Preceding car tracking using belief functions and a particle filter

John Klein, Christele Lecomte, Pierre Miche
University and INSA of Rouen, LITIS EA 4108
BP 12 - 76801 St Etienne du Rouvray, France
john.klein@etu.univ-rouen.fr

Abstract

This article presents a preceding car rear view tracking algorithm which utilizes a particle filter and belief function data fusion. Most of tracking applications resort to only one source of information, making the system dependent on the source reliability. To achieve more robust and longer tracking, multiple source data fusion is a solution. Belief functions are a powerful tool for data fusion. Using bridges between probability theory and belief function theory, data fusion information can be incorporated inside a particle filter. The efficiency of the proposed method is demonstrated on natural on-road sequences.

1. Introduction

Preceding car rear view tracking is a computer vision problem which is dedicated to intelligent transportation system applications. Given that input sequences are acquired on-road, this tracking problem has its own specific challenges:

- the scene background is dynamically modified, and consequently is difficult to model.
- strong illumination changes occur and disrupt pixel value based methods.
- the scenes are cluttered, *i.e.* many objects have identical visual properties.

Visual tracking quality mainly depends on two aspects: the tracking procedure and the object model. Concerning the tracking procedure, particle filters [5] have proven to be more robust to occlusions thanks to random sampling. In addition they do not put constraint on object modeling, therefore they were retained as the tracking procedure for our study. Object model can be enhanced by using more a priori information and learning techniques [1], or by extracting more information and performing data fusion [2]. This paper focuses on the latter solution.

In existing data fusion particle filter approaches, the fusion is obtained using probabilistic methods or by incorporating different sources at different steps of the filter [2, 5]. Instead of designing a new strategy, we propose to use another data fusion technique: belief functions.

Belief functions, also often referred as Dempster-Shafer theory, can specifically address data fusion issues. Concepts such as conflict between sources and ignorance can be designed. Information fusion is obtained using a combination rule, and is transformed into a probability distribution [6]. Using this distribution the particle filter is run classically, and the tracking performances are improved.

The paper is organized as follows: the first part presents a particle filter adapted to visual tracking. The second part introduces belief function theory and explains how fusion information is drawn. The third part is dedicated to results and method evaluation.

2. Particle filter based visual tracking

The particle filter (PF) used in this study is adapted to visual tracking and similar to that of Perez et al. [5]. It is not intended to propose a new PF, but to point out the interest of using belief function data fusion inside a PF, which constitutes the originality of our work.

2.1 Particle filters

Particle filters use a hidden state vector X_t representation of the problem. In our context, this vector fully describes the bounding box that is supposed to contain the tracked object. PFs aim at estimating X_t, knowing a second random vector Y_t, referred to as the observation vector. As in many inference problems, the estimation of X_t is obtained using the estimation of the posterior

density $p\left(X_t|Y_{1:t}\right)$, where $Y_{1:t}$ corresponds to all available observations at time t.
In particle filters, this posterior density is estimated using Monte Carlo approximation and importance sampling from a proposal density $q(.)$. The samples $X_t^{(i)}$ drawn from $q(.)$ are called particles with $i \in \{1, N\}$ the particle index. The chosen filter parametrisation is the following[1]:

- the state is defined as $X_t^T = (i_t, j_t, h_t, w_t)$, the target center being (i_t, j_t), and the box dimensions being (h_t, w_t).
- The observations Y_t correspond to a set of features drawn from various extraction methods. Using these features the likelihood $p\left(Y_t|X_t\right)$ can be estimated. Illumination changes, camera motion and clutter disrupt this estimation.
- $q = p\left(X_t|X_{t-1}\right)$, which is the state transition density. This density is itself chosen to be a random walk. In these conditions, the particle sampling is not restricted to some direction.

Disrupting events are unavoidable in natural scenes. To overcome this difficulty we claim that using multiple features drawn from several cues is the best solution. Eventually, to design the likelihood, data fusion is needed.

2.2 Particle filter and data fusion

There are many ways to perform data fusion inside a particle filter. This study is restricted to data fusion at the likelihood estimation step.
In most probabilistic data fusion PFs [2], additional random observation variables are introduced: $\left\{y_t^{S_j}\right\}_{j=1}^{M} = Y_t$, if M extractors, *i.e.* M sources are used. By alleging independence, the Bayes rule yields:

$$p\left(Y_t|X_t\right) = \prod_{i=1}^{M} p\left(y_t^{S_i}|X_t\right) \tag{1}$$

with $P\left(y_t^{S_i}|X_t\right)$ the likelihood drawn from source number i.
In terms of data fusion, probability theory has some drawbacks. It cannot explicitly take into account that a source may give wrong information, or may give no information at all for a period of time.
To avoid these matters, we propose to use a belief framework: the Transferable Belief Model (TBM) [6], which extends the probability theory. Using the TBM, some belief assigned to an event A does only give an opinion about the possibility of A, and make no further hypothesis about the non-event $\bar{A}$. It is then easier to:

- deal with conflicting sources, *i.e.* a source that gives credit to A and a second one to $\bar{A}$.
- describe the fact that no information is available from a given source.

The set of belief functions is broader than the set of probability distributions. In this new space more complex fusion processes can be designed, which justifies our approach. Once the fusion belief function is obtained from all sources, it has to be sent back on its corresponding probability set so as to fit in particle filters. Using the TBM, we contribute to obtaining a more robust likelihood estimation.
Faux et al. [**?**] have mixed belief functions and a particle filter. Belief functions are meant to procude a robust object model and do not explicitely interact with the filter. Due to the chosen model, only similar colour-based sources (colour components) can be aggregated. With our approach, different kinds of sources will be used.

3. Belief functions and likelihood estimation

3.1 Introduction to belief functions

Suppose a set Ω, containing K mutual exclusive hypotheses ω_k, $\Omega = \{\omega_1, .., \omega_K\}$. This set is named frame of discernment, and some belief can be assigned to any subset A of Ω. For a given source of information, the belief assigned to the subsets of Ω are represented by a basic belief assignment function (bba) defined on the set of all subsets of Ω, often written as 2^Ω. Let us denote such a bba $m\left[S_l\right](A)$, with S_l being the source number l. This function must be such that:

$$m\left[S_l\right](A) : 2^\Omega \rightarrow [0, 1] \tag{2}$$

$$\sum_{A \subseteq \Omega} m\left[S_l\right](A) = 1 \tag{3}$$

In the probabilistic framework, we would have: $\sum_{k=1}^{K} P\left(\omega_k\right) = 1$, which is a much stronger constraint than equation 3. As a consequence for exemple $m\left(A\right) = \mu$ does not imply $m\left(\bar{A}\right) = 1 - \mu$. In other words, in the TBM some belief is assigned to some subset only if some piece of evidence justifies it.
In our context of visual object tracking, the belief that the preceding car belongs to the observed image subwindow must be evaluated. The frame of discernment will be composed of three exclusive hypotheses $\Omega = \{\omega_1, \omega_2, \omega_3\}$:

[1] more details on this parametrisation are available in [5]

ω_1: *the subwindow contains the targeted car*
ω_2: *the subwindow contains a piece of the scene background*
ω_3: *the subwindow contains any other vehicle*
Two subsets of Ω have special roles:

- the belief given to Ω represents a part of ignorance. One cannot choose one hypothesis over another

- the belief given to $\emptyset$ represents a part of conflict. One believes that the actual solution is not in Ω.

We will now present how bbas are built from the observed features.

3.2 Bba models

Four feature extraction techniques, *i.e.* four sources $\{S_l\}_{l=1}^{4}$, are used to characterize the tracked car:

- a colour distribution represents the shadow beneath the car.

- the car body colour-texture is characterized by a colour co-occurrence matrix based method [3].

- shape information is exploited by symmetry cards drawn from an image of contours.

- the scene movement is analyzed using Lefaix et al. algorithm [4], so that independent movement from the background can be detected.

For each source S_l, and each subset A that can be given some credit, a distance d_l between a model and the observed subwindow corresponding to particle $X_l^{(i)}$ is computed. The model is a set of features computed on a few images acquired a priori. Using this distance, a Gaussian model is used to obtain bbas:

$$m_A\left[S_l\right](A) = Z_{l,A}\exp\left(-\left(d_{l,A}/\sigma_{l,A}\right)^2\right) \quad (4)$$

$$m_A\left[S_l\right](\Omega) = 1 - m_A\left[S_l\right](A) \quad (5)$$

with $Z_{l,A}$ a magnitude parameter and $\sigma_{l,A}$ the standard deviation of the Gaussian function. Bbas are defined for only a few subsets[2]:
$\{\omega_1\}$: *the car itself is detected.*
$\{\omega_2\}$: *a part of the background is detected (the road).*
$\{\omega_1,\omega_3\}$: *some vehicle is detected.*
For a given bba, subsets that have positive values are called **focal elements** of a bba.
The texture source produces two bbas, one dedicated to $\{\omega_1\}$ and another to $\{\omega_2\}$. All the other sources produce one bba related to $\{\omega_1,\omega_3\}$. Consequently there are five bbas to aggregate. Selecting subsets for sources is application dependent. All cars are moving similarly, have roughly speaking similar shapes and shadows, whereas its colour-texture is rarely perfectly identical, which justifies our choice.

3.3 Fusion process and likelihood estimation

Using the conjunctive rule of combination [6], a new bba, denoted $m_{\circledcap}$, can be obtained from M other bbas using the formula $\forall A \subseteq \Omega$:

$$m_{\circledcap}(A) = \sum_{B,C|B\cap C=A} m\left[S_1\right](B)\, m\left[S_2\right](C) \quad (6)$$

Let us give a worked out exemple of the above equation for the movement source S_1 and the shape source S_2 combination: these two sources have the same focal elements: $\{\{\omega_1,\omega_3\},\Omega\}$. Let us then calculate the combination of these two sources for the subset $\{\omega_1,\omega_3\}$. Focal elements couples whose intersection equals $\{\omega_1,\omega_3\}$ must be identified:

$$(\{\omega_1,\omega_3\},\Omega)\,;(\Omega,\{\omega_1,\omega_3\})\,;(\{\omega_1,\omega_3\},\{\omega_1,\omega_3\})$$

Then the equation 6 for $\{\omega_1,\omega_3\}$ is simply:

$$m_{1\circledcap 2}(\{\omega_1,\omega_3\})=m[S_1](\{\omega_1,\omega_3\})m[S_2](\Omega)+ m[S_1](\Omega)m[S_2](\{\omega_1,\omega_3\})+m[S_1](\{\omega_1,\omega_3\})m[S_2](\{\omega_1,\omega_3\})$$

The combination is easily extended to more than 2 sources using the rule associativity. The principle of the conjunctive rule is to extract common parts of pieces of evidence from all sources and retain the most valuable ones. The new bba $m_{\circledcap}$ contains the fusional information about the variable $Y_t|X_t$. It can be transformed into a probability distribution, denoted $BetP()$, using the pignistic transform [6]:

$$BetP\left(\{\omega_1\}\right) = \sum_{A\subseteq\Omega}\frac{|\omega_1\cap A|}{|A|}m_{\circledcap}(A) \quad (7)$$

Then by accepting $p\left(Y_t|X_t\right) = BetP\left(\omega_1\right)$, the tracking and fusion procedure is completed, and will be referred to as the Credal Data Fusion Particle Filter (CDFPF).

4. Results and discussion

Our approach was tested on a natural on-road video, that contains 1300 frames. The sequence corresponds approximately to a 6 minute and 11 kilometer long motorway drive. Previously mentioned disrupting events

[2] Note that one drawback of belief functions is the exponential cost of increasing the frame of discernment. Not only the proposed frame is only composed of three hypothesis, but also only a few subsets can be given credit which dramatically simplifies belief computation.

are encountered several times. Using one long sequence with strong dynamic changes is more difficult to process than several short seqeunces, since only one setting can be used. Figure 1 shows the 5 images used to build the object models, acquired from another sequence, and illustrating the car visual variability.

Figure 1. Car body being shadowed or brightened throughout the sequence

On figure 2 the estimated bounding box using our CDFPF are shown at regular time gaps.

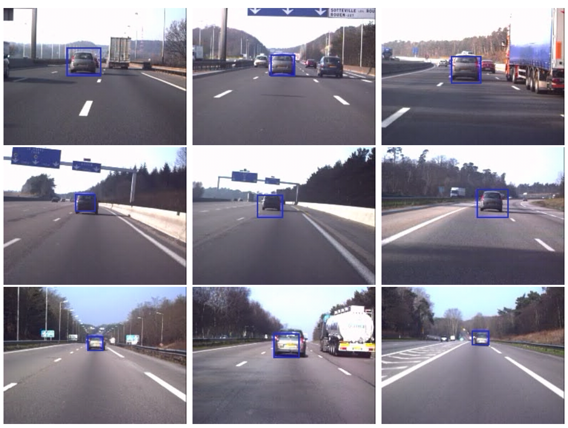

Figure 2. Tracking performance displayed every 150 frames.

CDFPF outperforms other approaches and never loses track of the preceding vehicle along the 1300 frames c.f. table 1.

Tracking method	Failures	Partial Tracking (50 to 80% detected)	Full Tracking (80 to 100% detected)
CDFPF	0 frame	475 frames	825 frames
Bayes data fusion PF	485 frames	170 frames	580 frames
Texture PF	550 frames	395 frames	355 frames

Table 1. Tracking quality comparison.

The best performing one-source approach is the texture PF. It maintains track up to frame 750, but the tracking quality is poor. The Bayesian data fusion PF produces a more accurate tracking, but completely loses track of the preceding car after frame 815. The experiments revealed that the classical Bayesian fusion approach is sensible to some lack of information. Indeed, if equation 1 is applied strictly, then only one zero-valued source likelihood implies a zero-valued global likelihood and cause a tracking failure. In practice such a source was discarded in the tests if $\max_i p\left(y_t^{S_l}|X_t^{(i)}\right)$ is less or equal to a given threshold. However the threshold is difficult to set since it has to be a compromise between losing information contained in small likehoods on the one hand and risking a computational failure on the other hand. Even if the threshold is efficient, the decision is riskier than CDFPF's one. CDFPF is free of these considerations and its performance proves the relevance of using belief functions data fusion.

5. Conclusion

In this article a new data fusion tracking algorithm was introduced. This algorithm uses sequential Monte Carlo techniques for tracking as well as the transferable belief model for data fusion. Thanks to belief modeling of the tracking problem introduced in this article, a more robust estimation of the likelihood was obtained. The CDFPF was tested on on-road sequences, and tracked a preceding car rear view with success. CDFPF gave a satisfying response to tracking issues related to car tracking. It shows the interest of using belief functions inside a particle filter. In further works, it is intended to make our approach cooperate with other particle filter data fusion strategies and to design more robust models using learning.

References

[1] Y.-S. C. Chih-Ming Fu, Chung-Lin Huang. Vision-based preceding vehicle detection and tracking. In *IEEE, ICPR*, volume 2, pages 1070–1073, Hong Kong, 2006.

[2] B. Han, S.-W. Joo, and L. S. Davis. Probabilistic fusion tracking using mixture kernel-bsaed bayesian filtering. In *IEEE int. conf. on computer vision*, Rio, september 2007.

[3] J. Klein, C. Lecomte, and P. Miche. Fast color-texture discrimination: application to car-tracking. In *IEEE ITSC*, pages 546–541, 2007.

[4] G. Lefaix, E. Marchand, and P. Bouthemy. Motion-based obstacle detection and tracking for car driving assistance. In *IEEE ICPR*, volume 4, Quebec (Canada), 2002.

[5] P. Perez, J. Vermaak, and A. Blake. Data fusion for visual tracking with particles. *Proc. IEEE*, 92(3):495–513, 2004.

[6] P. Smets and R. Kennes. The transferable belief model. *Artificial Intelligence*, 66(2):191–234, 1994.