MODEL LEVEL FUSION OF EDGE HISTOGRAM DESCRIPTORS AND GABOR WAVELETS FOR LANDMINE DETECTION WITH GROUND PENETRATING RADAR

Oualid Missaoui[1], Hichem Frigui[1], and Paul Gader[2]

[1]Multimedia Research Lab, CECS Dept. University of Louisville

[2]CISE Dept. University of Florida

Abstract

We propose combining heterogeneous sets of features for a continuous hidden Markov model classifier. We use a model level fusion approach and apply it to the problem of landmine detection using ground penetrating radar (GPR). We hypothesize that each signature (mine or non-mine) can be characterized better by multiple synchronous sequences that can capture different and complementary salient. Our work is motivated by the fact that mines and clutter objects can have different characteristics depending on the mine type, soil and weather conditions, and burial depth. Thus, ideally different sets of specialized feature extraction mechanisms, may be needed to achieve high detection and low false alarm rates. In order to fuse the different modalities, a multi-stream continuous HMM that includes a stream relevance weighting component is developed. In particular, we modify the probability density function that characterizes the standard continuous HMM to include state and component dependent stream relevance weights. We modify the maximum likelihood based Baum-Welch algorithm and the Minimum Classification Error/Gradient Probabilistic Descent (MCE/GPD) learning algorithms to include stream relevance weights. We generalize their objective functions and derive the necessary conditions to update all model parameters simultaneously. We use the proposed approach to build an HMM classifier that combines two sets of features. The first one, based on edge histogram descriptor, extracts edges in the time domain. The second set of features extracts edges in the frequency domain at multiple scales and orientations. Results on a large collection of GPR alarms show that the proposed model level fusion outperforms the baseline HMM when each feature is used independently and when both features are combined with equal weights.

1. INTRODUCTION

For complex classification systems, such as the landmine detection problem, data is usually gathered from multiple sources of information. Typically, the different sources or modalities do not have the same reliability or expressiveness of the original data. They interchangeably uncover the salient mine characteristics depending on soil type where the mine is buried, burial depth, mine type, and other unknown factors. Thus, the fusion of different features can potentially increase the overall performance of the landmine detector. Approaches toward combination of different sources of information can be divided into three main categories: feature level fusion or direct identification, decision level fusion or separate identification (also known as late integration) and model level fusion (early/intermediate integration) [1]. In feature level fusion, multiple features are concatenated into a large feature vector. This assumes that the different modalities have the same relevance. Since the modalities can have varying reliability, this might lead to a suboptimal behavior. In decision level fusion, the modalities are processed separately to build independent models [2]. This approach ignores the correlation between features.

In model level fusion and for an HMM-based classifier, an HMM model that is more complex than a standard one is sought. This additional complexity is needed to handle the correlation between modalities, and the loose synchronicity between sequences. Several HMM structures have been proposed for this purpose. Examples include factorial HMM [3], coupled HMM [4] and Multi-stream HMM [5]. Both factorial and coupled HMM structures allow asynchrony between sequences since a state sequences is assigned to each stream [6]. However, this is performed at the expense of approximate parameter estimation.

Multi-stream HMM (MSHMM) is an HMM based structure that handles multiple modalities for temporal data. It is used when the modalities (streams) are synchronous and independent. Since the streams are supposed to be synchronous, MSHMM assumes that for each time slot, there is a single hidden state, from which different streams interpret different observations. The independence of the streams means that their interpretation of the hidden state and their generation of the observations is performed independently.

An HMM is a model of doubly stochastic process that produces a sequence of random observation vectors at discrete times according to an underlying Markov chain. At each observation time, the Markov chain may be in one of N states. An HMM is characterized by three sets of probability density functions, the initial probabilities (π), the state transition probabilities **(A)**, and the observation probability density functions **(B)**. An HMM is called discrete if the observation probability density functions are discrete and continuous if the observation probability density functions are continuous. In the continuous HMM, $b_i(o_t)$'s are commonly defined by a mixture of Gaussian density where:

$$b_i(o_t) = \sum_{j=1}^{M} v_{ij}\varphi\left(o_t, \mu_{ij}, \Sigma_{\mathrm{ij}}\right) \text{ subject to } \sum_{j=1}^{M} v_{ij} = 1$$

where v_{ij} is the coefficient of the component j in state i, μ_{ij} and Σ_{ij} denote respectively the mean and the covariance matrix of the Gaussian mixture φ.

In this paper, we propose a multi-stream continuous HMM (MSCHMM) structure that integrates stream relevance weights. We generalize the objective function to include the stream relevance weights and derive the necessary conditions to update the parameters. In particular, we generalize the Baum-Welch [7] and the minimum classification error/generalized probabilistic descent (MCE/GPD) [8] training. We assume that we have L streams of information. These streams could have been generated by different sensors and/or different feature extraction algorithms. Each stream is represented by a different subset of features. Instead of treating the streams equally important or using user-specified weights, the proposed MSCHMM structure introduces a built-in component to learn a relevance weight to each stream. The stream relevance weights are encoded within the probability density function using:

$$b_i(o_t) = \sum_{j=1}^{M} v_{ij} b_{ij}(o_t) \quad subject\ to \quad \sum_{j=1}^{M} v_{ij} = 1$$

where

$$b_{ij}(o_t) = \sum_{k=1}^{L} \boldsymbol{w_{ijk}} \varphi(o_t, \mu_{ijk}, \Sigma_{ijk}) \quad subject\ to \quad \sum_{k=1}^{L} \boldsymbol{w_{ijk}} = 1$$

In the above, w_{ijk} denotes the relevance weight of each feature k.

This linear form of the probability density function allows the generalization of the standard Baum-Welch algorithm and to derive the necessary conditions for learning the relevance weights. In fact, it can be shown that the update equation for the stream relevance weight $\boldsymbol{w_{ijk}}$ is given by:

$$\boldsymbol{w_{ijk}} = \frac{\sum_{t=1}^{T} \gamma_t(i,j,k)}{\sum_{t=1}^{T} \gamma_t(i,j)}$$

where

$\gamma_t(i) = \frac{\alpha_t(i)\beta_t(i)}{\sum_{j=1}^{N} \alpha_t(j)\beta_t(j)}$ $\quad$ $\gamma_t(i,j) = \gamma_t(i) \frac{v_{ij} b_{ij}(o_t)}{b_i(o_t)}$, and $\gamma_t(i,j,k) = \gamma_t(i) \frac{v_{ij} \boldsymbol{w_{ijk}} \varphi(o_t, \mu_{ijk}, \Sigma_{ijk})}{b_i(o_t)}$

The parameters v_{ij} , μ_{ijk} , and Σ_{ijk} are updated as in the standard Baum-Welch.

We also generalize the discriminative training method based on classification error minimization using a steepest descent optimization algorithm [8] to learn the stream relevance weights. We optimize the following function to minimize the classification error over the training data:

$$L(\Lambda) = \sum_{r=1}^{R} \sum_{c=1}^{C} l_c(O_r, \Lambda) \delta(\mathrm{O_r} \in \mathrm{C_c})$$

where R is the size of the training data, C is the number of classes, $l_c(O_r, \Lambda)$ is a loss function that measures the loss correspondent to misclassifying the sequence O_r , and Λ represents the model parameters.

3. EXPERIMENTAL SETUP

As streams, we use two types of edge detection techniques that have been proposed and used independently. The first one characterizes edges in the frequency domain at multiple scales and orientations and are based on Gabor wavelets[9]. The second, edge histogram descriptors (EHD) [10], characterize edges in the spatial domain. Each set of these features represents a different interpretation of the raw data and aims at providing the best discrimination between mine and clutter signatures. They could be interpreted as multiple independent modalities or sources of information that characterize the original data.

The EHD and Gabor features are then extracted and used by the multi-stream CHMM structure to learn the parameters of the mine and the background HMM models. Preliminary results on large and diverse ground penetrating radar data show that the proposed method outperforms the basic CHMM where each stream is used independently and where all features are treated equally important.

REFERENCES

[1] V. Chatziz, A. Bors, and I. Pitas, “Multimodal decision level fusion for person authentication,” IEEE Trans. Syst. Man Cybern. A 29, 1999.

[2] A. Nefian, L. Liang, T. Fu, and X. Liu, “A Bayesian approach to audio-visual speaker identification,” Fourth International Conf. Audio and Video-based Biometric Person Authentication, 2003.

[3] Z. Ghahramani and M. Jordan, “Factorial Hidden Markov Models,” Neural Information Processing Systems (NIPS), 1995.

[4] A. Nefian, L. Liang, T. Fu, and X. Liu, “A Bayesian approach to audio-visual speaker identification,” Fourth International Conf. Audioand Video-based Biometric Person Authentication, 2003.

[5] S. Dupont and J. Luettin, “Audio-visual speech modeling for continuous speech recognition.” IEEE Transactions on Multimedia, Vol. 2, No. 3, September, 2000.

[6] P. Gerasimos, N. Chalapathy, L. Juergen, and M. Iain, “Audio-Visual Automatic Speech Recognition: An Overview,” Audio-Visual Speech Processing, E. Vatikiotis-Bateson, G. Bailly, and P. Perrier (Eds.), MIT Press, 2009.

[7] L. Rabiner, "*Tutorial on hidden Markov models and selected applications in speech recognition*". Proc. IEEE, 77:257–286, 1989.

[8] B.-H. Juang, W. Chou, and C.-H. Lee, “Minimum Classification Error Rate Methods for Speech Recognition,” Transactions on Speech and Audio Processing, VOL. 5, No.3, May 1997.

[9] H. Frigui, A. Hamdi, O. Missaoui, and P. Gader, "Landmine Detection using Mixture of Discrete Hidden Markov Models", SPIE, Detection and Sensing of Mines, Explosive Objects, and Obscured Targets XIV, 2009.

[10] H. Frigui and P. Gader, “Detection and Discrimination of Land Mines in Ground-Penetrating Radar Based on Edge Histogram Descriptors and a Possibilistic K-Nearest Neighbor Classifier,,” IEEE Transactions on Fuzzy Systems, Vol. 17, No.1,, 2009.