DIP

PROJECT PROPOSAL

Team Name - CONNECTED COMPONENTS

Project ID - 5

Github - https://github.com/Digital-Image-Processing-IIITH/dip-m22-project-connected-components

Team Members-

1)Chirag Parikh - 2022900005

2)Kethu Sesha Sarath Reddy - 2020102028

3)Aravapalli Dheeraj Murugan Sai - 2020102020

4)Neeraj Sai Ramana Veerla - 2021121008

Research paper link- http://people.csail.mit.edu/yichangshih/portrait_web/2014_portrait.pdf

1)Main goal of the project -

- Style transfer of head portrait (robustly transfer the local statistics) , by using a reference image while matching properties such as the local contrast and the overall lighting direction while being tolerant to the unavoidable differences between the faces of two different people.This can allow one to easily reproduce the look of renowned artists.
- Additionally, because artists sometimes produce entire headshot collections in a common style, we show how to automatically find a good example to use as a

---

reference for a given portrait, enabling style transfer without the user having to search for a suitable example for each input

2) Problem definition (What is the problem? How will things be done ?)

- **What is the problem? (Motivation):**
  - When we look at the photos of celebrities taken by professionals under perfect conditions ,we wish our head portraits to look similar to them in style, even professional photographers find it different to take many photos of a similar style.
  - Given an input unprocessed headshot and a model headshot by an artist, we describe an automatic algorithm to transfer the visual style of the model onto the input.

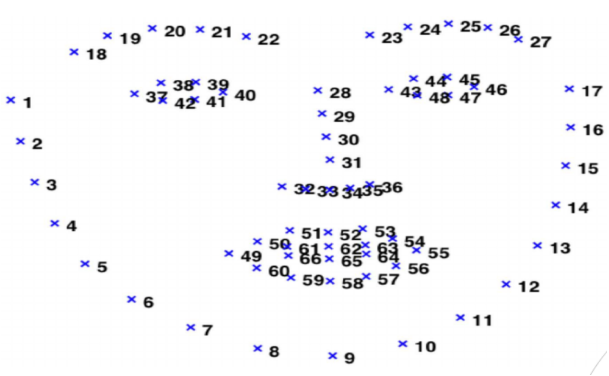
- **How will things be done ? (Process):**
  - We introduce a multiscale technique to transfer the local statistics of an image. We explain how to focus the transfer on a region of interest and how to cope with outliers.

#Part-1

Dense Correspondence of Example to Input (Warp):

1. **Template based facial landmarks localization:**
   We find 66 facial landmarks for a head image using an algorithm from the paper below.

   SARAGIH, J. M., LUCEY, S., AND COHN, J. F. 2009. Face alignment through subspace constrained mean-shifts. In *IEEE Conference on Computer Vision*, 1034–1041



2. **Affine transform based alignment of eyes and mouth:**
   We roughly align the eyes and mouth of the example with those of the input using an affine transform using an algorithm from the paper below.

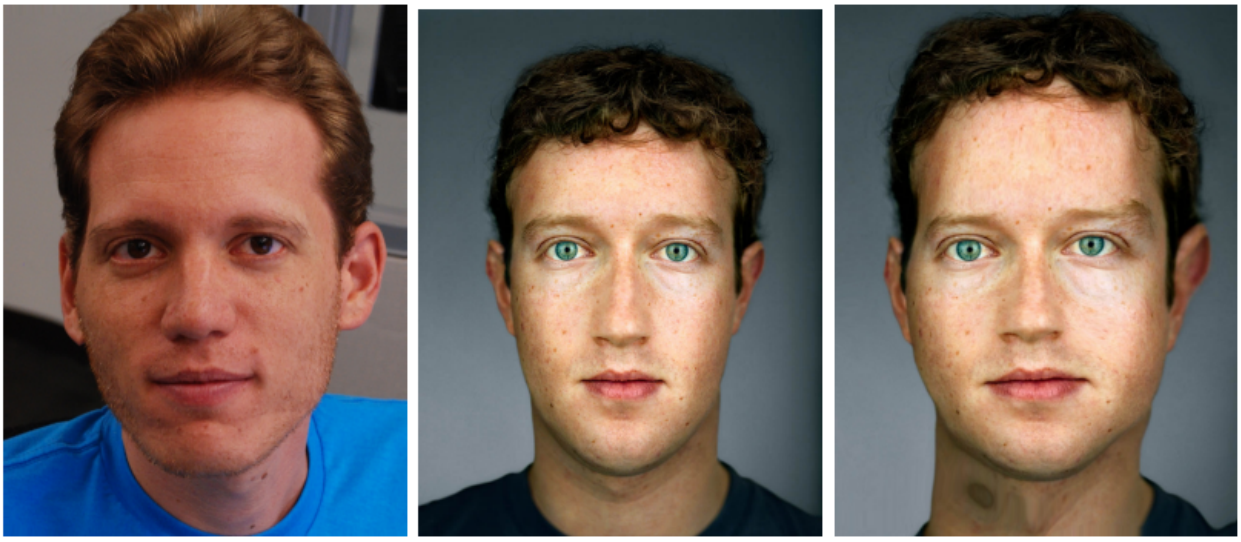
   JOSHI, N., MATUSIK, W., ADELSON, E. H., AND KRIEGMAN, D. J. 2010. Personal photo enhancement using example images. *ACM Transaction on Graphics (TOG) 29*, 2, 1–15.

3. **Face morphing using the segments on the face template:**
   We morph the example to the input using the segments on the face template using an algorithm from the paper below.

   BEIER, T., AND NEELY, S. 1992. Feature-based image metamor-phosis. In *ACM Trans. Graphics*, vol. 26.

4. **SIFT Flow :**
   Refine the correspondence using SIFT Flow using an algorithm from the paper below.

   LIU, C., YUEN, J., AND TORRALBA, A. 2011. Sift flow: Dense correspondence across scenes and its applications. *IEEE Trans-actions on Pattern Analysis and Machine Intelligence 33*, 5



(a) *Input* **(b)** *Example* **(c)** *Compute the dense correspondence between (a) and (b)*

#Part-2

Multiscale Transfer of Local Contrast-

1. Multi-scale decomposition (Laplacian) :

Decompose the input and example images into multiscale Laplacian stacks.

$L_\ell [I] = I - I \otimes G(2)$ if $\ell = 0$

$L_\ell [I] = I \otimes G(2\ell) - I \otimes G(2\ell+1)$ if $\ell > 0$

Example.

Input

2. Local Energy (Power) :

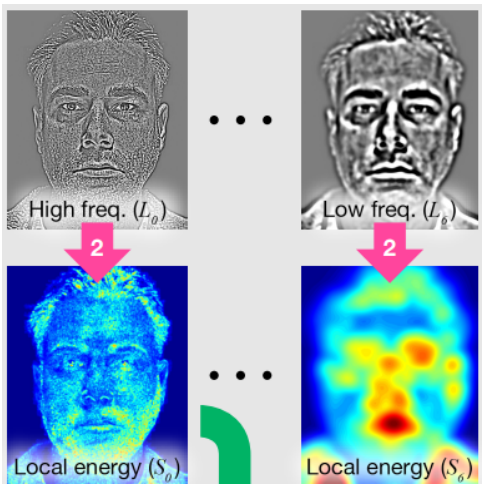
We estimate the local energy S in each subband by the local average of the square of subband coefficients.

$S_\ell [I] = L\,2\ell [I] \otimes G(2\ell+1)$

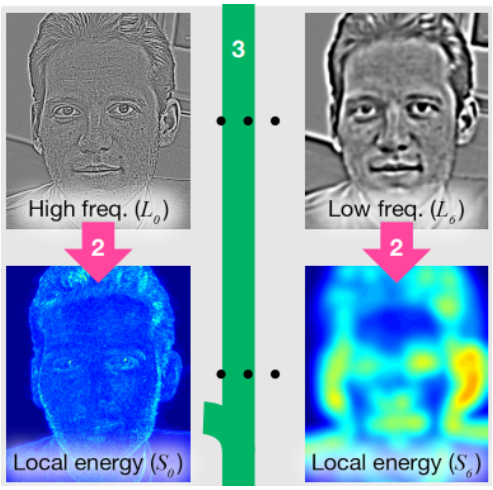
Now, we use the dense correspondence(wrap) on local energy.

$$\tilde{S}[E] = W(S_\ell[E])$$

Example



Input



3. Gain map (with robust impl.):

We modify the input subbands so that they get the same energy distribution as the example subbands.

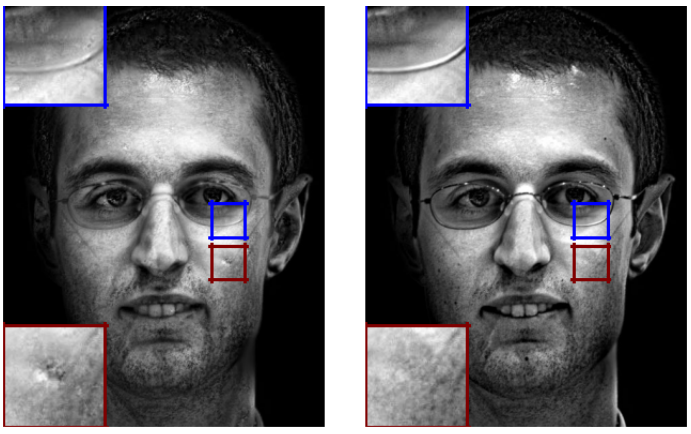
$$L\ell[O] = L\ell[I] \times \text{Gain}$$

$$\text{with Gain} = \sqrt{S\ell[E]/S\ell[I] + \epsilon}$$

It can introduce artifacts (moles or glasses in reference image) , to avoid it we robust implementation.

$$\text{RobustGain} = \max(\min(\text{Gain}, \theta h), \theta l) \otimes G(\beta 2\ell)$$

After this step we get style transfer in grayscale.



4. Dealing with colors:

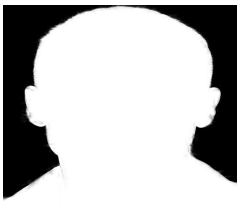
We work in the CIE-Lab color space as it is more confirming to human perception, and process each channel independently using the algorithm that we just described.

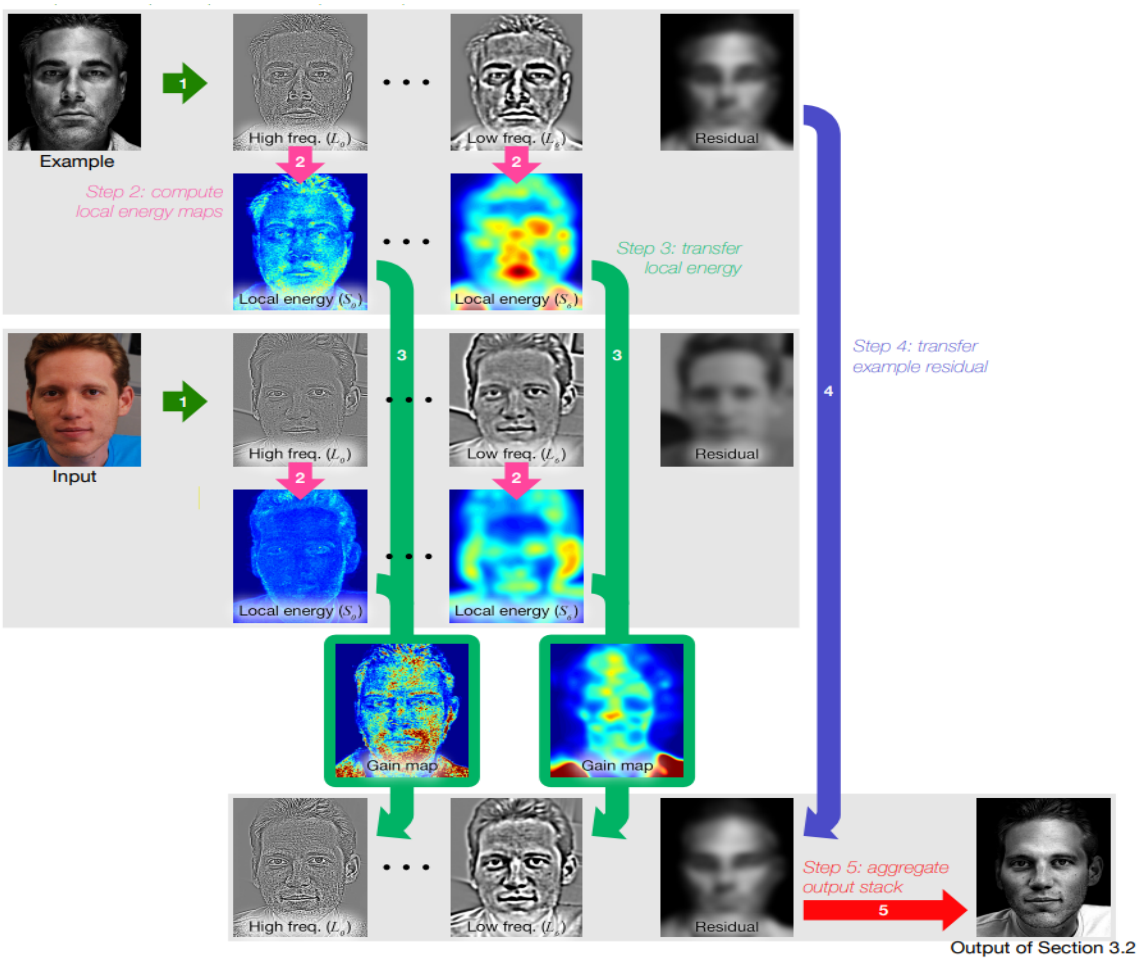
5. Using a face mask (GrabCut + Matting Laplacian) to preserve shape-

We truncate the Gaussian convolutions so that they only consider values within the mask.

- we run GrabCut [Rother et al. 2004] initialized with a face detection result to find a binary mask.
- we refine using the Matting Laplacian

$$Image \otimes G = (Image \times Mask) \otimes G \div Mask \otimes G$$





#Part-3

6. Additional Postprocessing:

Eye Highlights:

- **Locate the iris using circular arc detection:**
  We separate the specular reflection from the example eyeball and copy that onto the input's eyes by locating the iris using circular arc detection around the position using an algorithm from the paper below.

  DAUGMAN, J. G. 1993. High confidence visual recognition of persons by a test of statistical independence. IEEE Trans. Pattern Analysis and Machine Intelligence 15, 11, 1148–1161

- **K - means:**
  Then, we create an approximate segmentation into iris, highlight, and pupil by running a k-means algorithm on the pixel colors with k = 3.

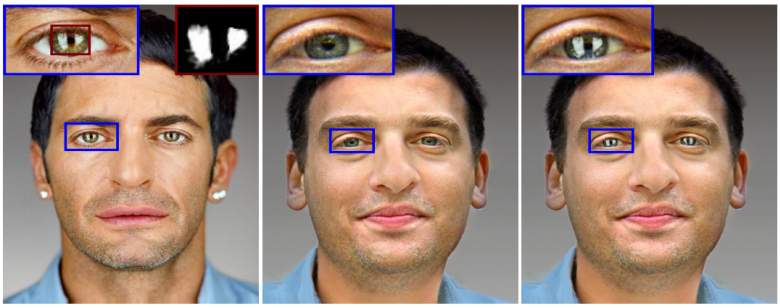
- **Alpha matting:**
  We refine the reflection mask using alpha matting from the paper below.

  LEVIN, A., LISCHINSKI, D., AND WEISS, Y. 2008. A closed-form solution to natural image matting. IEEE Trans. Pattern Analysis and Machine Intelligence 30, 2, 228–242

- **Existing highlights extraction via thresholding on L channel:**
  To find the highlighted regions of iris, we use a threshold of 60 on the L channel of CIE-Lab colorspace.

- **Replacement by inpainting:**
  Then, we erase the detected pixels and fill in the hole using inpainting. We will use the equivalent of griddata Matlab function.

- **Composing example highlights on input's eyes (with apt translation and scale adjustments):**
  We center them using the pupils as reference, and scale them in proportion to the iris radii for composing example highlights on top of input eyes.



(a) *Input* **(b)** *Without eye highlights* **(c)** *Adding eye highlights*

Background:

- **Inverted face mask based background replacement (inpainting if required to fill any holes):**
  We use the previously computed masks to extract the example background and replace the input background with it. If needed, we extrapolate the missing data using inpainting with equivalent griddata Matlab function.

Input

Eye highlights and Background

#Part-4

7. **Automatic Selection of the Example in a Collection:**

- It is an automatic algorithm to select a suitable example among a collection of consistently stylized headshots.
- We concatenate S_ℓ over all scales to get the feature vector representing a face image, and use the normalized cross correlation between the two feature vectors as the similarity function (paper claim it is better than L2).

(Employ warping also for more robust selection)

#Part-5 (If time permits)

8. **Extension to Video:**

- First frame based automatic example selection by finding the best candidate frame which matches the selected example.
- Optical flow based estimation of correspondence between the best candidate frame and the rest of the frames in video.
- Style transfer of the best candidate frame and propagating the transfer to rest of the frames using the computed correspondances (temporal coherence).

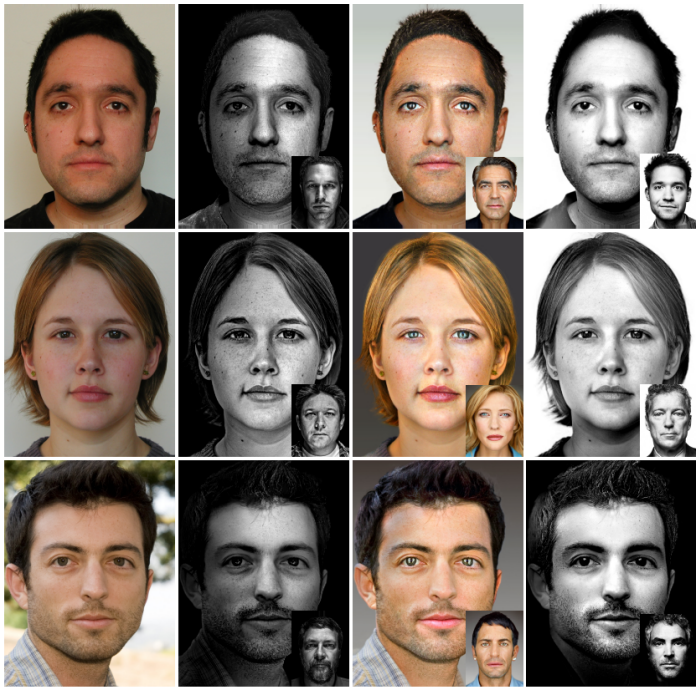
#Part-0

9. **Comparison with the state-of-the-art works and the implemented work itself. (Use implemented codes)**

We will compare our algorithm with the implemented code in matlab in each step and see the performance in speed and final output.

3) Results of the project (What will be done? What is the expected final result ?)

We selected different styles like, (black-and-white and colors, low key (i.e., dark) and high key (i.e., bright), soft and detailed). The algorithm should successfully transfer the style onto input photos using the reference image .



- **Result using with and without automated matching reference** image from dataset.

- Sometimes ,our local statistic matching does not reproduce the example global dynamic range,so to balance the local details and global range we can perform it by averaging the **local statistic matching result with and without histogram transfer** applied as a post-process.

4)What are the project milestones and expected timeline ?

Before Mid-Evaluation:

- **October 25 - November 2 : Dense Correspondence**
- **November 2 - November 10 : Multiscale Decomposition (grayscale, without face mask)**

After Mid-Evaluation:

- **November 10 - November 20 : Eye highlight, color image, automatic reference face matching.**
- **If time permits Optional: video style transfer**

5) Is there a dataset you need ? How do you plan to get it ?

The dataset data taken in a controlled environment as well as on a large set of photos downloaded from the Internet.

To Download Dataset and result from paper :

people.csail.mit.edu/yichangshih/portrait_web/

Uploaded the same to OneDrive from easy access :
https://1drv.ms/u/s!AufWkcWWKM7kgdRy1fkHVDrJNyfFTw?e=WOiCdp