Probability Fusion for Correlated Multimedia Streams

Pradeep K. Atrey and Mohan S. Kankanhalli
School of Computing
National University of Singapore
Singapore 117543
{pradeepk,mohan}@comp.nus.edu.sg

ABSTRACT

The fusion of multiple correlated observations of a multimedia system is a research problem arising in many multimedia applications. In this paper, we propose a novel framework for the probabilistic fusion of correlated multimedia observations. Assuming that each of the media stream has a priori probability of achieving the goal and their underlying correlations are known, our framework fuses the individual probabilities using the quantitative correlation based on a Bayesian approach. The simulation results show that fewer highly-positively-correlated observations better achieve a specified goal when compared to the use of a larger number of observations with low correlation.

Categories and Subject Descriptors

H.5.1 [**Multimedia Information Systems**]: Methodology

General Terms

Algorithms, Theory

Keywords

Correlated probability, Experiential sampling, Information fusion

1. INTRODUCTION

The need for utilization of the multiplicity as well as the correlation of multimedia streams is increasingly being felt since the real strength of multimedia systems lies in making of the appropriate use of its diverse information sources that are usually correlated [4, 5]. Any multimedia information processing system usually deals with multiple types of data such as video, audio, and text etc, each of which is spatio-temporal in nature and inherently correlated. Also, the each type of data stream possesses a tremendous volume with lot of redundancy. This gives rise to many interesting research issues such as:

MM'04, October 10-16, 2004, New York, New York, USA.

1. How do we find the relevant data from the huge volume of multimedia data to achieve a specified goal [1]?

2. Given that the each media stream is capable of *partially* achieving the goal, how do we best fuse the correlated information from the various media sources to attain the better probability of achieving the goal?

In this paper, we essentially focus on the issues listed above. We resolve (1) by using the experiential sampling approach [5] to find the region of attention. The experiential sampling technique provides an efficient way to derive the attention samples from the media (sensor) samples. Once we have the attention samples, the media processing is performed only on the attention samples instead of the entire media data.

To address the issue (2), we propose a Bayesian framework for the fusion of correlated probabilities. The correlated data fusion problem has been widely studied and applied in many decision fusion scenarios [1, 2]. With the assumption of sources being independent, it is trivial to perform probability-level fusion by using well established methods such as renormalized multiplication [3].

However, in multimedia information fusion applications, where the observations obtained from the multimedia sources are often correlated, this assumption usually does not hold. Therefore, it is necessary to use a more sophisticated approach that incorporates the correlation among the data streams.

Unlike the conventional meaning where the correlation implies a statistic representing how closely two variables covary, the correlation between the multimedia streams here refers to a measure of *agreement* between the streams. In a multimedia system, where observations obtained from the media streams usually contain some noise, the index of agreement between two media streams can be determined by knowing the evidence that each stream provides. The observations could either concur or disagree to different extents. The media streams having negative correlation provide contradictory observations, and the streams having positive correlation provide supportive observations. For example, consider a multimedia system (having two different media types - video and audio) that observes the state of an environment. Let the goal be to detect a person in a room by face detection in video and speech/footstep detection in audio. If both the video stream and audio stream detect the existence of a

[1]In the context of multimedia systems, the goal is the purpose of the current multimedia system's task [5]. For example, the goal of a multimedia system can be object/face detection, activity monitoring or event analysis

person, the two media streams will have positive correlation (or index of agreement). However, if only the video camera detects the existence of a face and audio sensor detects no sound; it suggests that the observations are contradictory, hence the media streams have negative correlation. In this example, we assume that a person is always associated with some sound.

The current methods for correlated probability fusion do not use the quantitative measure of correlation among the data [2]. Our proposed method is based on the quantitative correlation which refers to the degree of agreement among the observed data. Assuming that the probabilities of each media stream achieving the goal along with their correlations are known, our framework fuses the individual probabilities to output the combined probability (we call it the 'fusion probability') of achieving the goal. The positive correlation among the streams enhances their fusion probability. We also show that the fewer highly-positively-correlated observations contribute more in achieving the goal than a larger number of lesser correlated observations.

The paper is organized as follows. We begin the paper with the problem formulation in section 2. In section 3, we present our framework for the fusion of correlated probabilities. We present the simulation results in section 4. Finally, section 5 concludes the paper with a discussion on the future work.

2. PROBLEM FORMULATION

Let $\mathbf{M}^n(t) = \{M_1, M_2, \ldots, M_n\}$ be a set of n multimedia streams at time instant t.

We make the following assumptions:

A1 Let each media stream M_i, $1 \leq i \leq n$, based on its observation, has a *priori* probability $p_i = P(G|M_i)$ of achieving the goal G.

A2 For $1 \leq i, j \leq n$, the streams M_i and M_j provide *correlated* observations. Let the correlation at time t be represented by a set $\Gamma(t)$ of correlation coefficients. The $\Gamma(t)$ is expressed as -

$$\Gamma(t) = \{\gamma_{ij}(t)\} \tag{1}$$

where, the term $-1 \leq \gamma_{ij}(t) \leq 1$ connotes the correlation coefficient between the observations obtained from the media streams M_i and M_j at time instant t. Also, the correlation can evolve with the time. The cardinality of the set Γ is given by $\binom{n}{2}$.

A3 Probability of achieving the goal G with any i number of streams is greater than or equal to the probability of achieving the goal with any $i-1$ streams. i.e. $P(G|\mathbf{M}^i) \geq P(G|\mathbf{M}^{i-1})$.

The objective is to find the fusion probability $P(G|\Phi)$ of achieving the goal when a subset $\Phi \in$ (The power set of $\mathbf{M}^n$) of correlated media streams is used.

3. PROPOSED FRAMEWORK

Given the set $\mathbf{M}^n = \{M_1, M_2, \ldots, M_n\}$ of n media streams[2] in a multimedia system, using Bayes' theorem we obtain -

$$P(G|\mathbf{M}^n) = \frac{P(\mathbf{M}^n|G)P(G)}{P(\mathbf{M}^n)} \tag{2}$$

[2]In this paper we use the term 'media stream' and the 'observations from the media stream' interchangeably

where,
$P(G|\mathbf{M}^n)$ is the *posterior* probability of successfully achieving the goal G given that observations $\mathbf{M}^n$ have been obtained from media sources.
$P(\mathbf{M}^n|G)$ is the probability of the particular set of observations $\mathbf{M}^n$ being taken given that the goal is successfully achieved. This is also called the *likelihood* pool.
$P(G)$ is the *prior* probability of the goal G being successfully achieved.
$P(\mathbf{M}^n)$ serves as a normalization function, ensuring the posterior probabilities sum to one over the observation set $\mathbf{M}^n$.

The media stream model is based on $P(\mathbf{M}^n|G)$ by first fixing the G and then computing the probability density function for $\mathbf{M}^n$ i.e. in other words, we derive the likelihood of G. Then this stream model can be used to find the probability density function for the goal G.

Similar to [3], we assume that the observations obtained from different media sources are independent given the true underlying state of the world. The effectiveness of fusion relies on this assumption. In our case, the goal is intuitively analogous to the state and once the goal has been specified it is correspondingly reasonable to assume that the observations made are conditionally independent given the goal. For example, if the goal is to detect a human face, the observations made through various media streams can be considered independent given that the human face exists.

The assumption described above leads to the following -

$$P(M_1, M_2, \ldots, M_n|G) = \prod_{i=1}^{n} P(M_i|G) \tag{3}$$

The $P(M_i|G)$, $1 \leq i \leq n$ is the independent likelihood pool which is constructed for each media stream separately. One way of computing the likelihood pool is by using experiential sampling approach [5] as described in section 3.1.

We further expand (2) as -

$$P(G|\mathbf{M}^n) = [P(\mathbf{M}^n)]^{-1} P(G) \prod_{i=1}^{n} P(M_i|G) \tag{4}$$

Since we assume the prior $P(G)$ to be non-informative (i.e. $P(G) = 0.5$) and $P(\mathbf{M}^n)$ to be a constant relative to the likelihood pool [3], equation (3) can further be rewritten as-

$$P(G|\mathbf{M}^n) = \alpha \prod_{i=1}^{n} P(M_i|G) \tag{5}$$

where α is normalizing constant which can be given as -

$$\alpha = \frac{1}{\prod_{i=1}^{n} P(M_i|G) + \prod_{i=1}^{n} P(M_i|\overline{G})} \tag{6}$$

where, $P(M_i|\overline{G}) = 1 - P(M_i|G)$, be the likelihood that the observation M^i does *not* achieve the goal G.

3.1 Experiential sampling based likelihood pool computation

Using the experiential sampling technique [5], we compute the attention saturation $ASat_i$ for each media stream M_i. $ASat_i$ provides the measure of generalized attention in a given time slice and its value can range from 0 (lowest, no attention) to 1 (highest, full attention). The key idea is based on the assumption that the likelihood of achieving a goal is high when the attention is high, and vice versa. This

leads to -

$$P(M_i|G) \approx ASat_i \tag{7}$$

For example, let the goal be to detect a face. Using experiential sampling technique, the attention saturation value is higher in the region where the face is likely to be detected.

Note that the experiential sampling based approach is one possible way of computing the likelihood pool. However, one could also use any alternative method.

3.2 Recursive Bayesian updating

The Bayesian framework allows for incremental and recursive addition of new information. Let $P(G|\mathbf{M}^{i-1})$ denote the probability of the streams $M_1, M_2 \ldots M_{i-1}$ together achieving the goal G. The updated probability $P(G|\mathbf{M}^i)$ (i.e. the fusion probability after fusing the new observation obtained from the stream M_i) can be recursively computed as -

$$P(G|\mathbf{M}^i) = \frac{P(M_i|G)P(G|\mathbf{M}^{i-1})}{P(M_i|\mathbf{M}^{i-1})}$$

$$P(G|\mathbf{M}^i) = \alpha_i P(G|\mathbf{M}^{i-1})P(G|M_i) \tag{8}$$

where, α_i is again a normalizing constant which is given by-

$$\alpha_i = \frac{1}{P(G|\mathbf{M}^{i-1})P(G|M_i) + (P(\overline{G}|\mathbf{M}^{i-1})(P(\overline{G}|M_i)} \tag{9}$$

3.3 Fusion of correlation

The correlation among the streams improves the fusion probability. In correlation fusion, we show how the correlation between a *group* of streams and a *new* observation stream is computed. We denote the correlation coefficient between the sources $\mathbf{M}^{i-1}$ and M_i as $\overline{\gamma}_i$. To include the stream M_i, we first compute the $\overline{\gamma}_i$ as follows. As it is assumed (in section 2) that we know the correlation coefficients γ_{ki} for $1 \leq k \leq i-1$, $\overline{\gamma}_i$ is approximated by heuristically choosing the maximum of the correlation coefficients of i^{th} stream with the previously selected streams for fusion. This is computed as -

$$\overline{\gamma}_i = max(\gamma_{ki}) \tag{10}$$

where, $1 \leq k \leq i-1$. The fused correlation coefficient $\overline{\gamma}_i$ is used for combining M_i with $\mathbf{M}^{i-1}$.

3.4 Correlated probabilities fusion model

Using the assumption of conditional independence (equation 3), we propose a model for the fusion of correlated probabilities. The fusion of the correlated observations obtained from the two media sources $\mathbf{M}^{i-1}$ and M_i is modelled as -

$$P(G|\mathbf{M}^{i-1}, M_i) = f(P(G|\mathbf{M}^{i-1}), P(G|M_i), \overline{\gamma_i}) \tag{11}$$

where,
$P(G|\mathbf{M}^{i-1}, M_i)$ is the fusion probability.
$P(G|\mathbf{M}^{i-1})$ is the probability of the $i-1$ number of streams together achieving the goal G.
$P(G|M_i)$ is the probability of i^{th} media stream individually achieving the goal G.
$\overline{\gamma_i}$ is the correlation between $P(G|\mathbf{M}^{i-1})$ and $P(G|M_i)$.
f is a heuristic function that increases monotonically with respect to the correlation coefficient $\overline{\gamma_i}$.

The function f is defined in simpler notation as -

$$f(g,h,c) = \frac{g.h.e^{m.c}}{g.h.e^{m.c} + (1-g).(1-h)e^{-m.c}} \tag{12}$$

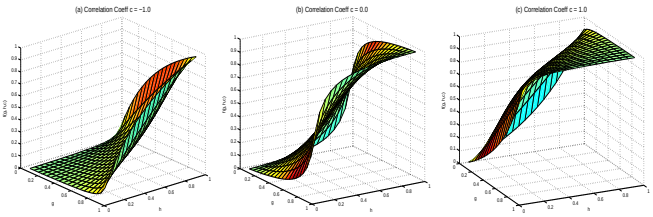

Figure 1: Behavior of the heuristic function (formulation (12)) used for probability fusion model

where, $g = 0 < P(G|\mathbf{M}^{i-1}) \leq 1, h = 0 < P(G|M_i) \leq 1$, $c = \overline{\gamma_i}$, e is the exponential function and $m \in (0, 1]$ is a model coefficient that denotes the weight of the correlation coefficient.

Note that the function f is one possible function that satisfies these boundary conditions and the monotonicity property, however any other function fulfilling these criteria can also be used.

If either of the media sources have the individual probability as zero, we prefer the conservative approach and choose the non-zero probability of successfully achieving the goal G. This approach suits to many real world applications including surveillance and monitoring. For example, if one media stream provides zero probability of detecting a vehicle going to into a no-entry area and other stream provides a probability greater than zero, the proposed algorithm chooses the second one to be on the safer side. Also, if the fusion probability is lesser than the individual probabilities of both the sources, we choose the maximum of individual probabilities (using the assumption **A3** given in section 2).

Figure 1 plots the behavior of the heuristic function f (equation (12)) used for the probability fusion model. The inputs are probabilities g and h on x and y axes, respectively, and output is the fusion probability $f(g,h,c)$ shown on the z-axis (vertical axis). Figures 1a, 1b and 1c show the plots with the correlation coefficients -1.0, 0.0 and +1.0, respectively, and with $m = 1$. It is clearly observable that the fusion probability sooner attains a value close to maximum with higher positive correlation (figure 1c) than with higher negative correlation (figure 1a). Also, as shown in figure 1c, the fusion probability $f(g,h,c)$ attains a value close to 1 even if g and h are well below 1 but have a high positive correlation ($c = +1.0$). It shows that the highly-positively-correlated streams with partial observations can achieve the goal with higher probability when used together.

3.5 Algorithm

In this section, we outline the algorithm for fusing the n observations obtained from the multimedia streams $\mathbf{M}^n$.

Inputs

$p_i, 1 \leq i \leq n$: Individual streams' probabilities.
Γ : The set of correlation coefficients.
m : Model coefficient.

Steps

1. $P = 0$, $P' = 0$
2. For $i = 1$ to n
3. if $(P = 0)$ or $(p_i = 0)$
4. $P = max(P, p_i)$
5. else
6. $\overline{\gamma}_i = 0$
7. For $k = 1$ to $i - 1$

8. if $\overline{\gamma}_i < \gamma_{ki}$
9. $\overline{\gamma}_i = \gamma_{ki}$
10. $P' = \frac{(P.p_i.e^{m.\overline{\gamma}_i})}{(P.p_i.e^{m.\overline{\gamma}_i})+(1-P)(1-p_i).e^{-m.\overline{\gamma}_i}}$
11. if $P' < max(P, p_i)$
12. $P' = max(P, p_i)$
13. $P = P'$
14. return P

Output

P : Fusion probability

4. SIMULATION RESULTS

We have simulated the fusion of observations of 100 media streams to study the behavior of the fusion framework. In figure 2, we show only up to 15 streams since after the fusion of 15 streams the fusion probability is close to the maximum in all cases (figure 2a-2d). To show how correlation affects the fusion, we assume that all the media streams are equiprobable of achieving the goal, and also there is uniform correlation coefficient among all the streams. The simulation is performed for four types of stream sets. The streams within each set have uniform probabilities which are 0.20, 0.40, 0.60, and 0.80 (figure 2a to 2d, respectively). For each set of streams, the stream probabilities are fused sequentially using the correlation coefficients -1.0, -0.5, 0.0, +0.5, +1.0. The value of model coefficient m is taken as 1.

Our observations from the graphs (in figure 2) are:

- The streams having lower probabilities (e.g. 0.20 or 0.40) can also achieve the goal if their correlation is high. The figures 2a & 2b show that 12 streams with correlation coefficient +1.0 (figure 2a), and 5 and 12 streams with correlation coefficients +1.0 and +0.5 (figure 2b), respectively, are sufficient to achieve the goal.
- Figure 2c shows that the negative correlation coefficients (-1.0 and -0.5) does not contribute in achieving the goal even if all the streams having moderate individual probabilities (i.e. 0.60) are used. However, a few streams (lesser than 5) with high correlation coefficients (+0.5 and +1.0) can attain the fusion probability close to maximum. It is also observed that, with zero correlation, around 15 streams having moderate probabilities can still achieve the goal.
- As shown in figure 2d, if the streams having high individual probabilities and high correlation coefficient are fused, even very few streams can achieve the goal. E.g. two streams with probabilities 0.80 and correlation coefficient +1.0 are adequate to achieve the goal.

These results suggest that streams having higher individual probabilities is better, but correlation also plays an important role in improving the overall fusion probability. This indicates that a few but highly correlated streams are better for achieving the goal. But this needs to be studied in detail under varying conditions and analytically proved rigorously. Due to space constraints, we have considered sequential fusion without worrying about any optimality criteria such as the cost of media sensors. Moreover, we have not considered the notion of redundancy which might be wasteful in terms of cost but important for the sake of reliability. We therefore intend to develop a formal notion of optimality in the selection of streams in our future work.

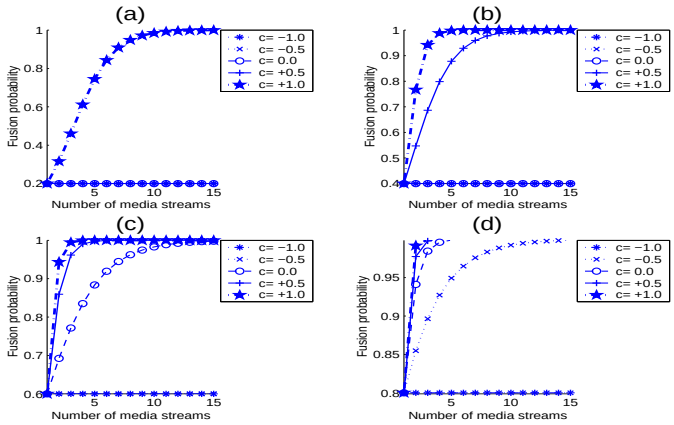

Figure 2: Fusion probability vs. Number of media streams (with uniform probabilities p for all streams) (a) $p = 0.20$ (b) $p = 0.40$ (c) $p = 0.60$ (d) $p = 0.80$

5. CONCLUSION

In this paper, we propose a framework for the probability fusion of correlated multimedia streams. Our method integrates the correlated observations based on the Bayesian approach and a heuristic function. The simulation results show that the correlated observations are advantageous in achieving the goal in a multimedia system environment. It also points out that fewer streams with higher correlations are better. But this needs to be studied in detail. We plan to use the framework in some real scenarios. We also plan to resolve the theoretical issues related to finding the optimal subset of streams for maximizing the probability of achieving the goal under various cost constraints.

6. ACKNOWLEDGEMENTS

We thank our colleague Ee-Chien Chang for providing technical insights and constructive comments during the writing of this paper.

7. REFERENCES

[1] M. Kam, Q. Zhu, and W. S. Gray. Optimal data fusion of correlated local decisions in multiple sensor detection systems. *IEEE Transactions on Aerospace and Electronic Systems*, 28(3):916–920, July 1992.

[2] J. O'Brien. Correlated probability fusion for multiple class discrimination. In *Proceedings of Information Decision and Control*, pages 571–577, Adelaide, Australia, February 1999.

[3] B. S. Rao and H. Durrant-Whyte. A decentralized bayesian algorithm for identification of tracked objects. *IEEE Transactions on Systems, Man and Cybernetics*, 23(6):1683–1698, November-December 1993.

[4] L. A. Rowe and R. Jain. ACM SIGMM retreat report on future directions in multimedia research, March 2004. URL-http://www.acm.org/sigmm/main/events/sigmm_retreat/sigmm-retreat03-final.pdf.

[5] J. Wang and M. Kankanhalli. Experience-based sampling for multimedia analysis. In *Proceedings of ACM Multimedia'03*, Berkeley, November 2003. URL-http://www.comp.nus.edu.sg/ mohan/ebs/.