
User Manual

for

CS 4ZP6 Capstone Project

Version 0.0

Prepared by Brendan Duke, Andrew Kohnen,
Udip Patel, David Pitkanen, Jordan Viveiros

Using Volere Template Edition 13

McMaster Text to Motion Database

February 26, 2017

Contents

0.1	Legal and Copyright Information	3
0.1.1	Need List of tables and figures	3
0.2	Introduction	3
0.2.1	Purpose	3
0.3	Roadmap	4
0.3.1	Roadmap: Web clients	4
0.3.2	Errors or overwriting Data	7
0.4	Roadmap: Service Providers	7
0.4.1	Installation	8
0.4.2	Installation of Dependencies	8
0.4.3	Troubleshooting	9

Revision History

Name	Date	Reason For Changes	Version
David Pitkanen	September. 25th, 2015	Initial Version	0.0

0.1 Legal and Copyright Information

Here we don't have anything copyrighted although we should mention the GNU. Legal concerns are that we are taking up peoples photos so they might be wondering what we are doing with their photos.

0.1.1 Need List of tables and figures

0.2 Introduction

This is a web application which can be used to extract a low dimensional representation of the human body dynamics present in a video or the static pose of a person that has been captured in a picture.

This web application will allow a user to upload a video or an image that contains at least one person in it. If the user uploaded an image a static description of the human pose will be extracted from the image and stored in a database. If the user uploads a video a representation of the human motion present in that video will be stored on the same database. The motion or pose that the application extracts from the video or image can be displayed by the application.

The motion or pose that is captured by the application is represented by thirteen (x,y) coordinates that reference the point pixels where 13 key body parts are located in the frame or picture of interest. The way these thirteen coordinates change from frame to frame can be used to model human motion. These points represent the positions of the head, shoulders, knees, wrists, elbows, hips, and feet (a total of 13 points) of the person on the 2D picture/video.

In addition the application offers a search function. If the media that is inputted includes a description of the motion or pose then other users may query the database. The database will perform a search for keywords and return motions or poses that match the keywords.

0.2.1 Purpose

The initial purpose for developing this tool was to generate a data source that would be useful to machine learning researchers and developers. Machine learning researchers are

interested in creating models from labelled data. A tool that can find the joint positions in images and videos is obviously a good method to generate data that is labelled with joint positions.

There are many large data sources (videos and images) which have been annotated with natural language describing the human motion. This motion/joint positions extracting tool could be used to enrich data sources annotated with natural language. These data sources would contain related natural language and motion descriptions and so would be good sources that create connections between these different domains.

However this project can be used as a tool for other applications as well since the task of finding the human motion in a video or pose in a picture is useful in its own right. Since it can keep track of human body positions on a screen it can also be used potentially for training people how to move or as an aid in games where motion needs to be kept track of.

Background Applications

Many applications use Microsoft human motion and video games such as the oui.

0.3 Roadmap

This text-to-motion software will have two broad classes of potential users. The first class will wish to use the web application as it is available on our website. These users will simply navigate to the site using a browser and take advantage of the framework provided. However the second class will wish to take the software we provide on github and host this software as their own website to collect their own data and provide their own services.

Both of these types of users will have their own challenges. Of course users who wish to become hosts will themselves have users of their own so the first class of potential users problems will be of interest to the second class.

We therefore break this section down into two streams. The first stream shows the challenges that are of interest to users of the web application once it has been installed. The second stream is for problems that will only be of interest to potential users who wish to get the software running on their own server.

0.3.1 Roadmap: Web clients

The web application we provide will have many pages that have different functionalities. The navigational structure of the application is shown in Figure 0.1.

The navigational structure of our site is very simple. The navigation starts at the Home page then any other page may be reached by clicking on a link on this page. The possible links are ImagePoseDraw, Login and Register.

Each end point on this graph represents a functionality that the application offers: create, search, details, contact, about, login, register, search and search results. To navigate to this functionality the navigational chart can be used in figure 0.2. A node

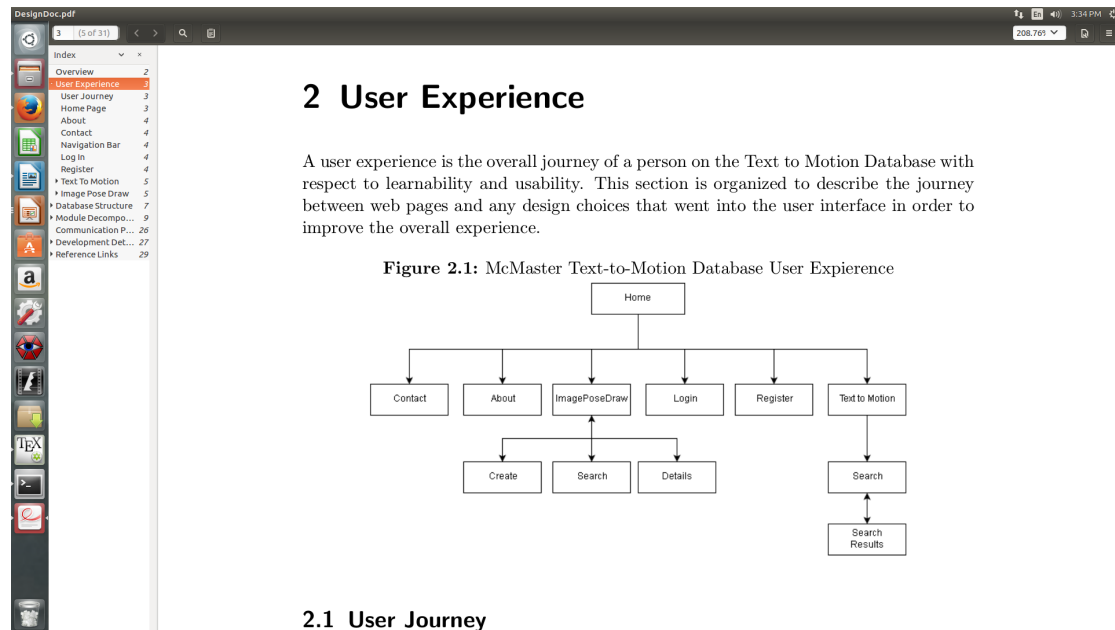


Figure 0.1: Web Application Navigation Structure

for a page lower in the structure can be reached from a node that is one level higher. To perform the navigation go to the more highly ranked page and then click on a link that is labelled with the lower nodes label.

For instance to get to the TextToMotion page start at the home page and then click on the link labelled TextToMotion.

The details, contact and about sections describe are actually pages that display important data. The contact page/function displays contact information to get in contact with the creators of the software. The about page/functionality displays the information that describes motivation for the web site.

For our site it is necessary to create a user profile to perform certain tasks. The login and register functions allow users to create profiles and login to their accounts.

Finally the search and search results allow users to search the data that is stored by the web application and to display the results of their searches.

Browser Requirements

Currently our website is hosted at the address address 159.203.10.112:80. By typing this address into a browser the Home page of our site should appear. The specific page that should be seen in figure 0.2

This website can be accessed by any browser that supportst the HTML5. Any version of Internet Explorer more recent than version 6 should work and updated verisons of Chrome and Firefox will work as well.

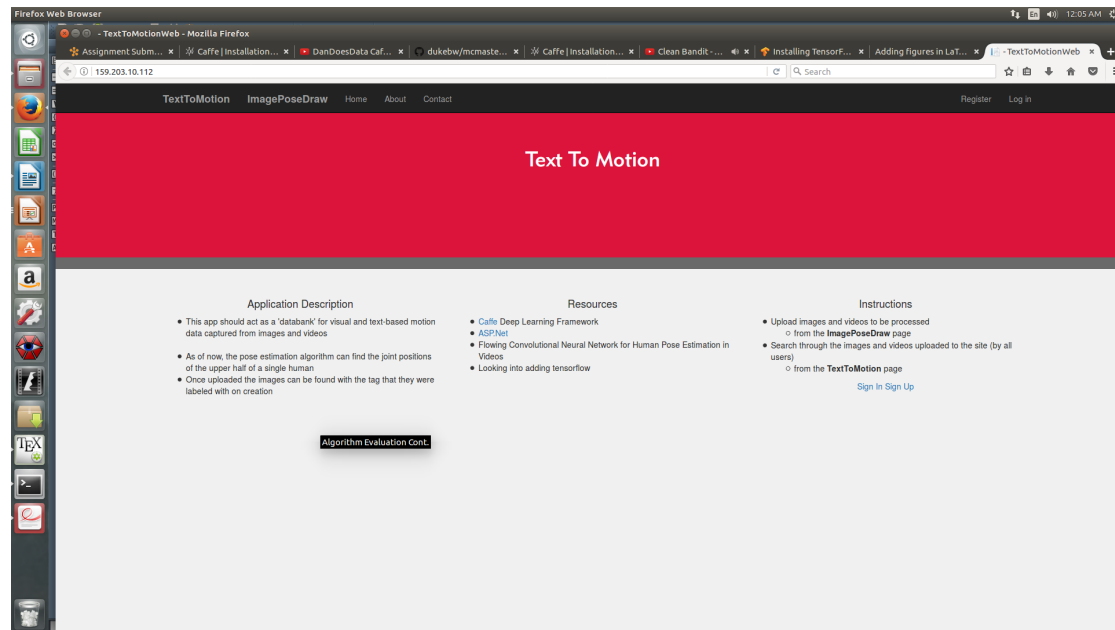


Figure 0.2: Web Application Navigation Structure

Log In/Sign Up

To use all of the features offered by the Text-To-Motion site users must create a profile. Without creating a profile data that is on the database may be searched and displayed however to enter new data a user must have a profile and be logged in.

To register as a user we can start at the home page and then click on the register link. Once this link has been clicked on the register page will appear. To register only an email address and password are necessary. Once these fields are filled out and the register button is clicked a user account will be created. Note that each user account must have a unique email address.

Data Input

The central feature of our application is in collecting and displaying data. To navigate to the page that performs data input and displays already collected data navigate to the home page and then click on the ImagePoseDraw link.

From this page data can be searched and inputted if the user is logged in. To input data simply click on the green button labelled "Add New Image".

After this button is clicked a new screen will appear that has 3 fields that need to be filled out: filename, filepath and description. The filepath is entered by clicking on a new button that opens a user interface for a file explorer which allows users to navigate the directories on their computer to the location of the file they wish to upload. Once the submit button is pushed the appropriate file will be uploaded to our database and the file will be labelled with the pixel locations that correspond to the joint positions.

Data restrictions are that the images file format must be in jpeg and peg formats. Also the website is set up for downloading large data samples but a limit of 2 minutes is set on the upload time. Because of this the sample will depend on the upload speed.

Data Search

Troubleshooting

One problem since the application provides a labelling service is that the labels may be assigned to the inputted data poorly. To improve on the data can anything be done pre-processing or what kinds of data does the application work best on?

The second type is the more technical in nature of problems of username, lost data and searching. Not returning the results that you want? How to best implement a search using this?

For entering usernames and passwords we require that passwords have at least 1 number, 1 letter, 1 capital and one non-alphanumeric characer. Uploading large files will only be allowed to take 2 minutes so that to upload larger files will have to be broken up into smaller segments.

0.3.2 Errors or overwriting Data

For users each user must create a unique username. If a user uses a username that is already in use a error message will be shown to the user and they will be asked to enter a new name. Multiple users can share the same password however.

For entering and videos no limit or restrictions are set for duplications. So the same image or video may be uploaded many times without causing overwriting to occur. However if the user wishes to avoid duplication they can search the database for specific pictures or videos using the search function.

0.4 Roadmap: Service Providers

As previously mentioned the software is for a web application which runs a computationally expensive image/video analysis on its backend. To store the large amounts of data a database server is needed for hosting. In addition we wish to query the database so a search library is also needed. Then to perform the calculation a machine learning library is used and finally to host a web framework is needed. All of these components need to be installed separately.

The specific software tools, packages and libraries that need to be available to the system to run our softwar are:

- Python version 3.x
- ASP.Net MVC
- MySQL

- Caffe
- TensorFlow
- Sphinx search library

0.4.1 Installation

The software project we have developed is available on github: <https://github.com/dukebw/mcmaster-text-to-motion-database> and can be downloaded by cloning our repository.

Instructions on using github are available at this address:

However if the user is running a linux system the repository can be downloaded by typing in the clone command from the terminal screen:

```
git clone --recursive https://github.com/dukebw/mcmaster-text-to-motion-database
```

However the software we have developed has several dependencies which fall under the three categories we have decomposed the project into: database, machine learning libraries, and the web framework. Python is a central language used in our project is in nearly all of these modules. We assume python version 3.5 or higher is installed.

0.4.2 Installation of Dependencies

Database Installation: MySQL and Sphinx

Machine Learnign Libraries Installation: TensorFlow and Caffe

Unfortunately for Caffe as a high level and efficient image analysis tool so it itself has other dependencies. The instructions for installing Caffe can be found here: <http://caffe.berkeleyvision.org/install apt.html>

As can be seen at this website in order to run caffe CUDA, BLAS and python are required. Instructions for installing CUDA and BLAS are shown at the website mentioned above. Within the ubuntu console these packages can be installed with the following commands:

```
sudo apt-get install libprotobuf-dev libleveldb-dev libsnappy-dev libopencv-dev libhdf5-serial-dev protobuf-compiler sudo apt-get install --no-install-recommends libboost-all-dev
```

Tensorflow is also used in are framework. This library can run both on python and C++ and in our program we use the python accessible version. To install Tensorflow again: <https://www.tensorflow.org/install/>. The install instructions are too numerous to mention here but it is available for all environments: linux, Mac OS and Windows.

Web Framework: ASP.NET MVC and Python

The web framework we are create has two separate modules that communicate with one another. To run them both Python and ASP.NET need to be installed. The python componen relies on socket library that is in the standard python library so no new installations are required for python.

To run ASP.NET the

0.4.3 Troubleshooting

Things they should look out for an possible errors?

Formats of files that they upload. Size capacity. Expected speed.