

Poster Abstract: A Wireless Pedestrian Tracking Network

Lun Jiang, Ankur Kamthe and Alberto E. Cerpa
Electrical Engineering and Computer Science
University of California - Merced
{ljiang2,akamthe,acerpa}@ucmerced.edu

Abstract

The ease of deploying wireless camera sensor nodes has grown with the reduction of manufacturing costs of low power, high resolution cameras. Although current wireless sensor network platforms have limited on-board resources for solving highly complex computer vision problems, we show that by splitting the processing costs between the sensor node and a powerful backend, we can achieve better classification results. Using such a distributed processing approach, we balance the computational and communication costs for achieving better detection performance while improving the system lifetime.

Categories and Subject Descriptors

C.2.1 [Computer-Communication Networks]: Network Architecture and Design—*Wireless communication*;
C.3 [Special Purpose and Application-Based Systems]: Real-time and Embedded Systems

General Terms

Algorithm, Design, Performance

Keywords

Wireless Camera Sensor Networks, Distributed Image Processing

1 Motivation

Wireless sensor network-based detection and tracking systems use either computationally lightweight algorithms when performing in-node processing of data or data aggregation techniques when transmitting raw data for centralized processing. Aslam et al. [1] describe a centralized framework for tracking a moving object by using a particle-filter for a binary sensor network. VigilNet [2] is a detection and classification system for tracking metallic objects using information from multiple magnetometer and acoustic sensors. In such approaches, the use of simple sensors for tracking necessitates dense deployment to localize an object or to infer its direction information.

In contrast, using camera sensor networks, we can infer position and direction information from a single sensor by identifying an object and tracking its position in successive images. A camera sensor network for detecting and tracking pedestrians is subjected to constraints such as extremely limited bandwidth for transmission and the relatively scarce processing power of camera sensors. Typical camera sensor network design boils down to two common paradigms. In the first design, the system requires the camera sensors to transmit every single bit of the captured images and relies on the central server to process the information. The drawback of this approach is obviously the overloading of the wireless channel. The second approach is to process all the image information on the sensor node and transmit only the answer to central server. The shortcoming of this method is straining the computational capacity of the sensor. While performing all the operation locally, the sensor will spend a majority of its time processing data, resulting in failure to sense the movement of people.

The common thread in both the approaches is that the node is required to transmit information to the central server. Another question that emerges is how accurate or redundant this meta-data should be. For certain purposes, e.g. detection of moving objects, this data can be very simple such as the number of objects. In this case, the server will be losing all the image details such as the shape and color of the moving objects since the transmitted information is so limited. For some other applications, such as classification of moving objects as human or non-human, tracking movement of person/s among several camera sensors, etc. the information needed for making a decision imposes a heavy computation load on the system. Teixeira et al. [3] used histogram features such as peak height, peak width of each object to track them. This kind of information is vague even for high precision tracking purposes, not to mention the task of classification and tracking. Kamthe et al. [4] used thresholding and foreground blob sizes for differentiating the moving objects as pedestrians instead of other objects. There are still a lot of cases in which the moving objects can have similar size as human beings.

In this work, we extended the work of SCOPES, by emphasizing balance between utilizing the computation and communication resources of the sensor network. In order to do so, we will explore what kind of meta-data, or features are

needed to detect a human being with high precision and incorporate that in a sensor network tracking framework. Our goal is to improve the detection performance while reducing the amount of information relayed to the central server.

2 System Design

2.1 General Description

We use the Imote2 platform combined with IMB400 sensor board which includes an OmniVision 7600 camera chip. The Imote2 captures images, preprocesses them and send the compressed information about the image back to server. The central server further on recovers this information and processes it with trained classifiers.

2.2 Processing Algorithms

2.2.1 Principal Component Analysis

Principal Component Analysis (PCA) is widely used in face recognition, pattern matching and object tracking. The idea behind PCA is to use a few orthogonal basis to represent a high dimension data point. PCA will sort its component vectors (basis) according to the data variation.

2.2.2 Support Vector Machine

Support Vector Machine (SVM) have been proved to be work well in both linear and non-linear classification applications. The intuition is to use a hyperplane to partition the data space according to different labels. SVM requires training process with labeled data. Oftentimes the dataset is projected non-linearly into a higher dimension space in order to separate the non-linear dataset. Intuitively SVM is trying to maximize the margin between different datasets.

2.3 Feature Extraction

In our approach, we perform PCA on the images captured by the camera and transmit only the principal component information to the server. This requires generating an optimized set of principal component basis based on the data set. In our case, we used 5000 images to train our basis. PCA basis is generated based on the covariance matrix of the data set, which represents the changes of the gray scale values per pixel. For our purpose, i.e. people recognition, images with people in them generally have similar features from the top view, a central blob of head, symmetric shoulders around them. These features can be viewed as the most obvious changes standing out from the background. PCA helps to identify these features and using it enables us to successfully recover the randomly projected samples from the Compressive Sensing results. The captured images are projected onto this set of basis vectors. This approach helps to compress the image information before the wireless transmission. Figure 1 shows an image captured by the camera and the corresponding recovered image when using 120 principal components.

2.4 Image Classification

In Figure 2, we show the classification results when all the 20 principal components are used to train the SVM. The SVM draws a contour fitting in very tightly with the data

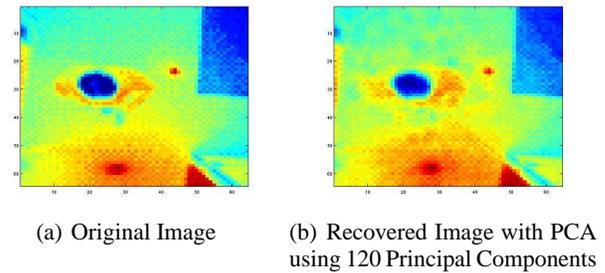


Figure 1. Recovered Images Using Different Number of Principal Components

set. We generated this contour by feeding a grid data trained model from SVM. Our initial experiments show a 93.7% success rate of differentiating a pedestrian from non-human objects. This performance is achieved when using a 66:1 compression ratio on the original image.

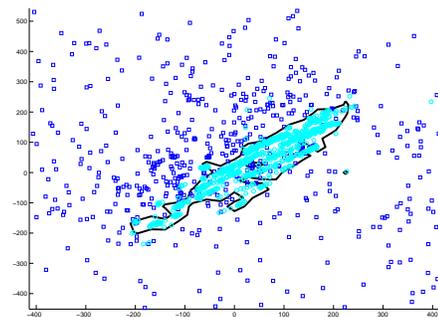


Figure 2. Contour generated by SVM to classify between the images without people (cyan circle) and images with people (blue square). The x axis is the first order principal component, y axis is the second order principal component.

3 References

- [1] J. Aslam, Z. Butler, F. Constantin, V. Crespi, G. Cybenko, and D. Rus. Tracking a moving object with a binary sensor network. In *SenSys '03*, 2003.
- [2] L. Gu, D. Jia, P. Vicaire, T. Yan, L. Luo, A. Tirumala, Q. Cao, T. He, J. A. Stankovic, T. Abdelzaher, and B. H. Krogh. Lightweight detection and classification for wireless sensor networks in realistic environments. In *SenSys '05*, 2005.
- [3] T. Teixeira and A. Savvides. Lightweight people counting and localizing in indoor spaces using camera sensor nodes. *ICDSC 2007*, September 2007.
- [4] A. Kamthe, L. Jiang, M. Dudys, and A. Cerpa. SCOPES: Smart Cameras Object Position Estimation System. In *EWSN '09*, 2009.