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

Brendan Duke
Andrew Kohnen
Udip Patel
Dave Pitkanen
Jordan Viveiros

March 20, 2017

Contents

1 Risks

1.1 test

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 are 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.

All of the above are risks in the sense that they pose a challenge to complete the items in time to be demonstrated.

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

The following set of deliverables will be completed for the proof of concept demonstration.

- A functional website, as an interface for running pose estimation.
- Said website should contain a database.
- The ability to upload images and videos, and to update the database with those uploaded data.
- The ability to run human pose estimation on any uploaded image and video. By human pose estimation we mean that for the uploaded media, the skeletons and joints of any humans in those media will be indicated visually.
- The ability to search for uploaded images and videos through some means, e.g. by tag or name.

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