ECE 472 Robotics and Vision Prof. K. Dana

Final Project: 3D Reconstruction, Deep Learning, Augmented Reality.

Submission Instructions for Final Project Submit a 2 page progress report and a 4-5 page final report. Explain all algorithms in paragraph form demonstrating your technical knowledge of the code. Use equations (latex is highly recommended).
Submit: 1) ProjectReport.pdf 2) Your Python Code 3) link to Images necessary to run the code in a compressed folder called *Images/*. (Do not use very high resolution images), 4) A list of dependencies required to run your code in a README file.
Additional Instructions You may use opensource code except where noted. Give specific credit in your report about the source of this code. All additional code should be your own. For Part1, all images for reconstruction should be your own. This is an individual project.
You are encouraged to discuss methods and issues with classmates, including discussion of opensource code and relevant tutorials; however, no copying of code, report text or images is allowed.
Use Piazza to discuss what works and doesn't work for you, especially in dealing with systems/software/library issues.

1. This semester you have learned algorithms for 3D reconstruction with two images (i.e. stereo reconstruction). However, 3D reconstruction with many images typically leads to much better results. Modern computer vision applications optimize 3D reconstruction over many views (e.g. from multiple cameras or from the video feed of a single camera). This process is called *structure from motion - SFM* or *multiview stereo* and the refinement of the estimate is called *bundle adjustment*.

   (a) With your own set of images of an object (minimum 4 images) write code to reconstruct the scene using multiview stereo. Show the original images and the point cloud with sufficient detail to convey the 3D shape. Include these images in the final report. The 3D shape should not be trivial (e.g. not planes, spheres or cylinders).

   (b) In your report, describe the algorithm that is implemented in the components of your code. Use your own words. Use equations that you format (latex is highly recommended). Insert snippets of code in the report for clarification.

   (c) Some useful resources may include:
      - `http://openmvg.readthedocs.io/en/latest/software/SfM/SfM/`
      - `https://blog.mapillary.com/update/2014/12/15/sfm-preview.html`
      - `http://scipy-cookbook.readthedocs.io/items/bundle_adjustment.html`
      - `https://github.com/snavely/bundler_sfm`
      - `https://bitbucket.org/devangel77b/python-sba`
      - `http://cdcseacave.github.io/openMVS/`
        Note: OpenMVS provides reconstructed surfaces, not just point clouds. Surfaces are not required for this assignment.

2. (a) **Fine tune a network for a 5 class classifier (not MNIST)** Using Pytorch, select a pre-trained network to conduct a recognition experiment. Obtain your images from an existing database.

(b) In your report, describe the network you use (1-2 paragraphs).

(c) In your report, report the accuracy, precision and recall for your experiment, for your each of the 5 classes.

(d) Use drop-out regularization and batch-normalization. Explain how these work in your report. Compare the results with and without normalization.

(e) In your report, show the confusion matrix for your experiments with the fine-tuned network. Describe what you see from the confusion matrix that you do not see with other metric.

(f) Some useful resources may include:

- `http://pytorch.org/tutorials/beginner/transfer_learning_tutorial.html`
- `https://github.com/Spandan-Madan/Pytorch_fine_tuning_Tutorial`
- `http://cs231n.github.io/transfer-learning/`
- `https://gist.github.com/jcjohnson/6e41e8512c17eae5da50aebef3378a4c`
- `https://flyyufelix.github.io/2016/10/03/fine-tuning-in-keras-part1.html`

3. Augmented Reality **Graduate Students**

(a) Do not use an augmented reality toolbox

(b) Find a plane in the images (e.g. a checkerboard). For this part you may use the same sequence as for 3D reconstruction, or you may use a different sequence.

(c) Associate a new coordinate frame with this plane

(d) Build a wireframe model in this frame (e.g. cube)

(e) Draw the wireframe model in the image (attached to the plane) (Remember you have the world coordinates and the Camera matrices, so you can render this synthetic object, in a similar way as you did in the hw assignments). Show three views of the attached cube.

(f) Map an image (e.g. of yourself!) to one of the facets of this wireframe model Show three views.

4. **Graduate Students** Devise a prediction network or action network using **one** of the following two concepts learned in class. Describe your goal, your evaluation and the quality of the results.

(a) LSTM or other recurrent network

(b) Reinforcement Learning

(c) Multimodal Deep Learning