MULTIMEDIA UNIVERSITY

Visual Information Processing
TDS 3651

Assignment

Sidharrth Nagappan
1181102313

Table of Contents

Abstract

In this task, a face segmentation algorithm that relies only on image processing techniques is designed. The absence of more sophisticated techniques like deep learning means we have to more closely select relevant features, define areas of interest and create the right filters. It is also obvious that no one mask fits all images and any pipeline will have limitations. In this task, a workable pipeline is established with an average precision of 0.6905, involving various colour spaces, denoising strategies, region of interest definitions and masking.

Introduction

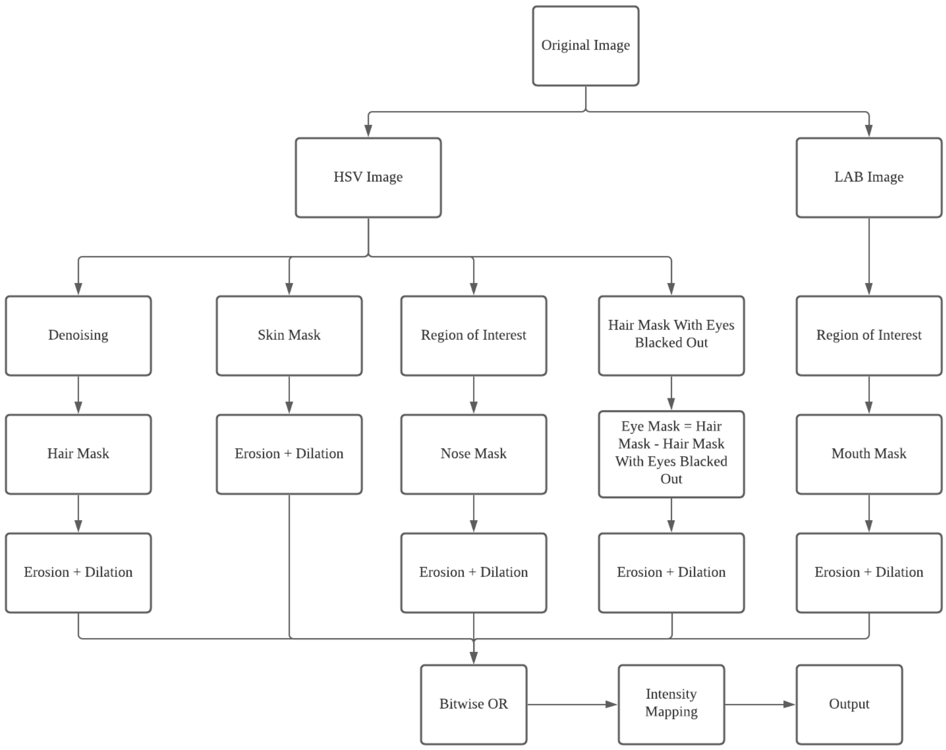
Before defining face segmentation specifically, it is important to understand the more general concept of semantic segmentation. Semantic segmentation assigns each pixel of an input image to an output class, essentially separating a visual field into specific entities.

Face segmentation splits a face into key elements; in this case, it is the background, skin, eyes, nose and mouth. A segmentation pipeline could either be based only on masks and filters, clustering or deep learning-based via an encoder-decoder architecture. UNet and MobileNet have produced remarkable results in this context. Uses include analysis of different facial regions, visual sentiment analysis, image reproduction, face recognition and beautification. Advances in computer vision have made these applications all the more straightforward.

This work will cover a naive approach that employs fundamental image processing techniques to segment out different parts of the human face.

Description of Methods

There are 6 separate elements to segment out of a face. Different colour spaces and methods work for different elements. Each will be masked out before having all masks applied to the input image to obtain the segmented output.



1. Background
   Binary thresholding is employed to separate the face (foreground) from the background. While this is simple, thresholding produces good results in detecting and labelling the background.

2. Hair/Eyebrows
   The hair uses the HSV colour space within a defined range. The mask is then eroded and heavily dilated to make the region complete. A shortcoming is different parts of the face with the same shade being classified as hair.

3. Mouth
   The mouth displays a bright green shade in the LAB colour space, that clearly accentuates it. This makes LAB a good candidate to extract the mouth. A region of interest is first defined to ensure that other areas of the face are not mislabelled; areas surrounding the mouth are covered with a black box.
   The image is converted into the LAB colour space and the range for the green shade is set to only mask out the mouth. This mask is then applied to the main HSV image.

4. Eyes
   Since the eyes and the hair are of the same shade, the hair mask often also segments the eyes. In order to only extract the eyes, an image with the eyes blacked out is subtracted from the hair mask. The result is an eye mask that will segment out only the eyes.

	-		=	
Hair Mask		*Hair Mask With Eyes Blacked Out*		*Eye Mask*

5. Nose

   The nose has a region of interest set in the HSV colour space. A lower and upper range is set before a mask is prepared. Applying the nose mask on the face results in a region filled with noise, because of the size of the region of interest. The noise is eroded away via the OPEN operation.

6. Skin

   The skin works in the HSV colour space and uses a set range of values to extract the part of the face. Since the colour range is set, the skin mask is not too effective for cases where the colour of the face or the shade is different.

After devising optimal masks for each part of the face, they are combined via a bitwise addition with the HSV version of the original image.

Original Test Dataset

The pipeline was applied to the test data and the additional data provided.
An average precision of 0.6905 is attained across 50 images. This may take a while to run.

IMAGE RESULTS

Image	Error	Precision	Recall	IoU
1	0.3497	0.8322	0.6533	0.5281
2	0.2817	0.7426	0.7399	0.5951
3	0.2931	0.7113	0.7498	0.571
4	0.3689	0.7445	0.6529	0.549
5	0.2434	0.7408	0.7968	0.6239
6	0.2215	0.8389	0.7749	0.657
7	0.3023	0.6723	0.7569	0.5797
8	0.3617	0.7682	0.6379	0.4966
9	0.3495	0.8051	0.6209	0.5424
10	0.3606	0.6887	0.6618	0.4991
11	0.2834	0.7452	0.7052	0.5931
12	0.4151	0.652	0.6216	0.4651
13	0.7149	0.4526	0.3572	0.2095
14	0.5367	0.535	0.4346	0.3575
15	0.311	0.6423	0.7795	0.5609
16	0.4005	0.7991	0.5965	0.466
17	0.2381	0.745	0.7951	0.636
18	0.4147	0.728	0.6259	0.4621
19	0.3296	0.7306	0.7144	0.5493
20	0.3513	0.6543	0.7119	0.537
21	0.3013	0.7543	0.7405	0.5701
22	0.3277	0.7116	0.7115	0.5506
23	0.4559	0.6158	0.5851	0.416
24	0.2558	0.7204	0.8147	0.6389
25	0.2525	0.7164	0.8019	0.6154
26	0.2364	0.8277	0.7501	0.6509
27	0.4626	0.6153	0.5691	0.4045
28	0.4662	0.6032	0.59	0.4171
29	0.2574	0.6858	0.869	0.6116
30	0.3635	0.6564	0.691	0.4907
31	0.3396	0.7197	0.6786	0.5339
32	0.3832	0.6362	0.7192	0.4747
33	0.3786	0.6799	0.6546	0.4735
34	0.3648	0.695	0.7322	0.4949
35	0.6233	0.4764	0.4966	0.2741
36	0.3299	0.6412	0.7686	0.5244
37	0.701	0.5311	0.3051	0.203
38	0.2936	0.7216	0.6966	0.5724
39	0.6053	0.6035	0.4339	0.2933
40	0.3709	0.7238	0.6893	0.4756
41	0.3156	0.7147	0.7049	0.5656
42	0.3486	0.8177	0.6647	0.5258
43	0.1458	0.8387	0.8776	0.7581
44	0.3091	0.7872	0.6646	0.5585
45	0.3052	0.7125	0.7126	0.5485
46	0.4514	0.768	0.585	0.4203
47	0.858	0.2647	0.1658	0.0875
48	0.3351	0.6805	0.7165	0.5146
49	0.3955	0.6167	0.6474	0.4691
50	0.3567	0.7601	0.5775	0.5083
All	0.3744	0.6905	0.66	0.5024

Part by part:

Part	Error	Precision	Recall	IoU
Background	0.2679	0.6955	0.8282	0.6119
Hair/Eyebrows	0.276	0.8537	0.6595	0.6159
Mouth	0.5518	0.746	0.3989	0.3115
Eyes	0.3317	0.5896	0.8846	0.5234
Nose	0.6556	0.4424	0.3237	0.2247
Skin	0.1633	0.8159	0.8652	0.727
All	0.0449	0.0829	0.0792	0.0603

A grid of the visual results on the test set:

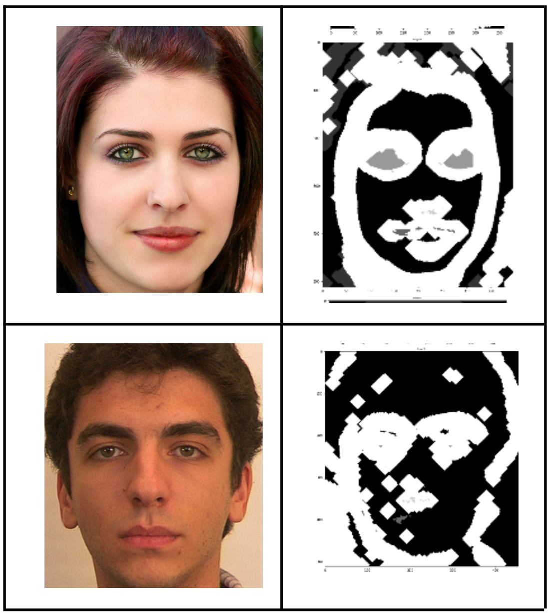
Analysis

It is observable that key parts of the face are at least located, even though they are not properly accentuated via segments. On the first inspection, the segmentation algorithm has managed to locate at least the generic regions. On most faces, the segmentation is clear besides random noise.

Two particularly difficult parts of the face to segment are the nose and mouth. The accent of the nose is very mild and does not particularly stand out in any colour space. Defining a region of interest did help improve the results slightly, but nevertheless, it is still a challenge. The mouth also suffers from the "*one size does not fit all*" problem with fixed masks. While it can be detected, it is difficult to decide how much dilation is required to properly emphasise it.

For the eyes, the algorithm does a good job at locating them but often misses out on the eyebrows. This is because the blacking out approach described earlier that estimates a region of interest for the eyes; in some circumstances, the eyebrows could accidentally be cropped out.

Additionally, the test data was generous in the sense that most faces were taken in the same environment, with similar lighting and shades, and most faces were Caucasian white. The skin mask operates in the HSV colour space and a problem with using a colour range means that applying this mask on a coloured person will not work. In the few faces, where there was a sort of yellow lighting or shadow or differently coloured hair, the skin mask completely fails; it treats the skin as part of the background.



The skin shades and colour tones of the above two photos are different from the rest. In the bottom picture, the man's background and skin is of the orange tone, resulting in both his skin and the background being marked as the background.

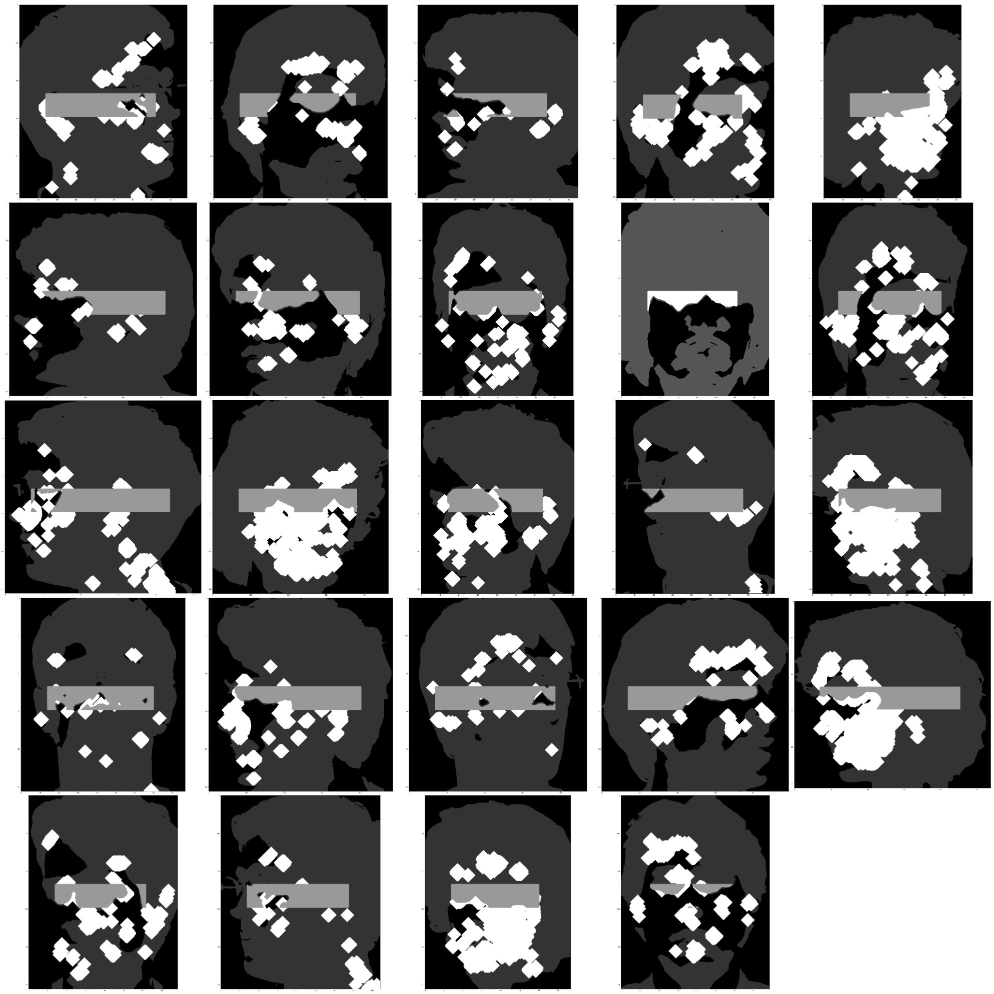
Additional Dataset

The additional dataset consisted of the side profile of faces and showed the shortcomings of the approaches taken in the original data.

An average precision of 0.3723 is attained:

All	0.6613	0.3723	0.459	0.2652

Visual results:



Analysis

Most notably, the region of interest definition failed. A specific region was defined for the eyes, nose and mouth, betting on the pattern displayed by the earlier images. When the face was turned, the nose and mouth were no longer in their expected locations. A grey rectangle (the region of interest for the eyes) is very obvious and the nose is almost not detected at all because it would traditionally appear at the centre of the face. The hair is often detected but fills large portions of the face that might have had the same colour range due to shadows, lighting or position.

The detection of skin was still effective in most cases, though there might have been insufficient or excessive dilation on some photos, resulting in that cubic outlook. Further dilation can fix this, at the risk of worsening the results on the previous set where skin detection is already sufficient. We can contend from here that there is no perfect number of erosions or dilations that will work for all faces and angles.

Suggestions for Improvement

The solution at this point is obviously not as robust. Manually defining regions of interest is not a scalable solution and this can possibly be replaced by a clustering algorithm that can group together pixels on the face. Clustering is not without shortcomings, but it should work for more face angles and shades. Felzenszwalb segmentation is an option.

Furthermore, using a colour range to extract more subtle features like the nose is insufficient. The algorithm can perhaps incorporate spatial elements and texture to understand the concavity of the region. This is in contrast to a nose mask that currently imprints random noise and is the most complex to detect.

Collaboration

Strategy and techniques were discussed with Vickey Tan, Pritesh Patel and Kan Eugene.