BISSIT 2022 Team project

Automated Mosaic Service Combining Multiple Models Development

Team : Mission i'm Possible

Team members : Da-Bin Lee, Jang-Hoon Park, Jong-Hoon An, Joy Muigai

1. Abstract

With the development of the media content market, the damage caused by portrait rights infringement is increasing. Currently, there are no legal regulations that punish portrait rights infringement, so there is a hassle that individuals have to file civil lawsuits, and as a countermeasure, specific platforms regulate portrait rights. However, video providers who do not use a specific platform continue to cause problems of portrait rights infringement. In order to prevent infringement of portrait rights, it is intended to provide a service that automatically processes mosaic using multiple models. In this work, we propose an automatic mosaic service using face detection and recognition. It is intended to solve the problem of portrait rights in the media market in the future by using the services implemented through this methodology.

2. Introduction

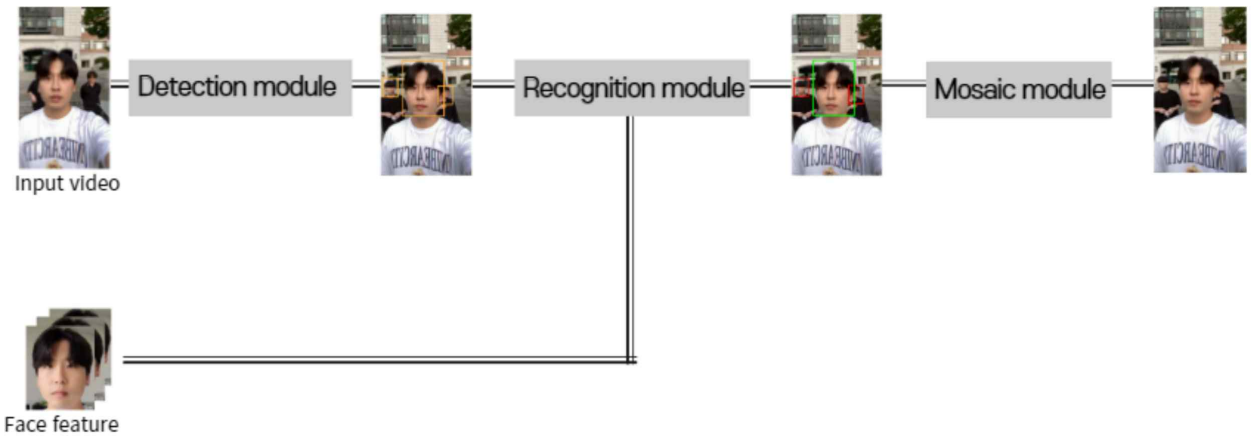
This project aims to deal with the problem of portrait rights infringement caused at the same time as the development of the media content market. The problem of infringement of portrait rights was greatly highlighted by the disaster reporting process of public broadcasting. Due to the nature of broadcasting, the site must be quickly delivered, so there have been cases in which the portrait rights of the characters appearing in the video were violated due to mistakes or errors in broadcasting. According to the current Article 5 of the Media Arbitration Act, the reasons for exemption for portrait rights infringement are defined, but this is limited to the media, Internet news services, and Internet multimedia broadcasting, so it does not apply to media platforms that provide videos. If the right to portrait is violated on the media platform, a claim for damages caused by civil illegal activities can be filed, but the process requires a lot of time and effort for individuals to proceed. In addition, there is a problem that if the backside is photographed, passes very quickly, or is difficult to recognize the person's face due to a distant background, it is not recognized as an infringement of portrait rights, and even if it is violated, the damage is insignificant and it is difficult to claim actual damages. Video is restricted by the platform's own rules for the current situation without clear legal standards. Accordingly, the video provider protects the portrait rights of others by mosaic using an image editing program. Not only is labor and time invested in this process, but there are limitations on resources in the editing process. In the case of YouTube, it is possible to mosaic other people's faces through the facial blurring function, but not only does it not detect all faces properly, but there is a problem with the mosaic range. Recently, a mosaic processing technology through face recognition has been devised in real time due to a corresponding problem, but it is difficult to apply the present technology

in real life due to a decrease in frame per second (FPS) according to a processing speed. Therefore, in this work, we propose a service that prevents unintended portrait rights infringement problems through post-processing automatic mosaic services through facial detection and recognition.

3. Methods

3.1. Overview

Our method consis of 3 module. the first module is “Detection module” which detect the faces and output the face box which consist of pixel position of face and width & height of the face. after that, the second module, “Recognition module”, take the first module's output and represent photographer's face feature which were generated in advanced and recognize the photographer's face box. for the last, the third module “Mosaic module” make the people's face blurred except the photographer's. Blueprint of our methods are like down below.



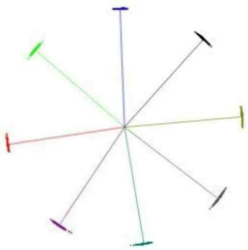
We approach in 2 different ways. the first one is method using nerual network. and the second one is method using classical computer vision algorithm. Before we noted the detail of the methods, in summary, we choosed the second one because of the task-specific problem (inference speed is way more important than accuracy)

And this is how we approached to handle the problem explained above.

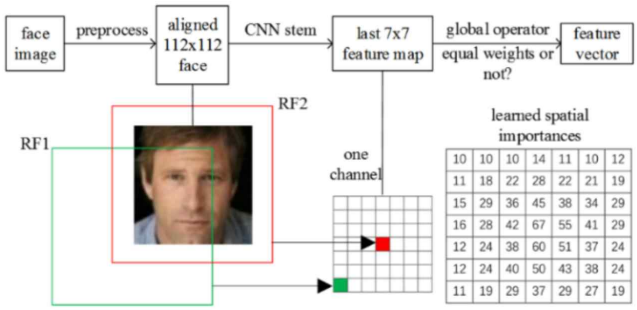
3.2. Detection & Recognition with Neural network

Neural network are leading the vision application these days. Neural network is data driven approach which is not fully explainable even now days. but with the proper network architecture and training data, this approach can lead the complicated results which rule based algorithm cannot make. so our first approach was making the detection module & recognition module with neural network. we've consisted our network with “Arcface” and “MobileFaceNet” with “K-Face” dataset and data augmentation.

“Arcface” and “MobileFaceNet” is face recognition model mainly consist of the convolutional layers and it shows highly accuracte and pretty fast infrence results. so we choosed this model to implement our application

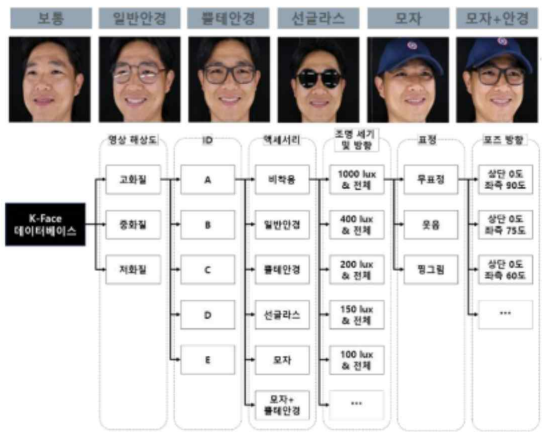


Arcface

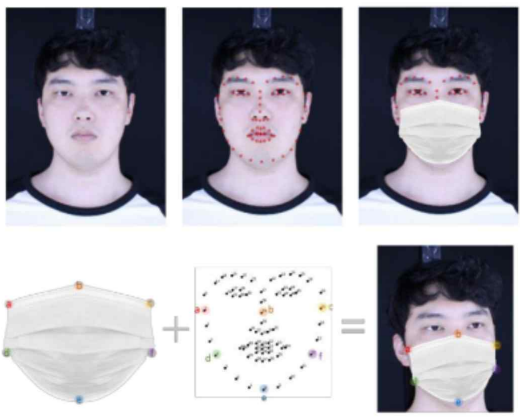


MobileFaceNets

"K-Face" dataset is the dataset that is consist of the many faces picture which contain different resolution, ID, accesory(like the cap, glass, etc...), light(intensity, angle), face expression, pose. with this dataset, we can recognize people invariant to the character explained above. and further more, to adapt our application to the Convid-19 pandemic era, we augment the data with Mask Perspective Transformation. We created an image by applying Perspective Transformation based on 6 points each in the coordinates of mask and facial feature points. with this data agumentation, we could improved the performance about 1 to 11%.



K-Face Dataset

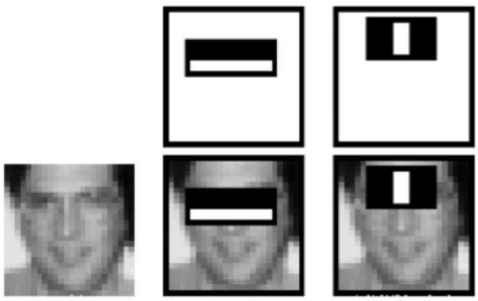


Mask
Perspective Transformation

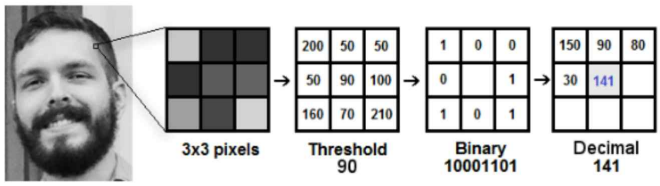
In this way, we implemented our application with neural network and achived the greate accuracy. howerver, as a trade-off, the inference speed slowed as the CNN require plenty of the matrix multiplication, which is easliy affected by the enviroment(especially when they can't use GPU) and is not generally applicable.

3.2. Detection & Recognition with Classical computer vision algorithm

Face recognition in this project consists of a total of two stages. In order to recognize the user's face in second approach, we used intensity based face detection algorithm, "Haar cascades" & histogram based face recognition algorithm, "LBPH algorithm".



Haar cascades for Face detection



LBPH algorithm for Face Recognition

In LBPH, the user's face generates each model. This has the advantage of devising a Re-ID (Re Identification) method, and if the model is assigned to each user, it can be extended without having to learn each additional person. However, there is a problem with the accuracy, which are important factors applied in the process. In order to improve the problem, the problem is improved through fine tuning and model lightening or model replacement.

3.3. Mosaic Methods

With the Gaussian kernel, to the target people's face pixels, we made the mosaic module that blurring the others faces.

4. Conclusion

Through this project, it was possible to reflect that off-line-based processing is more effective than real-time image processing. The proposed service allows producers of media to save human resources and time through automation rather than manual editing, and allows editing without resource constraints due to the progress of work through servers. In the future, it is expected to prevent potential single-person media producers or video editors from violating portrait rights. In addition, through future research, it is proposed to protect the portrait rights of surrounding figures and to use them for crime prevention and finding missing persons.