EBE: ELASTIC BLOB ENSEMBLE FOR COARSE HUMAN TRACKING

Daniel Węsierski, Patrick Horain [*]

Institut Mines-Télécom / Télécom SudParis
Département Electronique et Physique
9 rue Charles Fourier, 91011 Evry, France
{first.last}@telecom-sudparis.eu

Zdzisław Kowalczuk

Gdansk University of Technology
Department of Decision Systems
Narutowicza 11/12, 80-233 Gdansk, Poland
zdzislaw.kowalczuk@eti.pg.gda.pl

ABSTRACT

We propose a novel probabilistic tracking algorithm based on an *elastic blob ensemble* (EBE) which is applicable to track flexible objects. It outputs a coarse motion cue in the form of the object's location and orientation together with the location of the blobs. The main assumption is that the orientation of the whole object does not change much between neighboring frames. A discrete solution space is created in the current frame around the blobs' positions from the previous frame. Our model then promotes solutions whose orientations are close to the prior orientation, which match the modeled to the observed appearance well, and which follow modeled spatial configuration. It combines the strengths of three popular approaches to visual tracking : mean-shift tracker, particle filtering, and pictorial structures. As a result, the proposed framework tracks human upper body composed of head, torso, and hips at the rate of 40 fps on a regular desktop PC.

***Index Terms*—** Kernel collaboration, blob tracking, mean-shift tracking, pictorial structures

1. INTRODUCTION

Vision-based human tracking is a very important open problem. A successful tracking procedure will span numerous applications, for instance automated surveillance, gaming, or smart video indexing. Consequently, many human tracking approaches have been proposed. Among them blob tracking has become popular due to its general simplicity.

Numerous blob trackers exist, the space of the paper, though, does not allow for a proper review. One of the most popular blob tracking procedures is a mean-shift kernel tracker [1]. It is fast and demonstrates robustness due to the applied Bhattacharyya measure over kernel modulated histograms. However, the procedure has several significant limitations. First, it does not allow for search over kernel orientations. This problem can be solved by partitioning a blob into several smaller ones [2]. Second, it requires the kernel not to move between two consecutive frames further than its radial range - an assumption which is often violated in practice. It may be overcome by fusing mean-shift tracker with a particle filter [3], [4]. One starts mean-shift convergence from numerous positions instead of the previous position solely. Directly extending these approaches to multiple blobs increases exponentially the dimensionality of the searched space and hence the computational load. A further limitation lies in the assumed objects rigidity. Again, one can partition the object into less non-rigid parts and require a collaboration mechanism between the kernels overlaying the parts [5].

In this paper, we present a probabilistic tracking procedure which is able to react to and compute an orientation change while relaxing the object's rigidity and the aforementioned plane motion constraint. Namely, rather than using a single blob, we span the object with several small circular blobs and hence compute its orientation without explicit search. Unlike other approaches, we link the blobs kinematically under a probabilistic model to allow for object's elasticity. We therefore cast a blob collaboration in the framework of a pictorial structure [6]. Moreover, we track the blobs jointly by applying a particle filter to each blob. That is, we sample possible blobs' locations around their previous positions from a uniform distribution. Thus our procedure can track rapid motions. The framework is represented as a Markov Random Field (MRF), whose MAP inference is solved classically using dynamic programming.

The paper is organized as follows. In Section 2 we present our probabilistic model for blob collaboration with application to human upper body tracking. Next, our tracking procedure is described in Section 3. Experimental results are shown in Section 4. We summarize our tracking approach in Section 5.

2. ELASTIC BLOB ENSEMBLE MODEL

We define human tracking in a video sequence as a data association problem. Our goal consists in tracking blobs with a semantic meaning : human head, torso, and hips. We seek to estimate their position in the image. The orientation of the

[*]. This work was supported by European FP7 project 216487 CompanionAble.

human spine on which these body parts reside is also estimated. The spine is approximated as a line, which may bent in a constrained fashion. We call this configuration *Elastic Blob Ensemble* (EBE).

Graph definition. In our setting, a blob ensemble $B = \{b_i\}_{i=1}^K$ is a chain of K blobs admitting a fixed order across given video during tracking. The blob ordering is represented in the form of an MRF chain graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ with nodes $\mathcal{V}$ indicating the blobs and edges $\mathcal{E}$ indicating their particular configuration. The blobs b_i are modeled by their center location $l_i^t = [x_i^t, y_i^t]^{\mathrm{T}}$ in the current t-th frame I^t. The posterior of the elastic model over $L^t = \{l_i^t\}_{i=1}^K$ takes the known form of a pictorial structure [6] :

$$P(L^t|I^t) \propto \prod_{i=1}^{K} P(I^t|l_i^t) \prod_{(i,i+1)\in\mathcal{E}} P(l_i^t, l_{i+1}^t) \qquad (1)$$

Appearance term. The likelihood $P(I^t|l_i^t)$ corresponds to the local image evidence for the blob b_i. For computational efficiency, we do not search explicitly over the orientation of the blob. As proposed in [1], we model its appearance as a normalized m-dimensional $\texttt{RGB}$ color histogram $q_i = \{q_{i,u}\}_{u=1}^m$ weighted by Epanechnikov kernel. We choose isotropic kernels to ensure appearance invariance to orientation change. Now, the match between the model appearance q_i of the blob b_i with a candidate histogram $p(l_i^t) = \{p_u(l_i^t)\}_{u=1}^m$ extracted at l_i^t is then evaluated by the distance $D(p(l_i^t), q_i) = \sqrt{1 - \rho(p(l_i^t), q_i)}$, where $\rho(p(l_i^t), q_i) = \sum_{u=1}^m \sqrt{p_u(l_i^t) q_{i,u}}$ is the Bhattacharyya similarity measure. Thus we have :

$$P(I^t|l_i^t) \propto \exp(-D^2(p(l_i^t), q_i)) \qquad (2)$$

where D^2 is the squared distance, as in [2]. Other cues, e.g. obtained after the background subtraction assuming a still camera, can extend multiplicatively the appearance term in a straightforward fashion.

Spatial term. The spatial kinematic prior $P(l_i^t, l_{i+1}^t)$ between two neighbor blobs deserves our particular attention. In a typical pictorial structure framework, human body parts are parameterized additionally by orientation. This extra parameter is important as it allows for defining local coordinate systems for each part. On the other hand, this allows for an angular kinematic constraint within a pair of body parts (e.g. legs rather do not point up to the torso). Since the blobs are circular and parameterized merely by their center location, the kinematic constraint appears only in the form of undirected distance. Clearly, one might construct a local coordinate system between two neighbor blobs in order to constraint the angular motion of the blobs locally, e.g. to prevent them from flipping over between successive frames. Unfortunately, forming local coordinate systems between pairs of blobs is inefficient as the blob ensemble rolls quickly into *spaghetti*. Since we are generally interested in tracking human body parts which rest on the spine, we allow only for a limited amount of blobs elasticity. A remedy could be found by creating a graph with higher order cliques. This would, however, hamper the speed performance during inference. Consequently, we propose to model the *whole* blob ensemble by single orientation θ_B^t w.r.t. the image coordinate system. It is determined by first eigenvector of covariance matrix computed from L_{MAP}^t. As a consequence, our probabilistic blob ensemble collaborates locally w.r.t. the distance, and globally w.r.t. the orientation. Hence we define $P(l_i^t, l_{i+1}^t)$ as :

$$\mathcal{N}(\left\|l_i^t - l_{i+1}^t\right\|_2 ; \mu_{i,i+1}, \sigma_{i,i+1}^2)\mathcal{M}(\theta_{i,i+1}^t; \theta_B^{t-1}, k) \qquad (3)$$

where $\mathcal{M}$ denotes the von Mises distribution with k as angular stiffness [6]. The variable $\theta_{i,i+1}^t$ is defined as the angular displacement of pairs of blobs in the current frame I^t w.r.t. the image coordinate system. The parameters $\mu_{i,i+1}$ and $\sigma_{i,i+1}^2$ denote mean euclidean distance between locations of two neighbor blobs b_i and b_{i+1} and its variance, respectively.

In conclusion, our final pictorial structure model is defined by the set of appearance parameters $\Theta = \{q_i\}_{i=1}^K$, kinematic connections $\Psi = \left\{\mu_{i,i+1}, \sigma_{i,i+1}^2\right\}_{i=1}^{K-1}$, and angular constraints $\Phi = \left\{\theta_B^{t-1}, k\right\}$, where θ_B^{t-1} denotes orientation of the whole ensemble estimated in the previous frame.

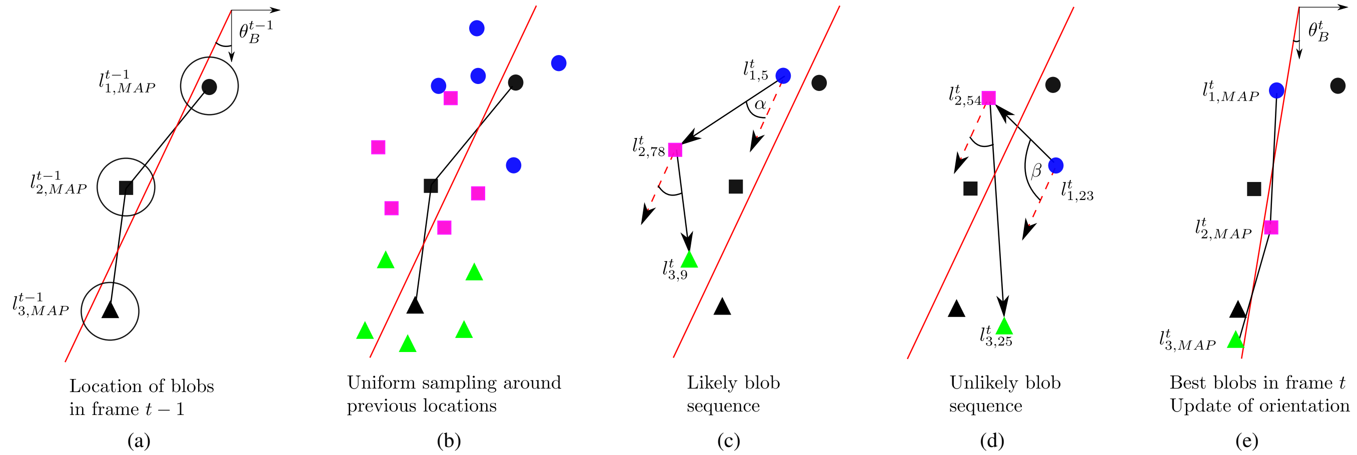
Inference. We localize the best blob configuration in the current frame I^t using dynamic programming. The inference corresponds to the MAP estimate of the pictorial model :

$$L_{MAP}^t = \underset{L^t}{\operatorname{argmax}} \; P(L^t|I^t) \qquad (4)$$

3. HUMAN TRACKING WITH EBE

Beneath we propose a tracking algorithm. In this paper, the tracker is initialized *manually* on human body parts lying on the spine : head, torso, hips. Alternatively, one might use an automatic initialization procedure, as in [7]. We assume a rather constant scale of the tracked human throughout the whole video sequence.

The tracking procedure is depicted in Fig. 1 and summarized in Algorithm 1. We track a blob ensemble by first generating H location hypotheses for each of K blobs b_i within a certain range R over the blob's location in the previous frame. We sample from a uniform motion prior over l_i to be able to deal with unexpected and fast motions (Fig. 1(b)). This also increases chances for the tracker to recover from a partially lost track caused by e.g. occlusions. To find the sequence of blobs which best explains the current image observation, the MRF chain is created with K nodes, each having H states determined by generated hypotheses. For each hypothesized location of each blob, we compute its appearance score using Eq. 2, which indicates how well the blob's modeled appearance matches the image. We also compute kinematic score between locations of pairs of blobs l_i, l_{i+1} using Eq. 3 given the orientation of the whole EBE computed in the previous frame θ_B^{t-1}. Specifically, in Fig. 1(c) a connection between hypothesized locations of the pair of blobs 1 and 2 has high



Location of blobs
in frame $t-1$

(a)

Uniform sampling around
previous locations

(b)

Likely blob
sequence

(c)

Unlikely blob
sequence

(d)

Best blobs in frame t
Update of orientation

(e)

Fig. 1: The tracking procedure on synthetic example. (a) Estimated location L_{MAP}^{t-1} and orientation θ_B^{t-1} of $K = 3$ blobs in frame $t-1$. (b) Uniform sampling around blobs' locations from previous frame to generate hypotheses on blobs' locations in frame t. (c) A blob sequence with high kinematic score (small α). (d) A blob sequence with low kinematic score (large β). (e) L_{MAP}^t of the blob ensemble inferred from MRF in frame t. The new blobs' locations are used, in turn, to compute new orientation θ_B^t of EBE.

score as the angular displacement α w.r.t. θ_B^{t-1} is small. On the other hand, Fig. 1(d) depicts a situation where a connection between blob 1 and 2 has very low score as the angular displacement β is very large. Finally, we infer on this model using Eq. 4 to get the MAP blob sequence L_{MAP}^t for frame I^t in $\mathcal{O}(KH^2)$ time. The object's location can be then computed as an unweighted mean of L_{MAP}^t.

Algorithm 1 EBE tracker.

```
Fix $K$-blobs, $H$-hypotheses, $R$-sampling radius length
**Initialization :**
Find $L^0 = \{l_i^0\}_{i=1}^K$ and compute $\{\Theta, \Psi, \Phi\}$ based on $L^0$
**Iterations :**
For frames $t = 1 \dots$ do
  For blobs $i = 1 \dots K$ do
    1. Generate $H$ hypotheses $l_{i,H}^t = \left\{l_{i,h}^t\right\}_{h=1}^H$
       by sampling uniformly around $l_{i,MAP}^{t-1}$ within $R$
  1. Set $L_H^t = \{l_{i,H}^t\}_{i=1}^K$ and build MRF chain graph :
    1.1. Compute appearance scores based on $\{\Theta, L_H^t\}$
    1.2. Compute kinematic scores based on $\{\Psi, \Phi, L_H^t\}$
  2. Solve $L_{MAP}^t$ with dynamic programming
  3. Update $\theta_B^t$ based on covariance computed from $L_{MAP}^t$
```

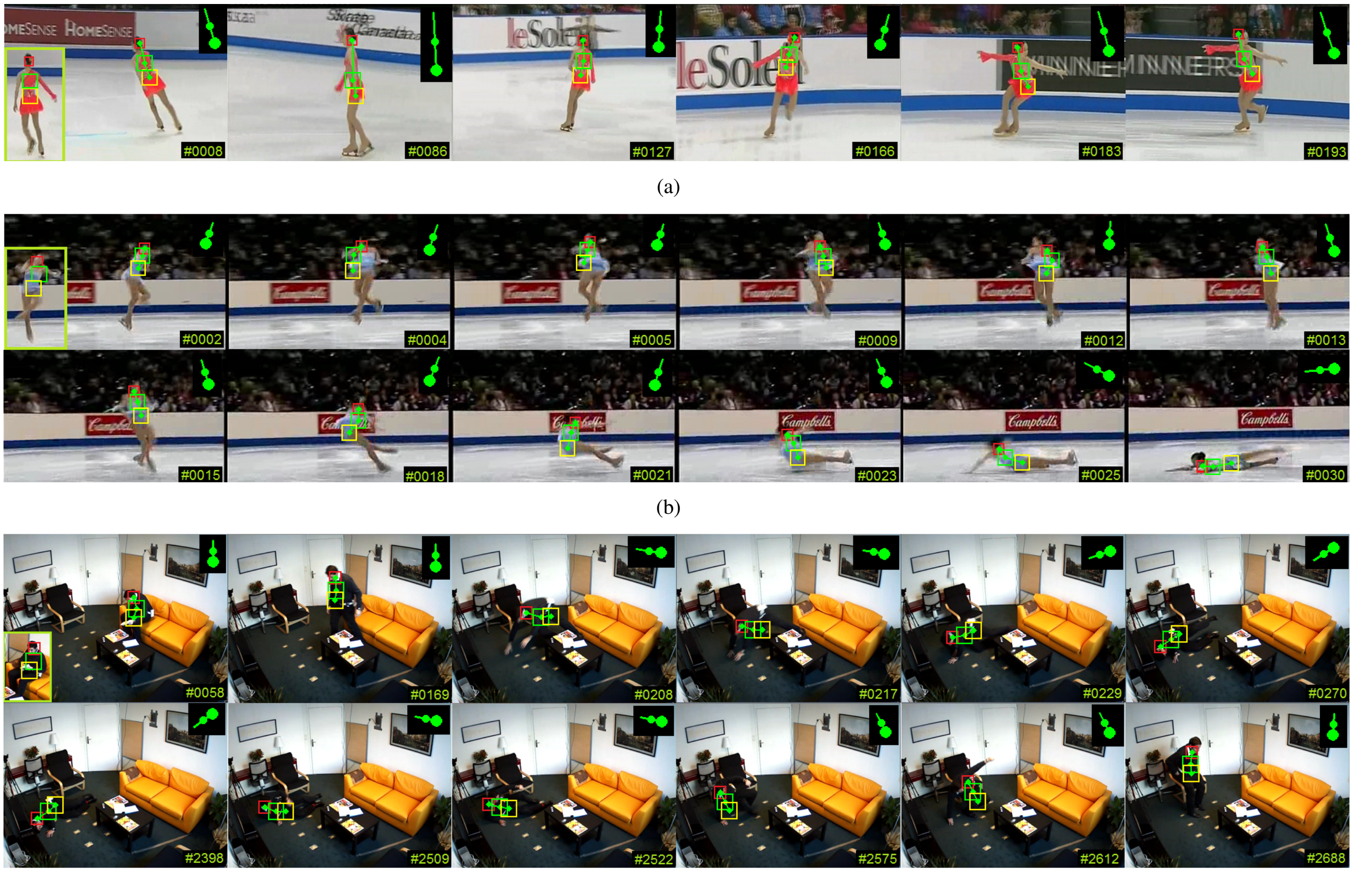
4. RESULTS

The EBE tracker ran on 2 GHz Pentium 4 processor which was connected to 3.5 GB RAM unit through 32-bit data bus with the support of 4 MB cache. The code was compiled with VC++ compiler under Windows 7 environment.

We fixed all the parameters in the experiments. The number of samples for each blob was set to $H = 100$. We sampled $H-1$ times, as one sample was the blob's location in the previous frame to explain a potential static behavior of the blob. As the main bottleneck is computing the color histograms and evaluating them, the speed of the algorithm depends on the number of both blobs K and hypotheses H. For the sequences shown in Fig. 2, the registered tracking speed was 40 fps. However, in case of tracking multiple people, possible speed-up might be achieved using an integral histogram [8].

We initialized the tracker manually for all video sequences roughly specifying locations and sizes of all $K = 3$ blobs, represented as 512-dimensional RGB color histograms. The mean distance between two consecutive blobs was defined as $\mu_{i,i+1} = \left\|l_i^0 - l_{i+1}^0\right\|_2$ and its standard deviation $\sigma_{i,i+1}$ as the average of two radii of blobs b_i and b_{i+1}. To be able to follow very fast motions of the upper body parts, we set the angular stiffness to $k = 0.4$, which approximately corresponds to the angular standard deviation of 90°, and the sampling radius R was set to two times the radius of each blob. Except for the orientation of EBE θ_B^t, we did not update any other parameters $\{\Theta, \Psi, \Phi\}$ online.

Fig. 2 shows the tracking results of our algorithm. In *Turning skater*, a skater is making turns from back, to side, to front, changing her appearance. She is tracked with correct estimates of position and orientation despite the change of scale (frame 86), which is currently not built into our model. In *Falling skater*, a skater is performing an unsuccessful jump and falling down rapidly. Despite the very high dynamics of such a motion, the tracker is able to follow the person precisely. In *Falling/Standing*, a person is standing up and falling down immediately afterwards. He lies on the floor for some time, and then stands up. Note a low contrast between the person and the background during falling down and standing up after the fall. The tracker was not perfectly initialized as the



(a)

(b)

(c)

Fig. 2: (a,b,c) show performance of our EBE tracker (best viewed in color). Initialization of appearance, position, and size of each blob is shown in the left bottom corner of the first image. Additionally, per frame orientation estimate is shown in the top right corner of each image. (a) *Turning skater* sequence. (b) *Falling skater* sequence. (c) *Falling/Standing* sequence.

person was holding an object in his hands. In spite of this, the tracker was correct. It partially lost track (wrong orientation) in frames 217 and 2575 but quickly recovered. The results show that the EBE tracker produces reliable motion trajectories despite self-occlusions, high motion dynamics, low image contrast, and partial scale change.

5. CONCLUSIONS

We have presented a novel human tracking procedure. It combines the strengths of mean-shift tracking, particle filtering, and pictorial structures. In particular, we have reformulated a classic tree-like pictorial structure into MRF chain and introduced a global orientation parameter, which jointly controls the amount of elasticity of the whole blob ensemble.

Despite the inevitable appearance changes of the tracked human, the proposed blob collaboration scheme allows the tracker to output human's orientation and position of its body parts reliably. This can be further utilized potentially by human fall detection algorithms.

6. REFERENCES

[1] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 564–575, 2003.

[2] P. Perez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in *ECCV*, 2002, pp. 661–675.

[3] E. Maggio and A. Cavallaro, "Hybrid particle filter and mean shift tracker with adaptive transition model," in *Proc. Int. Conf. Acoustics, Speech, and Signal Processing*, 2005, pp. 221–224.

[4] B. Han, D. Comaniciu, Y. Zhu, and L. S. Davis, "Incremental density approximation and kernel-based bayesian filtering for object tracking," in *CVPR*, 2004, pp. 638–644.

[5] Z. Fan, M. Yang, and Y. Wu, "Multiple collaborative kernel tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 7, pp. 1268–1273, 2007.

[6] P. F. Felzenszwalb and D. P. Huttenlocher, "Pictorial structures for object recognition," *International Journal of Computer Vision*, vol. 61, no. 1, pp. 55–79, 2005.

[7] D. Ramanan and D. A. Forsyth, "Finding and tracking people from the bottom up," in *CVPR*, 2003, pp. 467–474.

[8] F. M. Porikli, "Integral Histogram : A fast way to extract histograms in cartesian spaces," in *CVPR*, 2005, pp. 829–836.