12th International Conference on Information Fusion
Seattle, WA, USA, July 6-9, 2009

Information Fusion Based Decision Support via Hidden Markov Models and Time Series Anomaly Detection

Jon Barker, Richard Green, Paul Thomas
Sensors and Countermeasures
Dstl Porton Down
Salisbury, UK.
JEBARKER@dstl.gov.uk,
RJGREEN@dstl.gov.uk,
PATHOMAS@dstl.gov.uk

Gavin Brown
Information Management
Dstl Porton Down
Salisbury, UK.
GBROWN1@dstl.gov.uk

David Salmond
Air and Weapons Systems
Dstl Farnborough, UK.
DJSALMOND@dstl.gov.uk

Abstract – *An Information Fusion (IF) based Decision Support Tool (DST) is presented to aid the identification of a target, from a large set of candidates, carrying out a pattern of activity which could be comprised of a wide variety of possible sub-activities and chronologies of events. The overall activity can only be defined in terms of its impact and in some cases detectable signatures of sub-activities. Hidden Markov Models (HMMs) and time series anomaly detection methods process multi-modal sensor data which are then integrated by a novel, efficient Bayesian IF algorithm to provide a probability that each candidate under observation is carrying out the target activity. The DST has been developed to prototype status by implementing this framework using commercial off the shelf (COTS) software. The DST allows the decision maker to rapidly access current and historical situational awareness pictures quantifying the progress of the overall search. A range of geospatial visualization and data interrogation features available to the decision maker are described and their performance is qualitatively evaluated. Finally, planned future developments are outlined.*

Keywords: Decision support, Bayesian, Geospatial Information System, Hidden Markov Model, Situational awareness, Time series anomaly detection

1 Introduction

The task of identifying an unknown pattern of activity which may be being carried out by one or more individuals in a large set of candidates is a complex and challenging problem facing the military today. For example, candidates may refer to people, buildings or vehicles. In many cases only the output or impact of a target activity is known in advance meaning that there are no prescribed sub-activities which must take place for the overall activity to be successful. For example the pattern of activity carried out by an individual or group intending to threaten a military base may only be defined in terms of the form of the final attack. Overall activity success is often achievable using a large number of possible sub-activities which could be combined in a large number of possible chronologies. This means that many traditional pattern recognition techniques will fail due to the inability to accurately define a library of target patterns to monitor for. Sub-activities may have known detectable signatures but these are often temporally sparse and with low signal to noise ratio. When combined with low duty cycles for sensors this means that any single sensing solution is likely to have a low probability of detection. Furthermore the signatures of different sub-activities may be spread across multiple transmission modes for which there is no single cross-modal sensing technology. Benign activity being carried out by other candidates often introduces confounding and confusable signals which act to mask the presence of a target activity and hence reduce the probability of detection and introduce false alarms. These benign activities are often as ill-defined as the target activity. The compound effect of these challenges makes the fusion of a set of cross-modal sensors essential to detecting such an activity.

The Decision Support Tool (DST) presented in this paper was developed to support a decision maker in the search for such an activity. The high-level user requirements addressed during the design of the DST included:

1. Sensor processing algorithms capable of determining the relevance of the data collected by a set of cross-modal sensors to possible target activities.
2. An information fusion centre capable of combining the information output from the sensor processing algorithms to determine the overall belief that each candidate is carrying out the target activity.
3. A generic software framework implementing algorithms that are free of inbuilt hypotheses and assumptions that could reduce the detection sensitivity of the system.
4. A graphical user interface (GUI) capable of: visualising current and historical situational awareness pictures and sensor deployments; and, interrogating the underlying sensor data.

Hidden Markov Models (HMMs) are a natural choice for mathematically describing situations where there is a sequence of hidden states of the world observed only through noisy sensor measurements. Making an assumption of Markovian structure, the possible target and benign sequences of sub-activities that could be being carried out by each candidate can be modelled as HMMs. This allows the

Report Documentation Page

Form Approved
OMB No. 0704-0188

Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
JUL 2009		**06-07-2009 to 09-07-2009**

4. TITLE AND SUBTITLE	5a. CONTRACT NUMBER
Information Fusion Based Decision Support via Hidden Markov Models and Time Series Anomaly Detection	5b. GRANT NUMBER
	5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S)	5d. PROJECT NUMBER
	5e. TASK NUMBER
	5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) **Sensors and Countermeasures,Dstl Porton Down,Salisbury, UK, ,**	8. PERFORMING ORGANIZATION REPORT NUMBER
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSOR/MONITOR'S ACRONYM(S)
	11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION/AVAILABILITY STATEMENT
Approved for public release; distribution unlimited

13. SUPPLEMENTARY NOTES
See also ADM002299. Presented at the International Conference on Information Fusion (12th) (Fusion 2009). Held in Seattle, Washington, on 6-9 July 2009. U.S. Government or Federal Rights License.

14. ABSTRACT
An Information Fusion (IF) based Decision Support Tool (DST) is presented to aid the identification of a target, from a large set of candidates, carrying out a pattern of activity which could be comprised of a wide variety of possible sub-activities and chronologies of events. The overall activity can only be defined in terms of its impact and in some cases detectable signatures of subactivities. Hidden Markov Models (HMMs) and time series anomaly detection methods process multi-modal sensor data which are then integrated by a novel, efficient Bayesian IF algorithm to provide a probability that each candidate under observation is carrying out the target activity. The DST has been developed to prototype status by implementing this framework using commercial off the shelf (COTS) software. The DST allows the decision maker to rapidly access current and historical situational awareness pictures quantifying the progress of the overall search. A range of geospatial visualization and data interrogation features available to the decision maker are described and their performance is qualitatively evaluated. Finally planned future developments are outlined.

15. SUBJECT TERMS

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			
unclassified	**unclassified**	**unclassified**	**Public Release**	**8**	

likelihood that each candidate is carrying out the target activity to be calculated given an incomplete sequence of noisy sensor measurements. This process is formally described in Section 2.1.

For some sensors there are no signatures of sub-activities available that can instantaneously identify the activity; however the time series of some physical characteristic of a candidate carrying out the target activity may be expected to appear anomalous when compared to the time series of the same physical characteristic for all other (innocent) candidates over time. This makes the HMM approach unsuitable. In these cases a time series anomaly detection (TSAD) algorithm is used to produce an anomaly score for each candidate over time. Assuming the relationship between these anomalies and the possible target activities is known, this anomaly score can be used to calculate the likelihood that each candidate is carrying out the target activity. This process is formally described in Section 2.2.

Bayesian probability provides a natural information fusion framework when probabilistic information is available from the processing of a set of cross-modal sensors. In particular the combination of the probabilistic belief about the true activities carried out by each candidate based upon the output of HMMs and TSAD algorithms can easily be stated in terms of Bayesian probability. Furthermore, Bayesian probability naturally allows the incorporation of prior knowledge about the expected number of candidates carrying out the target activity. The highly efficient Bayesian algorithm that was designed and implemented for the DST is described in Section 2.3.

The DST is a concept demonstrator which has been developed to prototype status using two core COTS software applications customised to add bespoke additional functionality. The system architecture and GUI interfaces are discussed in Section 3. The visualisation options available to the decision maker to symbolise the probabilistic fusion outputs on top of imagery of the search area are described in Section 3.3. The bespoke data interrogation and sensor deployment tracking features that were developed are described in Section 3.3.

In Section 4 we summarise the performance of the DST and in Section 5 we discuss some aspirational future developments.

2 Core DST components

2.1 HMM

A HMM describes a system which at any time is in one of N distinct hidden states. At regularly spaced, discrete times, the system undergoes a change of state according to a set of state transition probabilities. These hidden states can only be observed indirectly through noisy sensor observations. In the case of a first order HMM, which we consider here, the state transition probabilities depend only on the preceding state, not the whole history of the hidden process. Given a possibly incomplete sequence of noisy observations the well-known forward algorithm ([2] pp. 203-206) is used to calculate the likelihood that the sequence of observations was produced by the underlying HMM. HMMs provide a natural framework in which to describe the remote detection of a target activity whose structure is uncertainly known.

More formally, a HMM is a quintuple (S, X, A, B, π), where $\Pi = (A,B,\pi)$ represents the set of model parameters, i.e. the state transition probability matrix, observation probability matrix and state prior probability vector. S denotes the set of possible hidden states and X denotes the set of possible observations. In the DST $S = \{SP, \neg SP\}$ denotes the hidden states *Signal present* and *Signal not present*, i.e. for a given sensor the activity being carried out by the candidate is either emitting or not emitting a signal that can be recognised by the sensor as being a signature of the target activity. Similarly, $X = \{SD, \neg SD\}$ denotes the observations *Signal detected* and *Signal not detected*. Let q_t and o_t denote the hidden state and observation at time t. All HMMs used in the DST have the structure shown in Figure 1 where the top four state transition probabilities are of the form $p(q_t \mid q_{t-1})$ and the bottom four detection probabilities are of the form $p(o_t \mid q_t)$. Note that $p(SD \mid SP)$ and $p(SD \mid \neg SP)$ denote the sensor probabilities of detection and false alarm respectively.

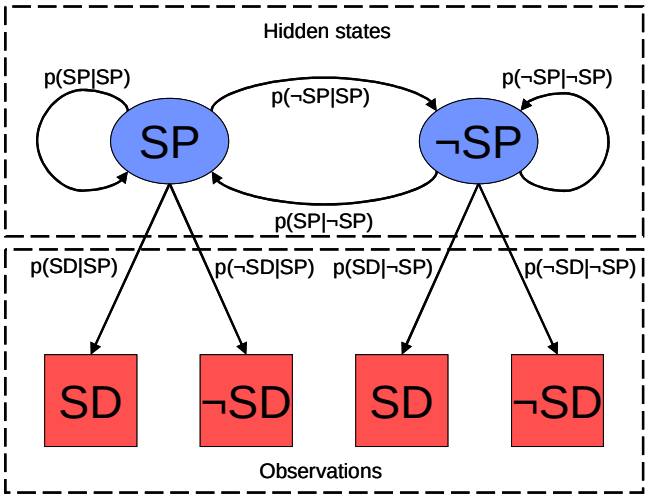

Figure 1: DST HMM structure

The state transition probability matrix A stores the probability that hidden state i follows hidden state j:

$$A = [a_{i,j}],\ a_{i,j} = p(q_t = s_j \mid q_{t-1} = s_i). \qquad (1)$$

Note that the transition probabilities are independent of time (Markov assumption); they will however depend on the sensor to which the HMM corresponds. The observation matrix B stores the probability of observation k being produced from state i and is again independent of time:

$$B = [b_i(k)],\ b_i(k) = p(o_t = x_k \mid q_t = s_i). \qquad (2)$$

The state prior probability matrix π stores the probability that at time $t=1$ the hidden state is s_j:

$$\pi = [\pi_i],\ \pi_i = p(q_1 = s_j). \qquad (3)$$

For a sequence of observations $O=o_1,o_2,\ldots,o_T$ the forward algorithm is used to compute the probability $p(O|\Pi)$. This problem can be viewed as evaluating how well the HMM predicts the given observation sequence.

The probability of the observation sequence O given a specific state sequence Q is:

$$\begin{aligned} p(O|Q,\Pi) &= \prod_{t=1}^{T} p(o_t|q_t,\Pi) \\ &= b_{q_1}(o_1)\times\cdots\times b_{q_T}(o_T). \end{aligned} \quad (4)$$

The probability of the state sequence is:

$$p(Q,\Pi) = \pi_{q_1} b_{q_1}(o_1) a_{q_1,q_2} b_{q_2}(o_2)\ldots a_{q_{T-1},q_T} b_{q_T}(o_T).$$

So we can calculate the probability of the observation sequence given Π as:

$$\begin{aligned} p(O|\Pi) &= \sum_{Q} p(O|Q,\Pi)P(Q|\Pi) \\ &= \sum_{q_1\cdots q_T} \pi_{q_1} b_{q_1}(o_1) a_{q_1,q_2} b_{q_2}(o_2)\ldots \\ &\quad \ldots a_{q_{T-1},q_T} b_{q_T}(o_T). \end{aligned} \quad (5)$$

Evaluating this probability directly would have computational complexity that is exponential in the number of time steps; therefore, an algorithm called the forward algorithm (a case of the Expectation Maximisation (EM) algorithm [4] pp. 124) is used to evaluate the probability recursively. We define

$$\alpha_t(i) = p(o_1 o_2 \ldots o_t, q_t = s_i \,|\, \Pi), \quad (6)$$

i.e. the probability of the partial observation sequence $o_1o_2\ldots o_t$ and the state s_i at time t given the model. After initialization α is calculated recursively as a sum over all states at the previous time step. The sum of all values of α at the final time step will equal the probability of obtaining the observation sequence given the model. More formally:

1. Initialisation: O may contain missing observations (for which we will use the MATLAB inspired notation NaN); in this case we sum over all possible observations

$$\alpha_1(i) = \begin{cases} \pi_i b_i(o_1) & \text{if } o_1 \neq \text{NaN} \\ \pi_i & \text{otherwise} \end{cases} \quad (7)$$

for $1 \le i \le N$.

2. Induction:

$$\alpha_{t+1}(j) = \begin{cases} [\sum_{i=1}^{N} \alpha_t(i) a_{i,j}] b_j(o_{t+1}) & \text{if } o_{t+1} \neq \text{NaN} \\ [\sum_{i=1}^{N} \alpha_t(i) a_{i,j}] & \text{otherwise} \end{cases} \quad (8)$$

for $1 \le t \le T-1,\ 1 \le j \le N$.

3. Termination:

$$p(O \,|\, \Pi) = \sum_{i=1}^{N} \alpha_T(i). \quad (9)$$

By calculating α as the sum over all states at the previous time step we reduce the complexity of the calculations involved from $2N^T T$ to $N^2 T$.

For a sequence of observations $O=o_1,o_2,\ldots,o_T$ from a specific single sensor input to the DST the forward algorithm is used to compute the probabilities $p(O|\Pi_R)$ and $p(O|\Pi_B)$ for two HMMs Π_R and Π_B. The subscripts R and B denote that the activity being carried out by the observed candidate is the target activity (referred to as RED) and or some other benign/innocent activity (referred to as BLUE). Π_R describes the transition of the hidden states (sub-activities) of the RED activity detectable by the sensor and the sensor performance in detecting these hidden states. Π_B performs the same role for an activity representing all possible BLUE activity. This problem can be viewed as evaluating how well each of the RED and BLUE activity models predict a given observation sequence. The output from the HMM processing of an observation sequence for a single building is the likelihood ratio

$$\frac{p(O|\Pi_R)}{p(O|\Pi_B)}. \quad (10)$$

2.2 TSAD

For some global sensor and target activity combinations the relationship between sensor observations and the target or background activities is poorly understood. For example the instantaneous trajectory of an individual intending to attack a military base may be indistinguishable from those of the innocent surrounding population; however, the route and movement track used by the individual over a period of time may appear anomalous when compared to the model of normality formed by monitoring all other individuals This means that traditional pattern recognition methods and the HMM method described in Section 2.1 are unable to calculate the likelihood that the sensor observations received are due to the presence of the target activity; however, it may be expected that the patterns of sensor observations of candidates carrying out target activities will appear anomalous when compared to the patterns of sensor observations of all other candidates over time.

For sensors in this category the DST takes as input for each candidate a multi-dimensional time series of independent statistics of measured physical emissions. For each pair of candidates the similarity between their time series is calculated using the Dynamic Time Warping (DTW) algorithm. DTW ([4] pp. 85) is a time series similarity measure that is able to recognise two time series as similar when one is merely a non-linear temporal warping of the other. This property makes DTW a suitable measure of similarity for detecting anomalous patterns of activity when two 'normal' candidates may be following non-linear temporal warpings of the same underlying pattern of activity. Based on their time series similarity to all other candidates the candidates are clustered. There are many possible clustering algorithms that can be used to cluster time series similarity measures [3]; we defer a complete description of the algorithm employed in the DST to a future publication. In essence, an anomaly score can be calculated for each candidate based on how much it is an outlier to each of the resulting clusters. Using historical data or expert knowledge, likelihood models can be constructed to calculate the likelihoods that the anomaly score obtained would be due to a RED activity being carried out or a BLUE activity being carried out by

the candidate. For each candidate, this can be expressed as the likelihood ratio

$$\frac{p(\text{anomaly score}|\text{RED})}{p(\text{anomaly score}|\text{BLUE})}. \quad (11)$$

These likelihood ratios constitute the TSAD input to the central fusion algorithm.

2.3 Bayesian fusion algorithm

Let $\{d_1,\ldots,d_N\}$ denote the set of *N* candidates under observation each of which is persistently in either the state RED (*R*) or BLUE (*B*); meaning a target activity either is or isn't being carried out by the candidate respectively. Let $\{v_1,\ldots,v_M\}$ denote the set of *M* observing sensors. At a given moment in time each sensor v_i processes all observations received for each candidate d_j and produces an array of continuous variables z_{ij}. Each z_{ij} could also be a vector, e.g. a time series of observations, and what follows holds directly for vectors. For each sensor v_i and each candidate d_j, z_{ij} is processed by either an HMM or the TSAD algorithm to produce the likelihood ratio

$$L_{ij} = \frac{p(z_{ij}|R)}{p(z_{ij}|B)}. \quad (12)$$

Note that $\forall i,j,\ L_{ij} \in [0,\infty]$; $L_{ij}=1$ indicates no information; $L_{ij} << 1$ indicates that candidate d_j is likely to be in state *B*; and, $L_{ij} >> 1$ indicates that candidate d_j is likely to be in state R. Furthermore, let

$$L_j = \prod_{i=1}^{M} \frac{p(z_{ij}|R)}{p(z_{ij}|B)} \quad (13)$$

denote the product of likelihood ratios for candidate d_j produced by all sensors v_i, $i=1\ldots M$. The persistent state of all *N* candidates can be described by a binary vector λ with *N* elements. For example, $\lambda = (10010\ldots)$ indicates that candidate 1 is *R*, candidates 2 and 3 are *B*, candidate 4 is *R* etc. There are a total of 2^N possible persistent states of the world. Furthermore, the number of distinct state vectors with exactly *n* elements in state R is

$$C(N,n) = \frac{N!}{n!(N-n)!}. \quad (14)$$

Any distribution may be assumed for the expected number *n* of candidates in state *R*. The prior probability of a particular state of the world λ is given by

$$p(\lambda) = \Sigma_n p(\lambda,n) = \Sigma_n p(\lambda|n)\,p(n) \quad (15)$$

where *n* is the number of candidates in state R and

$$p(\lambda|n) = \begin{cases} \text{constant} & \text{if } \lambda \text{ is compatible with } n \\ 0 & \text{otherwise} \end{cases} \quad (16)$$

Since, a priori, all λ with *n* candidates in state R are indistinguishable, and there are $C(N,n)$ of them we have

$$p(\lambda|n) = \begin{cases} \dfrac{1}{C(N,n)} & \text{if } \lambda \text{ is compatible with } n \\ 0 & \text{otherwise} \end{cases}. \quad (17)$$

The posterior probability of the true state of the world being a particular λ based upon the data observed by sensor v_i is given by

$$p(\lambda|Z_i) \propto p(Z_i|\lambda)\,p(\lambda) \quad (18)$$

where Z_i is the vector of sensor outputs for all candidates from sensor *i*, i.e. $Z_i = (z_{i1}, z_{i2},\ldots, z_{iN})$. If we assume the outputs from sensor *i* are mutually conditionally independent between candidates, then

$$p(Z_i|\lambda) = \prod_j p(z_{ij}|\lambda(j)) \quad (19)$$

where $\lambda(j)$ is the *j*-th element of the state vector. Furthermore, if we assume the outputs from all sensors are mutually conditionally independent for a given candidate, then

$$p(Z|\lambda) = \prod_i p(Z_i|\lambda) \quad (20)$$

where *Z* is the vector $Z = (Z_1, Z_2,\ldots, Z_M)$. Substituting this into Bayes rule and dividing through by $\Pi_i\Pi_j\,p(z_{ij}|B)$, we have

$$p(\lambda|Z) \propto \left(\prod_{j:\lambda(j)=1} L_j\right) p(\lambda). \quad (21)$$

Note that this is a product of the likelihood ratios of candidates in state *R* as specified by λ. Having calculated $p(\lambda|Z)$ we are in a strong position to obtain many useful probabilities. For example, the probability that candidate d_j is in state *R* (irrespective of the states of the other candidates) is given by the sum

$$\sum_{\lambda:\lambda(j)=1} p(\lambda|Z) \quad (22)$$

where the sum is over all λ for which candidate d_j is in state *R*.

For large candidate numbers it is not computationally feasible to implement this Bayesian framework in a "brute force" manner that explicitly calculates $p(\lambda|Z)$ individually for every possible λ. The time complexity of such an algorithm is $O(N2^{N-1})$. Therefore a novel and efficient algorithm with time complexity $O(N^4)$ was developed to greatly reduce the required number of computations to calculate the probabilities given by Equation (22). The algorithm is based upon the following proposition:

Proposition 1: Let **A** be the $N\times N$ matrix whose *ij*-th element (in the *i*-th row and *j*-th column) is given by

$$\begin{aligned} A_{i,1} &= L_i \\ A_{i,j} &= L_i \sum_{r=1}^{i-1} A_{r,j-1} \quad (j>1). \end{aligned} \quad (23)$$

and let $A_{N,*}$ denote the last row of **A**. Let P denote the vector $(P_1, P_2,\ldots, P_N)$ with P_m being the prior probability that a particular λ for which precisely *m* candidates are in

persistent state R represents the true state of the world. That is

$$P_m = \frac{1}{C(N,m)}\, p(m) \quad \text{from (17).} \tag{24}$$

Let $P \bullet A_{N,*}$ denote the scalar product of P and $A_{N,*}$. Then:

$$P \bullet A_{N,*} = \sum_{\lambda:\lambda(N)=1} p(Z|\lambda)\, p(\lambda) \tag{25}$$

Proof: Full proof omitted for brevity; the following is a sketch proof of this result. We first define some notation. Let $\Sigma_{\Pi(n,j)}$ denote the sum of all $C(n, j)$ possible products of the combinations of j elements of the set of likelihood ratios $\{L_1, L_2, \ldots, L_n\}$. For example,

$$\Sigma_{\Pi(3,2)} = L_1L_2 + L_1L_3 + L_2L_3 . \tag{26}$$

It can be shown that the ij^{th} element of **A** is equal to $L_i\Sigma_{\Pi(i-1,\, j-1)}$ which is clearly the sum of all $C(i, j)$ products of the combinations of j elements of $\{L_1, L_2, \ldots, L_i\}$ that include L_i. An inductive proof can be given that the sum of the m^{th} column of the first n rows of **A** is precisely the sum of all $C(n,m)$ products of likelihood ratios corresponding to combinations of m elements of $\{L_1, L_2, \ldots, L_n\}$. That is

$$\sum_{i=1}^{n} A_{i,m} = \Sigma_{\Pi(n,m)} . \tag{27}$$

From Equation (21) it can be shown that

$$p(Z \mid \lambda) = \prod_{j:\lambda(j)=1} L_j . \tag{28}$$

Given Equations (23), (27) and (28), it follows that

$$A_{N,m} = \sum_{\lambda:\lambda(N)=1,\ |\lambda|=m} p(Z \mid \lambda) \tag{29}$$

where $|\lambda|$ denotes the number of candidates in persistent state R in λ. The result given in Equation (25) is derived as follows:

$$\begin{aligned}
&\sum_{\lambda:\lambda(N)=1} p(Z \mid \lambda)\, p(\lambda) \\
&= \sum_{m} \sum_{\lambda:\lambda(N)=1,\ |\lambda|=m} p(Z \mid \lambda)\, p(\lambda \mid m)\, p(m) \\
&= \sum_{m} \sum_{\lambda:\lambda(N)=1,\ |\lambda|=m} p(Z \mid \lambda) \frac{1}{C(N,m)}\, p(m) \quad \text{from (17)} \\
&= \sum_{m} \frac{1}{C(N,m)}\, p(m) \sum_{\lambda:\lambda(N)=1,|\lambda|=m} p(Z \mid \lambda) \\
&= \sum_{m} P_m \sum_{\lambda:\lambda(N)=1,|\lambda|=m} p(Z \mid \lambda) \qquad \text{from (29)} \\
&= \sum_{m} P_m A_{N,m} = P \bullet A_{N,*}
\end{aligned} \tag{30}$$

□

The usefulness of Proposition 1 is in noticing that **A** can be calculated recursively with time complexity $O(N^3)$. If we cyclically permute $(L_1, L_2, \ldots, L_N)$ N-1 times and calculate **A** after each permutation we will have calculated Equation (25) for each target. This allows us to then calculate Equation (22) for each target as the normalization constant for Bayes rule will simply be the sum of the scalar product for all instances of **A** and P_0, the prior probability that there are no candidates in persistent state R. Full details of the proof of this proposition, the time complexity calculations and the avoidance of overflow issues in the algorithmic implementation will be given in a forthcoming publication by J. E. Barker and D. J. Salmond.

3 System architecture and visualisation

3.1 System overview

The DST presented here is a prototype concept demonstrator based on the HMM forward algorithm, TSAD and Bayesian fusion algorithm described in Section 2. The concept of use for the DST is as an aid to a decision maker leading a search for an activity, of the type described in Section 1, possibly being carried out by an unknown number of individuals within a large set of candidates. Through the course of the search the decision maker will task local sensors to remotely observe small numbers of candidates. The data collected by each local sensor will be processed after each tasking and the sequence of *Signal detected* and *Signal not detected* observations will be updated for each candidate. The decision maker may also task global sensors to remotely observe some physical emissions from all candidates over time; the processed data from each global sensor will be used to update a multi-dimensional real-valued time series stored for each candidate. On the presentation of new or updated data the DST processes the updated observation sequences or time series for each candidate using the HMM forward algorithm or TSAD. The updated likelihood ratios for all candidates are processed through the Bayesian fusion algorithm on a single sensor basis and across the complete set of sensors. The output of this is a set of updated probabilities that each candidate is carrying out the target activity based upon each of the individual sensors as well as the fusion of all sensors; these probabilities are stored within a central common data store as shown in Figure 2. The HMM forward algorithm, TSAD, Bayesian fusion algorithm and required data handling and storage functions are implemented in the COTS software The Mathwork's MATLAB. The probabilities stored in the common data store are then accessed by the ESRI Geographic Information System (GIS) ArcMap for visualisation.

On initiation of the fusion system the decision maker can choose to operate the system in one of two separate high-level components: the Standard Fusion system or the 'What-If' Fusion (WIF) system. The complete process described above constitutes the Standard Fusion system. The WIF system was developed to allow alternative hypotheses about the parameters of the target and background activities and sensor performance to be investigated during the search without altering the core Standard Fusion system. Essen-

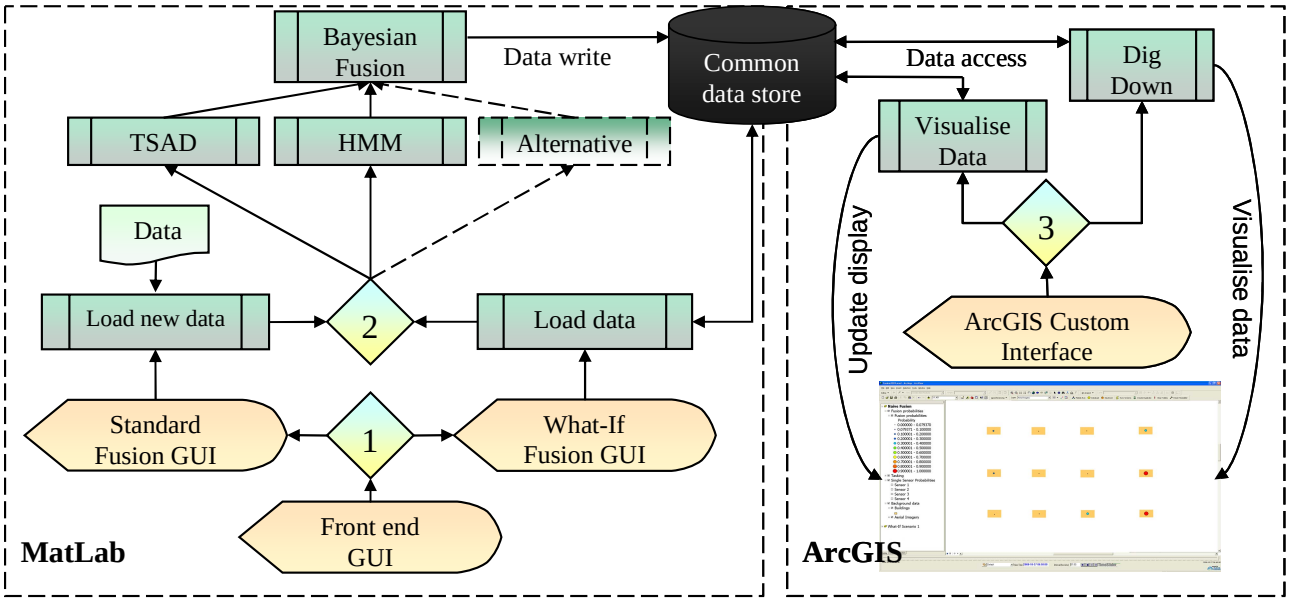

Figure 2: DST workflow diagram

tially the Standard Fusion system can be seen as utilising the best possible estimate of the HMM parameters and Bayesian prior probabilities given available historical data and expert knowledge of the target activity; whereas the WIF system allows the decision maker to explore the impact on the probabilistic outputs of changes to these assumed parameters. This option is represented by decision node 1 in Figure 2 and is made through the Fusion GUI shown in Figure 3. The GUI runs in MATLAB but is activated directly from ArcMap via a Visual Basic for Applications (VBA) script.

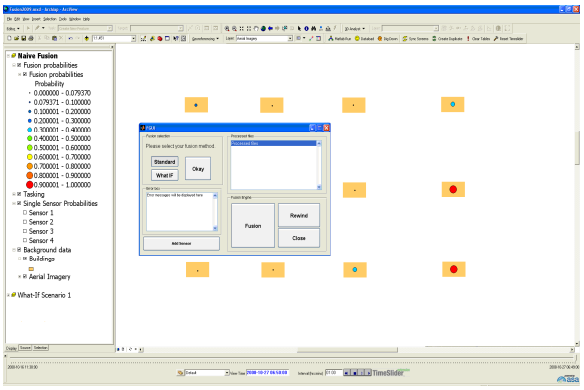

Figure 3: DST screenshot with GUI

3.2 Standard fusion

The Standard Fusion system has been designed to make no assumptions about the specific sensors which will be inputting observations to the system and as such can accommodate any sensor whose data can be meaningfully processed within the HMM or time series anomaly detection frameworks described in Section 2. Prior to initiating the search the decision maker provides a start-up file to the DST containing the following information:

- A list of Sensors;
- Each sensor's processing type (HMM or TSAD);
- HMM parameters;
- TSAD parameters including likelihood model specification;
- Candidate unique identifiers.

This information is used by the DST to automatically associate new sensor data files with the correct data processing algorithm, either the HMM or TSAD. This functionality, producing formatted data dependant on sensor type and the sensors associated processing parameters, is represented by decision node 2 in Figure 2. The data is then processed by the associated algorithm to produce a likelihood ratio of the form shown in Equation (12) for each candidate. These single sensor likelihood ratios are then processed through the Bayesian fusion algorithm to provide a probability that each candidate is carrying out the target activity based upon that single sensor's observations. The products all sensor likelihood ratios for each candidate (Equation (13)) are then processed through the Bayesian fusion algorithm to provide a probability that each candidate is carrying out the target activity given all sensor data. All calculated probabilities are output in a format that can immediately be symbolised by ArcMap when coupled with the candidate coordinates at the time of observation.

3.3 What-If fusion

In practice the HMM and TSAD parameters and Bayesian prior probabilities would be the best possible estimates based on historical data and expert knowledge of the target activity. Recognising that there may be error or uncertainty in the elicited probabilities the WIF system was developed

to allow alternative hypotheses about the parameters of the target and background HMMs and sensor performance to be investigated during the search without altering the core Standard Fusion system. This allows the decision maker to explore the impact on the probabilistic outputs of changes to the assumed parameters. These alternative hypotheses and corresponding parameters can be generated during the search through a dedicated WIF GUI and their impact seen without affecting the Standard Fusion system.

The WIF system allows testing of hypothesised scenarios that range from the simple, such as removing candidates or sensors from consideration by the DST, through to the more complex situation of hypothesising a specific target activity and re-processing the sensor data with the SME informing an alternative set of transition probabilities for the HMMs. The WIF system can also allow sensor data to be reprocessed using alternative algorithms to the HMM or TSAD, so long as the algorithms are compatible with both the Bayesian fusion framework and the data provided by the sensors.

The WIF GUI allows the decision maker to specify candidate removal and sensor subset hypotheses. The parameters required for data processing under these hypotheses are then specified through a Model Generation GUI. Upon specification of a sensor subset all data previously processed through the Standard Fusion GUI for those sensors is loaded from the common data store, the pre-formatted data is then restructured based upon any time constraints, e.g. data received out of chronological order. The newly structured data is then passed from the data restructuring process to the standard data processing described in Section 3.2.

3.4 Visualisation

Visualisation of the probabilistic outputs of the Standard Fusion and WIF systems is provided by a customised GIS application developed using the ESRI COTS software ArcMap. The visualisation application includes the following functionality:

- Display background geospatial information, e.g. aerial imagery, mapping and polygons representing the candidates;
- Visualise spatially and temporally referenced probabilistic output of the Standard Fusion and WIF systems;
- Temporal analysis of probabilistic information using a time analysis extension;
- Creation and visualisation of geographic representations of sensor tasking thus allowing temporal and spatial analysis of sensor deployments.

The foundation of the visualisation tool is the imported geospatial information; this provides the background imagery and underlying coordinate system over which the probabilistic information can be symbolised. The DST can be initialised with a single ortho-rectified, geo-referenced image and set of polygons representing the shape and spatial distribution of the candidates. The advantage of a GIS system however is in the depth of information which can be supported; additional information such as the infrastructure of the areas containing the set of candidates can easily be ingested into the DST to provide a rich source of background information.

The free ArcMap extension TimeSlider, from Applied Science Associates (ASA), was used to allow rapid changing between current and historical situational awareness pictures. Using this extension it is quick and easy for the decision maker to analyse the temporal trends in the spatially referenced probabilistic information.

The probabilistic information (with associated geospatial coordinates) is accessed by ArcMap from the common data store created by the MATLAB fusion systems. The visualisation tool uses VBA scripts to allow the decision maker to select and visualise a specific single sensor or fused probability file. Increasing probability that a candidate is carrying out the target activity is symbolised by a circular polygon of increasing radius and also a colour-map as shown in Figure 4. The ability to visualise single sensor as well as fused probabilistic outputs aid the decision maker in his search by allowing him to understand the relative contributions of each sensors evidence to the fused probabilistic output. This insight will also suggest possible 'What-it' hypotheses to generate.

Sensor taskings can be captured in the 2D environment using a semi-automated process where the deployment area is created by the user as a polygon and details such as sensor type, deployment time and deployment duration can be associated with it. This supports the decision maker in his search by allowing spatio-temporal analysis of sensor taskings to ensure limited sensing resources are being utilised optimally.

A data interrogation tool known as the 'Dig Down' tool has been to developed to allow the decision maker to quickly display plots of the sensor observations and time series data received for a particular candidate. This tool uses a VBA script to command MATLAB to plot data drawn from the common data store. This aids the decision maker in his search by presenting the underlying sensor data in such a way as to indicate the driving factors behind the probabilistic outputs.

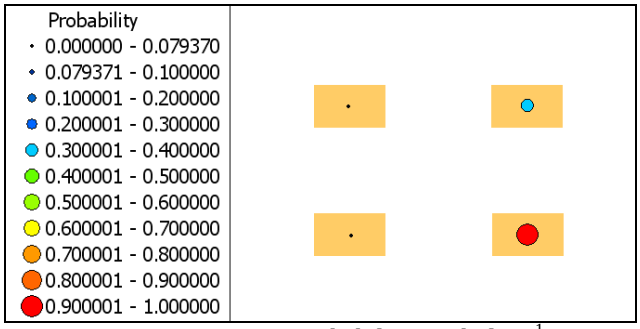

Figure 4: DST probability symbology[1]

[1] Candidate polygons shown have been chosen to be generic and not refer to any particular type of candidate.

4 Summary

The concept demonstrator DST presented in this paper has been tested through a series of evaluation scenarios supporting a real decision maker but based on synthetic data. The DST was found to be an effective and efficient way to collate a large number of sensor inputs over a large candidate target set and present the decision maker with a single current and historical situational awareness picture. This picture and the ability to track its evolution over time aided the decision maker in prioritising sensor deployments to effectively rule out some candidates and increase the focus on others. The DST acted not only as a record over time of the progress of the search but also as a record of sensor taskings and the volume of data collected against each candidate. The data interrogation features provided by the single sensor view and the 'Dig Down' tool were found to be essential in allowing the decision maker to understand the relative contributions of different sensors evidence to the overall belief attributed to each candidate.

5 Future Development

The current version of the DST is only able to process sensor observations and display the resulting probabilistic information at the individual candidate level. In practice, many sensors will return observations at different granularities of the candidate set, e.g. specific to sub-features of the candidates or only specific to a sub-set of all candidates. The process of associating all observations to individual candidates causes a loss of information between the raw sensor observations and the data maintained and processed by the DST. Research is now underway to understand how this information can be retained and exploited in a future version of the DST based on a hierarchical Bayesian fusion framework.

The current 2D visualisation tool has the limitation that only a single probability against each individual candidate polygon can be displayed at a specific time step. In practice it is often important to the decision maker to be able to display more specific information about the candidate such as its structure or status at the time of observation as these may be influence the relevance of the data collected to the search. Currently there is no mechanism for displaying such information. Options such as transitioning the DST to a 3D environment such as Esri's ArcScene are being investigated to provide these capabilities.

To support the large volumes of data that will be generated by the improved DST visualisation system a more intelligent geo-spatially referenced database system will be required. This database should allow data to be attributed at multiple levels of granularity as described above. This database will operate in a client-server environment that will allow alterations to be made by a number of different clients through specific applications. Applications will include: an interactive tool for sensor tasking queries which will provide information on data received and data requested; and, a sensor deployment tool that will allow mapping of sensor coverage and line of sight.

The HMM processing within the DST is currently applied to each sensor observation sequence separately. In practice this means that the likelihood of obtaining each sensor observation sequence is not calculated with an assumption of a single underlying sequence of hidden states which all sensors are observing. Ignoring this dependency between sensor observations means additional constraints on the possible underlying hidden state sequence may be being ignored. Work is underway to understand the maximum amount of information which can be extracted from the observation sequences across all sensors and the impact that this will have on the required sensor processing algorithms.

The current anomaly detection process makes an assumption that p(data|RED)=p(data|anomaly score). It was recognised during development of the DST that this is unlikely to be true in practice and the relationship between anomality and RED will be more complicated; for some sensor and target activity combinations it may in fact be the case that RED candidates do not appear anomalous at all. Understanding this relationship depends upon understanding the background activity taking place across the areas containing the set of candidates and the phenomenology of the sensor and target activity combination. For many applications this information may not available prior to initiating the search. To solve this complex problem we propose to separately track three hypotheses within the DST: non-target, target and anomalous where the target set and anomalous set are possibly intersecting subsets of the set of all candidates.

References

[1] Haiying Tu, J. Allanach, S. Singh, K. R. Pattipati "Information Integration via Hierarchical and Hybrid Bayesian Networks" *IEEE transactions on Systems, Man and Cybernetics, Part A* , Vol 36, No. 1, pp. 19-33, Dec. 2005

[2] S. Das, *High-level data fusion*, Artech House, Norwood, MA, 2008

[3] B. S. Everitt, S. Landau, M. Leese, *Cluster analysis* Fourth edition, Arnold, London, 2001

[4] H. B. Mitchell, *Multi-sensor data fusion: an introduction*, Springer, 2007

© Crown Copyright Dstl 2009