IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 4, 2019, accepted March 18, 2019, date of publication March 25, 2019, date of current version April 8, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2907256

Representation Learning and Nature Encoded Fusion for Heterogeneous Sensor Networks

LONGWEI WANG AND QILIAN LIANG, (Fellow, IEEE)
Department of Electrical Engineering, The University of Texas at Arlington, Arlington, TX 76019, USA
Corresponding author: Longwei Wang (longwei.wang@mavs.uta.edu)

ABSTRACT Target detection based on heterogeneous sensor networks is considered in this paper. Fusion problem is investigated to fully take advantage of the information of multi-modal data. The sensing data may not be compatible with each other due to heterogeneous sensing modalities, and the joint PDF of the sensors is not easily available. A two-stage fusion method is proposed to solve the heterogeneous data fusion problem. First, the multi-modality data is transformed into the same representation form by a certain linear or nonlinear transformation. Since there is a model mismatch among the different modalities, each modality is trained by an individual statistical model. In this way, the information of different modalities is preserved. Then, the representation is used as the input of the probabilistic fusion. The probabilistic framework allows data from different modalities to be processed in a unified information fusion space. The inherent inter-sensor relationship is exploited to encode the original sensor data on a graph. Iterative belief propagation is used to fuse the local sensing belief. The more general correlation case is also considered, in which the relation between two sensors is characterized by the correlation factor. The numerical results are provided to validate the effectiveness of the proposed method in heterogeneous sensor network fusion.

INDEX TERMS Heterogeneous sensor networks, multi-modal data fusion, representation learning, nature encoded fusion, belief propagation.

I. INTRODUCTION

Heterogeneous sensor networks [1]–[3] with multiple sensing modalities are gaining increasing popularity because they can provide several advantages for performance improvement in different realistic scenarios. Fusion of data from heterogeneous modalities, observing a certain phenomenon, has been shown to improve the performance of many surveillance and monitoring tasks. The key motivation is that sensors of different modalities will provide richer information than a single sensor, or even several sensors of the same modality. Take the audio-visual fusion [8] in human speech communication for example. Speech with the help of visual cues from the lip movements can enhance the intelligibility of speech. The aid of visual cues increase redundancy and makes the speech more robust to noise and interruptions.

A. RELATED WORKS

Two sensors are said to be heterogeneous if their respective observation models cannot be described by the same probability density function [6], [7]. This heterogeneity raises the question of how to integrate the data from such diversity of modalities. If the joint probability density function (PDF) under each hypothesis is known a prior, the optimal performance by the Neyman-Pearson rule for detection (binary hypothesis testing) can be easily obtained [4], [14]. However, in practice, this information may not be available. This usually happens when the dimensionality of the sample data is high and when there are not enough training samples to accurately estimate the joint PDF.

If the data from heterogeneous sensors are independent under certain hypothesis and the local individual sensing data is correctly received in the fusion center, the optimal fusion decision can be obtained by the conventional product rule. However, the problem becomes complicated when the condition that independence among the sensors does not hold [5], [10]. In this paper, the scenario that the joint distribution between the sensors being unknown is considered. This is commonly seen in heterogeneous sensor networks, i.e., sensors with disparate sensing modalities. For example, it is not immediately clear how one could model the joint distributions between data of an audio and a video sensor monitoring a common target of interest.

The associate editor coordinating the review of this manuscript and approving it for publication was Jinhwan Koh.

The heterogeneous data processing [9], [12] has been extensively studied in the literature. The problem of binary hypothesis testing with heterogeneous sensors has been considered in [7], where a parametric framework using the statistical theory of copulas is developed. The application of copula theory for fusing correlated decisions has been recently considered in [10]. It has been shown that there is diversity gain and redundancy loss in the detection problem [11] and the influence of statistical dependence has been characterized. Previous fusion methods include the linear weighting methods [13], majority voting methods and product methods for the totally independent sensors. In linear weighting methods, individual decisions are weighted according to the reliability of the detector and then a threshold comparison is performed to obtain the global decision. The previous methods may result in information loss in many practical scenarios which will degrade the fusion performance. Paper [15] considers the maximally informative projections learning based on mutual information for the speaker association with audiovisual fusion.

Since the data forms of different modalities are not the same, it is difficult to process and fuse them together. For example, the acoustic signal is one dimensional sequence, while the visual signals are two dimensional image flows. The data of different modalities may exhibit heterogeneous data forms and complexities, which makes it difficult for the processing in a single statistical model. Recently, representation learning [16] is widely investigated in the literature. Learning representations of the data makes it easier to extract useful information when building classifiers or other predictors [17]. In this paper, we try to learn a unified representations of heterogeneous modal data for the convenience of data fusion.

B. CONTRIBUTIONS AND ORGANIZATIONS

To overcome the data heterogeneity problem, a two stage fusion framework is proposed to deal with the multi-modal heterogeneous sensor data. Firstly, the multi-modality data is transformed into the same representation form by certain linear or nonlinear transformation. Since there is model mismatch among the different modalities, each modality is trained by an individual statistical model. In this way, the information of different modalities is preserved. Then the representation is used as the input of the probabilistic fusion. The probabilistic framework allows data from different modalities to be processed in a unified information fusion space. The inherent inter-sensor relationship is exploited to encode the original sensor data on a graph. Then iterative belief propagation is used to refine and fuse the local individual belief. Instead of estimating the joint PDF, we just need to abstract the local log-likelihood ratio from learned representations. Then the local log-likelihood ratio is sent to the fusion center for processing. We also consider the more general correlation case, in which the relation between two sensors is characterized by the correlation factor. The belief propagation provides intuitive insights as to how the probabilistic updates reinforce beliefs with the help of correlation factor.

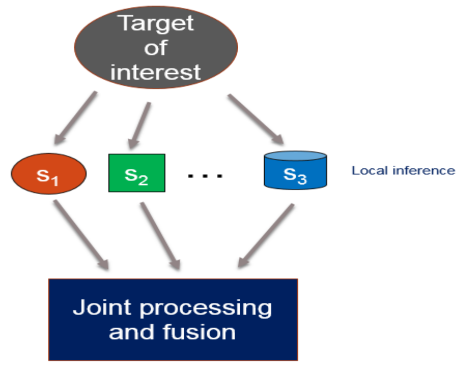

FIGURE 1. Heterogeneous sensors model.

In heterogeneous sensor network, due to the heterogeneity of the sensors, such as the cameras, audio and so on, they display different sensing capability, which would provide incorrect and conflict information sometimes. Also, sensors may suffer from external attacks which will provide the wrong information to the fusion centers. The iterative information fusion proposed in this work can be exploited to combat the conflict and attacks in heterogeneous sensor network.

The remainder of this paper is organized as follows. The two-stage fusion framework is briefly discussed in Section II. We provide in detail the representation learning in Section III. The nature encoded fusion and belief propagation, are presented in Section IV. The more general correlation case is considered in Section V. The numerical results and related discussions are provided in Section VI. Section VII finally concludes the paper.

II. TWO-STAGE FUSION FRAMEWORK

The data of different modalities are processed by two stages: representation learning and nature encoded fusion, as illustrated in Fig. 2.

A. REPRESENTATION LEARNING

Suppose that the data forms for the acoustic and visual signal are $\mathbf{s}^a$ and $\mathbf{s}^v$, respectively. $\mathbf{s}^a$ and $\mathbf{s}^v$ may be of different dimensions and lengths. Representation learning is trying to transform the original multimodal data into a unified representation space, such that the further fusion can be carried out based on the unified data space.

$$\mathbf{y}^a = f_1(\mathbf{s}^a; H^a) \tag{1}$$
$$\mathbf{y}^v = f_2(\mathbf{s}^v; H^v) \tag{2}$$

where $\mathbf{y}^v, \mathbf{y}^a$ are the representations for the two modalities. Since the binary detection problem is considered in this paper, $\mathbf{y}^v, \mathbf{y}^a \in \mathcal{R}^{2\times 1}$. And H^a is the parameters of the statistical transformation model f_1, H^v is the parameters of the statistical transformation model f_2,. Also different statistical models for the different modalities are selected, as shown with f_1 and f_2.

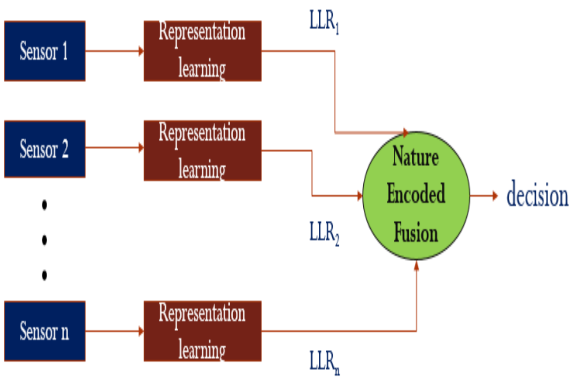

FIGURE 2. Two stage fusion frameworks: the first stage is representation learning, which transforms the heterogeneous data into the unified representation, and the second stage is nature encoded probabilistic fusion.

In this way, the information from heterogeneous modal can be maximally exploited in the heterogeneous sensor networks.

B. NATURE ENCODED FUSION

The learned representations $\mathbf{y}^v$, $\mathbf{y}^a$ are then used to calculate the probabilistic information of the target presence.

$$\mathrm{LLR}_i = \log \frac{p(C_0|\mathbf{s}^i)}{p(C_1|\mathbf{s}^i)} \tag{3}$$

which we refer to as the local belief of sensor $\mathbf{s}^i$ for the target.

The inter-sensor relations among the sensory signals is taken advantage of by modeling it as a nature encoding process. Then the belief propagation algorithm is applied to fuse the probabilistic information. The details of the fusion process is discussed in the following sections.

III. LEARNING REPRESENTATIONS AND TRANSFORMATIONS

Statistical models are developed to learn latent (one-layer or multiple-layer) representation for multi-modal data fusion. The first stage: learning representation, is discussed in this section. The multimodal data is transformed into the unified representation form such that the data fusion can be made possible to achieve improved detection performance. The learned representation provides the probabilistic information for the fusion stage.

A. TRANSFORMATIONS BASED ON DIFFERENT STATISTICAL MODELS

Since video signal data is one dimensional sequence, we can resort to the linear model $\mathbf{y}^a = f_1(\mathbf{s}^a; H^a)$, which can result in a significant computational savings. The transformation method for learning the representations of acoustic signal data is the linear function of the measurements, which is shown as follows,

$$\mathbf{y}^a = H^a \mathbf{s}^a \tag{4}$$

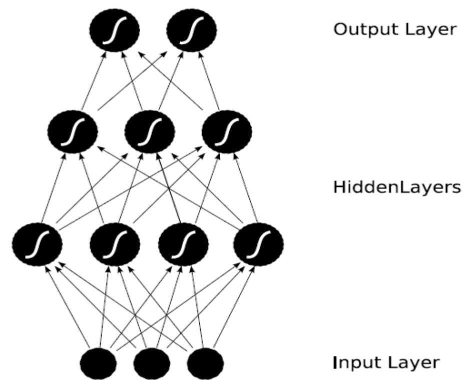

FIGURE 3. Multi-layer neural network based representation learning.

where $\mathbf{y}^a \in \mathcal{R}^{2\times 1}$, $\mathbf{s}^a \in \mathcal{R}^{N\times 1}$, and the binary detection problem, $\mathbf{H}^a \in \mathcal{R}^{2\times N}$ is considered here. This linear model can result in a significant computational savings for the acoustic data representation learning.

The visual signal is usually two dimensional image. The neural network [18] is an efficient and frequently used method for the two dimension visual images. The neural network can achieve better performance due to the shift invariance property and nonlinear transformation.

A simple neural networks is shown in Fig. 3. Consider the neural network [18] with I input units, activated by input vector $\mathbf{x}$. Each unit in the first hidden layer calculates a weighted sum of the input data. For hidden unit l, we refer to this sum as the network input to unit l, and denote it a_l. The activation function θ_l is then applied, yielding the final activation b_l of the unit. Denoting the weight from unit i to unit j as w_{ij}, we have

$$a_l = \Sigma_{i=1}^{I} w_{il} x_i \tag{5}$$
$$b_l = \theta_l(a_l) \tag{6}$$

The most common choices of activation function are the hyperbolic tangent

$$\tanh(x) = \frac{e^{2x}-1}{e^{2x}+1} \tag{7}$$

After calculating the activations of the units in the first hidden layer, the process of summation and activation is then repeated for the rest of the hidden layer units. The network output $\mathbf{y}_k^v$ is calculated by summing over the units connected to it,

$$\mathbf{y}_k^v = \Sigma_{i=1}^{H_m} w_{ik} x_i \tag{8}$$

for a network with m hidden layers, H_m is the number of units in the mth layer.

B. SOFTMAX FUNCTION TO OBTAIN ESTIMATES OF THE CLASS PROBABILITIES

For classification problems with K classes, the convention is to have K output units, and normalize the output activations

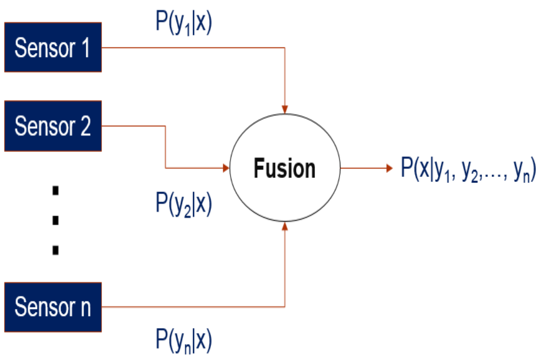

FIGURE 4. Nature encoded probabilistic fusion.

with the softmax function to obtain estimates of the class probabilities:

$$p(C_k|\mathbf{s}^v) = \frac{e^{\mathbf{y}_k^v}}{\sum_{k'=1}^{K} e^{\mathbf{y}_{k'}^v}} \tag{9}$$

The calculation of probabilities can be done in the same way for the acoustic signal case. Since the binary detection problem is considered in this paper, we assume here, $K = 2$. In this way, the learned unified representation is obtained and it provides the probabilistic information for the fusion stage.

IV. BELIEF PROPAGATION AND NATURE ENCODED FUSION

In this section, the representation outputs of the first stage are combined and fused under the probabilistic framework. The probabilistic approach theoretically provides a unified framework for data fusion in heterogeneous sensor networks.

A. CLASS PROBABILITIES AS THE INPUT OF THE PROBABILISTIC FUSION

Instead of estimating the joint PDF of the heterogeneous sensor observations, the local logarithmic likelihood ratio (LLR) of each sensors based on the class probabilities derived in the representation learning stage is calculated.

$$\text{LLR}_i = \log \frac{p(C_0|\mathbf{s}^i)}{p(C_1|\mathbf{s}^i)} \tag{10}$$

which is referred to as the local belief of sensor $\mathbf{s}^i$ for the target. The fusion framework is depicted in Fig. 4.

A nature encoded fusion method based on the belief propagation principle is proposed. The joint PDF of the sensors' observations is unknown. In the method, each sensor only perform local inference by calculating the local LLR. It is easily available since the local LLR can be estimated individually in each sensor by the representation learning. Belief fusion is an efficient way and provide an viable framework for information fusion in heterogeneous sensor networks. The belief updating process can provide useful insight as to how the heterogeneous sensors' information help reinforce the local beliefs in the data fusion.

B. INHERENT CONSTRAINTS AMONG THE SENSORS

The inter-sensor relations among the sensory signals are taken advantage of by modeling it as a nature encoding process. Then the LLR can be updated in the Bayesian way by posing some inherent constraints on the local inference of the heterogeneous sensors.

According to the Bayesian principle, the posterior LLR is the summation of likelihood LLR and the prior LLR. However, there is lack of prior information in the present model. In this paper, the prior of one sensor can be provided by the belief of other sensors through some inherent constraints among the sensors. Then the local belief can be propagated among the sensors and finally a better fusion results can be obtained.

$$\text{LLR}_{posterior} = \text{LLR}_{likelihood} + \text{LLR}_{prior} \tag{11}$$

When there is only one target, the sensing results should be the same for all the sensors. we can make use of this intermodal relation to fuse all the local belief. For any two of the sensors, we can generate a factor node by letting the logic summation of them be zero.

$$x_0 = x_1 = x_2 = \cdots = x_n$$

where x_n denotes the logic presence of the target for sensor n, which takes value 0 or 1.

The corresponding factors can be written as

$$f_1 = x_0 + x_1 = 0 \tag{12}$$
$$f_2 = x_1 + x_2 = 0 \tag{13}$$
$$\cdots \tag{14}$$
$$f_k = x_{k-1} + x_k = 0 \tag{15}$$

which we can write it in matrix form $\begin{bmatrix} 1 & 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$. Each row corresponds to a constraint and the entries with 1 mean that the sensor is involved in that constraint.

The constraints can be relaxed and let the summation of any even number of sensors be zero, such as

$$x_0 + x_1 + x_2 + x_3 = 0$$
$$x_1 + x_2 + x_3 + x_4 + x_5 + x_6 = 0$$

which will be evaluated in the simulation. The illustration for the factor graph is shown in Fig. 5.

C. PROBABILITY CALCULATION FOR THE FUSION PROCESS

For the modulo-2 addition, suppose there are two binary random variables x and y, with probability distributions $p(x) = \{p_0^x, p_1^x\}$ and $p(y) = \{p_0^y, p_1^y\}$, then the probability distribution for the modulo-2 addition [21] is

$$P(x \oplus y = 0) = \frac{1}{2}((p_0^x + p_1^x)(p_0^y + p_1^y) + (p_0^x - p_1^x)(p_0^y - p_1^y)) \tag{16}$$

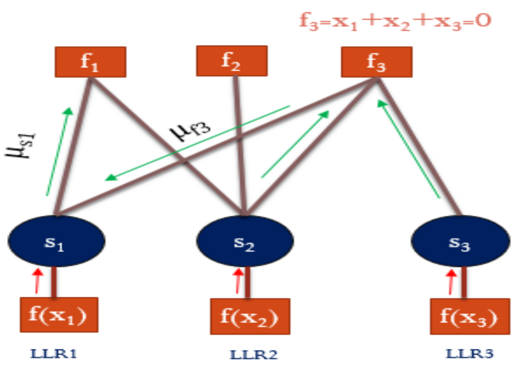

FIGURE 5. Factor graph for nature encoded fusion: graph relations between factor nodes and local sensor belief nodes.

The result can be extended to the infinite number of random variables x_i, $i = 0, 1, 2, \ldots, N$, then the distribution for the modulo-2 addition is

$$P(\sum_{i=1} \oplus x_i = 0) = \frac{1}{2}(\prod_{i=0}^{N}(p_0^{x_i} + p_1^{x_i}) + \prod_{i=0}^{N}(p_0^{x_i} - p_1^{x_i})) \quad (17)$$

$$P(\sum_{i=1} \oplus x_i = 1) = \frac{1}{2}(\prod_{i=0}^{N}(p_0^{x_i} + p_1^{x_i}) - \prod_{i=0}^{N}(p_0^{x_i} - p_1^{x_i})) \quad (18)$$

Then the log-likelyhood ratio (LLR) is defined as $L(x) = \ln(\frac{P(x=0)}{P(x=1)})$. We can write the corresponding probability in the LLR form [20] as follows:

$$P(x = 0) = \frac{e^{L(x)}}{1 + e^{L(x)}} \quad (19)$$

$$P(x = 1) = \frac{1}{1 + e^{L(x)}} \quad (20)$$

Then the LLR form of the modulo-2 variable $L(\sum_{i=1} \oplus x_i) = L(x_1 \oplus x_2 \oplus \ldots \oplus x_N)$ can be derived. According to (17)(18) and (19)(20),

$$L(\sum_{i=1} \oplus x_i) = \ln(\frac{\prod_{i=0}^{N}(e^{L(x_i)} + 1) + \prod_{i=0}^{N}(e^{L(x_i)} - 1)}{\prod_{i=0}^{N}(e^{L(x_i)} + 1) - \prod_{i=0}^{N}(e^{L(x_i)} - 1)}) \quad (21)$$

The equation can be rewritten as

$$L(\sum_{i=1} \oplus x_i) = \ln(\frac{1 + \prod_{i=0}^{N} \tanh(L(x_i))}{1 - \prod_{i=0}^{N} \tanh(L(x_i))})$$

here $\tanh(\frac{x}{2}) = \frac{e^x - 1}{e^x + 1}$.

Then let $\gamma = \prod_{i=0}^{N} \tanh(\frac{L(x_i)}{2})$, we have

$$\tanh(\frac{1}{2}L(\sum_{i=1} \oplus x_i)) = \frac{e^{\ln(\frac{1+\gamma}{1-\gamma})} - 1}{e^{\ln(\frac{1+\gamma}{1-\gamma})} + 1} = \frac{(1+\gamma) - (1-\gamma)}{(1+\gamma) + (1-\gamma)} = \gamma$$

So the following can be obtained

$$L(\sum_{i=1} \oplus x_i) = 2\tanh^{-1}(\prod_{i=0}^{N} \tanh(\frac{L(x_i)}{2})) \quad (22)$$

D. LLR FORM FOR THE BELIEF PROPAGATION

Belief propagation algorithm uses message passing over the factor graph [19]. There are two sets of variables over the graph: variable nodes and factor nodes. The variable nodes correspond to the sensor belief, while the factor nodes corresponds to the constraints in our settings. Now we will go to the details of applying belief propagation algorithm to the decoding of heterogeneous sensor networks. First, the one to one correspondence of message distribution and the LLR form is defined as

$$\lambda_{x_n \to f_m}(x_n) = L(\mu_{x_n \to f_m}(x_n)) \quad (23)$$

$$\lambda_{f_m \to x_n}(x_n) = L(\mu_{f_m \to x_n}(x_n)) \quad (24)$$

So it is straight forward to convert the updating rule from variable to factor message to the following form:

$$L(\mu_{x_n \to f_m}(x_n)) = L(\prod_{f_i \in f_n \backslash f_m} \mu_{f_i \to x_n}(x_n)) \quad (25)$$

$$\lambda_{x_n \to f_m}(x_n) = \sum_{f_i \in f_n \backslash f_m} \lambda_{f_i \to x_n}(x_n) \quad (26)$$

Then the updating from the factor to the variable message is considered. Since $f_m = \sum_{x_i \in x_m} \oplus x_i$ and the logic relation between x_n and f_m is

$$x_n = f_m \oplus \sum_{x_i \in x_m \backslash x_n} \oplus x_i \quad (27)$$

It is known that the updating rule for factor to variable is

$$L(\mu_{f_m \to x_n}) = L(\sum_{\mathcal{X}_m \backslash x_n} f_m(x_m) \prod_{x_i \in x_m \backslash x_n} \mu_{x_i \to f_m}(x_i)) \quad (28)$$

That is

$$\lambda_{f_m \to x_n}(x_n) = L(y_m \oplus \sum_{x_i \in \mathcal{X}_m \backslash x_n} \oplus x_i) \quad (29)$$

Based on the results (22), the following expression can be derived,

$$\lambda_{f_m \to x_n}(x_n) \quad (30)$$

$$= 2\tanh^{-1}(\tanh(\lambda_{f_m}) \prod_{x_i \in \mathcal{X}_m \backslash x_n} \tanh(\frac{\lambda_{x_i \to f_m}(x_i)}{2})) \quad (31)$$

V. GENERAL CORRELATION CASE AND THEORETICAL ANALYSIS OF THE FUSION PROBLEM

In this section, the more general correlation case is considered, in which the relation between two sensors is characterized by the spatial correlation factor λ_{ij}. A probabilistic graphical model is proposed to fuse the prior information from other sensors. The corresponding belief updating rule is developed and the performance is analyzed theoretically.

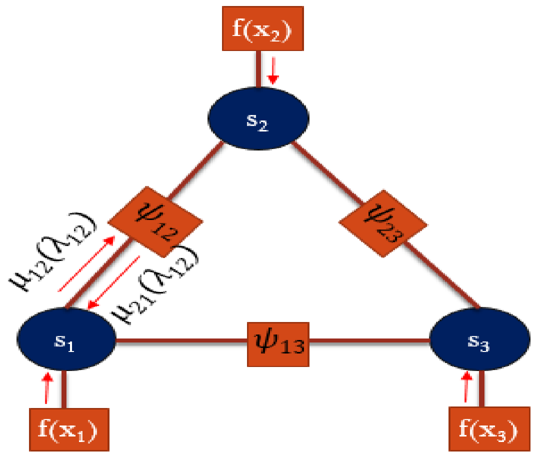

FIGURE 6. Factor graph for general correlated sensors.

A. CORRELATION MODELING

It is assumed that the spatial interactions between adjacent nodes are pairwise. The correlation between sensors x_i and x_j can be modeled as

$$\psi_{ij}(x_i, x_j) = \exp(\lambda_{ij} g(x_i, x_j)) \tag{32}$$

where $\lambda_{ij} \geq 0$ and it is proportional with the increased correlation between the sensors. And the indicator function $g(x_i, x_j)$ is defined as

$$g(x_i, x_j) \triangleq x_i x_j + (1 - x_i)(1 - x_j) \tag{33}$$

The joint posterior distribution can be given as

$$p(x|y) \propto \prod_{i=1} p(y_i|x_i) \prod_{j \in \mathcal{N}_i} \exp(\lambda_{ij} g(x_i, x_j)) \tag{34}$$

which is the product of the local likelihood and the correlation factor. The belief of the target for each sensor $p(x_n|y)$ is determined, given the observations and the correlations. The problem can be converted as the maximum a posteriori (MAP) estimation:

$$\hat{x} = \arg \max_{x \in \{0,1\}^n} p(x|y) \tag{35}$$

B. BELIEF PROPAGATION FOR SPATIAL CORRELATED SENSOR OBSERVATIONS

The spatially correlated sensors can be illustrated in a factor graph, as depicted in Fig. 6. The sensors are represented as circles and the inter-sensor correlations are represented as squares.

Now the sensors' information is fused based on belief propagation algorithm. Each sensor fuse the multi-prior information from other sensors through the inter-sensor correlations. The final belief of each sensor can be reached by iteration of messages among the nodes.

Here, the variable nodes correspond to the sensor belief, while the factor nodes corresponds to the inter-sensor correlation in our settings. It tries to reveal the distribution of variable nodes with the help of factor nodes. By performing message passing among each pair of the variable and factor nodes, the belief of the variable nodes can be updated. Usually, the updating rules [21] are defined as:

- The message from variable nodes to factor nodes:

$$\mu_{x_n \to f_m} = \prod_{f_i \in \mathcal{F}_n \backslash f_m} \mu_{f_i \to x_n}(x_n) \tag{36}$$

- Updating rule from factor to variable:

$$\mu_{f_m \to x_n} = \sum_{\sim x_n} f_m \prod_{x_i \in \mathcal{X}_m \backslash x_n} \mu_{x_i \to f_m}(x_i) \tag{37}$$

where the factor node f_m and the variable node x_n are connected. $\mathcal{F}_n$ is the set of factor nodes that are connected to variable node x_n and likewise, $\mathcal{X}_m$ is the set of variable nodes that are connected to factor node f_m.

Note that, there are two types of factor nodes connected to each variable node. The first one is the local sensing factor,

$$\mu_{f \to x_n} = f(x_n) \tag{38}$$

the other one is the inter-sensor correlation $f_m = \psi_{ni}$,

$$\mu_{f_m \to x_n} = \sum_{\sim x_i} f(x_n, x_i) \mu_{x_i \to f_m}(x_i) \tag{39}$$

$$= \sum_{\sim x_i} \psi_{ni}(x_n, x_i) \mu_{x_i \to f_m}(x_i) \tag{40}$$

After each iteration, the belief of each variable node is computed by the product of all the information from the factor node. And then the probability is normalized and pass it to its corresponding factor node.

C. KULLBACK-LEIBLER DIVERGENCE AND THEORECTICAL ANALYSIS

In the detection problem, when there are two hypothesis, the sensed data would exhibit distribution $f_0(x|H_0)$ under hypothesis H_0 and distribution $f_1(x|H_1)$ under hypothesis H_1. The Kullback-Leibler Divergence (KLD) [7] is usually used to evaluate the performance of detection, which is defined as

$$D(H_0 \| H_1) = \mathrm{E}_{H_0}[\log \frac{f(x|H_0)}{f(x|H_1)}] \tag{41}$$

where E_{H_0} is the expectation taken with respect to the joint distribution of x under hypothesis H_0. KLD can be interpreted as the error exponent in the Neyman-Pearson framework, which means that the probability of miss detection goes to zero exponentially with the number of observations at a rate equal to KLD.

If the sensor observations are independent of each other, then $p_1(x, y) = p_1(x)p_1(y)$, $p_0(x, y) = p_0(x)p_0(y)$. Then KLD can be written as

$$D(p_1(x, y) \| p_0(x, y)) = D(p_1(x) \| p_0(x)) + D(p_1(y) \| p_0(y)) \tag{42}$$

In the case of correlated modalities $p_1(x, y) \neq p_1(x)p_1(y)$, The KLD for the two hypothesis can be given according to the chain rule,

$$D(p_1(x, y)\|p_0(x, y)) = D(p_1(x)\|p_0(x)) + D(p_1(y|x)\|p_0(y|x)) \tag{43}$$

there is a lemma in [10] which proves that conditioning does not reduce relative entropy, which means that $D(p_1(y)\|p_0(y)) \leq D(p_1(y|x)\|p_0(y|x))$. In this paper, the two stage fusion framework is proposed, which exploits this kind of correlations between the sensor observations via belief propagation.

Also in [11], it gives a proposition that the KLD is non-decreasing with the increased number of sensors. For any $S' \subseteq S$, $D(S') \leq D(S)$. This means that the heterogeneity among the sensors can improve the detection performance.

VI. PERFORMANCE ANALYSIS AND NUMERICAL RESULTS

Numerical simulations are performed to illustrate our proposed fusion algorithm for the heterogeneous sensor network. To model the heterogeneous modalities, we make some abstractions and generate the sensing data based on different distributions. A sensor network with $N = 5$ sensors is considered. Each sensor generates the local belief of the target by representation learning. The training of the representation learning is performed individually based on the training data sets.

The correlation between sensors is quantified by the power exponential model.

$$\lambda_{ij} = e^{-d_{ij}^2} \tag{44}$$

where d_{ij} is the distance between node i and j. Each sensor's data follow Gaussian distribution with different mean values θ and variance σ^2. They can be adjusted to meet different SNR scenarios.

A. FUSION FOR THE CONFLICTED BELIEF

We first consider the robustness of the belief propagation based nature encoded fusion method. Two different initial LLR settings are evaluated, which are the $[1, -2, -1, -2, -1]$ and $[1, -1, -2, -5, -6]$. The convergence results for both cases are shown in Fig. 7 and Fig. 8. For both cases, the first sensor's initial belief is negative and is opposite to the rest of sensors. After a few iterations, all the sensors' belief become positive. The belief propagation forces the ambiguous belief to be correct thanks to the correlations among the sensors. After the fusion, we can apply the majority voting method and the detection performance would be robust to the error and uncertain noise.

B. IMPACT OF THE CORRELATION COEFFICIENTS

The scenario that the correlation among the sensors λ_{ij} goes large is considered in Fig. 9 and Fig. 10. The LLR values of all the sensors are forced to be the same belief in Fig. 9.

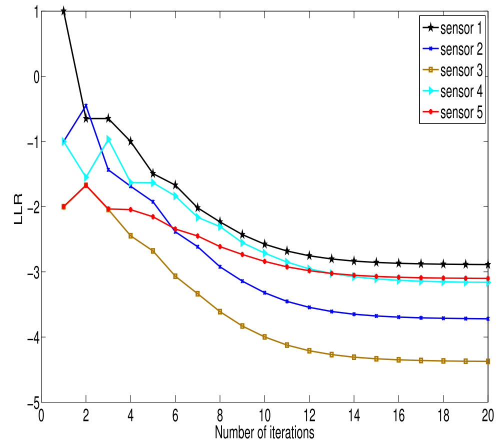

FIGURE 7. Iterations of belief propagation based fusion, LLR = [1, −2, −1, −2, −1].

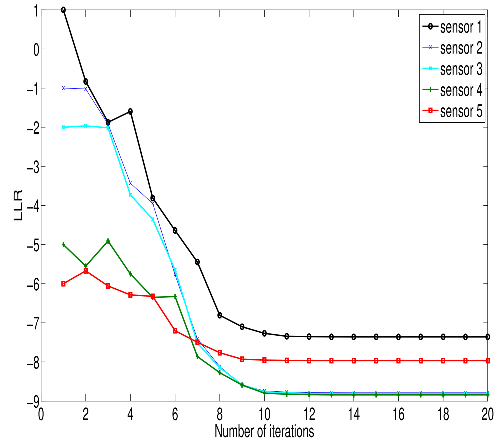

FIGURE 8. Iterations of belief propagation based fusion, LLR = [1, −1, −2, −5, −6].

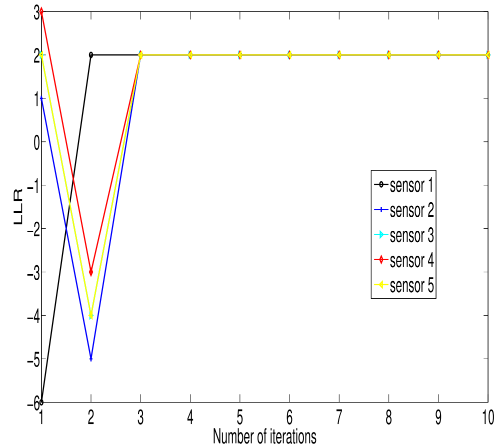

FIGURE 9. General correlation case: Initial LLR = [−6 1 2 3 2].

Even when the first sensor exhibits high negative LLR value, the iterative belief propagation can correct that belief and achieve a consensus fusion result. The reason why the case in Fig. 10 does not converge is that the factor graph contains loops in this example. Also we can see that the first three sensors' LLRs are negative and exhibit weak belief as to

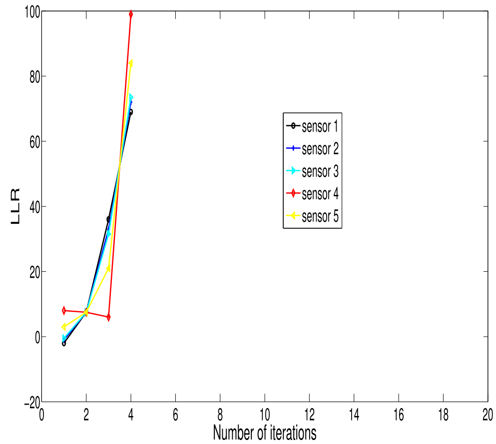

FIGURE 10. General correlation case: Initial LLR = [−0.2 −0.2 −0.5 8 3].

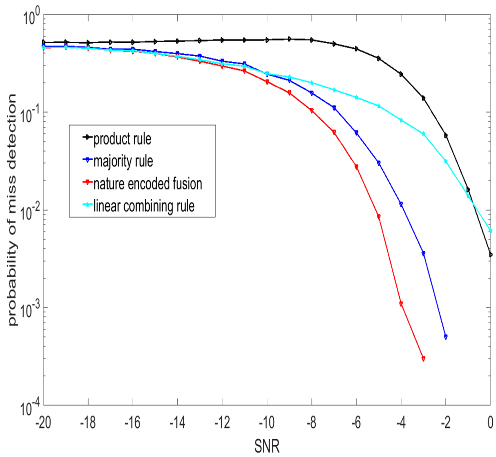

FIGURE 11. Performance comparison with product rule, linear combining rule, and majority voting.

whether the hypothesis is H_0. Still they can be refined to be positive thanks to the correlation among the sensors. If majority voting is applied after the refinement, we can achieve better detection performance. Correlation among the sensor observations makes some sensors' measurements redundant, which provides more robustness in the fusion process.

C. PERFORMANCE COMPARISON

The performance of proposed nature encoded fusion is compared with the majority voting and linear weighted combining method. The linear weighted combing is computed as

$$p_{fusion} = \frac{\sum_i w_i p_i}{\sum_i w_i} \tag{45}$$

Fig. 11 shows that the belief propagation based fusion can achieve superior detection performance compared with the other three fusion methods. It may result in information loss duo to hard processing of majority voting scheme. Also much training overhead is required to obtain the optimal weighting coefficient for the linear combining method. The product rule will cause error propagation which will degrade the fusion performance.

VII. CONCLUSIONS

In this paper, a two-stage framework for fusing information from heterogeneous sensors is proposed. The representation learning stage transforms the data into a unified data form. The nature encoded fusion allows data from different modalities to be processed in a unified probabilistic space. The inherent inter-sensor relationship is exploited and it can be seen as a nature encoded sensing with heterogeneous sensors. Then iterative belief propagation is used to refine and fuse the local individual belief. Instead of estimating the joint PDF, we just need to abstract the local log-likelihood ratio from each sensor. Then the local log-likelihood ratio is sent to the fusion center for processing. Further the more general correlation case is considered, in which the relation between two sensors is characterized by the correlation factor. The belief propagation provides intuitive insights as to how the probabilistic updates reinforce beliefs with the help of correlation factor.

REFERENCES

[1] Y. Tong, L. Liu, M. Zhao, J. Chen, and H. Li, "Adaptive fusion algorithm of heterogeneous sensor networks under different illumination conditions," *Signal Process.*, vol. 126, pp. 149–158, Sep. 2016.
[2] Z. Zhang, Z. Zhang, and S. Liu, "Cross domain boosting for information fusion in heterogeneous sensor-cyber sources," *Signal Process.*, vol. 126, pp. 180–186, Sep. 2016.
[3] M. Kushwaha, S. Oh, I. Amundson, X. Koutsoukos, and A. Ledeczi, "Target tracking in heterogeneous sensor networks using audio and video sensor fusion," in *Proc. IEEE Int. Conf. Multisensor Fusion Integr. Intell. Syst.*, Aug. 2008, pp. 14–19.
[4] L. Vergara, A. Soriano, G. Safont, and A. Salazar, "On the fusion of non-independent detectors," *Digit. Signal Process.*, vol. 50, pp. 24–33, Mar. 2016.
[5] X. Shen, P. K. Varshney, and Y. Zhu, "Robust distributed maximum likelihood estimation with dependent quantized data," *Automatica*, vol. 50, no. 1, pp. 169–174, 2014.
[6] V. V. Veeravalli and P. K. Varshney, "Distributed inference in wireless sensor networks. *Philos. Trans. Roy. Soc. A, Math., Phys. Eng. Sci.*, vol. 370, no. 1958, pp. 100–117, 2012.
[7] S. G. Iyengar, P. K. Varshney, and T. Damarla, "A parametric copula-based framework for hypothesis testing using heterogeneous data," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2308–2319, May 2011.
[8] P. Wu, H. Liu, X. Li, T. Fan, and X. Zhang, "A novel lip descriptor for audio-visual keyword spotting based on adaptive decision fusion," *IEEE Trans. Multimedia*, vol. 18, no. 3, pp. 326–338, Mar. 2016.
[9] R. Bramon *et al.*, "Multimodal data fusion based on mutual information," *IEEE Trans. Vis. Comput. Graphics*, vol. 18, no. 9, pp. 1574–1587, Sep. 2012.
[10] S. G. Iyengar, R. Niu, and P. K. Varshney, "Fusing dependent decisions for hypothesis testing with heterogeneous sensors," *IEEE Trans. Signal Process.*, vol. 60, no. 9, pp. 4888–4897, Sep. 2012.
[11] H. He and P. K. Varshney, "A coalitional game for distributed inference in sensor networks with dependent observations," *IEEE Trans. Signal Process.*, vol. 64, no. 7, pp. 1854–1866, Apr. 2016.
[12] N. Checka, K. Wilson, V. Rangarajan, and T. Darrell, "A probabilistic framework for multi-modal multi-person tracking," in *Proc. Conf. Comput. Vis. Pattern Recognit. Workshop*, vol. 9, Jun. 2003, p. 100.
[13] H. Glotin, D. Vergyr, C. Neti, G. Potamianos, and J. Luettin, "Weighting schemes for audio-visual fusion in speech recognition," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 1, May 2001, pp. 173–176.
[14] S. Kay, Q. Ding, and M. Rangaswamy, "Sensor integration by joint PDF construction using the exponential family," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 49, no. 1, pp. 580–593, Jan. 2013.
[15] H. Dilpazir, Z. Muhammad, Q. Minhas, F. Ahmed, H. Malik, and H. Mahmood, "Multivariate mutual information for audio video fusion," *Signal, Image Video Process.*, vol. 10, no. 7, pp. 1265–1272, 2016.

[16] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 8, pp. 1798–1828, Aug. 2013.
[17] N. Le and J.-M. Odobez, "Learning multimodal temporal representation for dubbing detection in broadcast media," in *Proc. 24th ACM Int. Conf. Multimedia*, Oct. 2016, pp. 202–206.
[18] M. Oquab, L. Bottou, I. Laptev, and J. Sivic, "Learning and transferring mid-level image representations using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1717–1724.
[19] H.-A. Loeliger, "An introduction to factor graphs," *IEEE Signal Process. Mag.*, vol. 21, no. 1, pp. 28–41, Jan. 2004.
[20] A. Montanari and R. Urbanke, "Course 2 Modern coding theory: The statistical mechanics and computer science point of view," *Les Houches*, vol. 85, pp. 67–130, 2007.
[21] D. J. Costello, *Error Control Coding: Fundamentals and Applications*. Upper Saddle River, NJ, USA: Prentice-Hall, 1983.
[22] T. J. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge, U.K.: Cambridge Univ. Press, 2008.

LONGWEI WANG is currently pursuing the Ph.D. degree in electrical engineering with The University of Texas at Arlington. His current research interest includes signal processing and wireless communications.

QILIAN LIANG (F'16) received the B.S. degree from Wuhan University, in 1993, the M.S. degree from the Beijing University of Posts and Telecommunications, in 1996, and the Ph.D. degree from the University of Southern California (USC), in 2000, all in electrical engineering. Prior to joining the faculty of The University of Texas at Arlington, in 2002, he was a Member of Technical Staff at Hughes Network Systems Inc., San Diego, CA, USA. He is currently a Distinguished University Professor with the Department of Electrical Engineering, The University of Texas at Arlington. He has published over 300 journal and conference papers and seven book chapters, and holds six U.S. patents pending. His research interests include radar sensor networks, wireless sensor networks, wireless communications, compressive sensing, smart grid, signal processing for communications, and fuzzy logic systems and applications. He received the 2002 IEEE Transactions on Fuzzy Systems Outstanding Paper Award, the 2003 U.S. Office of Naval Research (ONR) Young Investigator Award, the 2005 UTA College of Engineering Outstanding Young Faculty Award, the 2007, 2009, and 2010 U.S. Air Force Summer Faculty Fellowship Program Award, the 2012 UTA College of Engineering Excellence in Research Award, and the 2013 UTA Outstanding Research Achievement Award, and he was inducted into the UTA Academy of Distinguished Scholars, in 2015.