2010 International Conference on Pattern Recognition

Discriminant and Invariant Color Model for Tracking under Abrupt Illumination Changes

Jorge Scandaliaris and Alberto Sanfeliu
Institut de Robòtica i Informàtica Industrial, CSIC-UPC
jscandal,sanfeliu@iri.upc.edu

Abstract

The output from a color imaging sensor, or apparent color, can change considerably due to illumination conditions and scene geometry changes. In this work we take into account the dependence of apparent color with illumination an attempt to find appropriate color models for the typical conditions found in outdoor settings. We evaluate three color based trackers, one based on hue, another based on an intrinsic image representation and the last one based on a proposed combination of a chromaticity model with a physically reasoned adaptation of the target model. The evaluation is done on outdoor sequences with challenging illumination conditions, and shows that the proposed method improves the average track completeness *by over 22% over the hue-based tracker and the* closeness of track *by over 7% over the tracker based on the intrinsic image representation.*

Figure 1. Image sequence instances in the circuit.

1. Introduction

As pointed out by Yilmaz *et al*. [5], color is one the most widely used features for tracking, most probably because of its discriminant power and its apparent ease of use. However, except in scenarios where illumination conditions can be controlled or they don't change much, the apparent color from objects usually changes a lot. This fact is usually ignored [1], or color spaces with some invariant properties, such as HSV or rg, are used [1, 4] instead of RGB, and in other cases some adaptation strategies are adopted [3].

In this work we explicitly acknowledge the dependence of apparent color with illumination conditions, and particularly the conditions usually present in urban outdoor settings when tracking is done from a mobile platform. Our contribution is twofold: first, we show that by using the intrinsic image proposed by Finlayson *et al*. [2] as a feature we increase the robustness of the tracking results compared to using the hue component of the HSV color space. This image representation is based on a physical model of the image formation process. To our knowledge, it hasn't been assessed as a feature for tracking before. Second, we improve further on these results by noting that the intrinsic image representation has very good invariant properties at the expense of loosing discriminant power. We propose then a trade-off solution between invariance and discriminant power using the same image representation from where the intrinsic image is derived. The key idea is to allow the color distribution of the model to change in a principled way, following possible changes in the illumination.

2. Color based tracking

As a means for evaluating the impact of different color models on tracking we have chosen the mean shift algorithm [1]. The main motivations behind this selec-

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.454

tion were the fact that it is a very well known algorithm, most frequently used with color based features and well suited for tracking in real time. The mean shift algorithm is an efficient approach to tracking objects whose appearance is defined by histograms. This appearance is usually, but not limited to, color.

In this work we use three color based features: the hue component of the HSV color space, the intrinsic image representation [2], and a log-chromaticity representation paired with a reference target model adapted to illumination changes. From now on, we will refer to these three tracking methods as *hue*, *ii*, and *lc_me*, respectively.

3. Color Models

Below, a brief overview of the color models used is given, with an emphasis on their invariant properties against illumination changes.

HSV. HSV is an approximately perceptually uniform color space. Its hue component, in particular, is commonly used in tracking applications where some degree of illumination changes are expected. The hue component doesn't contain intensity information, and thus it is invariant to intensity changes in the illumination. Hue, however, is sensitive to light color changes.

Log chromaticity ratios and adapted target model. Finlayson *et al*. [2] use a transformation of the RGB color space that under certain assumptions on the illuminant, the camera sensitivities and surface reflectances, has some interesting properties. The assumptions adopted are a Lambertian model of image formation together with fairly narrow band camera sensitivities. The illuminant is restricted to be Planckian, and modelled with Wien's approximation to Planck's law. Under these assumptions, it can be shown [2] that forming the 3-vector chromaticities, c_k, by dividing each band by the geometric mean, $\sqrt[3]{R \times G \times B}$

$$c_k = \frac{R_k}{(\prod_{i=1}^{3} R_i)^{1/3}},\ k = 1, 2, 3 \tag{1}$$

and then calculating their logarithm, we arrive at a representation

$$\rho_k = \log(c_k),\ k = 1, 2, 3 \tag{2}$$

and in vector form

$$\underline{\rho} = \underline{s} + \frac{1}{T}\underline{e} \tag{3}$$

where R_k denote the sensor responses, $\underline{s}$ depends on surface and the camera, $\underline{e}$ is independent of surface, but which again depends on the camera, and T is the illuminant color temperature. All 3-vector $\underline{\rho}$ lie on a plane orthogonal to $\underline{u} = 1/\sqrt{3}(1, 1, 1)$. The redundant dimension is removed by transforming 3-vectors $\underline{\rho}$ into a coordinate system *in* the plane using a 2×3 matrix $\boldsymbol{U}$ (see [2] for details)

$$\underline{\chi} \equiv \boldsymbol{U}\underline{\rho},\ \underline{\chi}\ is\ 2 \times 1 \tag{4}$$

One way to interpret equation (3) is by noting how the transformed sensor responses, $\underline{\rho}$, change with different surfaces and illuminations. Surface properties affect only the first term, which can be seen as an offset with respect to the origin. Illumination color changes are modelled by the parameter T, color temperature, and they act as a scaling factor to $\underline{e}$. Because surface-related properties are concentrated on the first term and illumination is concentrated on the second, changes in illuminant color temperature result in shifts in the transformed sensor responses. Moreover, the shifts are in the same direction for all surfaces. This behavior is retained in (4). These conclusions are based on some restrictive assumptions. In practice, camera sensitivities are not exactly narrow band, and combination of Planckian illuminants do not yield another exactly Planckian illuminant. It has been shown [2], however, that this model is a good approximation in real situations.

Intrinsic image The invariant image representation [2] is obtained by projecting the log-chromaticity representation described above, $\underline{\chi}$, into the direction $\underline{e}^{\perp}$ orthogonal to $\underline{e}$, obtaining a single scalar

$$I^{'} = \chi_1 \cos\theta + \chi_2 \sin\theta \tag{5}$$

and to remove the effect of the logarithm, the last step in the derivation of the intrinsic image is to exponentiate

$$I = \exp(I^{'}) \tag{6}$$

In this 1-dimensional invariant, all points in the $\underline{\chi}$ log-chromaticity representation that are colinear in the direction of $\underline{e}$ are collapsed into a single point. As a result, this representation achieves near perfect invariance to illumination color change at the expense of loosing discriminant power. In figure 2 we see an example where two completely different surfaces, bricks from the floor and a blue bin, are indistinguishable in the intrinsic image.

4. Proposed method

The properties of the log-chromaticity space, $\underline{\chi}$, can be exploited for improving the robustness against illumination changes as follows. At any particular frame,

the target model is compared against target candidates at different locations and some similarity measure is maximized. Illumination color changes, intensity is already taken account for by normalization, will affect the similarity between the model and the candidates. In this representation, however, such changes will translate into shifts along the direction of $\underline{e}$. Without any a-priori knowledge of how the illumination will change, we assume that the illumination color temperature can both increase and decrease up to a finite amount. Then, to assure that we continue having a good similarity between the model and candidates, we enlarge the model in the direction of $\underline{e}$. In practice, given the particular model representation used in this case, we smooth the histogram of the target model by convolution with an anisotropic Gaussian filter. To simplify things, we rotate $\underline{\chi}$ using a rotation matrix $\boldsymbol{R}$ so the direction of $\underline{e}$ is coincident with the first axis, $\hat{\chi}_1$

$$\underline{\chi}^{'} = \boldsymbol{R}\underline{\chi} \tag{7}$$

and then we smooth the histogram of the target model with a gaussian filter

$$g(\chi_1^{'}, \chi_2^{'}) = \frac{1}{2\pi\sigma_1\sigma_2} \exp(\frac{\chi_1^{'}}{\sigma_1} + \frac{\chi_2^{'}}{\sigma_2}) \tag{8}$$

where $sigma_1$ controls the amount of smoothing in the direction of illumination change, and σ_2 can be used to account for model mismatches, i.e. the direction $\underline{e}$ has some dependency with surface reflectance.

The consequences of enlarging the initial model are that the invariant properties of the intrinsic image representation are retained, up to a given change in color temperature controlled by the amount of smoothing applied, while at the same time increasing its discriminant power.

5. Experiments

For evaluation, we acquired three sequences in an outdoor urban environment, at different times of the day. A camera mounted on a mobile platform moves together with a person around some raised garden beds describing a closed path. The distance and relative position of the person and the camera vary during the sequences, although the person is always within the field of view of the camera. The sequences can be characterized as having different and rapidly varying illumination conditions, and present cast shadows, over and under exposure during transitions from bright to dark regions and vice versa. Being a circular path, the sun position with respect to the camera also varies along the sequences. Each sequence has around 300 frames, for

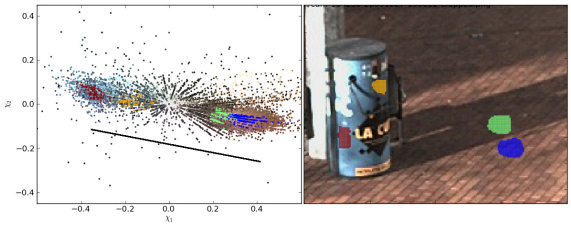

Figure 2. Image with two surfaces under two different illuminants (highlighted regions), and the corresponding log-chromaticity representation. The line corresponds to the direction of illumination change.

a duration of about a minute. All sequences were manually annotated with the position and scale of the person's upper-body.

Each method was tested with a set of different parameters. For the *hue* and the *ii* trackers, the only parameter to select is the number of bins used to represent the target and candidates model. We tried 180, 135 and 90 bins. For the *lc_me* tracker, we have to select additionally the amount of smoothing applied to the model. We only applied smoothing in the direction of the illumination change, to provide a better comparison with the *ii* tracker. We used 60×60, and 40×40 bins and a smoothing of 0.25, 0.2 and 0.15 expressed as a fraction of the number of bins. For each sequence, we initialized the tracking at three relative positions, start of the sequence, one third and two thirds of the sequence length. Because the sequences were acquired over a circular path, we could start the tracking anywhere in the sequence and let it run for the whole sequence. This gave us a total of nine runs for each method and parameter set, totalling 108 experiments.

For the quantitative evaluation of the results, we use some of the metrics defined by Yin [6]. *Track completeness*, c, measures the temporal overlap between a ground truth track and a system track, and *average track completeness*, $\overline{C}$, gives the same measure for a set of tracks. The *average closeness of track*, $\overline{a}_t$, is defined as the average spatial overlap between a ground truth track and a system track. The *closeness of track*, $\overline{A}$, and its standard deviation, σ_A, measure the average spatial overlap between ground truth tracks and system tracks for a complete video sequence.

For each method, we selected the parameters giving he best results and present them here. For the *hue* tracker, the best results corresponded to 180 bins, although all results were very close. The *ii* tracker also

Table 1. Performance evaluation metrics by runs.

	hue		ii		lc_me	
	$\overline{a}$	c	$\overline{a}$	c	$\overline{a}$	c
1	0.758	0.440	0.524	0.847	0.617	0.847
2	0.587	0.109	0.496	0.618	0.651	1
3	0.772	0.170	0.638	1	0.768	0.170
4	0.768	1	0.623	1	0.708	1
5	0.716	1	0.614	1	0.689	1
6	0.792	1	0.719	1	0.783	1
7	0.659	0.399	0.606	1	0.722	1
8	0.695	1	0.551	1	0.600	1
9	0.689	0.788	0.633	1	0.741	1

showed the best results with 180 bins. The *lc_em* tracker performed best with 60×60 bins and a smoothing parameter of 0.15. Table 1 shows the results by runs. None of the methods were able to complete successfully the nine runs. The *hue* tracker lost the target prematurely in 5 of the 9 runs. Both the *ii* and the *lc_em* trackers lost the target prematurely in 2 of the nine runs. The *hue* tracker has better discriminant power, reflected by the fact of having the highest average closeness of track in most of the runs. It is the most affected, however, by illumination changes, only being able to track completely 4 runs. Both *ii* and *lc_me* trackers are able to complete more runs and both show average closeness of track values lower than *hue*, although *lc_me* has consistently higher average closeness of track than *ii*. The results seem to indicate a trade-off between discriminant power and invariance to illumination conditions. Table 2 summarizes the results for the nine runs, that confirm that both *ii* and *lc_me* outperform *hue*, and that that *lc_me* improves over the target spatial localization from *ii*.

Both the *ii* and *lc_me* trackers were able to complete more runs successfully, seven against four, and covered successfully 22.5% more frames. The *lc_me* tracker improved, by over a 7%, over the *ii* tracker in the spatial localization of the target. Examination of the sequence frames that led the trackers to loose their target, showed significant amount of pixel clipping within the target, while having a background of similar color to the model. The hue based tracker, besides failing on frames similar to those causing failures to the other trackers, also failed at frames where there was pixel clipping in the region of the target person and dark background. The characteristics of the target and the background color distributions in these sequences don't expose the limitations of the intrinsic image representation, that is,

Table 2. Performance evaluation metrics calculated over the nine runs.

	$\overline{A}$	σ_A	$\overline{C}$
hue	0.731	0.098	0.655
ii	0.624	0.128	0.880
lc_me	0.695	0.111	0.888

its reduced discrimant power. Given different color distributions for the target and background, see figure 2 for such an example, we expect the *ii* tracker to perform much worst. The proposed method, on the other hand, would maintain its good performance, assuming that the distance, in the direction of the illumination change, between target and background in the log-chromaticity space is bigger than the smoothing applied to the target model.

6. Conclusions

We have evaluated the usefulness of the intrinsic image representation for tracking with challenging illumination conditions. The intrinsic image based tracker outperforms in all metrics the hue based tracker. Moreover, the proposed use of the log chromaticity representation combined with and enlargement of the model in the direction of the illumination change shows more discriminant power, as suggested by better average spatial localization of the target, while retaining the invariant properties.

References

[1] D. Comaniciu, V. Ramesh, and P. Meer. Kernel-based object tracking. *IEEE Trans. Pattern Anal. Machine Intell.*, 25(5):564–577, 2003.

[2] G. D. Finlayson, M. S. Drew, and C. Lu. Intrinsic Images by Entropy Minimization. In T. Pajdla and J. Matas, editors, *Proc. 8th European Conf. Comput. Vision*, volume 3024 of *Lect. Notes Comput. Sci.*, pages 582–595, Prague, May 2004. Springer-Verlag.

[3] S. J. McKenna, Y. Raja, and S. Gong. Tracking colour objects using adaptive mixture models. *Image Vision Comput.*, 17(3–4):225–231, 1999.

[4] R. Muñoz-Salinas, E. Aguirre, and M. García-Silvente. People detection and tracking using stereo vision and color. *Image Vision Comput.*, 25(6):995–1007, 2007.

[5] A. Yilmaz, O. Javed, and M. Shah. Object tracking: A survey. *ACM Comp. Surv.*, 38(4), 2006.

[6] F. Yin, D. Makris, S. Velastin, and J. Orwell. Quantitative evaluation of different aspects of motion trackers under various challenges. *Ann. BMVA*, 2009. Accepted.