Cardiff University

Prifysgol Caerdydd

School of Computer Science and Informatics

Initial Plan: Generation of facial cartoons

CM3203 Final Year Project
40 Credits

BSc Computer Science

Author:
Lewis Hemming

Supervisor:
Prof. Paul L Rosin

Moderator:
Dr. Yukun Lai

February 2020

1 Project Description

Over the past 15 years the use of social media has grown exponentially, rising from just 7% of American Adults actively using social media in 2005[3] to 70% in 2019[2].The global penetration of social media was 45% in 2019 which equates to an estimated 3.484 billion people actively using social media[2] making it a huge part of modern culture. One of the current trending features of social media is using apps to create facial effects from pictures, videos or live. However there are currently almost no social media applications that create accurate cartoon stylised faces of the user. This project aims to tackle this problem by creating a program that intakes an image of a frontal face and generates a cartoon stylised version. I will be using the google cartoon set;

https://google.github.io/cartoonset/index.html

as an initial library to generate facial cartoons.

At least in the first instance, this project will take a simple approach, in which facial features (eyes, mouth, hair, etc) are matched to this library to create an accurate cartoon representation of the face in the photo. Then the face can be stylized by compositing the matched features then refined by adapting them to better fit the input image and feature the same background.

Following instances may contain features such as; implementing a choice from multiple libraries to allow different cartoon and art styles to be chosen, being able to extract features from more natural images with noisier backgrounds then recreate the image in the new style, extracting features from faces at multiple angles, facing different directions, then refining the cartoon features to fit the original image.

2 Project Aims and Objectives

1. Edge Detection
   - *Aim:* To be able to create an accurate image of outlined facial features with minimal noise from the input image
   - *Objective:*
     I Use multiple edge detectors, such as Laplacian edge detection and the Sobel edge detector, to create multiple instances of the edge detection images
     II Form a convolution from the multiple instances to remove noise and show only the strongest edges
2. Feature Extraction
   - *Aim:* To be able to landmark the important features of the face within the image (eyes, mouth, chin, etc.)
   - *Objective:*
     I Using a method such as the Active appearance model the main features of a face (chin, nose, eyes, mouth and eyebrows) will be land-marked
     II Each feature will then be extracted so that they can be matched with library features
3. Hair Segmentation
   - *Aim:* To be able to partition the hair of the face from the rest of the input image
   - *Objective:*
     I Use a method such as Instance-level Human Parsing via Part Grouping Network[1] mixed with the aforementioned Feature extraction to segment the hair from the rest of the face.
     II Find the median hair colour by creating an organised list of all pixels within the hair segmentation
     III Save the extracted hair as a feature to be later matched to a library
4. Library Creation
   - *Aim:* To be able to create a library of facial features from a given data-set of cartoon images
   - *Objective:*
     I Using the aforementioned feature extraction and hair segmentation methods, extract all different features (eyes, hair, eyebrows, mouth, etc.) of each image and store them in separate sets.
     II Using the feature matching method described below and an appropriate threshold, each of the features within the set can be matched to the other features within the set to discover all the uniquely different features and delete the features that are the same to leave a data-set of unique features.
5. Feature Matching
   - *Aim:* To be able to match the features extracted from the input image to the most appropriate feature from the given library
   - *Objective:*
     I Using a method such as Euclidean distance, which is calculated by:

     $$D_2(j,k) = \sum_m \sum_n F(j+m)(k+n)T(m,n)$$

     we can calculate the similarity between the image features and each of the features within the library.

II Using thresholding and a comparison of the calculated similarities we can determine which feature within the library best matches the input image feature

6. Image Refining
   - *Aim:* To be able to warp the chosen features from the library to fit the image
   - *Objective:*
     I Adjust the pose (scaling, rotation, translation) to fit the feature to the same location and pose as that of the original image feature
7. Similarity Measurement
   - *Aim:* To be able to measure the similarity of the original image to that of the newly generated cartoon image so that the success of the results can be evaluated
   - *Objective:*
     I Use a mathematical formula or method, such as absolute distance, to compute a similarity measure of the output compared with the input and then evaluate the results from this data
     II Anouther option/method of data collection would be to ask a group of randomly selected people to match faces to the output result to see if people can still tell which face matches which cartoon stylised image, however this method would require ethics (*see part 4 of the initial plan*)

3 Work Plan

See Appendix: Figure 1 for Gantt chart of time plan

3.1 Program

1. Research:
   I believe an appropriate amount of time for further research would be one week as this will allow me time to make final decisions about the methods at which to approach each aim as well as to research new methods that i may not have read into previously.
2. System Design:
   Three days is enough for the system design as the preliminary research will have been completed by this point so all that will need to be done is too align which method shall be used to approach each aim and how the main program will run overall. This will also allow sometime to setup the I/O system and the image handling.
3. Feature Extraction:
   The feature extraction is one of the largest sections of code to tackle and is required to recognise faces in both the input image and the library creation. It is arguably the most important part of the code as it is the setup for both extracting data from the input and setting up the library. For this reason it has been assigned two weeks and four days so that it can be first implemented in a simple way then refined so that it is as accurate as possible and resistant to noise.
4. Feature Matching:
   Once the feature extraction is complete, the feature matching maybe begun. In theory the matching has been assigned three weeks however this is extremely dependent on the aforementioned section "Feature Extraction" so if the section is completed and refined enough prior to 02.03.2020 the time to complete the feature matching section could easily be extended. This will also allow time test the matching feature and ensure the average output is as accurate as possible.
5. Extra Libraries:
   The final week will give one week to find more open source data-sets to allow the users an option of what art style they wish the final image to be created in. As this is not a necessary feature but rather an extension this section will only be covered if the previous sections have been completed and refined.

3.2 Final Report

1. Introduction:
   As this is the initial section it has been allocated one week to complete as it will also cover the initial setup and style of the entire report. This will also cover time to set up the support sections such as the abstract, the glossary, acknowledgments, etc.
2. Background:
   As only one more section after the background section does not require a completed program this section has been assigned three weeks to complete. It will also allow plenty of time to cover all the preliminary research that's been completed to begin creating the program.
3. Implementation:
   Implementation should not take four weeks to complete however the further sections require the completion of the program so I cannot go past this section until that milestone is reached. This section will be gradually written as each aim is completed within the program as it will be impossible to write this section before parts of the program are complete and working.

4. Results/Evaluation:
   Once the program completion milestone has been reached this section maybe started. It's been assigned three weeks as it also includes the time to collect data and create surveys if necessary. However this time frame is the most flexible of the final report as it is entirely dependent on the completion of the program so if the program is completed earlier then this section maybe started sooner. On the other hand if there are too many unexpected errors this could push the date this section is started backwards meaning that there would less time to collect data and complete the evaluation of the results.

5. Future Work:
   This section has been assigned one week as it requires no further research but rather a reflection on what could have been added to the program had I had more time.

6. Conclusions:
   Conclusions has also been assigned one week as this will be an overall summary of the final report and all conclusions that can be drawn from the results. As by this point the main report will be mostly complete this conclusion week will also be used to make edits to the style and support sections so that the report is well presented.

7. Reflection:
   Reflection has been assigned one week, the final week, as it requires all previous sections to be completed. This will be reflection on what has been learnt from both the report and the program itself.

4 Ethics

As all data-sets used in both training and testing are open source no ethics are required for the use of facial images. However, if i choose to collect data from a group if people about the success of the program or similarity of the input image compared with the output cartoon stylised image then I will be required to meet Cardiff University's policy on ethics. This would involve both me and my supervisor (Prof. Paul L Rosin) to complete the "Research Integrity Online Training Programme". I will then be required to complete a Ethical Approval Form and send it to *comsc-ethics@cardiff.ac.uk* to be approved. Once this is completed I may begin data collection and Human involvment within my project.

References

[1] Ke Gong, Xiaodan Liang, Yicheng Li, Yimin Chen, Ming Yang, and Liang Lin. Instance-level human parsing via part grouping network. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 770–785, 2018.

[2] Simon Kemp. Digital 2019: Essential insights into how people around the world use the internet, mobile devices, social media, and e-commerce. *Kepios Pte Ltd., We Are Social Ltd. And Hootsuite Inc*, 2019.

[3] Andrew Perrin. Social media usage. *Pew research center*, pages 52–68, 2015.

5 Appendix

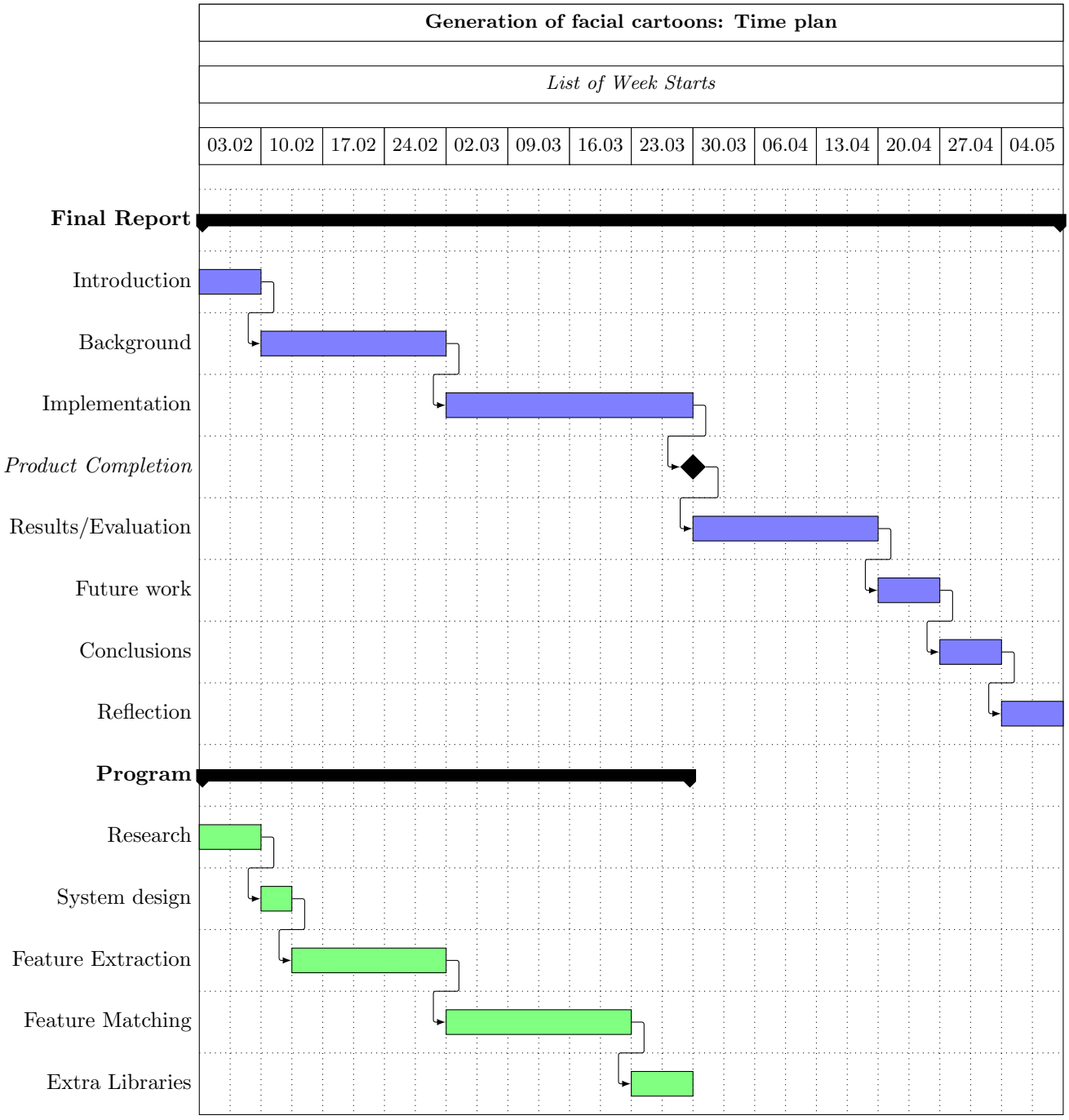

Figure 1: Gantt Chart of Planned Timeframes