Research Protocol

Peter Remøy Paulsen

2021
October

1 Synopsis

This experiment aims to demonstrate the quality of experience of a machine learning based silhouette extractor provided by the AdMiRe project.

A set of videos will be generated with the machine learning silhouette extractor developed by EPFL for the AdMiRe project. These videos will firstly be analysed by the objective measures pixel accuracy, Intersection over Union and Dice Coefficient, typically used in evaluation of semantic segmentation models. Afterwards, the videos will be rated subjectively by a set of participants. The participants will be asked about their satisfaction with the quality of the silhouette extraction, if they noticed any artefacts and if they thought the artefacts was annoying.

Lastly objective and subjective measures will be analysed and compared to see if there is a correlation between the measures.

2 Introduction

The last decades traditional television viewing numbers have steadily been declining [2]. People do not find TV as appealing as they once did. Today, more people seem to find more engaging and personal forms of entertainment elsewhere. Currently, the only form of doing, what one could call an engaging TV broadcast, is through the use of social media and hybrid broadcast broadband TV by incorporating comments, videos and audio from the audience into the TV broadcast. Sadly, this form of engagement is quite limited and does not give proficient results.

Wanting to innovate and create better experiences in this space, the AdMiRe [1] (Advanced Mixed Realities) project has been formed as a collaboration between Brainstorm, Disguise, NTNU, EPFL, UPF, NRK, Premiere, TVR and CSIC. The aim of the AdMiRe project is to make use of mixed reality solutions to enable audiences at home to be incorporated into live TV programs and interact with the other people in the TV studio.

Doing this using the available technology is hard because of the technical challenges. To make this easier AdMiRe is set out to develop and simplify key modules.

An important aspect of this technology is to make it look like the participant is in the studio, and to make it look like the participant is in the studio, the participant has to be extracted out of its own environment and inserted naturally into the studio environment. That's why we will take a look at the machine learning silhouette extraction module currently used in the AdMiRe project. We do this to evaluate the quality of experience and generally assess the quality of the technology by running some objective and subjective tests.

3 Hypothesis

RQ: Is there a correlation between the objective measures pixel accuracy, IoU and Dice Coefficient and the subjective measures of satisfaction, level of artifacts and level of annoyance in the machine learning based foreground extracted processed videos?

We form the following hypothesis expecting that the videos which gets a low score on the objective measures, also will receive a poor score subjectively.

$\mathbf{H_1}$: The videos with poor objective statistics will also receive poorer rating from the subjective testing.

Another interesting subject, is to see if the videos which received a good objective testing, will be matched with a good rating from the participants

$\mathbf{H_2}$: The videos with good objective statistics will also receive a good rating from the subjective testing.

4 Methodology and design

Video setup

We will create a system consisting of several parts. The basic construction will be set of test videos in front of a green screen (figure 1a), where we try to replicate situations which we assume would happen in a real world usage. These videoes will be created at the Sense-IT laboratory at NTNU. The resulting video will be edited using chroma key compositing to remove the background (figure 1b). Further on, the video will be inserted onto different kind of background videos (figure 1c). This video will finally be processed by our machine learning based foreground extractor, which will output our segmented silhouette extraction (figure 1d).

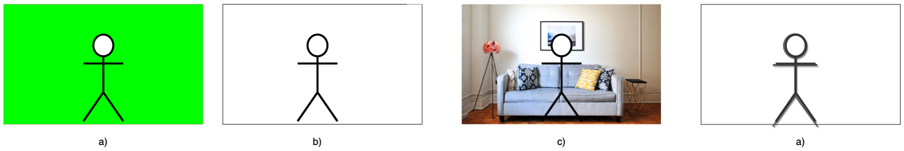

Figure 1: Phases of video system design

The list of backgrounds and foregrounds are as follows:

Backgrounds

- Simple white wall
- Complex wall with different colors and textures
- Background with windows to test for exposure difference and possible movements

Foregrounds

- Person counting ten fingers
- Person wearing a light clothing rocking back and forth
- Person wearing a dark clothing rocking back and forth
- Person displaying an object in their hands

By combining the foreground and background videos we will have a total of twelve (12) videos to run our objective and subjective analysis. Each video will have a duration of ten (10) seconds.

Objective measures

The resulting videos from figure 1b (chroma key video) and 1d (machine learning video) will be statistically compared and reviewed. We will be using Python to retrieve the following objective measures [4][7]:

- Pixel Accuracy
- Intersection-Over-Union [3]
- Dice Coefficient [6]

Subjective measures

A group of participants will be given a questionnaire. The questionnaire will map the demographic and they will be asked some questions for each video. The question will be as follows:

- How satisfied are you with the quality of the silhouette extraction?

- Did you notice any artefacts with the silhouette extraction?
- Do you think the artefacts were annoying?

The participants will be able to answer the questions with a five point Likert scale [5] and also be able to add additional qualitative feedback in the end if wanted.

The rating will be done using Google Forms, where the videos have been uploaded to YouTube (unlisted option). Using this setup, we will be able to test on a lot of participants, which hopefully will yield a more fair and even result.

Hardware

What	**Model**	**Specifications**	**Comment**
Video Camera Mobile Phone	Google Pixel 5	Resolution: 1920x1080 Codec: H.264, AAC, avc1 Color Profile: (5-1-6) Rec.601 (PAL)	Phone used to replicate the normal use case for the AdMiRe project
Editing machine	MacBook Air	1.1GHz 4-core i5, 16GB RAM	Editing software: Final Cut Pro 10.6
Machine learning machine	N.A.	3.6GHz 8-core i7-7700, RTX A6000	Running Ubuntu
Green Screen	Elgato	Extended: 148x180 cm	Feet get cut off
Ring Lights	Elgato	2900-7000K, 2500 lm, 45W	One for left and right side. To remove shadows from green screen
Tripods	Any	N.A.	One for each ring light, and one for camera

5 Results

The results from the questionnaire will be analysed and compared to the statistical measures. For the subjective measures we will look at the mean opinion score to minimise extremities from the answers. We will use Python to analyse and compare the objective and subjective data.

6 Timetable

Start	End	What	Comment
26.10.21	02.11.21	Approval of research protocol	
02.11.21	16.11.21	Develop test system	
16.11.21	30.11.21	Recruitment period	
30.11.21	20.12.21	Analysing results and writing paper	

References

[1] AdMiRe. URL: http://www.admire3d.eu/ (visited on 10/28/2021).

[2] *Halvparten av befolkningen dropper TV.* Statistisk Sentalbyrå. 2020. URL: https://www.ssb.no/kultur-og-fritid/artikler-og-publikasjoner/halvparten-av-befolkningen-dropper-tv (visited on 10/11/2021).

[3] *Jaccard index.* Wikipedia. URL: https://en.wikipedia.org/wiki/Jaccard_index (visited on 10/25/2021).

[4] Jeremy Jordan. *Evaluating image segmentation models.* 2018. URL: https://www.jeremyjordan.me/evaluating-image-segmentation-models/ (visited on 10/25/2021).

[5] *Likert scale.* Wikipedia. URL: https://en.wikipedia.org/wiki/Likert_scale (visited on 10/25/2021).

[6] *Sørensen–Dice coefficient.* Wikipedia. URL: https://en.wikipedia.org/wiki/S%C3%B8rensen%E2%80%93Dice_coefficient (visited on 10/25/2021).

[7] Ekin Tiu. *Metrics to Evaluate your Semantic Segmentation Model.* towards data science. 2019. URL: https://towardsdatascience.com/metrics-to-evaluate-your-semantic-segmentation-model-6bcb99639aa2 (visited on 10/25/2021).