Individual Lab Report #6

Sumit Saxena
Team F: Rescue Rangers

Teammates:

Juncheng Zhang

Karthik Ramachandran

Xiaoyang Liu

February 2, 2017

Contents

1. Individual Progress

1.1. Overview

During the past two weeks, I worked majorly on the following tasks:

1. Test plan for data collection using NEA payload
2. Learn to use TensorFlow to implement neural networks for image classification

In addition, during the vacation, I also did some literature study on:

3. Possible ways to fuse data/information from different sensing modalities

1.2. Plan for data collection using NEA payload

Unexpectedly, this task turned out to be pretty time-taking given the amount of detailing needed. The plan was to be directly shared with the Flight crew at NEA, and we were supposed to provide details about the flight scenario and the setting up of locations of interest in a very detailed manner so that the plan could be executed in our absence as well.

Figure 1 shows the three flight areas we have chosen. We had to plan for the following:

- Areas to cover and pattern:
    - We decided to cover three types of vegetation – mowed grass airstrip, bushy patch, and tree-covered
    - Lawn-mower pattern
- GPS locations for 'Locations of interest', and human signatures to be set up at each one of them
- Flight scenario: ground speed, altitudes, route spacing, no. of passes
    - Based on the thermal and RGB cameras' FOV and the kind of coverage we wanted: we planned the passes in such a way that we cover each object of interest with multiple angles of view.

Figure 1: Google Earth snapshot of the three flight areas we intend to cover with the locations of Locations of Interest

1.3. Learn to use TensorFlow to implement neural networks for image classification

Since human detection from aerial images is a complex problem, we plan to try out as many algorithms as possible to understand which algorithms work well for which cases. Then, we will come up with a fusion strategy to make optimal use of these of algorithms.

We decided to try neural networks as one of our algorithms. Given my familiarity with Python, I decided to go with it and explored a bit on which library to use for implementing a neural network. Since frameworks like TensorFlow and Torch are pretty popular these days for Deep Learning, I thought it would be good to get some hands-on experience with them. My plan is to use TensorFlow, start with a basic feed-forward neural network and explore the need of any complex neural networks.

I am taking a course on Udacity to quickly learn to work with TensorFlow. I have already implemented a multinomial logistic regression model on MNIST dataset and expect to soon be able to implement neural networks and then translate my learnings to our own datasets.

1.4. Possible ways to fuse data/information from different sensing modalities

Since we plan to use multiple algorithms to detect human signatures in data from multiple sensing modalities to effectively detect human signatures, we need to come up with techniques to combine their results in an effective manner. During the vacation, I did some literature study to understand how this is done in the field. I found some interesting but seemingly complex ways of how we could accomplish various kinds of fusion:

1. Fusing multiple vision algorithms:

[1] describes a head-tracking system that harnesses Bayesian modality fusion, a technique for integrating the analyses of multiple visual tracking algorithms within a probabilistic framework. Figure 2 shows a dynamic Bayesian model, presented in the paper, which can be used for this kind of fusion. The model defines some 'Reliability Indicators' for each of the modality. 'Ground Truth' represents the unknown state of the target. The ground-truth state causes the output from a visual modality. At run-time, the model is instantiated with a set of observations including the modality report and the status of reliability indicators. The reliability of the modality report is computed and the inferred reliability and report are considered in inferring a probability distribution over the ground-truth state of the target object.

Although a very interesting approach, I am not sure if it is practical for us to implement as establishing the ground truth and populating the conditional probability tables seemed challenging from the paper.

[1] "Bayesian Modality Fusion: Probabilistic Integration of Multiple Vision Algorithms for Head Tracking" by Kentaro Toyama, Eric Horvitz at Microsoft Research

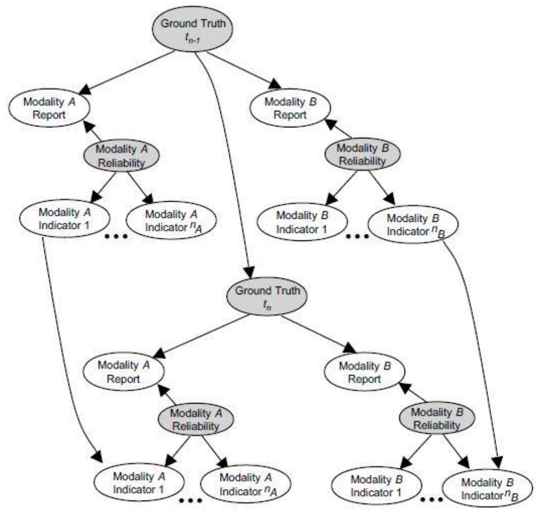

Figure 2: A dynamic Bayesian network model for integrating multiple visual processing modalities over time (from [1])

2. Fusing at a pixel-level:

[2] presents a method for optimally combining pixel information from an infra-red thermal imaging camera, and a conventional visible spectrum color camera, for tracking a moving target. The tracking algorithm rapidly re-learns its background models for each camera modality from scratch at every frame.

[3] presents a probabilistic model of the scene background where each pixel is represented as a multi-modal distribution with the changing number of modalities for both color and thermal input. Based on the background model, it introduces a pedestrian tracker designed as a particle filter. It also presents a number of informed reversible transformations to sample the model probability space in order to maximize our model posterior probability.

I found [2] to be much simpler than [3] and much more practical for our case. However, given that we have limited time, and multiple people to test and try different algorithms for different modalities, it does not make much sense for us to go with fusion at the pixel level.

It is highly likely that rather than implementing any of the algorithms described above, we may go with a simple scoring approach for different detection algorithms but it was interesting to see what kind of approaches people have developed and whether we could draw some

[2] "Bayesian fusion of thermal and visible spectra camera data for region based tracking with rapid background adaptation" by Rustam Stolkin, David Rees, Mohammed Talha, and Ionut Florescu; [3] "Pedestrian Tracking by Fusion of Thermal-Visible Surveillance Videos" by Alex Leykin, Riad Hammoud

inspiration to make our approach effective. I will present a more elaborate discussion in case we end up using any of the approaches mentioned above.

2. Challenges:

I faced following challenges during my work:

1. Using images of varied sizes for training/testing with TensorFlow:

   While I built some basic ML models using TensorFlow for MNIST dataset, I did not realize that MNIST had all the training set and test set images of the same size. When I tried to use some publically available thermal image dataset with training and test images of varied sizes, I realized that it was a problem. Now, I am trying to figure out if there are ways to treat varied size images in TensorFlow.

2. Algorithms for fusion:

   The papers on fusion algorithms I have come across seem to be almost impractical to implement. I am not sure if modeling of the sort I have described above is feasible for us. It would be great to get in touch with someone who is pretty used to this kind of work.

As a team, we are facing a challenge of not being able to collect data. We have shared the test plan with our sponsors but it could still take a few days depending on weather and the Flight crew/our availability.

3. Teamwork:

Work done by individual team members:

- Team:
    - Plan for data collection using NEA payload
- Juncheng Zhang:
    - Literature study on human detection in thermal images
    - Test HOG+SVM algorithm developed for RGB images on thermal images
- Sumit Saxena:
    - Learn to use TensorFlow to implement neural networks for image classification
    - Literature study on possible ways to fuse data/information from different sensing modalities
- Karthik Ramachandran:
    - Voice activity detection on sound samples
- Xiaoyang Liu:
    - Review and modification of previous human detection algorithms
    - Explore implementing Feed-forward neural network for RGB image classification

All of us have begun to be fully involved in developing the human signature detection algorithms. I and Xiaoyang are independently developing neural network algorithms for image classification.

4. Future plans:

Following are the tasks I plan to work on until the next PR:

1. Implement feed-forward neural network for image classification with satisfactory results; explore variations
2. Explore algorithms for image classification based on color
3. Work on the data we collect with NEA