Proof of Concept Plan for McMaster Text to Motion Database
CS 4ZP6

Brendan Duke
Andrew Kohnen
Udip Patel
Dave Pitkanen
Jordan Viveiros

October 26, 2016

1 Risks

The significant risks of this project are split into three major sections represented by the website, database, and deep learning network. The largest risk to the project is linking the three sections together so that the website can pull information from the database, and the deep learning network can use this information to run pose estimation.

Some significant risks are involved within each of these sections and will be elaborated on below:

- The website must use some form of database query in order to correctly return information that the user searched for.
- The database must contain all the required videos and text pairings from larger libraries like Charades.
- The deep learning network is going to use Caffe and requires the steep learning curve that is associated with deep learning.
- Generating the proper test cases to provide proof of correctness is difficult on large databases.

2 Datasets

Our primary mode of input data for the program will be uploading an image to the website from the user's computer. We will be largely using the "Charades" Database in order to act as input for our program.

3 Deliverables

As a proof of concept demo we will show a working website that will function as a base for running pose estimation. This is to show that the associated risks mentioned above can be overcome and fully worked out as the project matures.

To utilize the pose estimation the user will either have to upload an image, or search the database from a subset of the options. The prototype will have two aspects to the demonstration with the addition of a third if full text search if it is implemented on a subset of the given data. The first aspect will be a website. Once on the website running the pose estimation will be shown by either searching for an action or uploading a picture and displaying the position of the chin, left and right humerus, left and right radius/ulna, left and right femur, left and right tibia/fibula and the spine of the person found within the image.

- A functional website, as an interface for running pose esimtation.
- The ability to upload an image and update the database with that image.

4 Performance Metrics

There are a few key aspects to which we can measure the effectiveness of our proof of concept demonstration:

- Page loads should happen in real-time.
- Image query should take less than ten seconds.

5 Resources

- Caffe
- TensorFlow
- FFmpeg
- Sphinx