Inferential Measurements for Situation Awareness

Enhancing traffic surveillance by machine learning

Prapa Rattadilok
School of Computing Science and Digital Media
Robert Gordon University
Aberdeen, UK
p.rattadilok@rgu.ac.uk

Andrei Petrovski
School of Computing Science and Digital Media
Robert Gordon University
Aberdeen, UK
a.petrovski@rgu.ac.uk

***Abstract*—The paper proposes a generic approach to building inferential measurement systems. The large amount of data needed to be acquired and processed by such systems necessitates the use of machine learning techniques. In this study, an inferential measurement system aimed at enhancing situation awareness has been developed and tested on simulated traffic surveillance data. The performance of several Computational Intelligence techniques within this system has been examined and compared on the data containing anomalous driving patterns.**

Keywords—inferential measurement; situation awareness; machine learning; anomaly detection, unmanned aerial vehicles.

I. Introduction

Due to the fast pace of computing technology advancements in terms of both memory capacity and processing speed, more and more data is becoming available in real time. With the vast amounts of data, traditional data acquisition and data processing methods have become inefficient or sometimes inappropriate, especially in a real time environment. The term 'big data' has several definitions, but generally it refers to data that is in excess of existing managing capability of the organisation and/or the processing capability of the applications used to manipulate it. Big data can be high in volume, rate of acquisition and/or diversity, which makes it difficult to work with using conventional database management systems, statistics and visualisation packages.

It becomes more important than ever therefore to be able to effectively distil the large amount of data into meaningful information by using computational modelling, analysis and representation. Smart approaches to data processing are needed in order to enable gaining insight into the essence of a process represented by the data, to control/optimise this process, or to provide effective decision support to the operator.

The complexity in managing large amount of data increases exponentially when real or near-real time data mining and information delivery is of a concern. In many cases, the processed information is only useful until a particular point in time and excessive latency would render the information useless e.g. information regarding an engine failure is most useful before the failure takes place, but once it has happened this information becomes completely useless in the prevention or damage limitation sense [5]. Therefore, the significant issue of timely information rises.

Integration across heterogeneous data sources also adds complexity in the form of analytical challenges, especially when there exist time and/or cost differences in processing data from different sources. Selecting suitable data acquisition sources, from which the data that can be processed quickly in order to obtain representative samples, can help in time critical situations. Additional data acquisition sources that involve longer data processing but are more accurate or detailed, can be applied later to add a deeper focus on objects of interest identified in the representative samples. This process can include logistical challenges such as controlling additional equipment to intensify data capture of a particular identified object of interest, e.g. an unmanned aerial vehicle can be sent to track a particular object of interest, and its camera on-board can be used to zoom in for more detailed images of this object; various on-board sensors can also be applied to obtain other relevant information so that inferences about the object of interest can be made and utilised.

Therefore, an Unmanned Aerial Vehicle (UAV) can be considered as an autonomous sensor system that is used to acquire large amount of data about complex and dynamic environments, to perform interpretation and fusion of the data, and to present the information gathered or inferred in a synthetic and compact form highlighting the features of interest in the environment explored. The ultimate goal of this data gathering and processing activity is to enhance the situation awareness to support real time decision making for the human operator on the ground.

The rest of the paper is organized as follows. Section II proposes a generic methodology of building inferential measurement systems that uses a multi-tier framework of data-driven modelling. The proposed framework utilises several Computational Intelligence (CI) techniques that help in developing an anomaly detection tool described in Sections III and IV. In Section V the performance of several classification methods, including the CI-based ones, are compared on the simulated data obtained from the traffic simulation packages. Section VI summarizes the developed methodology and identifies directions for further research.

II. Inferential Measurement Framework

Inferential measurement systems (IMS) aim to model the relationship between primary characteristics that are difficult to measure directly and secondary variables that can be more easily monitored. Although inferential measurements are widely used in industry, only a few techniques for inferential model development have been examined in detail. In general, three different types of approaches to building inferential models have been suggested: mechanistic modelling (based on first principles), statistical regression and artificial intelligence modelling [11].

Mechanistic modelling methods are based on the laws of physics and take the form of differential or algebraic equations. These methods perform well on the basis of a clear and good understanding of the mechanisms of the process, which is rarely attainable in practice.

Statistical regression methods overcome the need to gain full understanding of often non-linear, complex and uncertain behaviour of the process under investigation for building a usable inferential measurement system. Multivariate statistical methods such as principal component regression (PCR) and partial least squares or projection to latent structures (PLS) have been successfully used to build good inferential models. PCR and PLS are capable of including all relevant process measurements in a model without the problem of ‘‘overfitting’’ that is present in ordinary regression methods [11]. In this way all the process information can be included in the model leading to more accurate predictions.

Artificial Intelligence (AI) modelling has become a versatile tool for enhancing the capabilities and efficiency of inferential measurement systems. This type of modelling utilises the computational capabilities of modern computing devices (smart sensors, DSP-based microcontroller, and microprocessors) to effectively process the acquired input and infer the desired information. The AI-based techniques are applicable at various layers of IMS – from the data acquisition (sensor) layer, through to the layer of instrument calibration and customisation, then to the layer of process modelling, control and optimisation, and finally to the knowledge acquisition layer. The wide spectrum of possible applications is due to the capabilities of an IMS to gain insight into the behaviour of complex dynamic systems by means of data-driven modelling, a systematic approach to which is described next.

A. Systematic approach to data-driven modelling

The conceptual framework proposed in this paper for implementing data-driven modelling in inferential measurement systems was inspired by the multi-tiered scheme that was suggested by Moya [7]. In these frameworks, each tier or layer is dedicated to certain data processing tasks, ranging from low-level data acquisition up to high-level data interpretation using either existing or acquired knowledge.

The four layers of the suggested conceptual model that will form the basis of the systematic approach to inferential measurement for situation awareness are shown in Fig. 1.

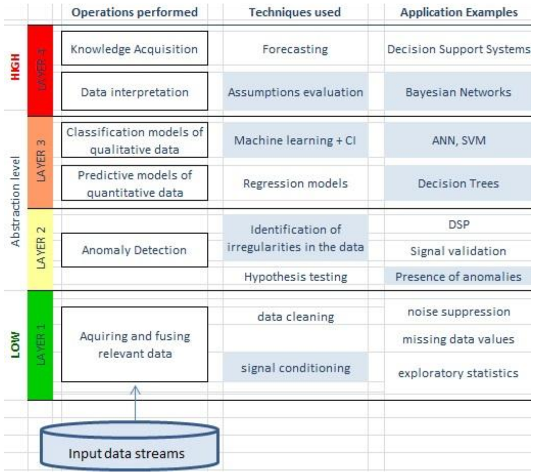

Fig. 1. A framework for building data-driven infential measurement systems

The layer with the lowest level of abstraction (LAYER 1) corresponds to the exploration of the available sensory data, including their visual representation, identification of the appropriate sampling periods, and data transformation (for example, differencing) for further analysis. The second layer (LAYER 2) deals with pre-processing of measured signals (e.g. identification of outliers, signal validation, etc.) and with detection of their salient features (e.g. the presence of anomalies). The main function of the second layer is to make necessary preparations for building data-driven models with good generalisation capabilities. Of particular interest to the authors are the models based on computational intelligence techniques – artificial neural networks, support vector machines, etc., built and tuned with the help of genetic algorithms, particle swarm optimization and artificial immune systems.

The remaining layers of the proposed conceptual model operate at a higher abstraction level. The third layer (LAYER 3) is responsible for building, evaluating, and correcting (if necessary) the data-driven models based on empirical data supplied by the lower layers. The final layer (LAYER 4) purports to examine the outputs of the models built at the previous layer in order to obtain or refine knowledge about the principles or rules that govern the dynamics of the processes under investigation.

In Section III we will exemplify the use of lower layers of the framework for detecting anomalies in the process of traffic surveillance.

B. Operating modes of an inferential measurement system

The inferential measurement systems (IMS) are expected to work in one of the following modes of operation [1]. For processes with high degree of stationarity, an IMS with a fixed structure and static parameters of the inferential algorithms used is usually appropriate. For processes exhibiting frequent, but non-fundamental, changes, the ways how information is inferred needs to continuously adapt to these changes. Generally, this is achieved by having a fixed structure of an IMS, but the parameters of its inferential algorithms are required to be dynamically tuned in response to the changes. Finally, for the processes undergoing fundamental changes, evolving IMS might be necessary, which are capable of changing their structure as well as adapting the algorithms' parameters.

In the present study, we will focus on building and simulating an IMS of the second type, and apply inference measurement in the context of anomaly detection in traffic surveillance.

III. Anomalies Detection

One possible way of selecting representative samples from big data is to identify potential anomalies. Anomalies can be defined as incidences or occurrences, under a given circumstances or a set of assumptions, that are different from the expectance. In many cases, anomalies are the indicators of possible problems, and thus a valuable information source for inferential measurement. By this definition, two main difficulties in identifying the possible problems arise - the expectance must be known *a priori*, and the set of assumptions must be valid.

Anomalies, by their very nature, are rare. They are incidences or occurrences that are first of all unexpected, and/or deviate excessively from the norm. Although the word anomaly implies some discrete event that is distinct from the surrounding background state, in fact anomalies are transitions from something that **is not** anomalous to something that **is**, and thus are far more continuous and involve subtle thresholding [3]. To make anomaly detection even more challenging, in a dynamic environment where an anomaly occurs too frequently, there is an argument for no longer classifying it as such.

Identification of anomalies can be viewed as outlier detection, i.e. a process of detecting patterns in a given data set that do not conform to an expectance. Promptly alerting users of those anomalies is a key requirement of real-time analysis of big data. Anomalies can also be used as evidence that the given set of assumptions or model might not hold true in practice – too many anomalies imply that the model adopted fails to accurately represent the process under investigation.

Association of anomalies with meaningful inferences can be very useful or even essential for some processes. Inferring future states of a process can lead to better understanding of potential causality, resulting thereby in a more predictive environment. By understanding the cause of anomalies, predictions of their future occurrences can be made more accurately or viably.

In dynamic real-time environments, significant changes in the values of important process characteristics (often referred to as primary variables) can influence process control procedures – for example, tracking or zooming in on objects of interest. The situation awareness of UAV can be considered as such a primary characteristic affected by numerous variables ranging from the levels of perception (e.g. visibility of objects of interest) to operational conditions (e.g. torque of the engine). This characteristic can be substantially enhanced by automatically adapting the control procedures on the basis of anomalies identified, as well as the prediction of their future occurrences.

In general, identification of anomalies improve the efficiency of working with big data by selectively obtaining representative samples that are the indicators of possible problems; identified anomalies can then be used to predict future occurrences of a certain event. The intelligent measurement system can then apply this information and optimise the controls of the processes accordingly.

Fig. 2 provides an example of the intelligent measurement system processing the data in three (potentially concurrent) steps leading firstly to identification of anomalies, then to their classification, and finally to prediction of likely outcomes.

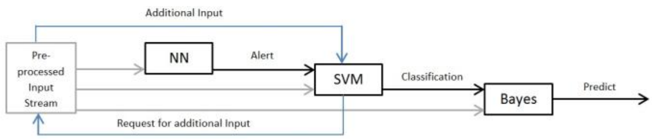

Fig. 2. System overview.

In the figure above, the pre-processed input stream exploits both fast and detail-rich data. The data is categorised as fast if it can be acquired with little or no delay, and its processing does not involve significant latency. The detail-rich data, on the other hand, does require greater processing power, and consequently more time-consuming or costly to obtain. The three Computational Intelligence (CI) techniques shown in the figure, which are responsible for identification, classification and prediction of anomalies, are artificial Neural Network (NN), Support Vector Machine (SVM) and Bayesian network (Bayes) respectively.

For instance, anomalies can be identified by applying a backpropagation-trained ANN operating on a fast data input stream [4]. Depending on the characteristics of the identified anomalies, either additional fast data may be required or detail-rich data may be processed to improve the accuracy of classifying the anomaly using an SVM approach. The Bayesian network builds a data-driven model that is capable of estimating the probabilities of future anomalies.

IV. Computational Intelligence And Big Data

The analysis of surveillance information in general, especially related to situation awareness, is a complex process that, given the amount and heterogeneous nature of data, is prone to data overload. This results in inability to support real-time processing and analysis of surveillance data, especially on board of mobile platforms, where datalink and bandwidth issues are significant [2, 8].

The data used in this research can be categorised as big data for two main reasons – due to its volume (in the order of 250,000 measurements per second) and the requirement of low latency.

In this study, the data to be acquired and processed by an inferential measurement system comes from various sensors on board of an UAV, such as radar, electro-optical/infra-red, GPS and Inertial Navigation Systems (INS).

Apart from on-board input streams, additional contextual input can also be taken into account. The choice of which contextual input to apply can be automatically tailored using the Computational Intelligence techniques. Fig. 3 illustrates an example of possible data sets that can be used during a surveillance mission. The complexity of the shown system increases with the number of additional dimensions of data.

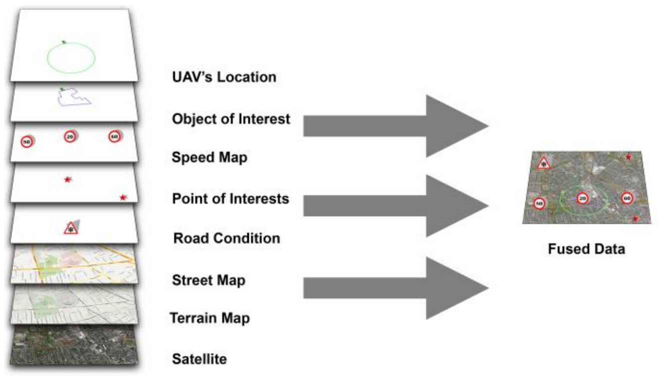

Fig. 3. Multiple data sources fused by an IMS

Therefore, a multi-tiered IMS that uses computational intelligence techniques should be able to enhance situation awareness of a UAV, especially in a real-time environment. Once anomalies are identified from fast data, additional data from both fast and detail-rich data sets can be added to improve the system classification and prediction performance.

V. Experimental Setup and Results

The data being used in this paper is simulated using the VISSIM [10] and VATIC [9] software packages, which are capable of generating simulated traffic and tracking the movement of individual vehicles from the simulation respectively. The aim of the experiments conducted was to test the design and basic functionality of the multi-tiered inferential measurement system in detecting anomalies and in exploiting additional input streams to enhance the UAV situation awareness. The simulated data is generated based on general traffic scenarios - each section on the roads is given a speed limit, permitted direction, overtaking permission, check points and traffic lights. An example snapshot of the road is shown in Fig 4.

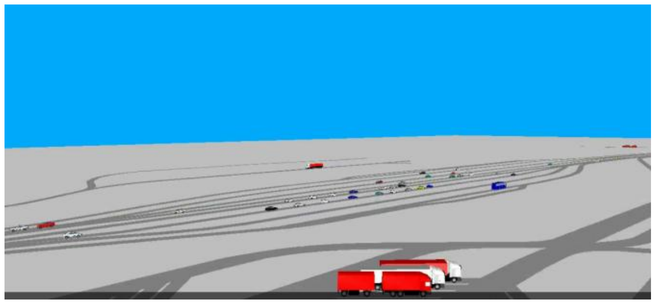

Fig. 4. Snapshot of a section of simulated roads and traffic.

Due to the fact that anomalies are not always discrete events, but rather transitions between state types, a single data point is often insufficient to identify certain anomalies that necessitate making a time series of measurements. The longer the time series of data, the easier it is to identify such an anomaly. However, during a typical surveillance mission, a mobile surveillance platform covers an area with many different vehicles (e.g. traffic monitoring over the M1) and purports to provide better support the gathering of information such as accidents, stranded car, or perhaps drank drivers.

The amount of data related to each vehicle within the area would vary depending on the distance and angle between it and the surveillance platform, as well as on the travelling direction and speed of the vehicle. It is also possible that the visibility of the vehicle may be affected by weather conditions, terrain, buildings, etc. To simulate such sporadic coverage, the data collected on each vehicle is limited in duration to two minutes of coverage.

Furthermore, vehicles vary in shapes, size, weight, inertia and drag. Each vehicle is probabilistically assigned a direction, in which to move, and a certain driving trait (e.g. overtaking, tailgating, etc.). The speed is also probabilistically determined depending on a given maximum, road and traffic conditions, as well as on surrounding objects (i.e. traffic lights, check point, and the like). Anomalies are specified as erratic driving behaviours, such as driving over the speed limit, illegal manoeuvre (e.g. driving in the wrong direction), large variance in speed, illegal overtaking, and a suspicious behaviour (e.g. avoiding check points).

The acquisition of traffic surveillance data, its processing and making inferences, as it could be carried out on board of an UAV, is simulated in this study using the abovementioned traffic simulation software. Video data collected on each vehicle is captured for no longer than two minutes at the rate of 10 frames per second. Once the sufficient number of frames have been captured, a set of statistics for each vehicle are calculated. These statistics are divided into three groups based on the complexity and amount of information required to calculate them. The first group makes up the largest number of attributes; in contrast, only one attribute was used for the other two groups.

In Group 1, the basic statistics for each vehicle are determined, including the driving direction, changes in direction, magnitude of lateral movements, variance in lateral movement, maximum and minimum speed, the mean value and variance of the speed. Group 2 characterises the traffic congestion for the entire surveillance area. These types of statistics require more advanced aggregation and storage of values to calculate metrics for a wider field of view – for example, the total number of vehicles in the area and the average speed of all vehicles. Group 3 includes statistics on local inter-object interaction. These statistics are the most processor-intensive, requiring evaluation and comparisons of complex data structures that characterise the vicinity of each vehicle, and aggregation of several time series. Some examples are the objects in close proximity to the front (i.e. in terms of travelling direction) of the vehicle of interest, and total number of other vehicles in its vicinity. Fig. 5 illustrates the graphical representations of the statistics used.

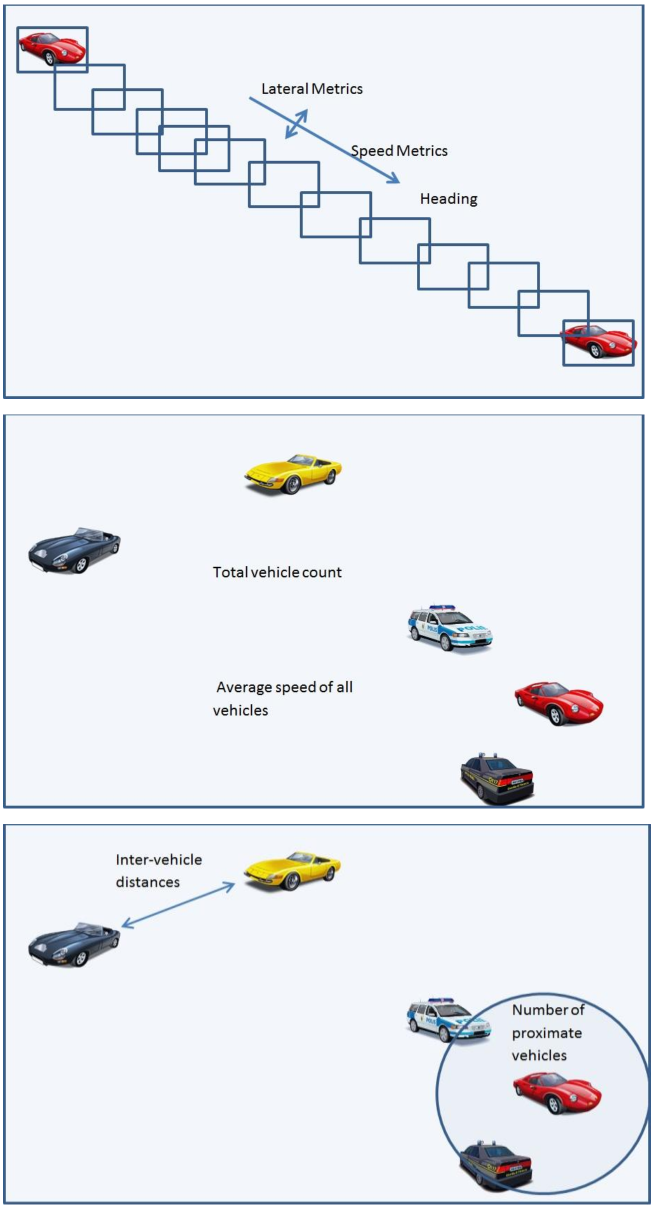

Fig. 5. Graphical representation of the three groups of statistics. Group 1 (top) – basic statistics. Group 2 (middle) – spatial congestion statistics. Group 3 (bottom) – local inter-vehicle interaction.

The task of the CI techniques used in the study was to distinguish, on the basis of the processed statistics, between two classes of vehicles – with benign and anomalous behaviour. The performance of various computational intelligence techniques used by the simulated IMS was evaluated on three sets of experiments. Firstly, the techniques were evaluated just using the vehicles' basic statistics. The second set of experiments combined these basic statistics with the addition of the congestion statistics. In this paper, the total number of moving vehicles is used to determine how congested the surveillance area is. Finally, the third set of experiments combined the statistics used in the first and second set of experiments with the addition of local inter-object interaction data, i.e. the detection of objects in close proximity to the front of the vehicle of interest.

The aim of the second and third sets of experiments is to investigate whether anomaly identification can be improved by awareness of traffic conditions (e.g. number of other vehicles in the area) and of the spatial data (e.g. gap distances). Hypothetically, the vehicles that were previously categorised as anomalies may be classified differently with the additional contextual information. The idea behind partially feeding the IMS with data is to minimise the size of initial data samples, but when necessary, acquire additional data either by analysing an additional input stream or by processing detail-rich data resources.

The reason for carrying out the three sets of experiments is to represent a scale of increasing workload on board of an UAV. The results of the experiments conducted are shown in Table 1 and represent the prediction accuracy of the CI techniques adopted by a virtual IMS on an unseen data set using a five-fold cross validation. The results shown are obtained using the WEKA data mining framework [12]. The CI techniques chosen from the WEKA package include Artificial Neural Network (ANN), Support Vector Machine (SVM), Bayesian network (Bayes); also, the Classification and Regression Tree (CART) algorithm was tested to evaluate the performance of statistical inference modelling.

TABLE I. PREDICTION ACCURACY OF COMPUTATIONAL INTELLIGENCE TECHNIQUES ON DIFFERENT DATA SETS

	ANN	**SVM**	**Bayes**	**CART**
Basic Stats	71.67%	74.17%	77.5%	79.16%
Basic Stats + Congestion	70.83%	73.33%	77.5%	79.16%
Basic Stats + Congestion + Local Inter-object Interaction Data	85.83%	85%	77.5%	83.33%

Using the basic statistics on their own (Group 1), or with the congestion data (Groups 1 and 2), yields similar results in terms of identification of vehicular classes in the range of 70.83% to 79.16%. The addition of local interaction data (Group 3) increases identification performance up to a maximum of 85.83% when attempting to identify vehicular objects not used in the training process. This level of accuracy would be quite difficult to achieve even for a human using a visual processing approach.

In the present study, the inaccuracies of the inferential models are due to either false positives (a vehicle is wrongly categorised as anomalous) or false negatives (a failure to identify an anomalous vehicle). While false positives may mean wasted effort, from the security perspective false negatives are usually more important. Fig. 6 illustrates the changes in false negatives as more data inputs are used. Again, the addition of the congestion statistic does not lead to the improvement of the false negative count (it even worsens the performance of ANN). Taking into account the local inter-vehicle interaction metric, on the other hand, significantly reduces the rate of false negatives. This is especially true for the SVM algorithm, which achieves the false negative value of about 11%.

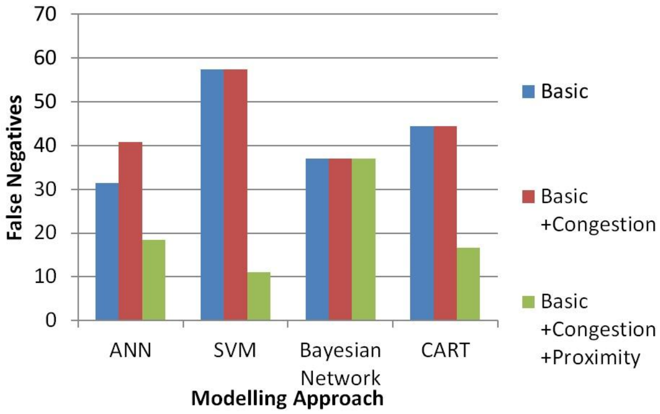

Fig. 6. Effect on false negatives of adding congestion and proximity data to the model.

A basic interpretation of these results suggests that the congestion data is of little practical value when modelling driver behaviour, whereas the local interaction metrics provide significant enhancement of situation awareness when added to the basic statistics. Given the relative simplicity of the simulated processes in comparison with the real ones, these results offer encouragement to continue the development of an IMS for the automated and real-time classification of traffic surveillance data.

VI. Conclusions

Situation awareness of an unmanned aerial vehicle involves acquiring and analysing large volumes of data from various on-board sensors. Furthermore, the processing and exchanging of data often needs to be performed in real time, necessitating the use of intelligent approaches to building a measurement system for such applications.

Computational intelligence techniques have been tested in this study to identify anomalies on simulated traffic surveillance data. The aim was to identify erratic driving behaviours, with and without additional contextual information, from the environmental data, such as the road and traffic conditions.

According to the results obtained, on a reduced data set the best performance in terms of prediction accuracy is achieved by a statistical technique of inference modelling – CART. Thus, the analysis of fast data within an inferential measurement system would be better performed by multivariate regression approaches.

With the addition of contextual information, however, the situation changes, and the use of computational intelligence becomes justified. The classification accuracy of traffic surveillance anomalies is better when detail-richer data streams are analysed. An artificial neural network appears to perform better in the experiments with additional contextual information. This conclusion confirms the hypothesis that once fitted with surrounding environment information, anomalies may no longer be categorised as such.

In the study, the duration of surveillance for each vehicle was limited to two minutes to simulate intermittent coverage. This did not obstruct the simulated IMS from making inferences regarding anomalous traffic; however, it would be interesting to see how the classification accuracy changes when the inferential measurement system operates on real surveillance data.

Acknowledgment

The authors would like to acknowledge the contribution of their industrial partner – Selex ES, a Finmeccanica Company – for providing funding and general support for the work on this research project.

References

[1] P. Angelov, A. Kordon, X. Zhou, "Evolving Fuzzy Inferential Sensors for Process Industry". In the Proceedings of 3rd International Workshop on Genetic and Evolving Fuzzy Systems, Witten-Bommerholz, Germany, March 2008. IEEE Xplore: 978-1-4244-1613-4/08.

[2] E. Bone, and C. Bolkcom, "Unmanned Aerial Vehicles: Background and Issues", Novinka Books, 2004.

[3] V. Chandola, A. Banerjee and V. Kumar, "Anomaly detection: a survey", ACM Computing Surverys, **41**(3), 2009.

[4] C. Evans, H. Chen, B. Battleson, J. K. Wölfel, J. Woelfel, "Unsupervised artificial neural networks for pattern recognition in multilingual text", Amherst, NY: RAH Press, 2010.

[5] A. Jacobs, "The Pathologies of Big Data", ACMQueue, 2009.

[6] T. Jan, "Neural network based threat assessment for automated visual surveillance". Neural Networks, 2004. Proceedings. 2004 IEEE International Joint Conference on (Vol. **2**, pp. 1309-1312). IEEE, 2004.

[7] R. R. Moya, "Multi-tier Framework for the Inferential measurement and Data-driven Modeling", PhD Thesis, Universitat Rovira i Virigili, 2007.

[8] K. P. Valavanis, "Advances in Unmanned Aerial Vehicles: State of the Art and the Road to Autonomy", Springer, 2007.

[9] VATIC (http://mit.edu/vondrick/vatic/)

[10] VISSIM (http://vision-traffic.ptvgroup.com/en-uk/products/ptv-vissim/trial-version/)

[11] K. Warne, G. Prasad, S. Rezvani, L. Maguiry, "Statistical and computational intelligence techniques for inferential model development: a comparative evaluation and a novel proposition for fusion". Engineering Application of Artificial Intelligence, Vol. 17, pp. 871-885. Elsevier, 2004. doi: 10.1016/j.engappai.2004.08.020.

[12] WEKA (http://www.cs.waikato.ac.nz/ml/weka/)