

Fusion of Color and Shape for Object Tracking under Varying Illumination

Francesc Moreno-Noguer, Juan Andrade-Cetto and Alberto Sanfeliu

Institut de Robòtica Industrial, Universitat Politècnica de Catalunya.
Llorens Artigas 4-6, 08028, Barcelona, Spain
{fmoreno, cetto, asanfeliu}@iri.upc.es

Abstract. In this paper a new technique to perform tracking in cluttered scenarios with varying illumination conditions is presented. The robustness of the approach lies in the integration of appearance and structural information of the object. The fusion is done using the CONDENSATION algorithm that formulates multiple hypothesis about the estimation of the object's color distribution and validates them taking into account the contour information of the object.

1 Introduction

Color represents a visual feature commonly used for object detection and tracking systems, specially in the field of human-computer interaction [1][5]. For such cases in which the environment is relatively simple, with controlled lighting conditions and an uncluttered background, color can be considered a robust cue. The problem appears when we are dealing with scenes with varying illumination conditions and confusing background. For example, in the upper row of Fig. 1 we can see some frames from a motion sequence of a Lambertian surface, in which the object of interest revolves around the light source. In the lower row, we show the corresponding color distributions, (in *RGB* color space) that belong to the reddish rectangle. Last image shows the path followed by the color distribution for the entire sequence.

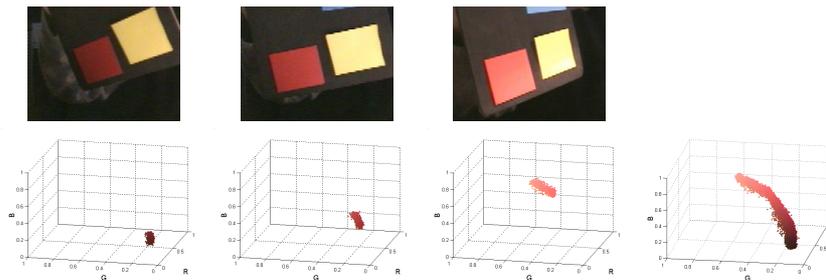


Fig. 1. Example frames of a time-varying color illuminant.

Thus, an important challenge for any color tracking system to work in real unconstrained environments, is the ability to accommodate variations in the amount of source light reflected from the tracked surface.

The choice of different color spaces like *HSL*, normalized color *rgb* ([10],[7]), or the color space $(B - G, G - R, R + G + B)$, can give some robustness against varying illumination, highlights, interreflections or changes in surface orientation (see [2] for an analysis of different color spaces). But none of these transformations is general enough to cope with arbitrary changes in illumination.

Instead of searching for color constancy, other approaches try to adapt the color distribution over time. In [6], for example, Gaussian mixtures models are used to estimate densities of color, and under the assumption that lighting conditions change smoothly over time, the models are recursively adapted. In [8], the color distribution is parameterized as a random vector and a second order Markov model is used to predict the evolution of the corresponding color histogram. These techniques perform much better than the mere change of color space representation, but have the drawback that they do not check for the goodness of the adaptation, which can lead to a failure.

The fusion of several visual modules using different criteria offers more reliability than methods that only use one feature. In this sense, the real-time head tracking system presented in [1], models the head of a person by an ellipse and uses intensity gradients and color histograms to update the head position over time. In [5], color histograms are fused with stereovision information in order to dynamically adapt the size of the tracked head. These real time applications are constrained to tracking of elliptical shapes.

In this paper, we present a new methodology that addresses the problems presented by the approaches described above, that results in a robust tracking system able to cope with cluttered scenes and varying illumination conditions. The robustness of our method lies in the fusion of color and shape information, which are iteratively adapted using the CONDENSATION algorithm [3].

Section 2 presents the main features of our method and the advantages of fusing color and shape. In Section 3 a detailed description of the method is given. Results and conclusions are presented in Sections 4 and 5, respectively.

2 Overview

Before entering into a detailed description of the proposed method we give a short glimpse of its main features:

- **Integration of color and shape information:** fusion of both vision modules makes our method appropriate to work in cluttered scenes. In Fig. 2 we can see an example of the power of this fusion in the tracking of a snail shell.

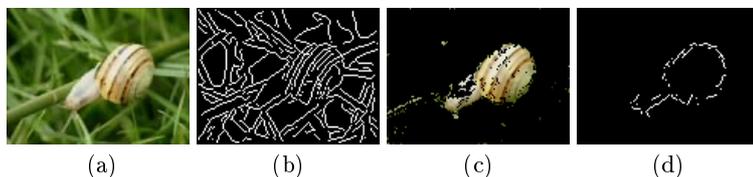


Fig. 2. Clutter in edge-map is reduced using the information from the object's color.

Fig. 2b illustrates the clutter of the scene difficulting the tracking procedure when only using edge information. If the color distribution of the shell is known, the image can be segmented via color histograms [9] (see Fig. 2c), and use this information to discriminate many false edges (Fig. 2d).

- **Ability to adapt shape deformation and varying illumination:** accommodation to varying illuminating conditions is needed to get a good color segmentation of the tracked object. As shown above (Fig. 2d), color segmentation is used to eliminate many false edges from the region of interest, simplifying a final stage of adapting a snake (maintaining affinity) to the contour of the object (assuming that the set of possible shapes of image contours do indeed form affine spaces). We introduce a restriction to the classical snake minimization procedure [4], to obtain affine deformations only. This feature makes our system robust to partial occlusions of the target.
- **Fusion of color and shape in a probabilistic framework:** the CONDENSATION algorithm offers the appropriate framework to integrate both color and contour information, and to perform tracking of the object color distribution in color space, and that of the object contour in image space, both simultaneously. That is, using the predictive filter, multiple estimates of the object color distribution are formulated at each iteration. These estimates are weighted and updated taking into account the object shape, enabling the rejection of objects with similar color but different shape than the target. Finally, the best color distribution is used to segment the image and refine the object's contour.

3 The Tracking Algorithm

In this section a detailed description of the steps used in the method is presented. For ease of explanation these steps are divided as in the CONDENSATION algorithm (Fig. 3 shows the one dimensional case):

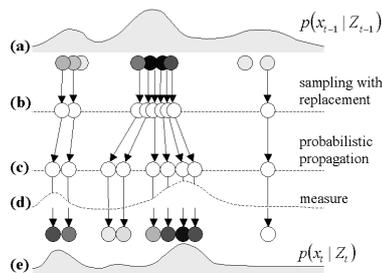


Fig. 3. One iteration of the CONDENSATION algorithm for the one-dimensional case. The weight of each sample is represented by its gray level.

3.1 Parameterization and *pdf* of color distribution

It has been pointed out that an interesting feature of the presented method is that tracking is performed simultaneously in both color and image spaces. In fact,

the element being directly tracked by the filter is the object color distribution \mathcal{C} , that at time t is the collection of all image pixel color values I_t that belong to the target, i.e. $\mathcal{C}_t = \{(R_i, G_i, B_i) \mid i = 1, \dots, M_t\}$, where M_t is the number of object points at time t , and $0 \leq R_i, G_i, B_i \leq 1$ (we assume without loss of generality, that the color space is RGB , but it is extensible to any color space). As the set of object points can be arbitrarily high, the state vector \mathbf{x}_t will be a parameterization of \mathcal{C}_t with components (adapted from [8]) $\mathbf{x}_t = [\mathbf{m}_t^\top, \boldsymbol{\lambda}_t^\top, \theta_t, \phi_t]^\top$, where $\mathbf{m}_t = [\bar{R}, \bar{G}, \bar{B}]^\top$ is the centroid of \mathcal{C}_t , $\boldsymbol{\lambda}_t = [\lambda_1, \lambda_2, \lambda_3]^\top$ are the magnitudes of the principal components of \mathcal{C}_t ; and θ_t, ϕ_t are the angles centered at \mathbf{m}_t that align the two most significant principal components of \mathcal{C}_t with respect to the principal components of \mathcal{C}_{t-1} (see Fig. 4).

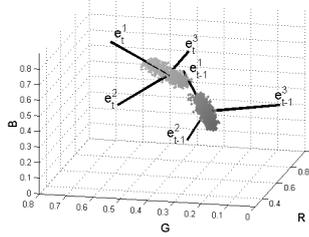


Fig. 4. Principal directions of two consecutive color distributions. The parameter θ_t is the angle that aligns \mathbf{e}_{t-1}^1 with respect to \mathbf{e}_t^1 . The parameter ϕ_t is the angle that aligns \mathbf{e}_{t-1}^2 with respect to \mathbf{e}_t^2 after having rotated \mathbf{e}_{t-1}^1 an angle θ_t around the axis $\mathbf{e}_{t-1}^1 \times \mathbf{e}_t^1$.

At time t , a set of N samples $\mathbf{s}_{t-1}^{(n)}$ ($n = 1, \dots, N$) of the form of \mathbf{x} , parameterizing N color distributions $\mathcal{C}_{t-1}^{(n)}$ are available (step (a) from Fig. 3). Each distribution has an associated weight $\pi_{t-1}^{(n)}$. The whole set represents an approximation of the *a posteriori* density function $p(\mathbf{x}_{t-1} | \mathcal{Z}_{t-1})$ (see Fig. 5), where $\mathcal{Z}_{t-1} = \{z_0, \dots, z_{t-1}\}$ is the history of measurements.

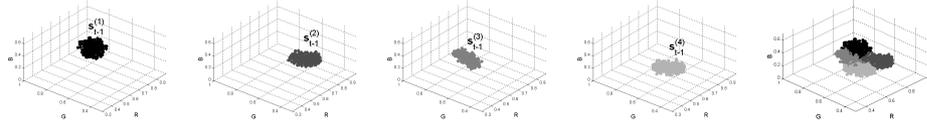


Fig. 5. Four samples of color distribution from the set $\{\mathbf{s}_{t-1}^{(n)}\}$ (in the last image, samples are shown together). Gray level is proportional to the sample weights. The set of all these distributions approximates the *a posteriori pdf* $p(\mathbf{x}_{t-1} | \mathcal{Z}_{t-1})$.

3.2 Sampling from $p(\mathbf{x}_{t-1} | \mathcal{Z}_{t-1})$

The next step in the estimation of $p(\mathbf{x}_t | \mathcal{Z}_t)$ consists of sampling with replacement N times the set $\{\mathbf{s}_{t-1}^{(n)}\}$, where each element has probability $\pi_{t-1}^{(n)}$ of being chosen (step (b) from Fig. 3). This, will give us a new set $\{\mathbf{s}_t^{(n)}\}$ of color distribution parameterizations. Those distributions having higher weights may be chosen

several times, so the new set can have identical copies of elements. On the other hand, those distributions having lower weights may not be chosen (see Fig. 6a).

3.3 Probabilistic propagation of samples

Each sample $\mathbf{s}_t^{(n)}$ of the set is propagated (see Fig. 3c and Fig. 6a) according to the following dynamic model:

$$\mathbf{s}_t^{(n)} = A\mathbf{s}_t^{(n)} + B\mathbf{w}_t^{(n)}$$

where A is the deterministic part, assigned as a first order model describing the movement of an object with constant velocity. $B\mathbf{w}_t^{(n)}$ is the stochastic component, with $\mathbf{w}_t^{(n)}$ a vector of standard normal random variables with unit standard deviation, and BB^T is the process noise covariance. The parameters A and B are estimated a priori from a training sequence.

Each predicted sample $\mathbf{s}_t^{(i)}$ represents the set of parameters defining the rigid transformations that will be used to warp the color distribution $\mathcal{C}_t^{(i)}$ associated with the sample $\mathbf{s}_t^{(i)}$, in order to obtain the new estimated distribution $\mathcal{C}_t^{(i)}$ (with parameters $\mathbf{s}_t^{(i)}$).

3.4 Measure and weight

In this step, each element $\mathbf{s}_t^{(n)}$ has to be weighted according to some measured features, and is precisely at this point where we integrate the structural information of the object's contour. From the propagated color distributions $\mathcal{C}_t^{(n)}$, we construct the color histograms $\mathcal{H}_t^{(n)}$ with $\mathcal{R} \cdot \mathcal{G} \cdot \mathcal{B}$ bins:

$$\mathcal{H}_t^{(n)}(r, g, b) = \# \left\{ (R, G, B) \in \mathcal{C}_t^{(n)} \mid \frac{r-1}{\mathcal{R}} < R \leq \frac{r}{\mathcal{R}}, \frac{g-1}{\mathcal{G}} < G \leq \frac{g}{\mathcal{G}}, \frac{b-1}{\mathcal{B}} < B \leq \frac{b}{\mathcal{B}} \right\}$$

and where $r = [1, \dots, \mathcal{R}]$, $g = [1, \dots, \mathcal{G}]$, $b = [1, \dots, \mathcal{B}]$, with $\{r, g, b, \mathcal{R}, \mathcal{G}, \mathcal{B}\} \in \mathbb{N}$. This histogram is used to generate a segmentation $S_t^{(n)}$ from the entire image I_t . That is, given a pixel $I_t(u, v)$ with color value (R, G, B) the corresponding value of the segmented image $S_t^{(n)}(u, v)$ will be assigned a value 1 if $\mathcal{H}_t^{(n)}(r, g, b) > 0$, where $r = \lfloor R \cdot \mathcal{R} \rfloor$, $g = \lfloor G \cdot \mathcal{G} \rfloor$ and $b = \lfloor B \cdot \mathcal{B} \rfloor$ (Fig. 6b).

The goal is to assign higher weights to the samples $\mathbf{s}_t^{(n)}$ generating "better" segmentations of the tracked object. To this end, simple morphological operations are performed on $S_t^{(n)}$ to extract a blob corresponding to the segmented object (Fig. 7a). After adjusting a snake along the contour of this blob, the weight assigned to $\mathbf{s}_t^{(n)}$ is computed according to the function:

$$\pi_t^{(n)} = e^{-\frac{\rho^2}{2\sigma^2}}$$

and $\rho = \mu_1(1 - \Phi_{affine}) + \mu_2(1 - \Phi_{area}) + \mu_3(1 - \Phi_{quality})$.

Functions Φ_{affine} , Φ_{area} and $\Phi_{quality}$ return a value in the range $[0, 1]$ and represent a measure of the following features:

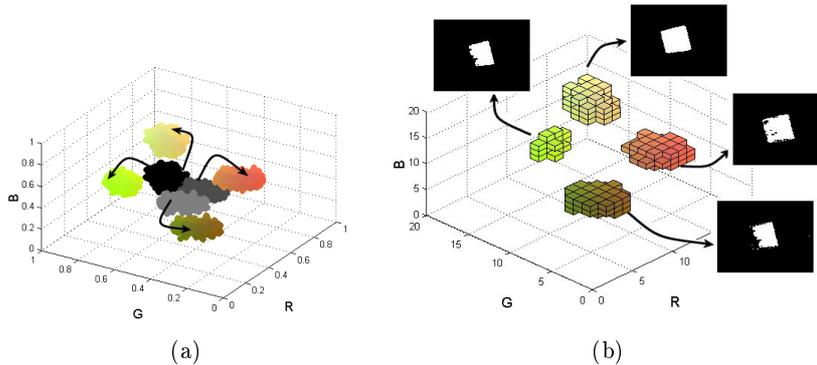


Fig. 6. (a) Sampling and probabilistic propagation from color distributions $\mathcal{C}_t^{(n)}$, of Fig. 5. Observe that the sample having the highest weight has been chosen two times, while another sample with lower weight has not been chosen. (b) Construction of the histograms $\mathcal{H}_t^{(n)}$ and results of the corresponding segmentations $\mathcal{S}_t^{(n)}$.

- **Affine similarity:** let $\nabla S_t^{(n)}$ be a binary image of the edges of $S_t^{(n)}$, and r_{t-1} a collection of image points along the snake adjusted to the contour of the object in the iteration $t-1$. r_{t-1} is used as initialization of an affine snake $r_t^{(n)}$ that is adjusted to $\nabla S_t^{(n)}$. Φ_{affine} , measures the similarity between $r_t^{(n)} = (u_t^{i,(n)}, v_t^{i,(n)})$ ($i = 1, \dots, N_r$), and $\nabla S_t^{(n)}$:

$$\Phi_{affine} = \frac{1}{N_r} \sum_{i=1}^{N_r} \nabla S_t^{(n)}(u_t^{i,(n)}, v_t^{i,(n)})$$

- **Congruent value of area:** another factor to take into account when evaluating the goodness of the segmentation $S_t^{(n)}$ is how close is the area $Area_t^{(n)}$ of the snake $r_t^{(n)}$ to the predicted area $\widetilde{Area}_t = Area_{t-1} + \mu(Area_{t-1} - Area_{t-2})$, where $Area_{t-i}$ is the area of the refined snake at iteration $t-i$ (see Sect. 3.5). This is considered in the function:

$$\Phi_{area} = |\widetilde{Area}_t - Area_t^{(n)}| / \max\{\widetilde{Area}_t, Area_t^{(n)}\}$$

- **Quality of the segmentation:** the function $\Phi_{quality}$ is introduced to penalize those segmentations of “low” quality that present some holes into the area of the segmented object. $\Phi_{quality}$ is a linear function of the Euler number of the processed $\mathcal{S}_t^{(n)}$.

Finally, the set of N weights $\pi_t^{(n)}$ associated to each of the samples $\mathbf{s}_t^{(n)}$, represents an approximation to the *a posteriori* density function $p(\mathbf{x}_t | \mathcal{Z}_t)$.

3.5 Contour updating

The last step of our algorithm, consists in refining the fitting of the object boundary, in order to obtain r_t . This is done by taking the contour of the segmented image corresponding to the best sample ($\nabla S_t^{(i)} \mid \pi_t^{(i)} \geq \pi_t^{(j)} \forall j \neq i$).

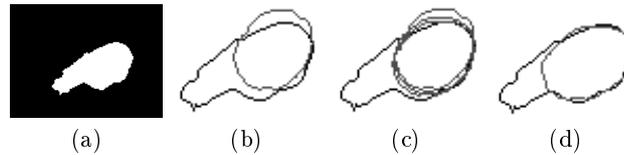


Fig. 7. (a) Segmented image $S_t^{(n)}$ after postprocessing operations (the example from Fig. 2) (b) The red curve is the initial snake (r_{t-1}). The black curve is the boundary ($\nabla S_t^{(n)}$) of the segmented image. (c) Intermediate steps of the affine snake fitting. (d) Final result of the snake fitting ($r_t^{(n)}$). Observe that if the deformations were not affine, the snake may have erroneously evolved to encompass the neck of the snail.

$i, 1 \leq j \leq N$), and instead of adjusting the snake r_{t-1} to $\nabla S_t^{(i)}$, it is adjusted to $\nabla I_t^* = \nabla I_t \cdot \text{dil}(\nabla S_t^{(i)})$, where ∇I_t is the gradient of I_t , and the function $\text{dil}(\cdot)$, refers to a morphological dilate operation. ∇I_t^* is in fact the original edge image, from which all the clutter and disturbing edges have been eliminated (see Fig. 2).

4 Experimental Results

In this Section four sets of sequence results are presented (summarized in Fig. 8) to illustrate the robustness of our system under different conditions. As the algorithm has been implemented in an interpretative language (MATLAB), speed results will not be analyzed. Attention will be focused on the effectiveness of the method. In the first experiment we show how our system is able to accommodate color by applying it over a synthetic sequence of circles moving around and changing randomly its color. In the upper left image of Fig. 8 the path of the color distributions for the tracked circle is shown. The second experiment (tracking of a colored rectangle) corresponds to the sequence introduced in Fig 2. It has to be pointed out that in the previous experiment we used the RGB color space, but in the present and subsequent experiments the color space used was the $(B - G, G - R, R + G + B)$ in order to provide robustness to specular highlights. The last two experiments, correspond to outdoor scenes, where although the change in illumination conditions is limited, they are useful to show that our method works with non-uniform shapes (third experiment of a beetle tracking), and in cluttered scenarios (fourth experiment of a snail tracking).

5 Conclusions

In this paper we have presented a new approach to the color object tracking under cluttered and varying illumination environments that dynamically accommodates the color distribution and shape of the object. The robustness of the method lies in the fusion of both modules in the probabilistic framework provided by the CONDENSATION algorithm. Results demonstrate the reliability of the tracking system in several experiments.

6 Acknowledgements

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.

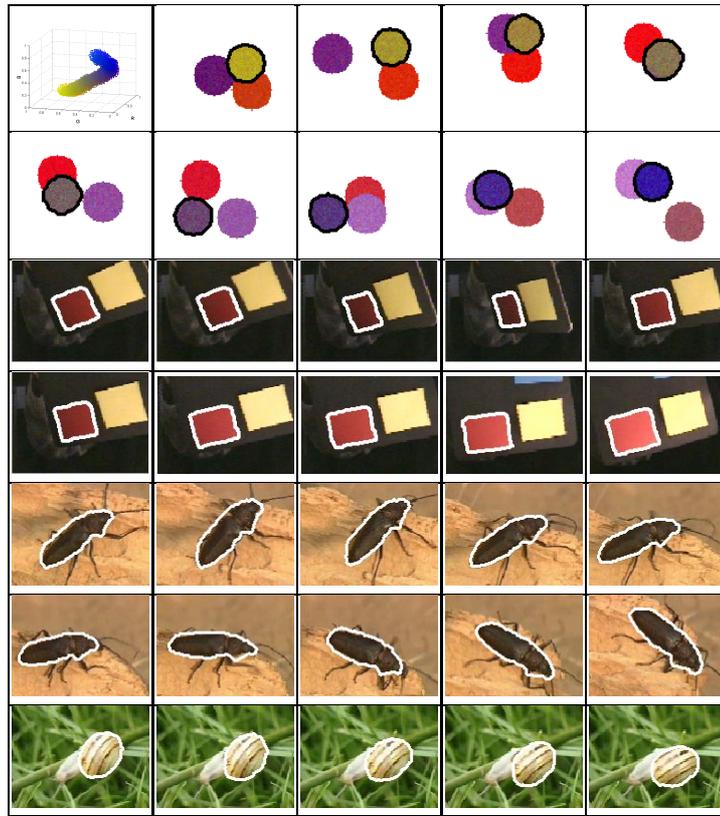


Fig. 8. Results of the 4 experiments.

References

1. Birchfield, S.: Elliptical head tracking using intensity gradients and color histograms. *CVPR*. (1998), 232–237.
2. Gevers, T., Smeulders A.: Color Based Object Recognition. *Pattern Recognition*. (1999), 32, 453–464
3. Isard, M., Blake, A.: Condensation-Conditional Density Propagation for Visual Tracking. *Int. J. Computer Vision*. (1998), 28(1), 5–28.
4. Kass, M., Witkin, A., Terzopoulos, D.: Snakes: Active Contour Models. *Int. J. Computer Vision*. (1987), 1(4), 321–331.
5. Moreno, F., Tarrida, A., Andrade-Cetto, J., Sanfeliu, A.: 3D Real-Time Head Tracking Fusing Color Histograms and Stereovision. *ICPR*. (2002), 368–371.
6. Raja, Y., Mckenna, S.J., Gong, S.: Tracking and Segmenting People in Varying Lighting Conditions using Colour. *AFG*. (1998), 228–233.
7. Schwerdt, K., Crowley, J.: Robust Face Tracking using Color. *AFG*. (2000), 90–95.
8. Sigal, L., Sclaroff, S., Athitsos, V.: Estimation and Prediction of Evolving Color Distributions for Skin Segmentation Under Varying Illumination. *CVPR*. (2000).
9. Swain, M., Ballard, D.: Color Indexing. *Int. J. Computer Vision*. (1991), 7(1), 11–32.
10. Yang, J., Lu, W., Waibel, A.: Skin-color Modeling and Adaption. *ACCV*. (1997), 687–694.