Integration of Shape and a Multihypothesis Fisher Color Model for Figure-Ground Segmentation in Non-Stationary Environments

Francesc Moreno-Noguer Alberto Sanfeliu
Institut de Robòtica i Informàtica Industrial, UPC-CSIC
Llorens Artigas 4-6, 08028, Barcelona, Spain
fmoreno,asanfeliu@iri.upc.es

Abstract

In this paper a new technique to perform figure-ground segmentation in image sequences of scenarios with varying illumination conditions is proposed. The color of the target and background are modelled with mixture of Gaussians, which optimum number is initialized automatically. Based on the 'Linear Discriminant Analysis' (LDA) a new colorspace that maximizes the foreground/background class separability is proposed. Moreover, there is no need to assume gradual change of the viewing conditions over time, because the method works with multiple hypothesis about the next state of the color distribution (some considering small changes and other more abrupt variations). The hypothesis that generates the best object segmentation and the shape information in the previous iteration are fused to accurately detect the object boundary, in a stage denominated 'sample concentration', introduced as a final step to the classical CONDENSATION algorithm.

1. Introduction

Color is a visual cue that is commonly used in computer vision applications such as object detection and tracking tasks. In environments with controlled lighting conditions and uncluttered background, color can be considered a robust and invariant cue, but when dealing with real scenes with changing illumination and confusing backgrounds, the apparent color of the objects varies considerably over time. In the literature, the techniques that cope with change in color appearance can be divided in two groups. On the one side, there are the approaches searching for color constancy (e.g. [2]); but in practice, these methods work on artificial and highly constrained environments. On the other hand, there are the techniques that generate a stochastic model of the color distribution, and adapt it over time ([7], [8]).

[0]This work was supported by CICYT projects DPI2001-2223 and DPI2000-1352-C02-01, and by a fellowship from the Spanish Ministry of Science and Technology.

The drawback in all these approaches is that they assume a smooth and slow color variation that can be predicted by a dynamic model based in only one hypothesis (usually by a weighting function). However, this assumption does not suffice to cope with general scenes, where the dynamics of the color distribution might follow an unknown an unpredictable path. In a previous work [6], we have suggested the use of a multihypothesis framework to track color objects in such situations, where the color could be approximated by an unimodal distribution. In the present work, we deal with multicolored objects.

An overview of the system is given in Section 2. In Section 3 we describe the Fisher colorspace. Sections 4 and 5 present the color and dynamical models. The complete tracking algorithm is depicted in Section 6. Results and conclusions are presented in Sections 7 and 8, respectively.

2. System Overview

In order to cope with unconstrained environments, we propose a system with the following main features, that represent contributions with respect to previous works:

- **Fisher color model:** Instead of using the classical RGB, rgb[1], XYZ or HSV color spaces, we propose the use of a color space efficient for the discrimination between foreground and background classes, based on the 2D projection of the R, G and B components on the plane obtained from a nonparametric LDA [4].
- **Multihypothesis framework:** The use of a particle filter formulation to predict the color distribution in following iterations, offers robustness to abrupt and unexpected changes in the color appearance.
- **Integration of color and shape:** The fusion of both vision modules is done in a final stage introduced to the classical CONDENSATION algorithm, and makes our method suitable to work in cluttered scenes.

[1]Lowercase letters represent normalized components

3. Fisher Colorspace

The selection of the colorspace is an important initial issue for any color-based figure-ground segmentation system. The typical selection criterion is based on the invariance of the color representation to illumination changes, and according to this idea, color is usually represented by two components of the rgb, HSV or xyz colorspaces. However, these representations are not robust enough to cope with abrupt illumination changes. In this paper we propose a different criterion and select a 2D colorspace that maximizes the separability of the object and background classes.

Let $\mathbf{x}$ be a 3D vector with the color value of image pixels in RGB space, which must be classified as foreground (ω_f) or background (ω_b). When we are dealing with multicolored objects, the parameterization of color distributions in 3D colorspace becomes very complex. To simplify, we reduce the dimensionality to 2D by projecting the data on a plane $\mathbf{\Phi} = [\phi_1, \phi_2] \in \mathcal{M}_{3\times 2}$, that is, $\mathbf{y} = \mathbf{\Phi}\mathbf{x}$, where $\mathbf{y}$ are the 2D linearly transformed coordinates used for classification. The most popular way to find the best linear features is the parametric version of the LDA method, where training data is used to construct the *within-class* S_w and *between-class* S_b scatter matrices, in the c-class problem defined as,

$$S_w = \sum_{i=1}^{c} P(\omega_i)\, E\left[(\mathbf{x}|_{\omega_i} - \mu_i)(\mathbf{x}|_{\omega_i} - \mu_i)^T\right] = \sum_{i=1}^{c} P(w_i)\, S_i$$
$$S_b = \sum_{i=1}^{c} P(\omega_i)\, E\left[(\mathbf{x} - \mu_o)(\mathbf{x} - \mu_o)^T\right] \tag{1}$$

where $P(\omega_i)$ is the prior of the ith class, μ_i and S_i are its expected vector and covariance matrix, μ_o is the overall mean and $\mathbf{x}|_{\omega_i}$ indicates that sample $\mathbf{x}$ belongs to ω_i.

A typical criterion of class separability is formulated by maximizing $J = trace\left(\left(\mathbf{\Phi}^T S_w \mathbf{\Phi}\right)^{-1}\left(\mathbf{\Phi}^T S_b \mathbf{\Phi}\right)\right)$, and seeks for the separation of the class means in the transformed Y-space (high S_b), while the classes remain compact (small S_w). The classical LDA maximizes J by constructing the columns of $\mathbf{\Phi}$ with the eigenvectors of $S_w^{-1} S_b$ with the highest eigenvalues.

As the maximum rank of S_b is $c-1$, this will be the maximum dimension of the projected Y-space. This limitation can be solved by the *nonparametric LDA* [4], that computes S_b using local information and the *k Nearest Neighbors* (KNN) rule. In the 2-class problem that we have in hands, this matrix (denoted Σ_b) is defined as,

$$\Sigma_b = \frac{1}{N}\sum_{i=1}^{N_f} w_i \left(\mathbf{x}_i|_{\omega_f} - M_b^k\left(\mathbf{x}_i|_{\omega_f}\right)\right)\left(\mathbf{x}_i|_{\omega_f} - M_b^k\left(\mathbf{x}_i|_{\omega_f}\right)\right)^T$$
$$+ \frac{1}{N}\sum_{i=1}^{N_b} w_i \left(\mathbf{x}_i|_{\omega_b} - M_f^k\left(\mathbf{x}_i|_{\omega_b}\right)\right)\left(\mathbf{x}_i|_{\omega_b} - M_f^k\left(\mathbf{x}_i|_{\omega_b}\right)\right)^T \tag{2}$$

where N_f and N_b are the number of samples of ω_f and ω_b, $N = N_f + N_b$, $M_j^k(\mathbf{x}_i)$ is the mean of the kNN in ω_j to a point $\mathbf{x}_i$, and w_i is a weighting function for deemphasizing samples far from the classification boundary (see [4]).

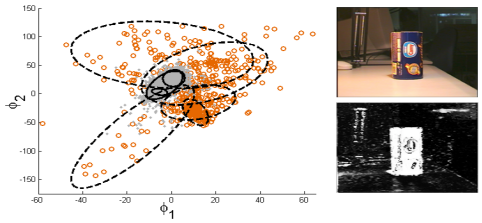


Figure 1. Gaussian mixture components of ω_f (the can) and ω_b. Left image: empty dots and dashed lines are ω_f data (projected on $\mathbf{\Phi}$) and Gaussian components, respectively. Filled dots and continuous lines are ω_b data and Gaussians. Lower right image: $p(\omega_f|\mathbf{y})$, where brighter points correspond to more likely pixels.

Given two sets $\{\mathbf{x}_1, \cdots, \mathbf{x}_{N_f}|\omega_f\}$, $\{\mathbf{x}_1, \cdots, \mathbf{x}_{N_b}|\omega_b\}$ of RGB pixel values used as training data, the optimum linear mapping is obtained with the following steps:

- Calculate S_w with eq. 1 and whiten the data with respect to it. That is, transform $\mathbf{x}$ to $\mathbf{z} = \Lambda^{-1/2}\Omega^T\mathbf{x}$, where Λ and Ω are the eigenvalue and eigenvector matrices of S_w.
- Select *k* and (in the Z-space) compute Σ_b using eq. 2.
- Select the two eigenvectors Ψ_1, Ψ_2 of Σ_b with the two largest eigenvalues.
- The optimum linear mapping from the original RGB space to the discriminant subspace (we call it *Fisher colorspace*) is given by $\mathbf{y} = \mathbf{\Psi}^T\Lambda^{-1/2}\Omega^T\mathbf{x}$.

In the Results Section it will be shown that with this colorspace we obtain better rates in pixel classification.

4 Color Model and Parameterization

The color distributions of fore and background transformed to the Fisher colorspace, are represented by a Gaussian mixture model. The conditional probability for a pixel $\mathbf{y}$ belonging to ω_f is expressed as a sum of M_f Gaussian components as $p(\mathbf{y}|\omega_f) = \sum_{j=1}^{M_f} p(\mathbf{y}|j)\, P(j)$. Similarly, the background color distribution will be represented by a mixture of M_b Gaussians (Fig. 1).

Next, the a posteriori probability that a pixel $\mathbf{y}$ belongs to the ω_f class is computed using the Bayes rule:

$$p(\omega_f|\mathbf{y}) = \frac{p(\mathbf{y}|\omega_f)\, P(\omega_f)}{p(\mathbf{y}|\omega_f)\, P(\omega_f) + p(\mathbf{y}|\omega_b)\, P(\omega_b)} \tag{3}$$

One of the problems when using mixture of Gaussians is the selection of the number of components that better adjust the data. We initialize this, with the modified *EM* algorithm proposed in [3], that is based on a *Minimum Message Length* criteria and iteratively fits and annihilates and initial and large number of components (introduced by the user).

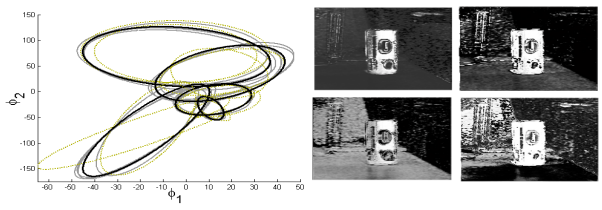


Figure 2. Left: Different predictions of the foreground mixture of Gaussians of Fig. 1. Continuous wider lines: mixture from previous iteration. Continuous narrower lines: mixtures predicted with a model accounting for smooth changes. Dashed lines: predictions done with a model accounting for abrupt changes. Right images: $p(\omega_f|\mathbf{y})$ for various configurations of Gaussian mixtures.

Once we have learnt the initial configuration of the Gaussian mixtures for ω_f and ω_b, they are parameterized by:

$$\mathcal{X}_\varepsilon = [\mathbf{p}_\varepsilon, \mu_\varepsilon, \lambda_\varepsilon, \theta_\varepsilon] \tag{4}$$

where $\varepsilon = \{f, b\}$, $\mathbf{p}_\varepsilon$ contains the priors for each component, μ_ε the centroids of each Gaussian , λ_ε the eigenvalues of the principal directions and θ_ε the angles between the principal axis of each component with the horizontal.

5 Learning the Dynamical Models

One of the stages of the tracking algorithm, consists of propagating the state vector 4 in order to generate multiple hypothesis about the future configuration of the mixture of Gaussians. To formulate these hypotheses we use the following dynamic motion model,

$$\mathcal{X}_{\varepsilon,t} = A_0\mathcal{X}_{\varepsilon,t-2} + A_1\mathcal{X}_{\varepsilon,t-1} + B_0\mathbf{w}_t \tag{5}$$

where the matrices A_0, A_1 represent the deterministic component of the model and $B_0\mathbf{w}_t$ is the stochastic component (learned a priori using the standard MLE algorithm described in [1]). In practice, we define two kinds of dynamical models, one taking in account smooth changes and another that considers more abrupt variations (see Fig. 2).

6 The Tracking Algorithm

The basic steps of our tracking algorithm follow the procedure of the particle filters, but we introduce a modification to the classical CONDENSATION algorithm (similar to the ICONDENSATION technique [5]), and in order to 'direct' the search for the next iteration we add a final stage that concentrates the future hypothesis on those areas of the state-space containing more information about $p(\omega_f|\mathbf{y})$. Moreover, in this final stage we fuse object color and shape information. Next, we present the steps of our algorithm:

- **pdf of color distribution:** At time t, a set of N samples $\mathcal{S}_{t-1}^{(n)}$ $(n = 1, \ldots, N)$ with the same structure than $\mathcal{X}$ (eq. 4), parameterizing N color distributions, are available from the previous iteration. Each sample has an associated weight $\pi_{t-1}^{(n)}$. The whole set represents an approximation to $p(\mathcal{X}_{t-1}|\mathcal{Z}_{t-1})$ where $\mathcal{Z}_{t-1} = \{z_0, \ldots, z_{t-1}\}$ is the history of the measurements. The goal of the algorithm consists on construct a new sample set $\{\mathcal{S}_t^{(n)}, \pi_t^{(n)}\}$ to estimate $p(\mathcal{X}_t|\mathcal{Z}_t)$.

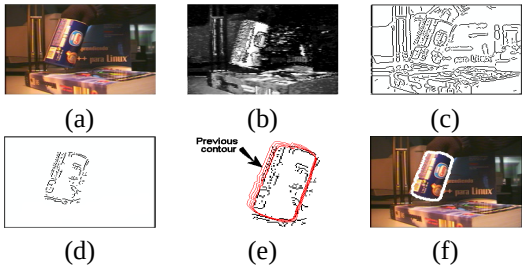


Figure 3. Steps to extract the exact position of the object fusing color segmentation and accurate adjustment by affine snakes (commented in the text).

- **Sampling from $p(\mathcal{X}_{t-1}|\mathcal{Z}_{t-1})$:** A sampling with replacement is performed N times on the set $\{\mathcal{S}_{t-1}^{(n)}\}$, where each element has probability $\pi_{t-1}^{(n)}$ of being chosen. This, will give us a set $\{\mathcal{S'}_{t-1}^{(n)}\}$.

- **Probabilistic propagation of the samples:** Each sample $\mathcal{S'}_t^{(n)}$ is propagated to $\tilde{\mathcal{S}}_t^{(n)}$, according to one of the learned dynamic models (eq. 5):

- **Measure and Weight:** Each element $\tilde{\mathcal{S}}_t^{(n)}$ has to be weighted according to some measured features. From the propagated samples $\tilde{\mathcal{S}}_t^{(n)}$ we construct the corresponding Gaussian mixtures, that are used to calculate $p(\omega_f|\mathbf{y})$ for the whole image in Fisher colorspace, using eq. 3. The weight assigned to each sample is $\pi_t^{(n)} = \frac{\sum_{\mathbf{y}\in W} p(\omega_f|\mathbf{y})}{N_w} - \frac{\sum_{\mathbf{y}\notin W} p(\omega_f|\mathbf{y})}{\overline{N_w}}$, where W is the interest region around the previous object position, and N_w, $\overline{N_w}$ are the number of image pixels in and out of this interest region.

- **Sample Concentration:** In the last stage of our algorithm, we concentrate the samples around the local maxima, so that in the following iteration the hypotheses are formulated around these more likely regions of the state space. In our case, this is absolutely necessary because the state vector $\mathcal{X}$ has high dimensionality, and if we let the samples move freely, uniquely governed by the dynamic model, the number of hypotheses needed to find the samples representing a correct color configuration, is extremely high.

  In this 'concentration' stage, firstly, the maximum from the set of weights $\{\pi_t^{(n)}\}$, is taken, and by morphologic operations on its $p(\omega_f|\mathbf{y})$ map (Fig. 3b), a

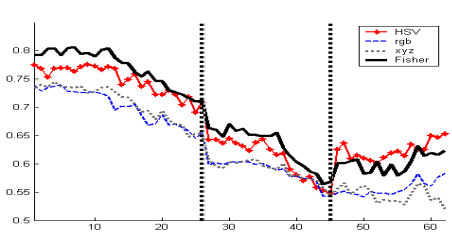


Figure 4. Classification results obtained with various colorspaces. Vertical bars, separate different sequences.

coarse approximation of the object shape is obtained. This, lets to eliminate noisy edges from the original image (Fig. 3c,d). Next, the contour of the object in the previous iteration, is used as initialization of an affine snake, that is adjusted (only by affine deformations) to the image of refined edges (Fig. 3e). The fusion of color and shape information increases the robustness of the system, because even when the color hypotheses give a highly rough estimation, they can be corrected using the contour information. Once the boundary of the object has been accurately detected (Fig. 3f), a Gaussian mixture is fitted to its color distribution (using the EM algorithm), giving a state vector $\mathcal{S}_t^*$. Samples $\{\tilde{\mathcal{S}}_{t-1}^{(n)}\}$ are 'concentrated' on this new distribution with the equation $S_t^{(n)} = (1-a)\widetilde{S}_t^{(n)} + a\mathcal{S}_t^*$, where the parameter a governs the level of concentration.

7 Results

We start this Results Section by comparing the class discrimination power of the Fisher colorspace with other colorspaces (two components of the rgb, HSV or xyz). To quantify the notion of class separability, a constant number of Gaussians are fitted to ω_f and ω_b distributions of hand segmented images, for each one of the colorspaces. Next, according to eq. refeq4 we segment the same images, assigning each pixel $\mathbf{y}$ to the class with maximal $p(\omega_f|\mathbf{y})$. This result is compared with the hand segmented ones. In Fig. 4 we show the results for three different sequences, where the vertical axis represents the percentage of pixels well classified. The best results are obtained with the Fisher colorspace (77.5% of correct classification), followed by the HV components of HSV (74.8%). Next, it is shown the performance of the tracking system in two different situations. Fig. 5 represents the results obtained on a sequence with a gradual change of illumination and object position. In the second experiment (Fig. 6), there is and abrupt change of both illumination and object position (Fig. 6a,b are the two consecutive frames). At least one of the multiple hypothesis of color distributions performs a good a posteriori probability map (Fig. 6c) that is used to fit the contour (Fig. 6d,e).

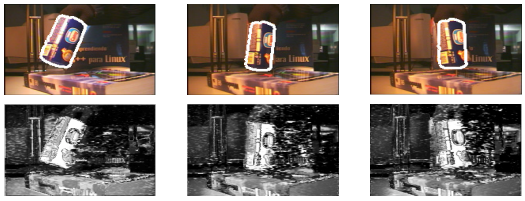
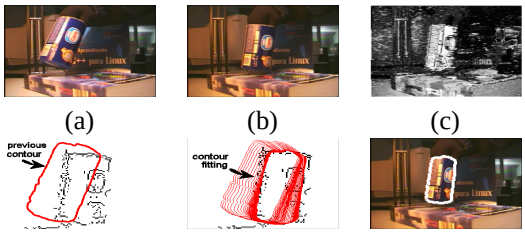


Figure 5. Contour fitted and $p(\omega_f|\mathbf{y})$ map on a sequence with gradual change of illuminant and object position.



(a) (b) (c) (d) (e) (f)

Figure 6. Results of an abrupt change of illuminant and object position (commented in text).

8 Conclusions

We have presented a new approach to figure-ground segmentation in non-stationary environments. It has been introduced the concept of Fisher colorspace, that has better object/background discrimination performance than typical colorspaces, and the fusion of shape and color information in the probabilistic multiple hypotheses framework offered by the CONDENSATION algorithm. This integration increases the robustness of the method, and is done in a last stage denominated 'sample concentration' that we have introduced to the classical CONDENSATION algorithm.

References

[1] A.Blake, M.Isard, Active contours. *Springer*,1998.
[2] G.D.Finlayson, B.V.Funt, K.Barnard. Color Constancy under Varying Illumination. *Proc. ICCV*, pp.720–725, 1995.
[3] M.A.T.Figueiredo, A.K.Jain. Unsupervised Learning of Finite Mixture Models. *Trans.PAMI*, Vol.24, num.3, pp.381-396, 2002.
[4] K.Fukunaga. Introduction to Statistical Pattern Recognition (2nd. Edition) Academic Press, 1990.
[5] M.Isard, A.Blake Icondensaton: Unifying Low-Level and High-Level Tracking in a Stochastic Framework. *Proc. ECCV*, Vol.1, pp.893-908, 1998.
[6] F.Moreno-Noguer, J.Andrade-Cetto, A.Sanfeliu. Fusion of Color and Shape for Object Tracking under Varying Illumination. *Proc.IBPRIA, LNCS 2652, Springer*, pp.580-588, 2003.
[7] L.Sigal, S.Sclaroff, V.Athitsos. Estimation and Prediction of Evolving Color Distributions for Skin. Segmentation under Varying Illumination. *Proc. CVPR*, Vol.2, pp.152–159, 2000.
[8] Y.Raja, S.McKenna, S.Gong. Colour Model Selection and Adaption in Dynamic Scenes. *Proc. ECCV*, Vol.1, pp.460–475, 2000.