Fusion of trajectory clusters for situation assessment

Lauro Snidaro, Claudio Piciarelli, and Gian Luca Foresti
Department of Mathematics and Computer Science
University of Udine
via delle Scienze 206, Udine, Italy
{snidaro, piccia, foresti}@dimi.uniud.it

Abstract - In this paper, we address the problem of identifying anomalous events in the context of a multi sensor surveillance system. Targets’ trajectories are analysed and compared to common patterns of activity represented as clusters of trajectories. Here we extend our previous work to cater for observations provided by multiple cameras observing the same scene. Data fusion is performed within the Dempster-Shafer theory of evidence framework. The proposed approach is validated through experimental results performed in the context of an automatic road traffic monitoring application.

Keywords: Trajectory clustering, multisensor data fusion, surveillance, situation assessment.

1 Introduction

The interest on behaviour understanding and situation assessment is rapidly gaining momentum in the last few years. In particular, the data fusion research community has witnessed a gradual shift of attention from observations to entailments. Building a situational picture of the ongoing activities in a given area is the main concern of automatic surveillance systems [1], which are meant for detecting anomalous activities in real-time, eventually sending a warning message to a remote operator.

Generally, these systems have to operate in crowded and complex environments where anomalous events have to be distinguished from the normal course of actions. What constitutes an anomalous event can be automatically detected as a deviation from common patterns of activity. Then how to represent and manage the knowledge about normal behaviour becomes a key point.

One way to infer the behaviour of the observed objects is by analysing their movements [1]. Similar trajectories repeated by many different objects constitute a common pattern of activity and should therefore be grouped together into a unique structure called cluster.

As described in our previous work [2], in opposition to much of the literature (see for example [3, 4, 5]), we developed an on-line clustering algorithm which is able to group the trajectories in real-time, as data is acquired. This was designed in order to avoid the classical two-step clustering (data collection and off-line processing), since we want to exploit the information on clusters for on-line behaviour understanding. The reader is also referred to [6] for a brief survey of relevant literature.

Here, we further develop our work to cater for uncertainty in building the clusters. Within the Dempster-Shafer theory of evidence framework [7], a trajectory’s belonging to a given cluster is postulated based on the evidence accrued through observations. In this way, each hypothesis is expressed with an associated level of belief while possibly retaining a certain level of uncertainty. Once the evidence collected so far is deemed sufficient to support a given hypothesis, the system commits itself by making the decision of assigning the trajectory under observation to a cluster or to flag it as an anomalous event. Again, this extension allows us to treat the multi-sensor system case where the observations coming from different sources observing the same scene can be fused to increase the confidence in the hypotheses under consideration. In this paper, we will deal with video sensors but the clustering algorithms and the fusion process are independent of the type of sensors employed as long as they are able to provide positional data on the targets observed.

This clustering algorithm should not be considered a fully fledged behaviour understating module, but an intermediate step of computation that can be functional to higher level situation assessment routines driven by the application’s goals and ontology. In essence, the approach presented here serves as a way to organize and encode positional data so that basic behavioural information can be extracted and further processing, possibly with the aid of other information sources, is facilitated in order to reach a higher level of situational awareness.

Experiments are shown in the context of an automatic road traffic monitoring application. The system under study is part of the research activities conducted within the FP6 European project MISS which aims to increase citizens’ road safety by enabling an organic surveillance and communication network between the road operators. In particular, the theoretical advances and experimental results presented here are part of the investigation geared to detect traffic anomalies in real-time that could lead to dangerous situations by analysing the video streams coming from multiple cameras installed on the road network.

1 Monitor Integrated Safety System, (TST4-CT-2004-516235)
The paper is organized as follows: the following Section gives an overview of the logical architecture so to collocate the role of the clustering algorithm in the processing flow. Section 3 introduces the formal representation of trajectory clusters and key notation. Section 4 represents the core of the paper describing the process of fusing trajectory information across multiple sensors. Experimental results are shown in Section 5, and conclusions are drawn in Section 6.

2 System architecture

This section provides an insight on the organization of the processing steps. Figure 1 sketches the processing flow in a general way independent of the system architecture and sensor types.

![Logical Architecture Diagram](image)

Three main modules are shown corresponding to different levels of abstraction. From the bottom, the first module deals with the actual detection of the targets. In the case of video sensors, this step mostly involves image processing techniques used to identify the moving objects in the monitored scene [1]. Detected moving objects can also undergo a classification step if different types are expected (e.g. cars, bikes, pedestrians, etc.). Targets can be observed and located by multiple sensors according to their own coordinate frames. It is generally useful to reference the targets according to a world coordinate frame, for example a map of the monitored area. This step is here called mapping and is mandatory in the case of multisensor systems. Finally, targets must be tracked as they move in the scene, solving the association problem when there is no clear correspondence between the objects detected at time $t$ and those detected at time $t + 1$, and finally recording their trajectory.

Next, the clustering algorithms, described in [2] and extended here to the multisensor case, roll in. The main function of this module is to conveniently organize data collected at the previous stage in order to obtain meaningful information for a high-level analysis able to automatically detect anomalies or dangerous behaviour. In a surveillance application, objects’ trajectories can be particularly relevant. By identifying the patterns of activity corresponding to normal behaviour, anomalies can be detected as deviations from these patterns. Trajectory clusters are versatile and adaptive representation to manage this kind of data. To clarify the discussion, the drawing in Figure 2 anticipates the structures that will be described in the following sections. The sketch illustrates several trajectories (a) and the clusters that have been identified (b). Common prefixes naturally induce a tree structure as shown in (c). Referring back to Figure 1, the image processing module and the data organization module can both be considered Level 1 algorithms according to the JDL model [8].

The behaviour analysis and situation assessment routines are typically Level 2 and are meant to understand the behaviour of the single target and to provide a complete picture of the current situation involving all the observed objects respectively. At this level, anomalous, dangerous, or forbidden events can be detected. In the road traffic context, U-turns, wrong-side driving, driving on the safety lane, stopping in forbidden areas are examples of anomalous events. The set of possible events is different according to the type of object. In the case of pedestrians, jaywalking in the middle of the lane is certainly a dangerous behaviour. These algorithms are thus devoted to the semantic understanding of the scene, describing complex activities from the raw data collected by the sensors.

Finally, after each update to the situational picture, the danger associated to the ongoing activities is estimated (JDL Level 3). If the observed course of actions is deemed potentially dangerous, then actions are taken according to the application. Possible actions in the road traffic example are alerting a remote operator, or controlling the traffic lights.
3 Cluster representation

The formal definition of a trajectory cluster given in [2] will be here adapted to support the extensions proposed in this paper. Let us define a trajectory $t$ as a list of vectors $v_j$ representing the spatial position of the object along the $x$ and $y$ axes:

$$ t = \{v_j : 1 \leq j \leq n(t)\}, \quad v_j = (x_j, y_j) \quad (1) $$

where $n(t)$ is the number of elements in the list $t$. Note that the position data also implicitly code temporal information, since position vectors $v_j$ are acquired at fixed time intervals. Clusters are again represented as a list of vectors:

$$ c = \{w_j : 1 \leq j \leq n(c)\}, \quad w_j = (x_j, y_j, \sigma_j^2) \quad (2) $$

where $\sigma_j^2$ is an approximation of the local variance of the cluster, and $n(c)$ is the number of elements of the cluster.

In order to check if a trajectory matches a given cluster, a distance or similarity measure must be defined. Given a trajectory $t$ and a cluster $c$ the distance measure adopted is defined as:

$$ D(t, c) = \frac{1}{n(t)} \sum_{j=1}^{n(t)} d(v_j, c) \quad (3) $$

where

$$ d(v_j, c) = \min_i \left( \frac{\text{dist}(v_j, Hw_i)}{\sqrt{\sigma_i^2}} \right) \quad (4) $$

$$ i \in \{\lfloor (1 - \delta)j \rfloor \ldots \lceil (1 + \delta)j \rceil\} $$

where $H$ is a $2 \times 3$ matrix projecting the first two components of vectors $w_i$, and $\text{dist}()$ is the usual Euclidean distance. The distance of a trajectory from a cluster is thus a mean of the normalized distances of every point of the trajectory from the nearest point of the cluster found inside a sliding temporal window centered in $j$. The temporal window is of increasing size proportional to the parameter $\delta$, in order to permit matching under accumulating temporal differences. The window is clipped in the range $[1 \ldots n(c)]$ and, if it completely falls outside this range, the distance $D$ is set to $\infty$. Note that the values $d(v_j, c)$ can be computed as the track develops, and can be used to detect if the track is moving away from the cluster.

4 Multisensor fusion

If multiple sensors are observing the same scene, redundant information may be available on a target’s trajectory. As a consequence, redundant information is also available to build and update the trajectory clusters. Figure 3 exemplifies the case of five clusters observed by two cameras.

However, readings resulting from the observation of the same target by different sensors are most likely to have different accuracies. This may be due to a number of reasons, for example, different placement, resolution, and performance (i.e. detection capabilities under changing environmental conditions). To fuse the evidence collected by different sources we have decided to adopt the Dempster-Shafer theory of evidence framework as follows.

4.1 Dempster-Shafer framework

Clusters and trees are here managed and maintained within the Dempster-Shafer theory of evidence [7]. This theory is more flexible than classical Bayesian theory when dealing with uncertain and incomplete knowledge and allows the fusion of multiple sources of information.

Let $\Theta$ be the set of all atomic hypotheses, and $2^\Theta$ be the set of all possible subsets of $\Theta$. A function $m$ is a Basic Probability Assignment (BPA) if:

$$ m : 2^\Theta \rightarrow [0, 1], \quad m(\emptyset) = 0, \quad \sum_{A \subseteq \Theta} m(A) = 1 \quad (5) $$

The value $m(A)$ represents the exact belief in the proposition depicted by $A$. The functions $\text{Bel}$ (degree of belief) and $\text{Pl}$ (degree of plausibility) are derived from the BPA:

$$ \text{Bel} : 2^\Theta \rightarrow [0, 1], \quad \text{Bel}(A) = \sum_{B \subseteq A} m(B) \quad (6) $$

$$ \text{Pl} : 2^\Theta \rightarrow [0, 1], \quad \text{Pl}(A) = \sum_{B \cap A \neq \emptyset} m(B) \quad (7) $$

$\text{Bel}(A)$ indicates the total degree of belief supporting $B$, whereas $\text{Pl}(A)$ is the degree of belief not contradicting $B$. From this definition, it follows that $\text{Bel}(A) \leq \text{Pl}(A), A \subseteq \Theta$. If $m_1$ and $m_2$ are BPAs, their combination through the Dempster combination rule or orthogonal sum is written as $m = m_1 \oplus m_2$. The actual computation is performed via the following general formula:

$$ m(A) = M^{-1} \sum_{i, A_i = A} \prod_i m_i(A_i) \quad (8) $$

$$ M = \sum_{i, A_i \neq \emptyset} \prod_i m_i(A_i), \quad m(\emptyset) = 0, \quad A \neq \emptyset $$

Figure 3: Two cameras with partially overlapping fields of view and observed clusters.
4.2 Cluster matching

The theory just described is instrumental in the process of deciding whether a trajectory matches a known cluster or if it constitutes a possible anomalous event. It also provides a solid framework for fusing data from multiple sensors. As the observed target moves, each sensor detecting it collects evidence as new positional data arrives, and updates its current beliefs on the trajectory.

Let analyse first the case of a single sensor $s$. Consider at time $k$ the frame of discernment $\Theta^k = \{\theta_i : 0 \leq i \leq |\Theta^k|\}$, where $\theta_i$ is the hypothesis “cluster $c_i$ is matched” for $1 \leq i \leq |\Theta^k|$, and $\theta_0$ corresponds to the hypothesis “no known cluster is matched”. When a new position vector $v_j$ of trajectory $t$ arrives, it is considered a new observation that has to be interpreted as new evidence. This is performed through BPAs functions. Therefore, for each atomic hypothesis $\theta_i$, the sensor $s$ produces at time $k$ a BPA function $m_{s,i}^k$ to express the belief associated to it. BPAs are here defined by (9) when $\Theta^k$ is not a singleton, and $d$ is the distance function defined by (4). In the degenerate case of $|\Theta^k| = 1$, the only hypothesis should assume-trivially- all the probability mass.

For each new observation, $|\Theta^k|+1$ BPA functions are thus created. Each $m_{s,i}^k$ represents a separate declaration, made by sensor $s$, in favour of atomic hypothesis $i$ with an associated level of ignorance expressed by the probability mass assigned to $\Theta$. In the first case of (9), a probability mass inversely proportional to the distance between the current position vector $v_j$ and the cluster $c_i$ is assigned to hypothesis $\theta_i$. The second case regards the mass assigned to $\Theta^k$ by each single BPA, that is, all that is remaining when the mass of the focal hypothesis $\theta_i$ has been subtracted from 1. The third condition indicates the mass that has to be assigned to $\theta_0$ when no cluster is matched. This happens when the distance between $v_j$ and all the clusters is greater or equal to 1. In this case, the probability mass is proportional to the distance between $v_j$ and the closest cluster.

All these functions must be fused together into a single BPA $m_{s}^k$ according to (8):

$$m_{s}^k = m_{s,1}^k \oplus \ldots \oplus m_{s,|\Theta|}^k$$

(10)

4.3 Temporal-spatial fusion

As described in [9], fusion can be performed in a temporal-spatial fashion as follows. Each new observation made by sensor $s$ at time $k$ must be temporally fused with the cumulative beliefs accrued so far. More specifically, observation $m_{s}^k$ should be fused with cumulative information $m_{s}^{k-1|k-1}$. Note that the superscript $k$ denotes a quantity at time $k$, while the superscript $k|k$ indicates that a quantity includes information collected from time 1 up to time $k$. Applying again the orthogonal sum:

$$m_{s}^{k|k} = m_{s}^k \oplus m_{s}^{k-1|k-1}$$

(11)

In the case of a multisensor system, where $N$ sensors are sharing a common frame of discernment $\Theta$, spatial fusion of the cumulative information collected by each sensor can be performed:

$$m_{s}^{k|k} = m_{s,1}^{k|k} \oplus \ldots \oplus m_{s,N}^{k|k}$$

(12)

The process of actually deciding which hypothesis best represents the current status of things is explained in the following section.

4.4 Decision making

At each time instant $k$, pignistic probabilities are computed for each atomic hypothesis $\theta_i$ as follows:

$$BetP_{m}^{k|k}(\theta_i) = \sum_{X \in \Theta^k} \frac{|X \cap \theta_i|}{|X|} m_{s}^{k|k}$$

(13)

Pignistic probability [10] is used for decision making and offers a trade-off between the credibility (6) and the plausibility (7), which are considered too pessimistic and too optimistic respectively.

In the trajectory clustering context we use the pignistic probability to check if a trajectory matched a cluster. If $BetP(\theta_i)$ > $BetP(\theta_j), \forall j \neq i$ and $BetP(\theta_i)$ > $t_h$, where $t_h$ is predefined threshold, then hypothesis $\theta_i$ (this is, a trajectory is matching cluster $c_i$) is considered verified.

4.5 Cluster update

As described in [2], clusters can be organized in a tree structure for anomaly detection via probabilistic reasoning. Depending on the number of trajectories that have traversed a given cluster, a probability is assigned to it. A low probability path in the tree can be related to an uncommon trajectory and possibly to an anomalous event.

In real world scenes, several maintenance operations of the tree should be accounted for. For example, the size of a cluster can vary with the passing of time to account for the characteristics of the traffic in it. New patterns of activity can be discovered, therefore adding new clusters and varying the structure of the tree. Since this paper mainly deals with the multisensor case, the reader is referred to [2] for additional information on anomaly detection and tree maintenance.
5 Experimental results

The system’s performance has been tested on a real world environment, consisting in a road junction monitored by two colour CCD cameras, as shown in figure 4. Each camera is connected to a processing unit performing the low-level object detection and tracking tasks, as well as the cluster matching step based on the temporal data fusion approach described in this paper. The data computed by the local processing units are then analysed by a central node in order to perform the spatial fusion step and the cluster updating routines.

![Figure 4: Environment for experimental results. (a) the scene seen by camera 1; (b) a map of the scene with the position and field of view of the two cameras.](image)

Figure 5 shows the clusters detected by the analysis of vehicle trajectories coming from the left part of the map over a period of 1 hour. As can be seen, three clusters organised in a tree-like structure were detected. The first cluster $c_1$ models the initial, common part of the trajectories shared by all the moving objects, while the two children clusters $c_2$ and $c_3$ represent the possible directions that can be taken after leaving $c_1$. The fusion technique proposed in this paper has been applied to robustly match the correct cluster for objects leaving $c_1$.

A trajectory matching cluster $c_2$, drawn in bold in Figure 5(a), is here analysed to exemplify the proposed matching approach. Tables 1 and 2 show the pignistic probabilities of a match for clusters $c_2$ and $c_3$, as computed both locally by the single processing units using the temporal fusion procedure, and globally using the spatially fused data gathered by the two sensors.

![Figure 5: Detected clusters for vehicles coming from the left zone of the map. Figures (a) and (b) show two examples of detected trajectories, respectively representing a normal situation in which $c_2$ is matched and an anomalous one in which no cluster is matched.](image)

<table>
<thead>
<tr>
<th>#</th>
<th>Camera A</th>
<th>Camera B</th>
<th>Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.4361</td>
<td>0.4638</td>
<td>0.5374</td>
</tr>
<tr>
<td>5</td>
<td>0.5250</td>
<td>0.5741</td>
<td>0.6779</td>
</tr>
<tr>
<td>10</td>
<td>0.5602</td>
<td>0.6169</td>
<td>0.7270</td>
</tr>
<tr>
<td>15</td>
<td>0.5793</td>
<td>0.6400</td>
<td>0.7524</td>
</tr>
<tr>
<td>20</td>
<td>0.5921</td>
<td>0.6555</td>
<td>0.7690</td>
</tr>
<tr>
<td>25</td>
<td>0.6018</td>
<td>0.6669</td>
<td>0.7811</td>
</tr>
<tr>
<td>30</td>
<td>0.6094</td>
<td>0.6760</td>
<td>0.7906</td>
</tr>
<tr>
<td>35</td>
<td>0.6157</td>
<td>0.6835</td>
<td>0.7983</td>
</tr>
<tr>
<td>40</td>
<td>0.6211</td>
<td>0.6898</td>
<td>0.8047</td>
</tr>
<tr>
<td>45</td>
<td>0.6257</td>
<td>0.6953</td>
<td>0.8102</td>
</tr>
<tr>
<td>50</td>
<td>0.6290</td>
<td>0.6992</td>
<td>0.8141</td>
</tr>
</tbody>
</table>

Table 1: Pignistic probabilities for the trajectory shown in figure 5(a) matching cluster $c_2$.

The results are depicted in Figure 6, representing the fused pignistic probabilities of the two clusters being matched as the data are acquired. The figure clearly shows how cluster $c_2$ is being given a constantly increasing evidence, while the probabilities of $c_3$ and not-found hypotheses decrease.

Figures 7 and 8 compare the performances of the proposed approach with the simple temporal data fusion as locally computed by each camera. The graphs show how, for both clusters, camera B is providing more evidence than camera A, and how the spatial fusion procedure integrates both the information, thus giving a more accurate result.

Figure 5(b) shows an example of anomalous trajectory, in which the not-found node is used. In this case the system initially gives more evidence to cluster $c_2$, since it is the nearest cluster to the detected trajectory.
Table 2: Pignistic probabilities for the trajectory shown in figure 5(a) matching cluster $c_3$.

<table>
<thead>
<tr>
<th>#</th>
<th>Camera A</th>
<th>Camera B</th>
<th>Fusion</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>0.3143</td>
<td>0.2914</td>
<td>0.2750</td>
</tr>
<tr>
<td>5</td>
<td>0.2935</td>
<td>0.2519</td>
<td>0.2201</td>
</tr>
<tr>
<td>10</td>
<td>0.2832</td>
<td>0.2349</td>
<td>0.1961</td>
</tr>
<tr>
<td>15</td>
<td>0.2771</td>
<td>0.2252</td>
<td>0.1826</td>
</tr>
<tr>
<td>20</td>
<td>0.2727</td>
<td>0.2184</td>
<td>0.1733</td>
</tr>
<tr>
<td>25</td>
<td>0.2693</td>
<td>0.2133</td>
<td>0.1662</td>
</tr>
<tr>
<td>30</td>
<td>0.2665</td>
<td>0.2092</td>
<td>0.1559</td>
</tr>
<tr>
<td>35</td>
<td>0.2642</td>
<td>0.2057</td>
<td>0.1519</td>
</tr>
<tr>
<td>40</td>
<td>0.2621</td>
<td>0.2028</td>
<td>0.1484</td>
</tr>
<tr>
<td>45</td>
<td>0.2603</td>
<td>0.2002</td>
<td>0.1484</td>
</tr>
<tr>
<td>50</td>
<td>0.2590</td>
<td>0.1984</td>
<td>0.1460</td>
</tr>
</tbody>
</table>

Figure 6: Fused pignistic probabilities for $c_1$, $c_2$ and not-found clusters matching the trajectory shown in figure 5(a).

When the distance from $c_2$ increases too much, the system starts giving more evidence to the not-found node, thus dropping to zero the pignistic probabilities of both $c_2$ and $c_3$, as depicted in figure 9. The not-found node can thus be used by behaviour analysis modules for anomaly detection, or for updating the tree’s structure as soon as a new pattern of activity justifying the new node is confirmed. Figures 10 and 11 show the local and spatial fusion results respectively for clusters $c_2$ and $c_3$.

Figure 8: Pignistic probabilities for cluster 3 matching the trajectory shown in figure 5(a).

The test consisted of 451 trajectories detected over a 1 hour period. The system correctly matched 97% of the given patterns; the majority of matching errors were due to errors in the tracking module.

6 Conclusions

In this paper, we have extended our previous work on trajectory clustering to the multisensor case. The Dempster-Shafer theory of evidence framework has been employed in order to fuse the observations in a...
principled way, explicitly managing the uncertainty inherent in the process. The matching of a trajectory with a cluster is postulated by a hypothesis with associated degree of belief based on collected evidence. The work here presented represents lays the basis for further processing geared to the understanding of a scene being observed by a surveillance system. Promising experimental results have been shown in the context of a road traffic monitoring application.

Acknowledgement

This work was supported by the European FP6 Project MISS “Monitor Integrated Safety System” (TST4-CT-2004-516235).

References


