algorithm, based on GW, that can be used for real-time colour-based image segmentation which is more robust to big sudden illuminant colour changes than the popular rg-normalised algorithm.

* Work funded by CICyT under project number TIC1999-1021

2 Grey World-based colour constancy

Colour constancy is the perceptual ability to assign the same colour to objects under different lighting conditions. The goal of any colour constancy algorithm is to transform the original $[RGB]$ values of the image into constant colour descriptors. In the case of Lambertian surfaces, the colour of an image pixel (ij) can be modelled by a lighting geometry component s_{ij}, which scales the $[rgb]$ surface reflectances of every pixel independently, and three colour illuminant components (α, β, γ), which scale respectively the red, green and blue colour channels of the image as a whole [6]. The lighting geometry component accounts for surface geometry and illuminant intensity variations, while the colour illuminant components account for variations in the illuminant colour. According to this model, two pixels $I(ij)$ and $I(kl)$ of an image would have the following $[RGB]$ values: $[s_{ij}\alpha r_{ij}, s_{ij}\beta g_{ij}, s_{ij}\gamma b_{ij}]$, $[s_{kl}\alpha r_{kl}, s_{kl}\beta g_{kl}, s_{kl}\gamma b_{kl}]$, where $[r_{ij}, g_{ij}, b_{ij}]$ and $[r_{kl}, g_{kl}, b_{kl}]$ represent surface reflectances; i.e. real object colour, independent of the illuminant.

The GW algorithm proposed by Buchsbaum [2] assumed that the average surface reflectances in an image with enough different surfaces is grey. So, the average reflected intensity corresponds to the illuminant colour, which can be used to compute the colour descriptors. This algorithm was refined in [8] by actually obtaining an average model of surface reflectances and proposing a procedure to compute the average image reflectance.

Let us define the *image average geometrical reflectance*, $\bar{\mu}$, as

$$\bar{\mu} = \frac{1}{n} \sum_{ij \in I} [s_{ij} r_{ij}, s_{ij} g_{ij}, s_{ij} b_{ij}],$$

where n is the number of image pixels. It represents the average $[RGB]$ image values, once we have eliminated the colour illuminant component.

If we assume that the average geometrical reflectance is constant over the image sequence, then the image average $[RGB]$ variation between two images is proportional to the illuminant colour variation. On the basis of this, a colour normalisation invariant to illuminant colour changes can be devised:

$$\left[\frac{I_r(ij)}{\frac{1}{n}\sum_{ij\in I} I_r(ij)}, \frac{I_g(ij)}{\frac{1}{n}\sum_{ij\in I} I_g(ij)}, \frac{I_b(ij)}{\frac{1}{n}\sum_{ij\in I} I_b(ij)}\right] = \left[\frac{s_{ij} r_{ij}}{\bar{\mu}_r}, \frac{s_{ij} g_{ij}}{\bar{\mu}_g}, \frac{s_{ij} b_{ij}}{\bar{\mu}_b}\right], \quad (1)$$

where, if x represents the colour channel ($x \in [r, g, b]$), $I_x(ij)$ is the value of the channel x for pixel $I(ij)$, and $\bar{\mu}_x$ is the image channel x average geometrical reflectance.

The previous normalisation is what we call basic GW algorithm. It is robust to illuminant colour variations, but it only works for sequences with constant image average geometrical reflectance. In consequence, basic GW fails when a new object appears in the image or when the illuminant geometry changes. In the next section we propose an extension to the basic GW algorithm that solves these problems using redundant information available in an image sequence.

3 Face tracking using Dynamic Grey World

In this section we present a colour-based face tracking algorithm. First we will briefly describe how to track a coloured patch using simple statistics, afterwards the Dynamic GW (DGW) algorithm is presented.

3.1 Face segmentation and tracking using a skin colour model

Given a sequence of colour images, building a face tracker is straight forward if we have a reliable model of the image colour distributions. Let I_{rgb} be the $[RGB]$ channels of image I, and let $p(I_{rgb}|skin)$ and $p(I_{rgb}|back)$ be the conditional colour probability density functions (pfds) of the skin and background respectively (we assume that background is anything that is not skin). Using the Bayes formula, the probability that a pixel with colour I_{rgb} be *skin*, $P(skin|I_{rgb})$, can be computed as follows:

$$P(skin|I_{rgb}) = \frac{p(I_{rgb}|skin)P_s}{p(I_{rgb}|skin)P_s + p(I_{rgb}|back)P_b},$$

where P_s and P_b are the a priori probabilities of *skin* and *background*. The transformation $\mathcal{T}(I_{rgb}) = 255 \times P(skin|I_{rgb})$ returns an image whose grey values represent the probability of being skin (see first column in Fig. 3). Face tracking on this image can be performed with a mode seeking algorithm, like [4], by computing the position and orientation of the face colour cluster in each frame [3].

The problem now is how to make the previous statistical model invariant to variations in the scene illumination. In most real-time systems this invariance is achieved by working in a rg-normalised chromaticity space. As we previously mentioned, this method fails when there is a sudden change of the illuminant colour. In our segmentation algorithm we propose using the GW colour space, $[\hat{r}\hat{g}\hat{b}]$, defined in section 2:

$$I_{\hat{r}\hat{g}\hat{b}}(ij) = n \times [\frac{I_r(ij)}{\sum_I I_r} \frac{I_g(ij)}{\sum_I I_g} \frac{I_b(ij)}{\sum_I I_b}].$$

We model the *skin* GW colour distribution with a continuous Gaussian model. As can be seen in Fig. 1, $p(I_{\hat{r}\hat{g}\hat{b}}|skin)$ is approximately Gaussian. On the left are shown the Chi-square and Gaussian plots of the $I_{\hat{r}}$, $I_{\hat{g}}$ and $I_{\hat{b}}$ marginals and the $I_{\hat{r}\hat{g}\hat{b}}$ multivariate distribution. From the analysis of these plots we can verify that the assumption $p(I_{\hat{r}\hat{g}\hat{b}}|skin) \sim N(\bar{m}_s, \Sigma_s, I_{\hat{r}\hat{g}\hat{b}})$ can not be rejected. On the other hand, it is not possible to find an analytic model for the *background* pdf, so we will model it with a uniform distribution, $h_b(I_{\hat{r}\hat{g}\hat{b}})$. Other authors have indicated different preferences for modelling the colour distributions. In [11] Gaussian mixture models, whereas in [5] and [12] pure histogram-based representations are chosen. In our experiments we found that using a a continuous model yields better results because of the high space dimensionality (3D).

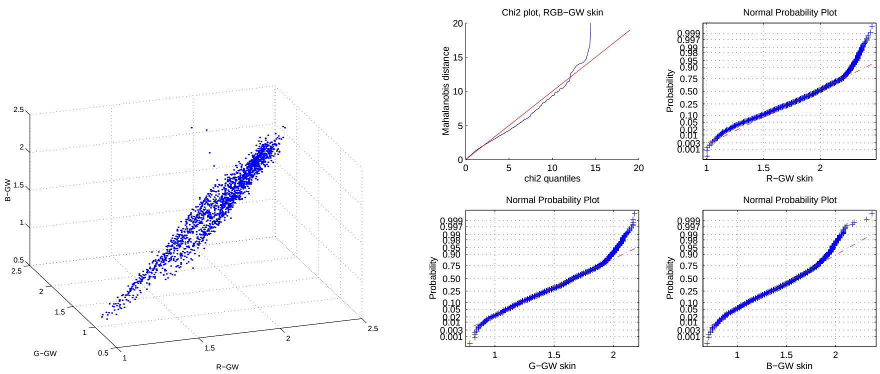

Fig. 1. Skin colour pdf in GW space. On the left is shown the skin colour cluster in GW colour space. On the right are shown the Chi-square plot for the multivariate distribution and the Normal plots for the marginals.

If we approximate the priors $P_s \approx n_s/n$ and $P_b \approx n_b/n$, where n_s and n_b are respectively the number of the *skin* and *background* pixels, then

$$P(skin|I_{rgb}) = \frac{n_s\, N(\bar{m},\, \Sigma, I_{rgb})}{n_s\, N(\bar{m},\, \Sigma, I_{rgb}) + n_b\, h_b(I_{rgb})}.$$

3.2 The Dynamic Grey World algorithm

The main problem of the basic GW algorithm is that it was conceived for static images; i.e. it fails when there is a big change in the image average geometrical reflectance. In this section we propose a dynamic extension to GW (DGW) which will detect this situation and update the GW model (see Fig. 2).

In the following we assume that there exists a partition of the image sequence into a set of image subsequences such that the image average geometrical reflectance is constant over each subsequence; i.e. the basic GW algorithm can be used as a colour constancy criterion over each subsequence. We will use the first image of each subsequence as a *reference image*. The other images of the subsequence will be segmented using the colour descriptors of the reference image.

Let I^r_{rgb}, I^t_{rgb} and I^{t-1}_{rgb} be respectively the reference image, the present and the previous image, $F^r_{\hat{r}\hat{g}\hat{b}}$ be the face pixels in GW space, $\bar{\mu}^{I^t}_{rgb}$ be the average value for each colour channel in I^t_{rgb}, $\bar{\mu}^{F^r}_{\hat{r}\hat{g}\hat{b}}$ and $\bar{\mu}^{F^t}_{\hat{r}\hat{g}\hat{b}}$ be the average GW descriptors for the face pixels in the reference and present image, and $\bar{m}_s$, Σ_s, h_b, P_s, P_b be the GW colour descriptors statistical distribution for the reference image.

The problem now is how to segment each reference image and how to detect a change of subsequence. Reference images can be segmented with the average [RGB] values of the previous image ($\bar{\mu}^{I^{t-1}}_{rgb}$), provided that the change in average

Initialisation
  /**Initialise the reference image model using motion segmentation and a precalculated colour model**/
  $[\bar{m}_s, \Sigma_s, P_s, P_b, \bar{\mu}^{F^r}_{\hat{r}\hat{g}\hat{b}}]$ = InitTracking();
While (true)   /* *tracker main loop* */
  $\bar{\mu}^{I^t}_{rgb}$ = Mean(I^t_{rgb}); /* *image mean rgb values* */
  $I^t_{\hat{r}\hat{g}\hat{b}} = \frac{I^t_{rgb}}{\bar{\mu}^{I^t}_{rgb}}$ /* *GW normalisation* */
  $F^t_{\hat{r}\hat{g}\hat{b}}$ = ProbabilisticSegment($I^t_{\hat{r}\hat{g}\hat{b}}, \bar{m}_s, \Sigma_s, P_s, P_b$); /* *segment img* */
  $\bar{\mu}^{F^t}_{\hat{r}\hat{g}\hat{b}}$ = ComputeAvgFaceGW($F^t_{\hat{r}\hat{g}\hat{b}}$); /* *face avg GW descriptors* */
  If $\|\bar{\mu}^{F^r}_{\hat{r}\hat{g}\hat{b}} - \bar{\mu}^{F^t}_{\hat{r}\hat{g}\hat{b}}\| > \Delta$ then /* *change of subsequence* */
    $I^t_{\hat{r}\hat{g}\hat{b}} = \frac{I^t_{rgb}}{\bar{\mu}^{I^{t-1}}_{rgb}}$ /* *GW normalise with previous mean* */
    $F^t_{\hat{r}\hat{g}\hat{b}}$ = ProbabilisticSegment($I^t_{\hat{r}\hat{g}\hat{b}}, \bar{m}_s, \Sigma_s, P_s, P_b$); /* *segment image* */
    $I^r_{\hat{r}\hat{g}\hat{b}}$ = $I^t_{\hat{r}\hat{g}\hat{b}}$ /* *update reference image* */
    $\bar{\mu}^{F^r}_{\hat{r}\hat{g}\hat{b}}$ = ComputeAvgFaceGW($F^t_{\hat{r}\hat{g}\hat{b}}$); /* *face avg GW descriptors* */
    $[\bar{m}_s, \Sigma_s, P_s, P_b]$ = ColourDistrib($F^t_{\hat{r}\hat{g}\hat{b}}$); /* *ref. colour distrib* */
  end /* *if* */
end /* *while* */

Fig. 2. Dymanic Grey World Algorithm

geometrical reflectance is caused mainly by the appearance of new objects in the scene.

A change of subsequence is detected just by detecting a change in the average geometrical reflectance. This can not be accomplished on the basis of analysing $\bar{\mu}^I_{rgb}$, as $\bar{\mu}^I_{rgb}$ also changes with the illuminant colour. We solve this problem by monitoring the average GW descriptors of the face pixels. As they are invariant to illuminant colour changes, a change in these descriptors is necessarily caused by a change in average geometrical reflectance.

4 Experiments

In our experiments we used a VL500 Sony colour digital camera at 320×240 resolution, iris open, no gain, no gamma correction. Images were taken with regular roof fluorescent lights and variations in illumination colour were obtained using a controled tungsten light, a green color filter, and turning on and off roof fluorescent lights.

In the first experiment we validate the DGW algorithm hypothesis: variations in the average geometrical reflectance can be detected, and the reference image of each subsequence can be segmented. We acquired a sequence of 200 images with a green object appearing at one point and illuminant geometrical variations taking place at different moments. The result of this experiment is shown, from left to right, in Fig. 3: the first image of the sequence (image 1), a change in the illuminant (roof lights turned off) (image 26), and the appearance and disappearance of an object (images 88 and 139). In this experiment the system detects three subsequences (1 to 87, 88 to 138, and 139 to 200). This is clearly visible in the plot at the bottom of Fig. 3. In image 26 the roof fluorescent lights are turned off. This geometrical illumination variation can be perceived again in the face GW desciptors plot. In this case the segmentation is good. This is an example of "worst case" test. In similar situations with stronger variations in the illuminant geometry, the system may not be able to segment the image and eventually may loose the target.

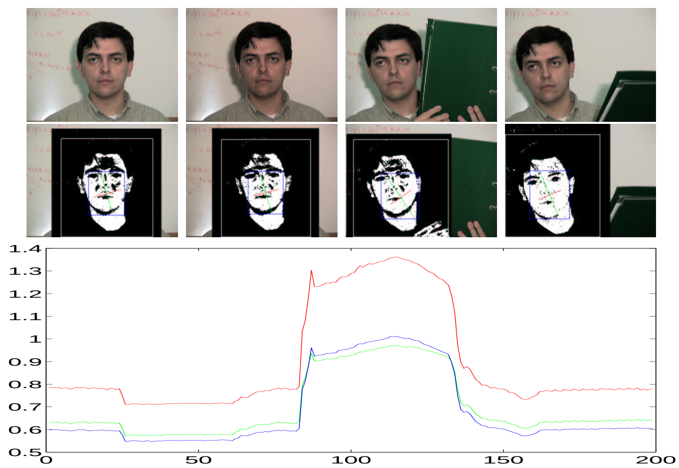

Fig. 3. Hypothesis validation experiment. On the first row four images of a sequence are shown. Their segmentation with the DGW algorithm is presented on the second row. The average r,g and b face GW descriptors (in red, green and blue color respectively) are shown in the third row.

The goal of the next experiment is to check that the dynamic extension to GW is necessary; i.e. to see what would happen if we segment the previous sequence with the basic GW algorithm. In Fig. 4 are shown the same images as in Fig. 3. We can clearly perceive that without the dynamic extension, the initial colour model is invalid when a change in the image average geometrical reflectance (caused by the appearance of an object) takes place. The initial model gradually becomes valid again as the object disappears (see last column).

In the following experiment we compare the performance of the DGW algorithm with the rg-normalised algorithm. We have created a sequence with a set of images with "difficult" background (i.e. brownish door and shelves to distract the segmentation). In Fig. 5 four frames of the sequence are shown in each colum representing: initial image, red object appears, tungsten frontal light turns on,

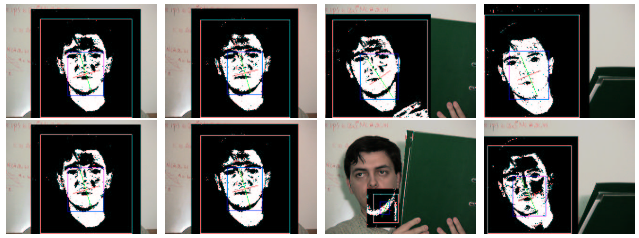

Fig. 4. DGW algorithm versus basic GW. DGW algorithm segmentation results are shown in first row and basic GW in the second one.

green filter is introduced. DGW segmentation results are shown in the second row and rg-normalised results in the third one. Visual inspection of these results show that both algorithms have similar results in the least favourable cases for the DGW algorithm (second and third columns) and a clear success of the DGW compared to the rg-normalisation when the illuminant colour abruptly changes (fourth column).

Fig. 5. Comparison of DGW and RG-normalisation colour constancy for face tracking.

5 Conclusions

We have introduced the Dynamic Grey World algorithm (DGW) a colour constacy algorithm based on an extension of the well known GW algorithm. It was designed to work in real-time with sequences of images with varying environmental conditions. In the experiments conducted it performed better than the rg-normalised algorithm when sudden changes in the illuminant colour take place. The least favorable case for our algorithm occurs when changes in the illuminant geometry take place. In this paper we have analysed some of the weak points of the rg-normalised algorithm. The DGW algorithm is not perfect either, as its performance can be seriously affected by strong and fast changes in the illuminant geometry. In spite of these limitations, colour-based trackers are good as a fast initial estimate or follow-up verification of face location.

References

1. D. Berwick and S.W. Lee. A chromaticity space for specularity-, illumination color- and illumination pose invariant 3-d object recognition. *Proc. of the Int. Conf. on Computer Vision*. Bombay, India. 1998.
2. G. Buchsbaum. A spatial processor model for object colour perception. *Journal of the Fanklin Institute*, 310: 1-26, 1980.
3. G. Bradski. Computer Vision face tracking for use in a perceptual user interface. *Proc. of Workshop on applications of Computer Vision, WACV'98*, 214–219, 1998.
4. Y. Cheng. Mean shift, mode seeking and clustering. *IEEE Trans. on Pattern Analysis and Machine Intelligence*, 17: 790-799, 1995.
5. J. L. Crowley and J. Schwerdt. Robust tracking and compression for video communication. *Proc. of the Int. Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time* (RATFG'99), 2–9, Corfu. Greece. 1999.
6. G.D. Finlayson, B. Shiele and J.L. Crowley. Comprehensive colour normalization. *Proc. European Conf. on Computer Vison (ECCV). Vol. I*, 475–490, Freiburg, Germany. 1998.
7. G.D. Finlayson and G. Shaefer. Constrained dichromatic colour constancy. *Proc. ECCV. Vol. II*, 342–358, Dublin. Ireland. 2000.
8. R. Gershon, A.D. Jepson and J.K. Tsotsos. From [R,G,B] to surface reflectance: Computing color constant descriptors in images. *Proc. Int. Joint Conf. on Artificial Intelligence*, 755–758, 1987.
9. G.J. Klinker, S.A. Shafer and T. Kanade. A physical approach to color image understanding. *International Journal of Computer Vision*, 4: 7–38, 1990.
10. H. Lee. Method for computing the scene illuminant chromaticity from specular highlights. *Journal of the Optical Society of America A*, 3: 1694–1699, 1986.
11. Y. Raja, S.J. McKenna, S. Gong. Colour model selection and adaptation in dynamic scenes. *Proc. ECCV. Vol. I*, 460–474. 1998.
12. M. Soriano, B. Martinkauppi, S. Huovinen, M. Laaksonen. Skin detection in video under changing illumination conditions. *Proceedings of the Int. Conference on Automatic Face and Gesture Recognition* (FG'00), 839–842, Grenoble. France. 2000.
13. M. Störring, H.J. Andersen and E. Granum. Estimation of the illuminant colour from human skin colour. *Proceedings of the Int. Conference on Automatic Face and Gesture Recognition* (FG'00), 64–69, Grenoble. France. 2000.
14. K. Toyama. Prolegomena for robust face tracking. MSR-TR-98-65. Microsoft Research, November 1998.
15. Y. Wu, Q. Liu and T.S. Huang. Robust real-time hand localization by self-organizing color segmentation. *Proceedings RATFG'99*, 161–166. 1999.
16. J. Yang, W. Lu, A. Waibel. Skin-color modeling and adaptation. *Proceedings Third Asian Conference on Computer Vision Vol. II*, 142-147. 1998.
17. M. D'Zmura and P. Lennie. Mechanisms of colour constancy. *Journal of the Optical Society of America A*, 3: 1662–1672, 1986.