A PROBABILISTIC COOPERATION BETWEEN TRACKERS OF COUPLED OBJECTS

Ido Leichter, Michael Lindenbaum and Ehud Rivlin

Faculty of Computer Science
Technion - Israel Institute of Technology
Technion City, Haifa 32000, Israel

ABSTRACT

Much work has been done in the field of visual object tracking, yielding a wide range of trackers, including ones aimed for multiple objects. In many cases, there may be a coupling between simultaneously tracked objects, e.g., the locations of some person's eyes. In such cases, tracking each object independently, or using any multi-target tracker ignoring this coupling, will be suboptimal. This paper addresses these cases, and takes advantage of the coupling between the tracked objects to enhance the tracking performance. An analytically justified, probabilistic framework for cooperating between the individual trackers is suggested. The framework is fairly general, allowing to cooperate between any two trackers which output a probability density function of the tracked state, even when the objects are tracked in different state spaces. The framework is successfully tested on two different kinds of trackers, showing the benefit gained from the coupling exploitation.

1. INTRODUCTION

Much work has been done in field of visual object tracking, yielding a wide range of tracking algorithms, including ones suited for multiple objects (e.g., [1], [2], [3] and other contributions in [4]). In cases where the simultaneously tracked objects are coupled, a large enhancement in tracking performance may be gained if the tracker cooperation will exploit this coupling. In these cases, each state bears information on the other states. Thus, an improvement in the performance of each tracker may be gained by taking into consideration the estimates of other trackers.

This paper tackles this kind of cases and develops an analytically justified, probabilistic framework for cooperating between two trackers in such cases. The framework is fairly general, allowing cooperation between any two trackers which output a PDF (probability density function) of the tracked state, even when the objects are tracked in different state spaces. Another advantage of the suggested framework is its treatment of the individual trackers as "closed boxes" (or almost as such), easing the implementation from the software aspects.

This approach is related to the framework described in [5], focusing on the combination of different tracking algorithms for tracking a single common object.

2. FRAMEWORK

We assume the following regarding the individual trackers:

1. The trackers provide a PDF estimate of the tracked state, sequentially for each image. Such trackers are very common. For example, any tracker using Kalman filtering explicitly provides a Gaussian PDF of the tracked state (e.g., [7]). Other trackers employing a general discrete probability distribution for tracking are described in [8], [9] and [10]. Exemplar-based trackers are another example [11].

2. A (possibly approximated) probability distribution of the tracked state of an object, conditioned on the state of the other object, is supplied. That is, denoting the state space of tracker A_i by $\mathcal{S}_i$ ($i = 1, 2$), then $f_{\mathcal{S}_1 \to \mathcal{S}_2}(\mathbf{x}_2|\mathbf{x}_1)$ and $f_{\mathcal{S}_2 \to \mathcal{S}_1}(\mathbf{x}_1|\mathbf{x}_2)$ where $\mathbf{x}_i \in \mathcal{S}_i$, are used by the cooperation process.

3. The trackers are *conditionally independent*, i.e., each tracker relies on features which, given the state of one of the tracked objects, are conditionally independent of the features used by the other tracker. That is, if tracker A_1 uses features $\mathbf{z}_1$ and tracker A_2 uses features $\mathbf{z}_2$, then we assume for every state $\mathbf{x}$ of any of the tracked objects that

$$f(\mathbf{z}_1, \mathbf{z}_2|\mathbf{x}) = f(\mathbf{z}_1|\mathbf{x}) \cdot f(\mathbf{z}_2|\mathbf{x}). \tag{1}$$

Note that since each tracker relies on features which are related only to the object it tracks, each tracker observes different pixels. Thus, the conditional independence assumption is very conservative and regularly does not pose any restrictions.

3. TRACKER COOPERATION

3.1. Fusing PDFs

Before approaching the goal of tracker cooperation, let us deal with the simpler task of fusing two conditionally independent PDF estimates of a common tracked state. The fusion developed hereinafter is then used in Section 3.2 for cooperating between the trackers.

0-7803-8554-3/04/$20.00 ©2004 IEEE.

Consider the case where we are given two conditionally independent PDF estimates of a tracked state $\mathbf{x}$ at time t: $f_t(\mathbf{x}|\mathbf{z}_1, f_{t-1})$ and $f_t(\mathbf{x}|\mathbf{z}_2, f_{t-1})$, where $\mathbf{z}_1$ and $\mathbf{z}_2$ are the features used by the two estimates, respectively. Note that both PDFs are conditioned also on the PDF at the previous time f_{t-1}. $f_t(\mathbf{x}|\mathbf{z}_1, \mathbf{z}_2, f_{t-1})$ – the state PDF using both $\mathbf{z}_1$ and $\mathbf{z}_2$, may now be derived. Applying Bayes' rule [12] and assumption (1), and noting that $f_t(\mathbf{z}|\mathbf{x}, f_{t-1}) = f_t(\mathbf{z}|\mathbf{x})$, it may be shown that (see [6])

$$f_t(\mathbf{x}|\mathbf{z}_1, \mathbf{z}_2, f_{t-1}) = k \cdot \frac{f_t(\mathbf{x}|\mathbf{z}_1, f_{t-1}) f_t(\mathbf{x}|\mathbf{z}_2, f_{t-1})}{f_t(\mathbf{x}|f_{t-1})}, \quad (2)$$

where $k = \frac{1}{\int_{\mathcal{S}} \frac{f_t(\mathbf{x}'|\mathbf{z}_1, f_{t-1}) f_t(\mathbf{x}'|\mathbf{z}_2, f_{t-1})}{f_t(\mathbf{x}'|f_{t-1})} d\mathbf{x}'}$ does not depend on the state $\mathbf{x}$ ($\mathcal{S}$ denotes the space of the tracked state). Note that the multiplication by k in (2) is equivalent to scaling $\int_{\mathcal{S}} f_t(\mathbf{x}|\mathbf{z}_1, \mathbf{z}_2, f_{t-1}) d\mathbf{x}$ to unit. Now we see that the two PDF estimates $f_t(\mathbf{x}|\mathbf{z}_1, f_{t-1})$ and $f_t(\mathbf{x}|\mathbf{z}_2, f_{t-1})$ may be easily fused by their multiplication, followed by division by the PDF $f_t(\mathbf{x}|f_{t-1})$, and scaling to have a unit integral.

The PDF $f_t(\mathbf{x}|f_{t-1})$ is usually referred to as the *prior* PDF – the PDF of the tracked state predicted prior the measurements at time t. In our experiments, we make a worst case simplification and assume that no knowledge regarding the tracked object dynamics is given in the cooperation (the individual trackers, however, do use a motion model). Thus, we set the prior PDF to a uniform one, which allows treating the individual trackers as "closed boxes".

One important special case is the case of normal densities. In such a case, it can be shown (see [6]) that the fused PDF remains normal $N(\mu, C)$, with the covariance matrix and mean being

$$\begin{aligned} C^{-1} &\triangleq C_1^{-1} + C_2^{-1} - C_3^{-1} \\ \mu &\triangleq C(C_1^{-1}\mu_1 + C_2^{-1}\mu_2 - C_3^{-1}\mu_3), \end{aligned} \quad (3)$$

where $N(\mu_1, C_1)$ and $N(\mu_2, C_2)$ are the two original PDF estimates, and $N(\mu_3, C_3)$ is the prior PDF.

3.2. Cooperating trackers

We now turn to the cooperation of A_1 and A_2, two individual trackers, tracking a pair of coupled states $\mathbf{x}$ and $\mathbf{y}$. Denote by $\mathcal{S}_1$ and $\mathcal{S}_2$ the corresponding state spaces, and denote by $\mathbf{z}_1$ and $\mathbf{z}_2$ the features used by A_1 and A_2, respectively. See illustration in Fig. 1.

Since the two tracked states are coupled, each individual tracker may be improved by taking into consideration the PDF estimate of the other tracker. In order to combine the trackers' estimations, each provided PDF has to be translated into a PDF on the other state space. This may be accomplished given the PDFs of the state on one space conditioned on the state in the other space – $f_{\mathcal{S}_2 \to \mathcal{S}_1}(\mathbf{x}|\mathbf{y})$ and $f_{\mathcal{S}_1 \to \mathcal{S}_2}(\mathbf{y}|\mathbf{x})$, where $\mathbf{x} \in \mathcal{S}_1$ and $\mathbf{y} \in \mathcal{S}_2$ (as was

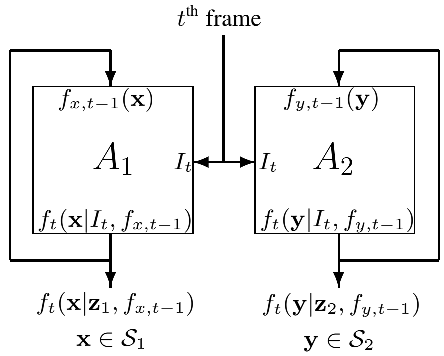


Fig. 1. Two individual trackers, tracking the states $\mathbf{x}$ and $\mathbf{y}$.

outlined in Section 2). Using these PDFs we can compute $f_t(\mathbf{x}|\mathbf{z}_2, f_{\mathbf{y},t-1})$ and $f_t(\mathbf{y}|\mathbf{z}_1, f_{\mathbf{x},t-1})$, where $f_{\mathbf{x},t-1}$ and $f_{\mathbf{y},t-1}$ are the PDFs of $\mathbf{x}$ and $\mathbf{y}$ at frame $t-1$, respectively. These PDFs are used as approximations to $f_t(\mathbf{x}|\mathbf{z}_2, f_{\mathbf{x},t-1})$ and $f_t(\mathbf{y}|\mathbf{z}_1, f_{\mathbf{y},t-1})$, respectively. Note that if the cooperation does not assume any dynamic model, then the tracked states are treated independently of the past, making the approximations of the last two PDFs exact. This was the approach taken in the cooperation process of our experiments. After having the last two PDFs, we can fuse the trackers' PDF estimates using (2), achieving the goal of tracker cooperation. The resulting cooperation is illustrated in Figure 2.

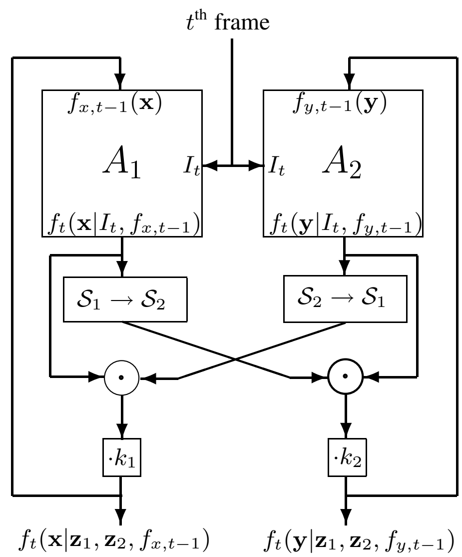


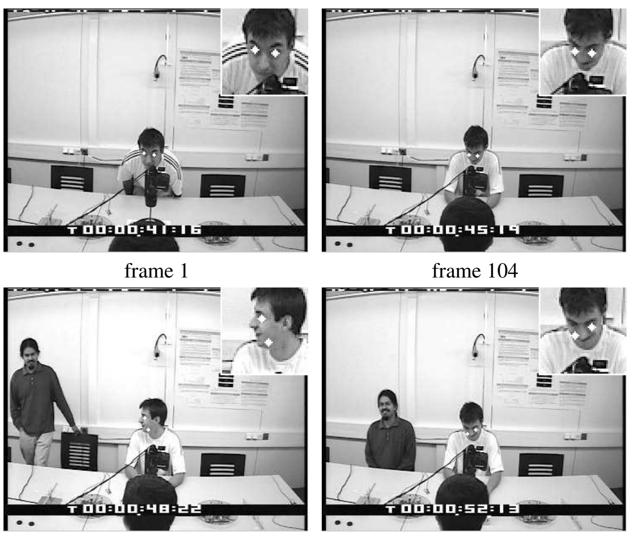
Fig. 2. Cooperating two trackers, tracking the pair of coupled states $\mathbf{x}$ and $\mathbf{y}$.

4. EXPERIMENTS

We applied the suggested framework on two different types of basic trackers: Kalman filter-based trackers and probabilistic exemplar-based trackers [11]. In both cases, the cooperation resulted in a great improvement of the tracking performance.

4.1. Experiment I: Kalman filter-based trackers

In this experiment we demonstrate the cooperation of two trackers, tracking two separate, albeit coupled, objects. The experiment consisted of tracking the eye centers of a person taking part in a "smart meeting". We used a 1:2 downsampled version of an image sequence from the fourth PETS workshop [13] (PETS-ICVS, scenario A, camera 1). We first tracked each of the eyes independently, using a template-based tracking algorithm, aided with a background subtraction scheme to locate the left and right margins of the person's head: A location of high correlation with the eye's template was considered correct with high confidence only if the location also fulfilled requirements regarding the position of the eye within the head boundary. The state space of the trackers is composed of four parameters: The 2D coordinates of the eye center in the current frame and in the previous frame. The state is initialized manually, and propagated using a Kalman filter. A few of the frames with the tracking results imposed are shown in Figure 3. It is seen that after the person rotates his head to his right, the tracker of the right eye fails without recovering.



frame 1   frame 104

frame 182   frame 273

Fig. 3. A few frames of the sequence used in Experiment I, with the tracking results of the two individual trackers imposed (marked by dots). After the head rotates to the right, the right eye becomes occluded, causing an irrecoverable failure.

Since the two tracked states are coupled, their estimations may be combined to enhance robustness. We combined the two trackers using only the outputs as described in Section 3.2. The PDF provided by the tracker of the person's left eye state is translated into a PDF estimating the person's right eye state, by defining the PDF of the person's right eye center, conditioned on a location of the person's left eye, to be a Gaussian centered a few pixels (8 for this sequence) to the left of the person's left eye center. The PDF-translation in the other direction is performed symmetrically. The translated PDFs remain Gaussian, making their fusion with the originally provided Gaussian PDFs Gaussian too (see (3)). Thus, the feedback to the Kalman filters is feasible. The combined algorithm is now more robust, recovering from head rotations to both sides. A few representative frames are shown in Figure 4.

frame 182   frame 273

Fig. 4. Results of Experiment I using the cooperating trackers. The trackers recover now from failures caused by the head rotations.

4.2. Experiment II: Exemplar-based trackers

This experiment demonstrates the application of the framework to cooperate between the exemplar-based trackers of [11]. First, we implemented the tracker version which had been used for the mouth tracking in [11], and used two instances of it for separately tracking the states of a person's left eye, x^{L}, and right eye, x^{R}. The two trackers picked the exemplars ($\{\tilde{x}_k^{\mathrm{L}}\}_{k=1}^{32}$ and $\{\tilde{x}_k^{\mathrm{R}}\}_{k=1}^{32}$ for the left and right eye, respectively) and learned the M^2 kernel parameters and dynamics from two respective training sequences, taken simultaneously. The trained trackers were tested on two new test sequences, one of each eye, taken simultaneously as well. However, the tracker of the right eye was challenged by replacing frames of its test sequence by noise for a couple of sections. The two trackers managed to track the eye states by their exemplars, but as expected, the tracker of the right eye failed during the disturbed time periods. The results at a few representative frames are shown in Figure 5.

As discussed above, the two eye states are regularly very coupled, as was the case in the training and test sequences here. Therefore, combining the two trackers has the potential of overcoming the disturbances, and the corresponding failures. Each exemplar-based tracker provides a probability distribution on its set of exemplars, thus making it feasible to combine the two trackers using the suggested framework. The set of exemplars of each tracker constitutes its state space. In order to translate the probability distribution from the left eye's exemplars set to the right eye's exemplars set, we set

$$\Pr(x^{\mathrm{R}} = \tilde{x}_j^{\mathrm{R}} | x^{\mathrm{L}} = \tilde{x}_k^{\mathrm{L}}) \sim f(I_{\mathrm{R}} | x^{\mathrm{R}} = \tilde{x}_j^{\mathrm{R}})$$

($j, k = 1, 2, \ldots, 32$), where I_{R} is the frame of the right eye which was taken simultaneously with $\tilde{x}_k^{\mathrm{L}}$ in the training se-

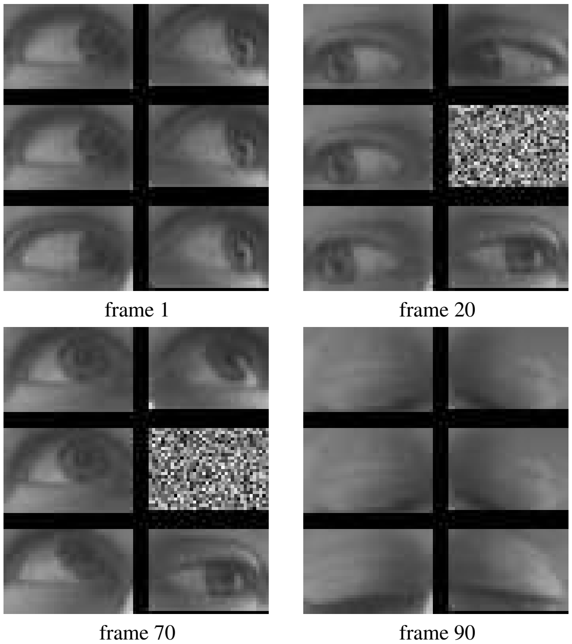


Fig. 5. Results of Experiment II using the two trackers separately. The two upper images in each frame are original test frames, the two middle images are the test frames feeded to the trackers, and the two bottom images are the exemplars approximated by the trackers as most probable. (The exemplars were taken from the training sequences.) The tracking succeeds during the undisturbed time periods, but fails whenever the image of the right eye is replaced by noise.

quence. The PDF f is estimated according the learnt M^2 kernel parameters for the right eye. The translation of the probability distribution in the opposite direction was similarly performed. Using the same test sequence, we found that the cooperating trackers were powerful enough to overcome the disturbances (see Figure 6). (Note that those disturbances would not have been overcome by simply using a single exemplar-based tracker, the exemplars of which are a spatial concatenation of the left and right eye images. The reason is that in this approach, the disturbed images are enormously distant from all exemplars.)

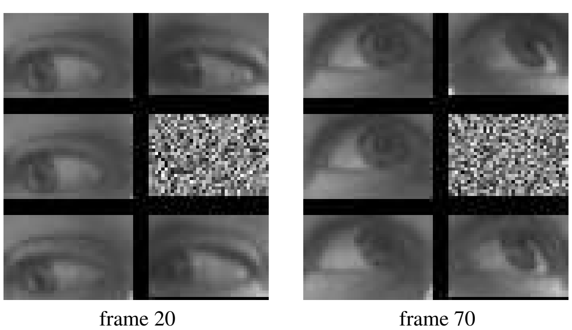


Fig. 6. Results of Experiment II after cooperating between the two trackers. Now the tracking succeeds also when the image of the right eye is replaced by noise.

5. CONCLUSION

An analytically justified, probabilistic framework for cooperating between two individual trackers of coupled objects was developed. The framework is fairly general, allowing cooperation between any two trackers which output a PDF estimate of the tracked state. The framework was successfully tested on two kinds of trackers, exhibiting the benefit gained from the coupling exploitation.

We are currently investigating the extension of the framework to cooperate between more than two trackers, as well as between CONDENSATION-based trackers [14].

6. REFERENCES

[1] J. MacCormick and A. Blake, "A probabilistic exclusion principle for tracking multiple objects," *IJCV*, vol. 39, no. 1, pp. 57–71, 2000.

[2] C. Hue, J.P. Le Cadre, and P. Perez, "A particle filter to track multiple objects," *WOMOT*, 2001.

[3] H. Sidenbladh and S.L. Wirkander, "Tracking random sets of vehicles in terrain," *WOMOT*, 2003.

[4] *WOMOT*, 2001, 2003.

[5] I. Leichter, M. Lindenbaum, and E. Rivlin, "A probabilistic framework for combining tracking algorithms," To appear in *CVPR* 2004.

[6] I. Leichter, M. Lindenbaum, and E. Rivlin, "A probabilistic framework for combining tracking algorithms," CIS-2004-2, Comp. Sci. Department, Technion – Israel Inst. Tech., 2004.

[7] A. Blake, R. Curwen, and A. Zisserman, "A framework for spatio-temporal control in the tracking of visual contours," *IJCV*, vol. 11, no. 2, pp. 127–145, 1993.

[8] Y. Rosenberg and M. Werman, "A general filter for measurements with any probability distribution," *CVPR*, pp. 654–659, 1997.

[9] Y. Rosenberg and M. Werman, "Representing local motion as a probability distribution matrix applied to object tracking," *CVPR*, pp. 106–111, 1997.

[10] T. Baker and M. Strens, "Representation of uncertainty in spatial target tracking," in *ICPR 98*, 1998, pp. 1339–1342.

[11] K. Toyama and A. Blake, "Probabilistic tracking with exemplars in a metric space," *IJCV*, vol. 48, no. 1, pp. 9–19, 2002.

[12] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, 3rd edition, 1991.

[13] `http://www.visualsurveillance.org` and related links.

[14] M. Isard and A. Blake, "CONDENSATION - conditional density propagation for visual tracking," *IJCV*, vol. 29, no. 1, pp. 5–28, 1998.