OPPORTUNISTIC SENSING FOR OBJECT RECOGNITION — A UNIFIED FORMULATION FOR DYNAMIC SENSOR SELECTION AND FEATURE EXTRACTION

Zhaowen Wang[†], *Jianchao Yang*[†‡], *Nasser Nasrabadi*[§], *Jiangping Wang*[†], *and Thomas Huang*[†]

[†]Beckman Institute, University of Illinois at Urbana-Champaign, Urbana, IL 61801, USA
[‡]Adobe Systems Inc, San Jose, CA 95110, USA
[§]U.S. Army Research Laboratory, Adelphi, MD 20783, USA
{wang308,jwang63,huang}@ifp.uiuc.edu, jiayang@adobe.com, nasser.m.nasrabadi.civ@mail.mil

ABSTRACT

A novel problem of object recognition with dynamically allocated sensing resources is considered in this paper. We call this problem opportunistic sensing since prior knowledge about the correlation between class label and signal distribution is exploited as early as in data acquisition. Two forms of sensing parameters – discrete sensor index and continuous linear measurement vector – are optimized within the same maximum negative entropy framework. The computationally intractable expected entropy is approximated using unscented transform for Gaussian models, and we solve the problem using a gradient-based method. Our formulation is theoretically shown to be closely related to the maximum mutual information criterion for sensor selection and linear feature extraction techniques such as PCA, LDA, and CCA. The proposed approach is validated on multi-view vehicle classification and face recognition datasets, and remarkable improvement over baseline methods is demonstrated in the experiments.

Index Terms— opportunistic sensing, view selection, feature extraction, objection recognition

1. INTRODUCTION

Recently, the demand for large scale data analysis has grown rapidly with the huge amount of multimedia information generated every day from various sources such as internet and surveillance network. However, in many cases, the acquisition of such large data corpus is not only impractical due to prohibitive cost, but also unnecessary because of the noisiness, irrelevance and redundancy in the data. Therefore, an alternative data acquisition paradigm, called *opportunistic sensing*, is considered in this paper to obtain only the most essential part of data to our task with limited sensing resources.

In opportunistic sensing, data pieces are collected sequentially so that sensing resources can be dynamically allocated based on the prior knowledge of data distribution and existing observations, in the hope that data to be captured in the next observation can bring the highest rewards to the current status of task. Obtaining sequential observations of the same physical object or event is possible in many ways. The most common one is via multiple sensors, where we get one observation from one sensor at each time. In this setting, opportunistic sensing becomes the well-known sensor selection problem which has found applications in wireless sensor network [1], target tracking [2], multimedia fusion [3], *etc*.

Another way to observe an object sequentially using a single sensor is to make each observation with a different sensing parameter. For the widely used linear observation of multidimensional signals, the measurement (or projection) vectors are the sensing parameters to be adjusted. It has been demonstrated that adaptively designed measurement vectors can improve reconstruction accuracy [4] or reduce the required number of samples [5] for sequential compressive sensing. In pattern recognition community, appropriate linear projections have been sought for feature extraction and dimensionality reduction. Principle Component Analysis (PCA), Linear Discriminant Analysis (LDA), and Canonical Correlation Analysis (CCA) are among the most popular ones. However, the conventional feature extraction methods only work after data have been fully observed, and therefore are not effective when applied to the problem of opportunistic sensing.

In this paper, we propose to consider both discrete sensor selection and continuous feature extraction in a unified framework for opportunistic sensing, with the aim of robust object recognition using constrained number of data observations. The problem is formulated using a dynamic Bayesian graphical model (Section 2), and the negative expected entropy of class label posterior is used as the metric for selecting data observation most contributive to the ultimate goal of recognition (Section 3). We use unscented transform to approximate the computationally intractable entropy of mixture of Gaussian distribution, and find its gradient to optimize the measurement vector (Section 4). The maximum negative entropy principle is shown theoretically identical to maximum mutual information criterion employed in sensor selection [6], and it also enjoys all the strengths of PCA, LDA, and CCA (Sec-

The work is supported by the U.S. Army Research Laboratory and U.S. Army Research Office under grant number W911NF-09-1-0383.

tion 5). The ability of our algorithm to select the most informative sensors and measurement projections is demonstrated in both multi-view vehicle classification and face recognition tasks, and our approach outperforms existing methods significantly when the number of observations is limited (Section 6). We give concluding remarks in the end (Section 7).

2. FORMULATION OF OPPORTUNISTIC SENSING

In opportunistic sensing, our goal is to find a target's class label x by sequentially extracting the most informative observations from a set of M sensors. At each time step t, we assume a scalar observation y_t of the target can be obtained as the linear projection of the underlying multi-dimensional signal $\mathbf{v}^{z_t}$ perceivable to the sensor with index z_t using measurement vector $\tilde{\mathbf{p}}_t$:

$$y_t = \tilde{\mathbf{p}}_t^T \mathbf{v}^{z_t} + \epsilon, \tag{1}$$

where ϵ is a noise of normal distribution with variance σ_s^2: $\epsilon \sim \mathcal{N}(\epsilon; 0, \sigma_s^2)$. To be more compact, we define concatenated data observation $\mathbf{v} = [\mathbf{v}^1; ...; \mathbf{v}^M]$, and have

$$p(y_t|\mathbf{v}, \mathbf{p}_t) = \mathcal{N}(y_t; \mathbf{p}_t^T \mathbf{v}, \sigma_s^2), \tag{2}$$

where $\mathbf{p}_t = [\delta(z_t - 1) \cdot \mathbf{I}; ...; \delta(z_t - M) \cdot \mathbf{I}]\tilde{\mathbf{p}}_t$ is the concatenated measurement vector containing all the information about sensing condition. Note that $\mathbf{p}_t$ also serves as the operator for linear feature extraction in our opportunistic sensing problem, which entails the joint consideration for data acquisition and recognition. Eq. (2) is called sensing model.

Given class label $x = c$, $\mathbf{v}$ is also assumed to follow a Gaussian distribution described by our data generating model:

$$p(\mathbf{v}|x^c) = \mathcal{N}(\mathbf{v}; \boldsymbol{\mu}_c, \boldsymbol{\Sigma}_c), \tag{3}$$

where $\boldsymbol{\mu}_c$ and $\boldsymbol{\Sigma}_c$ are the mean and covariance matrix of class c, respectively. We use Gaussian distributions here just for simplicity, although our formulation can be easily extended to sensing or generating model of any other form.

Starting from the initial belief of x given by class prior $p(x)$, we update the posterior $p(x|y_{1:t})$ sequentially at each time t by taking an action A_{t+1}, which determines the sensing condition for the next observation through execution model $p(\mathbf{p}_{t+1}|A_{t+1}, \mathbf{p}_t)$. This problem can be represented graphically as in Fig. 1. Our goal is to get a reliable and confident inference of the target label x by dynamically selecting an action sequence $\{A_{1:T}\}$, where T is the total number of available observations. A new action A_{t+1} is selected by maximizing its expected utility function U with the following form:

$$\begin{aligned} & U(A_{t+1}|y_{1:t}, A_{1:t}) \\ = & \int \sum_c U_0(x^c, y_{1:t+1}, A_{1:t+1}) p(x^c|y_{1:t+1}, A_{1:t+1}) \\ & \times p(y_{t+1}|y_{1:t}, A_{1:t}, A_{t+1}) dy_{t+1}. \end{aligned} \tag{4}$$

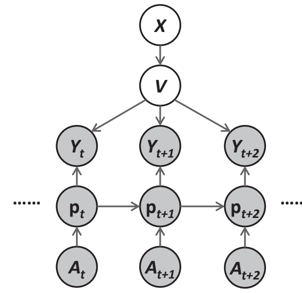

Fig. 1. The graphical model for opportunistic sensing. Shaded nodes are observable.

where $U_0(x, y_{1:t+1}, A_{1:t+1})$ is the utility function for time step $t+1$ if the true class label is x. Only the reward of next one step is considered in the utility function above, which corresponds to greedy decision making. Rewards of future time steps can also be included recursively as in [7].

The utility function U_0, which determines what kind of solution (putting aside whether it is solvable at all) we can get from Eq. (4), should be carefully designed for the purpose of specific application. For active recognition problems, U_0 should reflect the confidence in class label x as well as the cost of taking actions $\{A_t\}$. Here we assume that any action can be taken with equal cost and executed exactly without error. Although not realistic, this assumption leads to a simplified deterministic execution model and focuses our attentions on the performance of recognition.A more sophisticated treatment is possible if the motion dynamics of sensing platform and the conditions for image formation are considered as extra sensing constraints.

3. MAXIMUM NEGATIVE ENTROPY

It has been suggested in many works [8, 9] to use the expected Kullback-Leibler divergence between class posterior and prior as a utility function for classification, which is equivalent to selecting U_0 in Eq. (4) as the log-ratio of posterior and prior. Formally, the joint sensor selection and feature extraction problem can be stated as:

$$\begin{aligned} \mathbf{p}_{t+1}^* = & \arg\max_{\mathbf{p}_{t+1}} \int \sum_c \log \frac{p(x^c|y_{1:t+1}, \mathbf{p}_{1:t+1})}{p(x^c|y_{1:t}, \mathbf{p}_{1:t})} \times \\ & p(x^c|y_{1:t+1}, \mathbf{p}_{1:t+1}) p(y_{t+1}|y_{1:t}, \mathbf{p}_{1:t}, \mathbf{p}_{t+1}) dy_{t+1}, \end{aligned} \tag{5}$$

where actions A_t's are replaced by measurement vectors $\mathbf{p}_t$'s due to the deterministic execution model assumption. This utility function measures the dynamic information gain of our belief in label x after obtaining a new observation y_{t+1} via measurement vector $\mathbf{p}_{t+1}$. The dynamic prior $p(x|y_{1:t}, \mathbf{p}_{1:t})$ is independent of $\mathbf{p}_{t+1}$ and can be integrated out, leading to the negative entropy definition of utility function:

$$\mathbf{p}_{t+1}^{*} = \arg\max_{\mathbf{p}_{t+1}} \mathrm{E}\left[-H(x|y_{1:t+1}, \mathbf{p}_{1:t+1})|y_{1:t}, \mathbf{p}_{1:t+1}\right], \quad (6)$$

where $H(x|y_{1:t+1}, \mathbf{p}_{1:t+1})$ denotes the entropy of distribution $p(x|y_{1:t+1}, \mathbf{p}_{1:t+1})$, and the expectation is taken with respect to $p(y_{t+1}|y_{1:t}, \mathbf{p}_{1:t+1})$. The class posterior can be evaluated recursively as

$$p(x^c|y_{1:t+1}, \mathbf{p}_{1:t+1}) = w_c \mathcal{M}_c(y_{t+1}) / \mathcal{M}(y_{t+1}), \quad (7)$$

where

$$w_c = p(x^c|y_{1:t}, \mathbf{p}_{1:t}), \quad (8)$$

$$\mathcal{M}_c(y_{t+1}) = p(y_{t+1}|x^c, y_{1:t}, \mathbf{p}_{1:t+1}) = \mathcal{N}(y_{t+1}; \mu_{\mathcal{M}c}, \sigma_{\mathcal{M}c}^2), \quad (9)$$

$$\mathcal{M}(y_{t+1}) = p(y_{t+1}|y_{1:t}, \mathbf{p}_{1:t+1}) = \sum_c w_c \mathcal{M}_c(y_{t+1}). \quad (10)$$

$\{w_c\}$ in Eq (8) is the posterior distribution of x at time t. In Eq. (9), $\mu_{\mathcal{M}c} = \mathbf{p}_{t+1}^T[\boldsymbol{\mu}_c + \mathbf{K}_{c,t}(\mathbf{y}_t - \mathbf{P}_t^T \boldsymbol{\mu}_c)]$, $\sigma_{\mathcal{M}c}^2 = \mathbf{p}_{t+1}^T(\mathbf{I} - \mathbf{K}_{c,t}\mathbf{P}_t^T)\boldsymbol{\Sigma}_c \mathbf{p}_{t+1} + \sigma_s^2$, $\mathbf{K}_{c,t} = \boldsymbol{\Sigma}_c \mathbf{P}_t(\sigma_s^2 \cdot \mathbf{I} + \mathbf{P}_t^T \boldsymbol{\Sigma}_c \mathbf{P}_t)^{-1}$, and we define $\mathbf{y}_t = [y_1, ..., y_t]^T$, $\mathbf{P}_t = [\mathbf{p}_1, ..., \mathbf{p}_t]$. Plugging Eq. (7) into Eq. (6), we have

$$\begin{aligned} &\mathrm{E}\left[-H(x|y_{1:t+1}, \mathbf{p}_{1:t+1})|y_{1:t}, \mathbf{p}_{1:t+1}\right] \\ =&\sum_c w_c \int \mathcal{M}_c(y_{t+1}) \log \mathcal{M}_c(y_{t+1}) dy_{t+1} \\ &+ \sum_c w_c \log w_c - \int \mathcal{M}(y_{t+1}) \log \mathcal{M}(y_{t+1}) dy_{t+1}, \end{aligned} \quad (11)$$

Detailed derivations for the above can be found in the supplemental material.

4. OPTIMIZING MEASUREMENT VECTOR

The first term in Eq. (11) is a linear combination of negative Gaussian entropy, and the second term is independent of measurement vector. With these observations, we can rewrite the optimization problem in Eq. (6) as

$$\begin{aligned} &\max_{\mathbf{p}_{t+1}} U(\mathbf{p}_{t+1}) \\ =& -\frac{1}{2} \sum_c w_c \ln\left[\mathbf{p}_{t+1}^T(\mathbf{I} - \mathbf{K}_{c,t}\mathbf{P}_t^T)\boldsymbol{\Sigma}_c \mathbf{p}_{t+1} + \sigma_s^2\right] \\ &- \sum_c w_c \int \mathcal{M}_c(y_{t+1}) \log[\sum_{c'} w_{c'} \mathcal{M}_{c'}(y_{t+1})] dy_{t+1}, \\ s.t.\ & \mathbf{p}_{t+1} = [\delta(z_{t+1} - 1) \cdot \mathbf{I}; ...; \delta(z_{t+1} - M) \cdot \mathbf{I}]\tilde{\mathbf{p}}_{t+1}, \\ & ||\tilde{\mathbf{p}}_{t+1}|| = 1, \quad z_{t+1} \in \{1, ..., M\}, \end{aligned} \quad (12)$$

where we use the fact that the entropy of a Gaussian distribution with covariance $\boldsymbol{\Sigma}$ is $\frac{1}{2}\ln|\boldsymbol{\Sigma}|$ plus a constant. The unit norm constraint of measurement vector is placed to keep finite signal noise ratio.

The measurement vector $\mathbf{p}_{t+1}$ is defined by continuous projection $\tilde{\mathbf{p}}_{t+1}$ and discrete sensor index z_{t+1}. To solve Eq. (12), we enumerate all the possible z_{t+1}'s and use projected gradient ascent method to find the optimal $\tilde{\mathbf{p}}_{t+1}$. The gradient of the objective function $U(\mathbf{p}_{t+1})$ can be evaluated (detailed in the supplemental material) as

$$\begin{aligned} &\mathbf{g}(\mathbf{p}_{t+1}) \\ =& -\sum_c \frac{w_c(\mathbf{I} - \mathbf{K}_{c,t}\mathbf{P}_t^T)\boldsymbol{\Sigma}_c \mathbf{p}_{t+1}}{\mathbf{p}_{t+1}^T(\mathbf{I} - \mathbf{K}_{c,t}\mathbf{P}_t^T)\boldsymbol{\Sigma}_c \mathbf{p}_{t+1} + \sigma_s^2} \\ &- \sum_c w_c \int \frac{\partial \mathcal{M}_c(y_{t+1})}{\partial \mathbf{p}_{t+1}} \log\left(\sum_{c'} w_{c'} \mathcal{M}_{c'}(y_{t+1})\right) dy_{t+1} \end{aligned} \quad (13)$$

with

$$\begin{aligned} \frac{\partial \mathcal{M}_c(y_{t+1})}{\partial \mathbf{p}_{t+1}} =& \frac{\mathcal{M}_c(y_{t+1})}{\sigma_{\mathcal{M}c}} \left[(y_{t+1} - \mu_{\mathcal{M}c})\left(\frac{\partial \mu_{\mathcal{M}c}}{\partial \mathbf{p}_{t+1}} \frac{1}{\sigma_{\mathcal{M}c}}\right.\right. \\ &\left.\left. + \frac{\partial \sigma_{\mathcal{M}c}}{\partial \mathbf{p}_{t+1}} \frac{y_{t+1} - \mu_{\mathcal{M}c}}{\sigma_{\mathcal{M}c}^2}\right) - \frac{\partial \sigma_{\mathcal{M}c}}{\partial \mathbf{p}_{t+1}}\right], \end{aligned} \quad (14)$$

$$\partial \mu_{\mathcal{M}c} / \partial \mathbf{p}_{t+1} = \boldsymbol{\mu}_c + \mathbf{K}_{c,t}(\mathbf{y}_t - \mathbf{P}_t^T \boldsymbol{\mu}_c) \quad (15)$$

$$\partial \sigma_{\mathcal{M}c} / \partial \mathbf{p}_{t+1} = \sigma_{\mathcal{M}c}^{-1}(\mathbf{I} - \mathbf{K}_{c,t}\mathbf{P}_t^T)\boldsymbol{\Sigma}_c \mathbf{p}_{t+1} \quad (16)$$

It is still impossible to analytically evaluate the integration in the second term of Eq. (13). However, by plugging Eq. (14) into Eq. (13), we realize that this integration is the expectation of a random variable y_{t+1} with normal distribution $\mathcal{M}_c$ under a nonlinear mapping. Such expectation can be numerically approximated by unscented transform as in [10]. In one-dimensional case, unscented transform approximates $\mathcal{M}_c$ by deterministically sampling 3 sigma points:

$$y_{c,k} = \mu_{\mathcal{M}c} + k\sqrt{1+\kappa} \cdot \sigma_{\mathcal{M}c}, \quad k \in \{-1, 0, 1\}, \quad (17)$$

with weights $u_0 = \kappa/(1+\kappa), u_{\pm 1} = 1/2(1+\kappa)$, where κ is a positive number. The expectation of any finite function $f(y_{t+1})$ can be approximated by

$$\int \mathcal{M}_c(y_{t+1}) f(y_{t+1}) dy_{t+1} \approx \sum_{k=-1}^{1} u_k f(y_{c,k}). \quad (18)$$

Therefore, Eq. (13) can be approximated (detailed in the supplemental material) as

$$\begin{aligned} &\mathbf{g}(\mathbf{p}_{t+1}) \\ \approx& -\sum_c \frac{w_c(\mathbf{I} - \mathbf{K}_{c,t}\mathbf{P}_t^T)\boldsymbol{\Sigma}_c \mathbf{p}_{t+1}}{\mathbf{p}_{t+1}^T(\mathbf{I} - \mathbf{K}_{c,t}\mathbf{P}_t^T)\boldsymbol{\Sigma}_c \mathbf{p}_{t+1} + \sigma_s^2} \\ &- \sum_c \frac{w_c}{\sigma_{\mathcal{M}c}} \sum_{k=-1}^{1} u_k \left[k\sqrt{1+\kappa} \frac{\partial \mu_{\mathcal{M}c}}{\partial \mathbf{p}_{t+1}} + \right. \\ &\left.(k^2(1+\kappa) - 1)\frac{\partial \sigma_{\mathcal{M}c}}{\partial \mathbf{p}_{t+1}}\right] \log\left(\sum_{c'} w_{c'} \mathcal{M}_{c'}(y_{c,k})\right). \end{aligned} \quad (19)$$

The value of $U(\mathbf{p}_{t+1})$ can be approximated in a similar way.

The proposed method for joint sensor selection and feature extraction is summarized in Algorithm 1. Observations are iteratively obtained until the maximum allowable number T is reached. The gradient ascend method is initialized using PCA with the covariance matrix $\mathbf{\Sigma_v}$ for $p(\mathbf{v})$, and updates with step size α. The updating is terminated when the change in $\mathbf{p}_{t+1}$ is less than ε, which usually occurs within 20 iterations.

5. INTERPRETATIONS OF THE MAXIMUM NEGATIVE ENTROPY PRINCIPLE

Maximizing the negative classification entropy is an intuitive way to improve recognition confidence. Nevertheless, it can be interpreted from different perspectives.

First, the expected entropy in Eq. (11) can be reorganized as a measure of Shannon information:

$$\begin{aligned} & \mathrm{E}\left[-H(x|y_{1:t+1}, \mathbf{p}_{1:t+1})|y_{1:t}, \mathbf{p}_{1:t+1}\right] \\ = \; & \mathrm{E}_{p(x|y_{1:t}, \mathbf{p}_{1:t})}\left[-H(y_{t+1}|x, y_{1:t}, \mathbf{p}_{1:t+1})\right] \\ & -H(x|y_{1:t}, \mathbf{p}_{1:t}) + H(y_{t+1}|y_{1:t}, \mathbf{p}_{1:t+1}) \\ = \; & -H((y_{t+1}|x)|y_{1:t}, \mathbf{p}_{1:t+1}) + H(y_{t+1}|y_{1:t}, \mathbf{p}_{1:t+1}) \\ & -H(x|y_{1:t}, \mathbf{p}_{1:t}) \\ = \; & I(y_{t+1}; x|y_{1:t}, \mathbf{p}_{1:t+1}) - H(x|y_{1:t}, \mathbf{p}_{1:t}), \end{aligned} \tag{20}$$

where $I(\cdot;\cdot)$ denotes the mutual information between two random variables. Note that the second term in Eq. (20) is a constant independent of $\mathbf{p}_{t+1}$. Therefore, maximizing negative entropy is equivalent to maximizing the mutual information between class label x and new measurement y_{t+1}, which is also observed in [6] for sensor selection problem.

The maximum negative entropy principle can also be understood as a feature extraction criterion in optimizing the measurement vector $\mathbf{p}_{t+1}$. After some manipulations on the first term of the objective function in Eq. (12), we notice that a large value for the following expression is favoured:

$$-\mathbf{p}_{t+1}^T \mathbf{\Sigma}_c \mathbf{p}_{t+1} + \mathbf{p}_{t+1}^T \mathbf{C}_{\mathbf{v},\mathbf{y}_t} \mathbf{C}_{\mathbf{y}_t,\mathbf{y}_t}^{-1} \mathbf{C}_{\mathbf{y}_t,\mathbf{v}} \mathbf{p}_{t+1} \tag{21}$$

where $\mathbf{C}_{\mathbf{v},\mathbf{y}_t} = \mathbf{C}_{\mathbf{y}_t,\mathbf{v}}^T = \mathbf{\Sigma}_c \mathbf{P}_t$ is the cross-covariance matrix for $\mathbf{v}$ and $\mathbf{y}_t$, and $\mathbf{C}_{\mathbf{y}_t,\mathbf{y}_t} = \sigma_s^2 \cdot \mathbf{I} + \mathbf{P}_t^T \mathbf{\Sigma}_c \mathbf{P}_t$ is the covariance matrix for $\mathbf{y}_t$. Large value for the first term in Eq. (21) for any class c is similar to small within-class variances required in LDA. The second term in Eq. (21) differs from CCA only in that the variance of $\mathbf{v}$ is not normalized. As such, both the variance of new observation and its correlation with previous observations are promoted, which is a combination of PCA and CCA. In addition, the second term in Eq. (12) is the entropy of a mixture of Gaussian. Maximizing the entropy effectively separates the class centers $\mathbf{p}_{t+1}^T[\boldsymbol{\mu}_c + \mathbf{K}_{c,t}(\mathbf{y}_t - \mathbf{P}_t^T \boldsymbol{\mu}_c)]$, leading to a large between-class variance as required in LDA. Therefore, the strengths of PCA, LDA and CCA are unified in an adaptive way through maximum negative entropy principle.

Algorithm 1 Dynamic sensor selection and feature extraction for opportunistic sensing.

initialize $w_c = p(x^c)$, $\mathbf{K}_{c,0} = 0$, $\mathbf{P}_0 = []$, $\mathbf{y}_0 = []$, $t = 0$
for $t = 1...T$ **do**
  set $U^* = -\infty$
  for $m = 1...M$ **do**
    set $i = 0$
    initialize $\mathbf{p}_{t+1}^{(0)} = \arg\max_{\mathbf{p}} \mathbf{p}^T \mathbf{\Sigma_v} \mathbf{p} \quad s.t. \quad \mathbf{p} = [\delta(m-1) \cdot \mathbf{I}; ...; \delta(m-M) \cdot \mathbf{I}]\tilde{\mathbf{p}}, \mathbf{p} \perp \mathbf{P}_t, ||\mathbf{p}|| = 1$
    while true **do**
      find $\mu_{\mathcal{M}c}, \sigma_{\mathcal{M}c}$ using $\mathbf{p}_{t+1}^{(i)}$ for all c
      set $\mathbf{p}_{t+1}^{(i+1)} = \mathbf{p}_{t+1}^{(i)} + \alpha \cdot \mathbf{g}(\mathbf{p}_{t+1}^{(i)})$ by Eq. (19)
      project $\mathbf{p}_{t+1}^{(i+1)}$ to the feasible region in Eq. (12)
      if $||\mathbf{p}_{t+1}^{(i+1)} - \mathbf{p}_{t+1}^{(i)}|| < \varepsilon$ **then**
        break
      end if
      set $i = i + 1$
    end while
    if $U(\mathbf{p}_{t+1}^{(i)}) > U^*$ **then**
      set $\mathbf{p}_{t+1}^* = \mathbf{p}_{t+1}^{(i)}$, $z_{t+1}^* = m$
      set $U^* = U(\mathbf{p}_{t+1}^{(i)})$
    end if
  end for
  get observation y_{t+1} with measurement vector $\mathbf{p}_{t+1}^*$
  update posterior $\{w_c\}$ according to Eq. (7)
  set $\mathbf{P}_{t+1} = [\mathbf{P}_t, \mathbf{p}_{t+1}^*]$, $\mathbf{y}_{t+1} = [\mathbf{y}_t; y_{t+1}]$
  set $\mathbf{K}_{c,t+1} = \mathbf{\Sigma}_c \mathbf{P}_{t+1}(\sigma_s^2 \cdot \mathbf{I} + \mathbf{P}_{t+1}^T \mathbf{\Sigma}_c \mathbf{P}_{t+1})^{-1}, \forall c$
end for
return $\hat{c} = \arg\max_c w_c$

6. EXPERIMENTS

Recognition performances of opportunistic sensing are reported on two datasets in this section. We use parameters $\alpha = 1$, $\varepsilon = 0.01$, $\kappa = 1$ in the following tests, and set observation noise variance σ_s^2 equivalent to SNR of 20dB.

6.1. Vehicle Classification

Our method is first applied to the 10-class military vehicle classification problem on the MSTAR [11] dataset, which contains noisy airborne X-band SAR images as shown in Fig. 2. 4785 images with depression angles 17° and 30° are used for training, and 4351 images with depression angles 15° and 45° are used for testing. The vehicle targets are captured in various poses and we quantize them into 12 discrete poses as captured by 12 sensors deployed in different directions. 12 images from the same class with different poses are randomly selected and combined as a test instance, and we repeat the testing of each class for 120 trials with the maximum number of observations up to $T = 50$. Observations are simulated sequentially as the linear projections of raw image pixels.

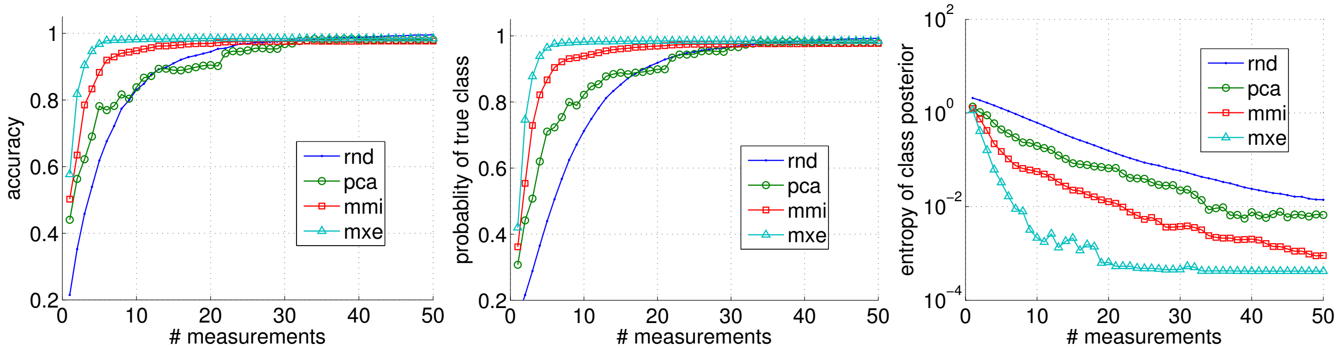

Fig. 3. Performance of opportunistic sensing algorithms on the MSTAR test set under constraints of different measurement numbers. From left to right: classification accuracy; correct class posterior probability; class posterior entropy.

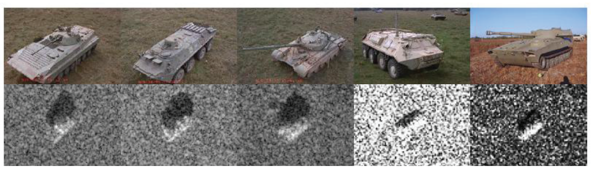

Fig. 2. Sample images for 5 out of 10 classes in the MSTAR dataset. First row shows the illustrative real life images and second row shows the SAR images from the dataset.

Table 1. Average number of measurements required to reduce class posterior entropy below 0.001 on the MSTAR test set.

method	#measurements (std.)
rnd	28.85 (±2.09)
pca	18.04 (±1.51)
mmi	9.58 (±1.24)
mxe	5.14 (±0.65)

We compare the proposed maximum negative entropy method ("mxe") with the baseline approaches which select sensor and measurement vector randomly ("rnd"), by the order of PCA components ("pca"), and by maximum mutual information ("mmi") [6]. As shown in Fig. 3, our "mxe" method achieves the highest classification accuracy, the highest probability of correct class and the lowest class posterior entropy under different constraints of total available measurement number up to $T = 50$. Remarkable performance gain is attained especially when a small number of measurements is allowed ($T < 10$).

Instead of using up all the available measurements, we can also terminate the sensing procedure when the confidence of recognition reaches a certain level, which can be characterized by a threshold on the entropy of class posterior. Table 1 shows that our method can achieve the same level of recognition performance with significantly fewer data measurements than other methods.

The selected sensors and projections for one test instance are plotted in Fig. 4 for detailed inspection. This test instance corresponds to the class of the first column in Fig. 2. The features detected by the projections using "mxe" method are more meaningful than using "rnd", and tend to focus on discriminative local part (*e.g.*, the second projection basis selected by "mxe") as opposed to the global reconstructive features detected by "pca" and "mmi".

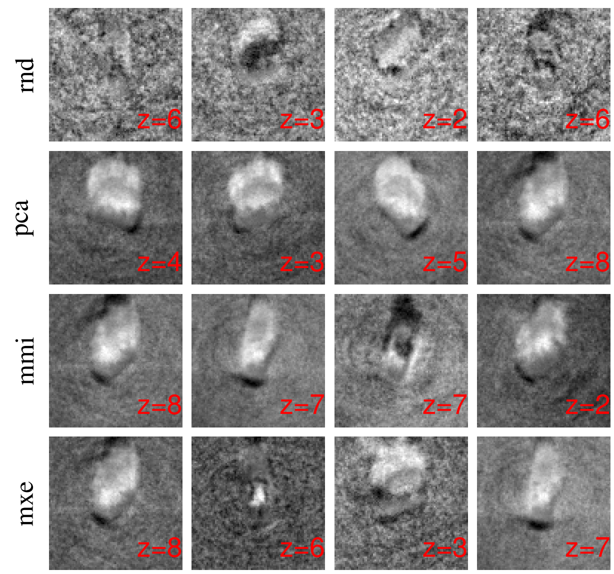

Fig. 4. Examples of selected sensor (in red text) and projection basis (in gray image) using different methods (from top to bottom) on the MSTAR data. The first 4 time steps (from left to right) in opportunistic sensing are shown.

6.2. Face Recognition

In the second experiment, we test the opportunistic sensing algorithm for face recognition on the CMU Multi-Pie dataset [12]. We use 15480 images of 129 subjects collected in 3 sessions as training data, and use 5160 images of another session for testing. Face images with poses $-45°$ and $75°$ are used as data from two different views, and a total number of observations up to $T = 40$ are allowed. The large number of

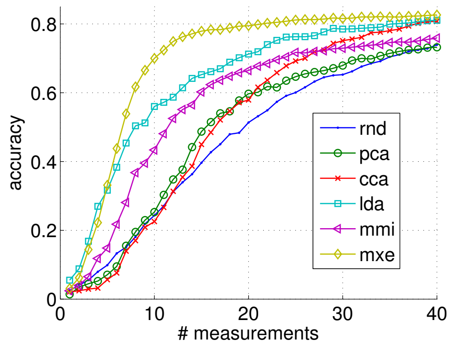

Fig. 5. Classification accuracy for the Multi-Pie test set under constraints of different measurement numbers.

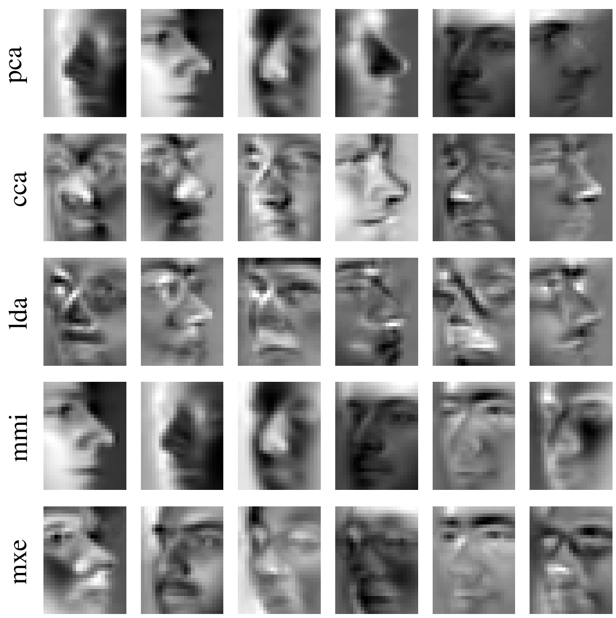

Fig. 6. Examples of selected view and projection using different methods (from top to down), for the first 6 time steps (from left to right) on the Multi-Pie data.

classes and high within-class illumination variation make the problem more challenging than the previous experiment.

The proposed method ("mxe") is compared with the random ("rnd"), PCA ("pca"), CCA ("cca"), LDA ("lda"), and maximum mutual information ("mmi") methods. As shown in Fig. 5, both "mxe" and "lda" achieve much higher accuracy than other methods under the same number of measurements; while "mxe" further outperforms "lda" by a large margin for allowed measurement number $5 < T < 25$. Some selected sensor projection basis are visualized in Fig. 6.

7. CONCLUSION

A new paradigm of data acquisition named opportunistic sensing is proposed for recognizing object class with a limited number of dynamically observed data samples. Both sensor selection and linear projection are under our optimization for negative class posterior entropy maximization. Our approach is shown theoretically connected to many common sensor selection and feature extraction methods, and experimentally superior to all these competitive algorithms.

We will consider more realistic sensor configurations and navigation costs in our future work. Incorporating non-linear sensing/classification models and time varying states are also possible extensions for practical applications.

8. REFERENCES

[1] Feng Zhao and Leonidas Guibas, *Wireless Sensor Networks: An Information Processing Approach*, Elsevier/Morgan-Kaufmann, 2004.

[2] Volkan Isler and Ruzena Bajcsy, "The sensor selection problem for bounded uncertainty sensing models," in *Proceedings of the 4th international symposium on Information processing in sensor networks*, 2005.

[3] Pradeep K. Atrey, Mohan S. Kankanhalli, and John B. Oommen, "Goal-oriented optimal subset selection of correlated multimedia streams," *ACM Transactions on Multimedia Computing, Communications, and Applications*, vol. 3, no. 1, 2007.

[4] Jarvis Haupt, Richard Baraniuk, Rui Castro, and Robert Nowak, "Sequentially designed compressed sensing," in *IEEE Statistical Signal Processing Workshop*, 2012.

[5] Dmitry M. Malioutov, Sujay R. Sanghavi, and Alan S. Willsky, "Sequential compressed sensing," *IEEE Journal of Selected Topics in Signal Process*, vol. 4, no. 2, pp. 435–444, 2010.

[6] Emre Ertin, John W. Fisher, and Lee C. Potter, "Maximum mutual information principle for dynamic sensor query problems," in *2nd International Workshop on Information Processing in Sensor Networks*. 2003, pp. 405–416, Springer-Verlag.

[7] Leslie Pack Kaelbling, Michael L. Littman, and Anthony R. Cassandra, "Planning and acting in partially observable stochastic domains," *Artificial Intelligence*, vol. 101, pp. 99–134, 1998.

[8] D. V. Lindley, "On a measure of the information provided by an experiment," *The Annals of Mathematical Statistics*, vol. 27, pp. 986–1005, 1956.

[9] M. Stone, "Application of a measure of information to the design and comparison of regression experiments," *The Annals of Mathematical Statistics*, vol. 30, pp. 55–70, 1959.

[10] Jacob Goldberger and Hagai Aronowitz, "A distance measure between GMMs based on the unscented transform and its application to speaker recognition," in *Proceedings of Interspeech*, 2005, pp. 1985–1989.

[11] Timothy D. Ross, Steven W. Worrell, Vincent J. Velten, John C. Mossing, and Michael L. Bryant, "Standard SAR ATR experiments using the MSTAR public release data set," *Proc. SPIE Algorithms for Synthetic Aperture Radar Imagery V*, vol. 3370, pp. 566–573, 1998.

[12] Ralph Gross, Iain Matthews, Jeffrey Cohn, Takeo Kanade, and Simon Baker, "Multi-pie," *Image and Vision Computing*, vol. 28, pp. 807–813, 2010.