An adaptive mixture color model for robust visual tracking

A. Lehuger & P. Lechat
France Telecom R&D
4, rue du Clos Courtel
35512 Cesson-Sévigné, France

P. Pérez
IRISA/INRIA Rennes
Campus Universitaire de Beaulieu
35042 Rennes Cedex, France

Abstract

Global color characterization is a very powerful tool to model in a simple yet discriminant way the visual appearance of complex objects. A fixed reference model of this type can be used within both deterministic and probabilistic sequential estimation frameworks to track targets that undergo drastic changes of detailed appearance. However, changes of illumination as well as occlusions require that reference model is updated while avoiding drift. Within the particle filtering framework, we propose to address this adaptation problem using a dynamic mixture of color models with two components which are respectively fixed and rapidly updated. The merit of this approach is demonstrated on tracking players in team sport videos.

1 Introduction

Visual tracking of objects in videos can rely on a detailed modeling of the object’s appearance (typically at the pixel level), on a more global description (for instance of the color) or on a combination of both. Former models can be very discriminant but, as a downside, they must be updated online (unless a precise knowledge of the object is available for off-line appearance learning). This is tricky since rapid adaptability makes model prone to drift as well as to loss after occlusions.

Global color characterization with histograms is also very powerful [2, 3]. Color can indeed be very persistent despite extreme changes of detailed appearance due to pose changes, deformations, etc. Tracking players in team sport videos, which is our context of application, makes this type of tools extremely appealing. Excellent tracking can already be obtained with no adaptation of the reference and a deterministic search at each instant. Further improvements, especially regarding occlusions, have been obtained by probabilistic extension with particle filtering [8, 6]. In this context, there are still situations though, where one fills the need of adapting the color model specified at initialization time. Only very simple linear updates have been used so far [7, 4, 1]. But, as with adaptive detailed models, they can cause drift problems as well as loss during occlusions.

In an approach which is reminiscent to the online exemplar gathering of [9], we propose a new adaptive mixture color model where the original reference model is kept unchanged, but competes constantly with another model that is rapidly updated. This new approach makes use of a binary switch which is easily accommodated in a particle filter. Results on rugby games are presented.

2 Non-parametric color modeling

Reference and candidate histograms

The powerful approach introduced by Comaniciu et al. [2] relies on modeling the appearance of the target with a simple low-dimensional shape (an ellipse) and a global color histogram. Tracking then proceeds by looking in the current image for regions in the same class of shapes whose associated color histogram is similar to the reference one. Similarity is measured by the Battacharyya coefficient (other measures, such as information divergence, have also been proposed). In [2], a kernel (Gaussian or Epanechnikov), whose support defines the region of interest, is used to build the reference and candidate histograms. The use of such a differentiable kernel gives access to deterministic search of the best location in the new image thanks to a particular gradient ascent, the “mean-shift” algorithm, started at the location in the previous image or from some prediction based on it. Scale and orientation parameters, however, are not so easily searched, and one usually resorts to greedy techniques for them.

In probabilistic approaches based on the same type of modeling [8, 6], the use of a differentiable kernel is not required anymore, and simple bin counting in a rectangular region of interest for instance can be used. This is what we chose to use here. For a given partitioning of the selected color space into $B$ bins, we shall denote $q_0 = \{q_{0,u}\}_{u=1}^B$ the normalized reference color histogram associated to the region selected in the first frame of the sequence. If $p$ is the color histogram of a candidate region in some latter im-
age of the sequence, its Battacharyya coefficient with the reference is defined by
\[ \rho(q_0, p) = \frac{1}{2} \sum_{u=1}^{B} (q_{0,u} p_{u}) \]

Obviously, the definition of the reference color model is crucial both for the precision and the robustness of the tracking. With a manual initialization and a shape model as simple as an ellipse or a rectangle, it is likely that this model gets contaminated by a fraction of background included in the selection. This is especially true in our applicative context where players in the image occupy small but complex image portions surrounded by a consistently colored background (the green field).

Hence, following [3], we modify reference and candidate histograms using a simple color characterization of the surrounding. More precisely, we compute, at initialization time, the foreground color histogram \( q_0^f \), associated to the initial box, and the background color histogram, \( q_0^b \), associated to the band of vertical and horizontal widths amounting to the width and height of the rectangle. The final reference color histogram \( q_0 \) is obtained by “removing” background components from foreground model according to:
\[
q_{0,u} \propto \frac{q_{0,u}^f}{\max(q_{0,u}^b, q_{0,u}^b_{\min})}, \quad u = 1 \ldots B,
\]

where \( q_{0,u}^b_{\min} \) is the smallest non-zero component of \( q_0^b \), and the \( q_{0,u}^f \)'s are normalized to one (Fig. 1). Any candidate histogram \( p \) is modified similarly, using background model \( q_0^b \).

**Figure 1.** Extraction at initialization time of a foreground color histogram (in the bounding box around the player), of a background color histogram (in the band surrounding the box) and combination of both to get the modified reference histogram used for tracking.

**Adaptive mixture color model** Despite its surprising ability to permit robust tracking of objects whose detailed appearance changes dramatically over time, a fixed reference color model gathered at initialization might become insufficient if illumination conditions changes or if the colorimetric layout of the tracked object changes with pose (new differently colored parts appearing). For this reason, a simple linear update of the reference color histogram has been introduced in [7]. The reference model at time \( t \) being now denoted \( q_t \), this adaptation amounts to
\[
q_t = (1 - \gamma)q_{t-1} + \gamma \hat{p}_{t-1},
\]

where \( \gamma \) is an adaptation speed parameter and \( \hat{p}_{t-1} \) is either the histogram associated to the point estimate of the target region at time \( t \) or the point histogram estimate (i.e., with particle filters, either the histogram of the mean particle [7] or the average of particles’ histograms [4]).

This adaptation is limited in two ways. First it requires to set a fixed mixing rate, whereas the adaptation extent should varies over time. More importantly, as well known with adaptive pixel-wise appearance models, adaptation with exponential loss of memory often causes drift problems. Indeed, a succession of slightly misplaced results, which is likely to occur, will cause the reference model to get gradually polluted by background, until is does not allow any longer the tracking of the target of interest. The problem is even more acute when temporary occlusions occur, resulting in the reference model to learn fast the color of the occluder on which the tracker will eventually lock.

To circumvent these problems, while being able to depart momentarily from the original reference model when required, Collins et al. [1] propose to anchor the model on the original one by using \( q_0 \) instead of \( q_{t-1} \) in (2), leading to
\[
q_t = (1 - \gamma)q_0 + \gamma \hat{p}_{t-1}.
\]

We extend this approach by resorting to a dynamic mixture of the form:
\[
q_t = \begin{cases} q_0 & \text{if } s_t = 0, \\ \hat{p}_{t-1} & \text{if } s_t = 1, \end{cases}
\]

where \( s_t \) is an auxiliary binary switch between the original reference model and another one which is very rapidly refreshed. In our experiments, \( \hat{p}_{t-1} \) is simply gathered at the estimated location in previous frame (including colour model \( q_{t-1}^b \) used to remove background contamination). As we shall see, the definition of an appropriate Markovian dynamical prior on this discrete process will permit to rely on the original model most of the time, while switching (statistically) to the rapidly updated one over short time intervals.

### 3 Bayesian filter

**Model ingredients** For a fixed aspect ratio, the candidate bounding box around the target at time \( t \) is defined by its center location \( (x_t, y_t) \) and its scale \( \alpha_t \), which form the continuous part \( x_t = (x_t, y_t, \alpha_t) \) of the state vector. As explained above, we introduce a binary switch \( s_t \in \{0, 1\} \) pointing either to the original reference model \( (z_t = 0) \) or to the rapidly updated one \( (z_t = 1) \). The compound state is then \( X_t = (x_t, s_t) \).
Unless geometric information that allow mapping of image coordinates at each instant to field positions are available (which requires preliminary calibration and remains nonetheless problematic with moving cameras), it is difficult to tune precise and physically meaningful dynamics in our applicative context. Hence, as in a number of other works where detailed dynamics are also not available, we resort to a simple random walk models on each component of the continuous state $x_t$, with respective variances $\sigma_x^2$, $\sigma_y^2$ and $\sigma_s^2$ forming diagonal matrix $\Sigma$.

The discrete switch is itself driven by an independent first-order homogeneous Markov chain with transition matrix

$$T = \begin{bmatrix} \alpha_0 & 1 - \alpha_0 \\ 1 - \alpha_1 & \alpha_1 \end{bmatrix}. \quad (4)$$

Since the rapidly updated model will be, by definition, often closer to candidate histograms than the reference model is, the prior on $s_t$ must strongly compensate for this bias, such that the original reference model remains the privileged one and drift is avoided (especially after occlusion). To this end, we chose $\alpha_0 = 0.9$ and $\alpha_1 = 0.7$ in our experiments.

As for the observation model, we follow [8, 6], with the required adaptation to our two-fold color modeling. If $p_t(x_t)$ denotes the color histogram of the rectangle defined by $x_t$ in the image at time $t$, modified by $q_{0t}^b$, $p_t(x_t)$ the same modified by $q_{0t}^b$ instead, and $z_t$ stands of the image measures at this instant, the hidden state likelihood is chosen as:

$$p(z_t|X_t) \propto \exp \lambda \rho[(1 - \delta_{s_t})q_{0t} + \delta_{s_t}p_{t-1}(\tilde{x}_{t-1})p_t(x_t)] \quad (5)$$

whit $\delta$ the Kronecker delta-function. Because of the hybrid nature (continuous-discrete) of the compound state-space, we define the point estimate $\hat{X}_t = (\hat{x}_t, \hat{s}_t)$ the following way:

$$\hat{s}_t = \arg\max_{s_t} p(s_t|z_{1:t}), \quad \hat{x}_t = \mathbb{E}(x_t|\hat{s}_t, z_{1:t}). \quad (6)$$

**Particle filter implementation** The combination of a highly non-linear and multi-modal observation model with a discrete-continuous state space makes the use of particle filtering especially appealing. We resort here, as in [8, 6], to the simple bootstrap filter [5], where the proposal density function according to which samples are moved at current time coincides with the dynamical model defined over the state space. Given the set $(X^{(n)}_{t-1}, w^{(n)}_{t-1})_{n=1\ldots N}$ of $N$
weighted samples that approximate the filtering distribution \( p(\tilde{x}_{t-1}\mid z_{1:t}) \) at time \( t-1 \), the overall procedure at time \( t \) is as follows:

- **Prediction**: for \( n = 1 \cdots N \), draw \( \tilde{x}_t^{(n)} \) from \( X_t^{(n)} \sim N(x_{t-1}^{(n)}; x_{t-1}, \Sigma) \), \( \Pr(s_t^{(n)} = s_{t-1}^{(n)}) = \alpha_s^{(n)} \).

- **Weight update**: for \( n = 1 \cdots N \),
  \[
  w_t^{(n)} = \frac{w_{t-1}^{(n)} \exp \lambda \delta(1 - \delta(s_t^{(n)}))q_0 + \delta(s_t^{(n)})p_{t-1}(\tilde{x}_{t-1}), p_t(\tilde{x}_t^{(n)})}{\sum_n w_t^{(n)}} = 1.
  \]

- **Resampling**: for \( n = 1 \cdots N \), draw \( a_n \) from \( \Pr(a_n) = w_t^{(n)} \) and set \( (X_t^{(n)}, w_t^{(n)}) \leftarrow (X^{(a_n)}, \frac{1}{N}) \).

- **Monte Carlo point estimation**
  \[
  \hat{s}_t = \arg \max_{s_t \in \{0,1\}} \sum_{n: a_n = s_t} w_t^{(n)}
  \]
  \[
  \hat{x}_t = \frac{\sum_{n: a_n = s_t} w_t^{(n)} x_t^{(n)}}{\sum_{n: a_n = s_t} w_t^{(n)}}
  \]

4 Results and perspectives

We report sample results of tracking players in a rugby game broadcast. For each experiment, the initialization is manual, particle filters are run with \( N = 300 \) particles and color histograms are formed by concatenating 64-dimensional marginal histograms in R, G, B and H (hue) channels. Histograms are thus of dimension 256. Using marginal histograms instead of a joint one allows computational savings (our method runs in real time).

In the first example (Fig.2) the field of view is large and players are just a few pixels high, whereas, in the second example (Fig.3), the field of view is narrower with larger players in the image. In both cases, the tracked player undergoes temporary occlusions (behind an opponent or a pole) and transition from a brightly lit portion of the field to a shaded one.

These examples demonstrate the improved ability of the adaptive mixture color model to deal with illumination changes and occlusions. In the latter cases, the rapidly adapted model captures the appearance of the occluder as the occlusion proceeds, but, contrary to more traditional adaptive models, our model is able to rapidly switch back to the original color model when the target re-appears.

As for perspectives, the proposed model could be extended in two different ways. Similar to the exemplar catalog of Rahimi et al. [9], more than two color models could be used. Secondly, a more sophisticated update mechanism could be devised for the rapidly adapted color model.

References


