COLOR INVARIANT DENSITY ESTIMATION FOR IMAGE SEGMENTATION AND OBJECT TRACKING

Theo Gevers and Frank Aldershoff

Faculty of Science, University of Amsterdam, The Netherlands

Abstract

In this paper, we formulate a novel density estimation scheme derived from color invariants for image segmentation and object tracking. The advantage of color invariants is that they are robust against varying illumination. However, color invariants are ill-defined when the intensity or saturation is low.

Therefore, to achieve robust density estimation, computational methods are presented to estimate the amount of sensor noise through these color invariant images. The obtained uncertainty is subsequently used as a weighting term in the density estimation process to achieve robust image segmentation and object tracking.

Experiments are conducted on image sequences recorded from complex 3D scenes. From the experimental results it is shown that the proposed method successfully segments and finds objects robust against illumination and noisy data.

1. INTRODUCTION

Kernel density estimation is a widely used technique in statistics and recently in computer vision. For example, for the purpose of image segmentation and object tracking, kernel density estimation is a powerful technique to estimate the probability density function of an empirical distribution [2]. However, estimating the density of RGB images will depend on illumination and scene geometry as the data points are sensitive to these phenomena. Therefore, the RGB data should be transformed into color invariant spaces, e.g. rgb, H and $\theta_1\theta_2$ color models [3]. However, these color space transforms are unstable for noisy data. For example, normalized colors rgb and $\theta_1\theta_2$ are ill-defined near the black point while H is ill-defined along the achromatic axis. Further, another drawback of kernel density estimation is that the bandwidth is fixed for all data points and often chosen in an ad-hoc manner. Obviously, a more principled way is required to compute the bandwidth of the kernel based on the propagation of sensor noise through color invariants [4]. The contribution of this paper is to extend this principle and to use the variable kernel density estimation to achieve robust image segmentation and object tracking. We focus on object tracking in the context of player tracking in sports scenes.

Therefore, in this paper, we aim at variable density estimation to segment and track objects in video. Computational methods are presented to estimate the amount of uncertainty of color invariants. Instead of always assigning a fixed kernel to a datapoint, the support of the measurement is steered by the amount of uncertainty. As a result, noisy and unstable color information will contribute less to the underlying density estimation than reliable color information yielding more robust object segmentation and tracking. The paper is organized as follows. In section 2, variable kernel density estimation for color invariant images is presented. Image segmentation and object tracking is proposed in sections 3 and 4.

2. KERNEL DENSITY ESTIMATION OF COLOR INVARIANT IMAGES

In this section, density estimation is proposed for color invariant images. First, in section 2.1, color invariant features will be presented. Noise propagation through the color invariant transformations will be discussed in section 2.2. Variable kernel density estimation is proposed in sections 2.3 and 2.4.

2.1. Color invariant features

In order to obtain a representation that is robust under varying imaging conditions, an invariant color feature space is required. Although other color invariant models can be taken, we focus on $\theta_1\theta_2$ which is insensitive to surface orientation, illumination direction and illumination intensity [3]:

$$\theta_1 = \arctan\left(\frac{R}{G}\right), \tag{1}$$

$$\theta_2 = \arctan\left(\frac{R}{B}\right). \tag{2}$$

As with all sensor data, a video stream contains noise. Noise can be introduced by the sensor (the video camera), the data transport (including storage methods) and the encoding of the data (compression). Additive Gaussian noise is widely

0-7803-8554-3/04/$20.00 ©2004 IEEE.

used to model thermal noise and is the limiting behavior of photon counting noise and film grain noise. Therefore, in this paper, we assume that sensor noise is normally distributed.

2.2. Sensor noise propagation

When the RGB values are converted to color invariant values, in our case from RGB to $\theta_1\theta_2$, this is usually a non-linear operation. This means that small variations due to noise may lead to very large changes in the new color invariant space. Noise propagation through arbitrary functions is as follows.

For an indirect measurement, the true value of a measurand u is related to its N arguments, denoted by u_j, as follows:

$$u = q(u_1, u_2, \cdots, u_N). \tag{3}$$

The measurand u can be estimated by a measurement $\hat{u}$ by substituting $\hat{u}_j$ for u_j. Then $\hat{u}_1, \hat{u}_2, \cdots, \hat{u}_N$ are measurements and we get:

$$\hat{u} = q(\hat{u}_1, \hat{u}_2, \cdots, \hat{u}_N), \tag{4}$$

with corresponding standard deviations $\sigma_{\hat{u}_1}, \sigma_{\hat{u}_2}, \cdots, \sigma_{\hat{u}_N}$. Then, for N arguments, it follows that if the uncertainties in $\hat{u}_1, \cdots, \hat{u}_N$ are independent, random, and relatively small, the predicted uncertainty in q is given by:

$$\sigma_q = \sqrt{\sum_{j=1}^{N} \left(\frac{\partial q}{\hat{u}_i} \sigma_{\hat{u}_i} \right)^2}. \tag{5}$$

Although (5) is deduced for random errors, it is used as an universal formula for various kinds of errors.

Substitution of (1) and (2) in (5) gives the uncertainty for the normalized coordinates

$$\begin{aligned} \sigma_{\theta_1} &= \sec^2 \frac{RG\sigma_R^2 - R^2\sigma_G^2}{G^3}, \\ \sigma_{\theta_2} &= \sec^2 \frac{BG\sigma_B^2 - B^2\sigma_G^2}{G^3}. \end{aligned} \tag{6}$$

Assuming normally distributed random quantities, the standard way to calculate the standard deviations σ_R, σ_G, and σ_B is to compute the mean and variance estimates derived from a homogeneously colored surface patches in an image under controlled imaging conditions. From the analytical study of (6), it can be derived that $\theta_1\theta_2$ becomes unstable around the black point $R = G = B = 0$. Hence, it can be analytically derived that normalized color is unstable at low intensity.

2.3. Kernel density estimation

For the simplest case of one-dimensional data, the kernel density estimate $\hat{f}(y)$ of a distribution is obtained based on a kernel function $K(x)$ and a bandwidth h as the average:

$$\hat{f}(y) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{y - y_i}{h} \right). \tag{7}$$

Here, kernel K is a function satisfying $\int K(x)dx = 1$. Further, n is the number of pixels with value y_i in the image, h is the bin width and y the range of the data. The difference between a histogram and kernel density estimation is that a kernel is placed at each datapoint convolving the surrounding data with the kernel function. Note that the kernel has a finite width meaning that not all data points contribute to the convolution. The bandwidth controls the support and hence the number of points contributing to the estimate. In general, circular symmetric kernels are considered whose radius is controlled by the bandwidth. Further, the bandwidth is fixed for all datapoints. From section 2.2 this is obviously not the case for color invariant values, as the amount of uncertainty of color invariants depends on the color transform. Therefore, the kernel width is in nature not symmetrical i.e. the uncertainty of normalized color at a pixel is different in each color dimension. Therefore, the uncertainty of color invariants should be taken into account. Instead of always assigning a fixed kernel to a datapoint, the support of the measurement is steered by the amount of the uncertainty. The more certain a datapoint is, the more peaked the kernel will be. In the next section, we propose the variable kernel density estimation, where the amount of certainty steers the bandwidth in a principled way.

2.4. Variable kernel

In the *variable* kernel density estimator, the single h is replaced by n values $\alpha(y_i), i = 1, \cdots, n$. This estimator is of the form

$$\hat{f}(y) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\alpha(y_i)} K\left(\frac{x - y_i}{\alpha(y_i)} \right). \tag{8}$$

The kernel centered on y_i has associated with it its own scale parameter $\alpha(y_i)$, thus allowing different degrees of smoothing. To use variable kernel density estimators for color images, we let the scale parameter be a function of the RGB-values and the color space transform. We are now left with the problem of determining the scale and shape of the kernel. Assuming normally distributed noise, the distribution is approximated well by the Gauss distribution [6]

$$K(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp^{\frac{-x^2}{2}}. \tag{9}$$

Then the variable kernel method for the bivariate normalized $\theta_1\theta_2$ kernel is given by:

$$\hat{f}(\theta_1, \theta_2) = \frac{1}{n} \sum_{i=1}^{n} \sigma_{\theta_{1_i}}^{-1} K\left(\frac{\theta_1 - \theta_{1_i}}{\sigma_{\theta_{1_i}}} \right) \sigma_{\theta_{1_i}}^{-1} K\left(\frac{\theta_1 - \theta_{1_i}}{\sigma_{\theta_{1_i}}} \right), \tag{10}$$

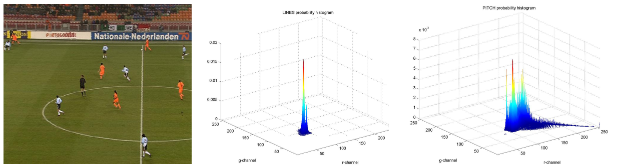

Fig. 1. *Discrete versions of kernel density estimation for lines (1.b) and players from one team (1.c) of image (1.a).*

Fig. 2. *Classification of color areas with* grass *(2.b) and players from one team (2.c) based on kernel density classification from (2.a).*

where the width of the kernel is the standard deviation of the Gaussian kernel denoted by $\sigma_{\theta_{1_i}}, \sigma_{\theta_{2_i}}$ derived according to (6). In conclusion, to reduce the effect of sensor noise during density estimation, we use variable kernels where the normal distribution defines the shape of the kernel. Further, kernel sizes are steered by the amount of uncertainty of the color invariant values.

3. IMAGE SEGMENTATION

A drawback of color classifiers, that do not account for the image formation process [5], is that the values of the color features, on which the classification relies, depend on the geometry of the object, the viewpoint of the camera and on the illumination conditions. As a consequence, the obtained classification results may be affected negatively by shadows, shading and illumination. In this paper, variable kernel density estimation is used for image segmentation based on Bayes' theorem. In this case, for a K class problem, variable kernel density estimation is produced $\hat{f}_j(y)$, $j = 1, ..., K$ separately for each of the classes. Further, when estimates of the class priors $\hat{\pi}_j$ are obtained (e.g. corresponding to the sample proportions), then:

$$\hat{P}(C = j|y = y_0) = \frac{\hat{\pi}_j \hat{f}_j(y_0)}{\sum_{k=1}^{K} \hat{\pi}_k \hat{f}_k(y_0)}. \quad (11)$$

For image segmentation, in this paper, we will consider the problem of classifying objects in football images. Variable kernel density estimation is used to distinguish background from players. Further, the background is divided into field (grass), lines, ball and remaining (billboards, public etc). First, color invariant statics are obtained by selecting representative samples (i.e. grass, player and background) by hand. Then, we fit kernel density estimates separately for each of the classes according to eq. (10). For example, in Figure 1, a discrete version of kernel density estimation is shown for lines and players. Finally, kernel density classification is performed according to eq. (11).

In fact, eq. (11) can be seen as density back-projection which is a way to relate density estimation back to the spatial domain. Density estimation is used to assign to every pixel in the image the probability that it is part of the object that is search for. A density estimation for all the pixels that are *not* part of the target object is created similarly. Consider the image of a football game shown in figure (2.a). The image is clearly contaminated by shadows, shading and illumination changes. Figures 2.a and 2.b show how grass and players (from one team) have been classified. The classification results are robust against low SNR, shading and illumination variations. Processing of the density classification scheme took no more than 1.2 seconds.

4. OBJECT TRACKING

To apply kernel density estimation in the context of object tracking, in this section, we will focus on tracking players in football videos.

A target, in our case a player, is represented by a pdf in the color invariant feature space. Further, the density estimation of an object is obtained for the $\theta_1 \times \theta_2$ feature space. Tracking consists of computing the similarity of a pdf of a region with the target pdf. In this paper, the measure between two distributions is the Bhattacharyya coefficient

$$\rho\,[p, q] = \int \sqrt{p(u), q(u)}du. \quad (12)$$

A larger ρ indicates more similar pdf's, so we use the distance measure

$$d = \sqrt{1 - \rho\,[p, q]}, \quad (13)$$

which is called the Bhattacharyya distance.

Object localisation is now a search for the point $\mathbf{y}$ where the distance between the pdf of the candidate object $p_{\mathbf{y}}$ and the target object $p_{\mathbf{q}}$ is minimal. Several methods have been proposed to efficiently perform this search. There is of course the brute-force approach in which all points in the area are evaluated, but with kernel based tracking more efficient solutions are possible. In this paper, the mean-shift is used to achieve efficient tracking.

Another example is shown in Fig 4. Despite severe occlusion and varying imaging conditions, the tracker performs satisfactory and keeps track of the player.

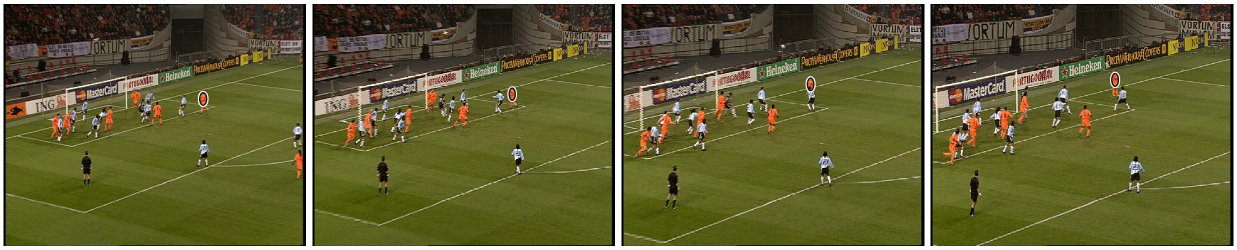

Fig. 3. First example of tracking a player. An ellipse is drawn over the player, the size of the ellipse indicates the size of the tracked player.

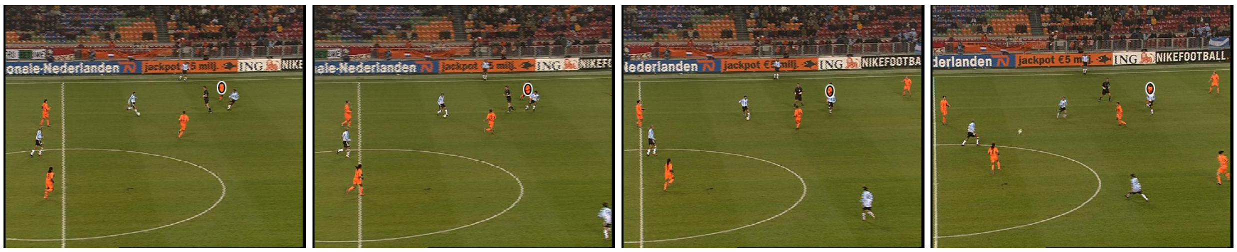

Fig. 4. Second example of tracking a player. An ellipse is drawn over the player, the size of the ellipse indicates the size of the tracked player.

4.1. Player tracking

The application is a monitoring system for players of football matches. The information obtained could be used for tactical information to improve the training of players. Furthermore, for football graphical analysis, visualisation tools are used to show the trajectories of players and their interactions. To this end, a player has been tracked for about 20 seconds, see Fig 3. The size of the images are 256x256 with 8 bits per color. The images show a considerable amount of shadows, shading, and illumination variations. These images can be seen as images of snap shot quality, a good representation of views from everyday life as it appears in home video, the news, and consumer digital photography in general. Although the low-quality of the video (low SNR) and varying imaging conditions, the tracker did not loose the object even in the presence of severe occlusion. The speed of the tracking scheme was 10 frames per second on average.

5. CONCLUSION

Segmenting and tracking of objects in video is of great importance for video-based encoding, surveillance and retrieval. However, the inherent difficult of object segmentation and tracking is to distinguish changes in objects and their displacements from disturbing effects such as noise and illumination changes.

In this paper, we have formulated kernel density estimation for color invariant images. To achieve robust density estimation, computational methods have been presented to estimate the amount of sensor noise through these color invariant images. The associated uncertainty is used to derive the parameterization of the variable kernel for the purpose of robust pdf construction.

Experiments have been conducted on image sequences recorded from complex 3D scenes. Two different applications have been chosen: image segmentation and object tracking. From the experimental results it has been shown that the proposed method successfully segment and finds objects robust against illumination and noisy data.

6. REFERENCES

[1] D. Comaniciu and P. Meer, Mean Shift Analysis and Applications, IEEE ICCV, Greece, pp. 1197-1203, 1999.

[2] A. Elgammal, R. Duraiswami, L.S. Davis, Efficient Kernel Density Estimation using the Fast Gauss Transform with Applications to Color Modeling and Tracking, IEEE Trans. PAMI, vol. 25, no. 11, 2003.

[3] Th. Gevers and Arnold W.M. Smeulders, Color Based Object Recognition, Pattern Recognition, 32, pp. 453-464, March, 1999.

[4] Th. Gevers, Robust Histogram Construction from Color Invariants, ICCV, 2001.

[5] J. Liu and Y-H. Yang, Multi-resolution color image segmentation, IEEE Transactions Pattern Analysis and Machine Intelligence, 16(7):689–700, 1994.

[6] J. R. Taylor, *An Introduction to Error Analysis*, University Science Books, 1982.