Austin Hester  
CS 6420 – Image Processing  
UMSL, SS2021  
Prof. Sanjiv Bhatia  
Survey Paper

**Artistic Style Transfer**

Computer-assisted artistic style transfer gives animators a tool to produce rotoscope-like videos with relative ease. Automatic artistic style transfer can transform a simple home movie into something straight out of the artist's imagination, and they must only provide one example style frame, provided the subject does not change dramatically. Artistic style transfer softwares allow ideas to flow more freely in a short-film medium by providing the tools to apply an artistic style and background to the entire frame sequence, given only a single example frame [6 Fišer]. Attractive constraints of the style transfer can range from a total makeover of the entire frame to simply an addition of mustaches to every person in the picture. The focus of style transfer shifts to facial animations, such as the "filters" seen on many popular social media apps. The use of technology for fun became more commonplace in 2020, and tools to help people utilize their imagination are increasingly valuable. <https://github.com/ahester57/style_transfer>

**1. History**

In the late 19th century, Eadweard Muybridge’s horse sequences were painted on glass discs. These discs were used in a special projector called a zoopraxiscope, which intakes a glass disc. Light is projected onto these discs through slits and projected via a series of mirrors and lenses.

Also in the late 19th century, Claymation appears. Claymation provides an artist to provide an animated scene with accurate lighting and shading. The artist also has a more concrete scene in which to let his imagination run. The amount of time and precision Claymation requires makes this a rarely chosen medium.

In 1915, Max Fleischer invents Rotoscoping, known back then as the “Fleishcher Process,” which inspires many future animators. Rotoscoping makes animation efficient by providing an example frame from a live recording to trace onto a frame of the animated video. Rotoscoping provides more natural motion since movement is traced from a real person. Fleischer's animated series Out of the Inkwell (1918-1927) starring Koko the Clown showcases the powers of rotoscoping by conveying animated imitations of popular performing artists of the time. In 1934, Fleischer’s patent expires. Other studios begin using rotoscoping and a flood of animated works hit the silver screen.

**2. Challenges**

With a focus on transferring an example style frame onto a video of a person's face.

**2.1 Preserve Artistic Style of Template Image**

First and foremost, the goal is to apply the color, texture, contrast, and style of the example frame to a normal frame sequence of a target's facial animation.

* Preserve artistic style's fine details.
* Preserve artistic style's mood, which can be difficult to for an algorithm to capture.
* Preserve artistic style's textures.

**2.2 Preserve Identity of Target Image**

The target's identity must be preserved to an extent. Given an example frame as the Mona Lisa, and a target video of Bill Gate, the output video should still be recognizable as Bill Gates.

* Preserve the identity of the target.
* Preserve attributes such as the shape, size, and position of facial features.
* Preserve eye normalcy.

**2.3 Video Only: Maintain Temporal Cohesion**

The output video should maintain some temporal cohesion, so the target's features do not dramatically change from one frame to the next. Temporal cohesion is when an object maintains its identity from frame-to-frame. Optical Flow, offered by OpenCV, aims to identify and track objects throughout a sequence of frames. However, in *Color Me Noisy* [4 Fišer], reason to stay below full temporal cohesion gives the output a sought-after hand-drawn animation style in digital renderings.

**3. My Attempt**

My attempt is nothing to write home about. It is capable of averaging quadrants of two images together, up to eight recursive levels of splitting the images into quadrants. With three levels of quadrant depth, a 500x500 pixel image takes under 1 second using the quadrant-mean method on my i5-3570k @ 4.1 GHz. F

Both methods provide parameters for applying weights for each HSV channel of the style and target images. Both also allow a *quadrant depth* parameter, but this is defective in blend mode. The effect of quadrant depth on mean mode can be found in "doc/MEAN\_RESULTS.md" in the project repository. Similar, more disturbing results can be found for blend mode in "doc/BLEND\_RESULTS.md".

Pre-processing each image can be summarized as:

1. Apply GuassianBlur.
2. Convert to HSV.
3. [optional] Split into quadrants.

Next, the process of style transfer has two modes: *mean* and *blend*. The following shows the process for both, with a placeholder for each quadrant:

1. Register the template image to the target image.
2. Split HSV into planes.
3. For each quadrant:
   1. Extract ROIs for hue, saturation, and value of template and target.
   2. *mean* or *blend*.
4. Merge output planes together.
5. Convert back to RGB.

For *mean* mode:

1. Compute mean of template's and target's quadrant.
2. Plug in each mean, independently weighed, into output image planes via system of equations.

For *blend* mode:

1. Resize template quadrant to target quadrant.
2. Blend each channel, independently weighed, into output image planes via system of equations.

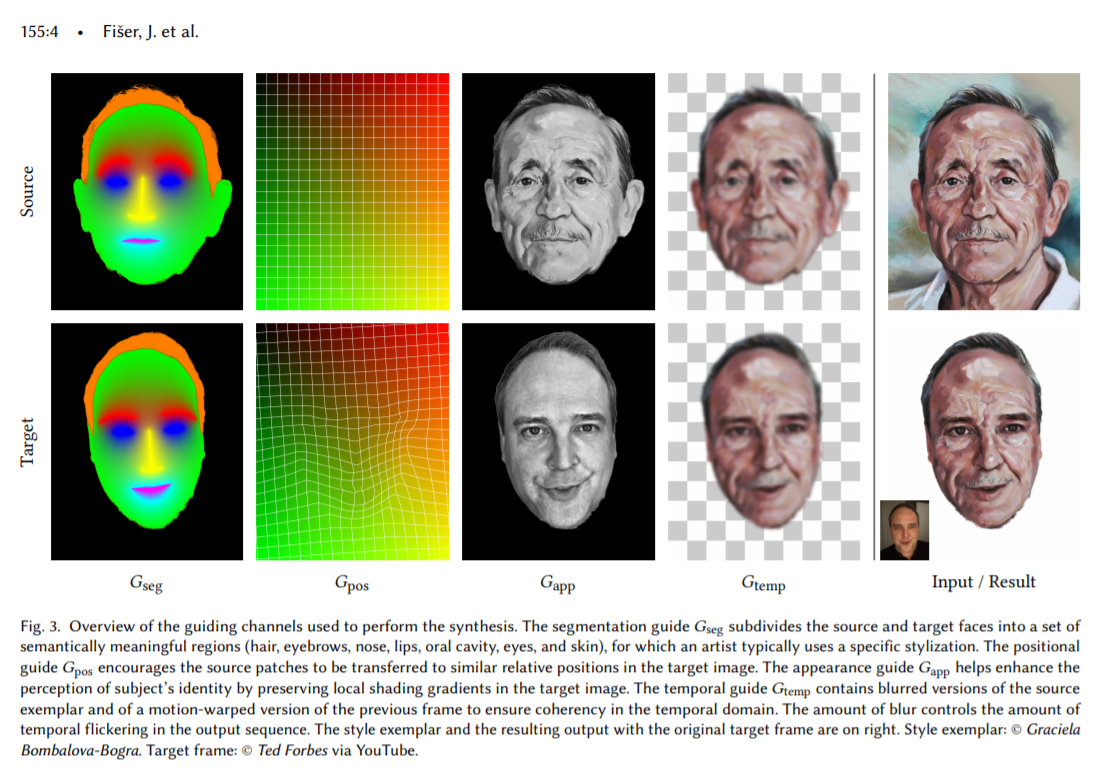
There are many things I could have done to make this better. Nevertheless, I learned a great deal and built a project with good code reusability so I can pursue ideas more easily in the future.

I have yet to attempt style transfer of video, but the following section outlines a refined method of example-based style transfer of video while maintaining the artistic style of the template, the identity of the target image, and temporal cohesion throughout the frame sequence [5 Fišer].

**4. Example-based Style Transfer**

Why not neural net? Topic. Bullshit. Facts. More facts. bullshit. facts. topic again. reason topic is right.

other bullshit. facts to support the other bullshit. topic 2 because of other bullshit.

[5 Fišer]

FaceBlit is a technology that aims to provide real-time facial style transfer, running on a smartphone CPU.

In order to make real-time work, they use a 3D lookup table for values of Red, Green, and Gray.

Uses the same positional guides as Example-Based Synthesis…, but generates these guides:

Using a linear approximation in order to support real-time results. [7 Texler]

**5. Usages**

Definitely this section should be about 80% bullshit.

**6. Conclusion**

The use of technology for fun became more commonplace in 2020, and tools to help people utilize their imagination are increasingly valuable.

**References**

[1] Barnes, C et al. PatchMatch: A randomized correspondence algorithm for structural image editing. *(ACM Transactions on Graphics 28, 3 2009)*.

[2] Cavalier, Stephen. *The World History of Animation*. ISBN:9780520261129. (*Aurum 2011*). <https://www.ucpress.edu/book/9780520261129/the-world-history-of-animation>

[3] Futschik et al., STALP: Style Transfer with Auxiliary Limited Pairing (*Eurographics May 2021*). <https://dcgi.fel.cvut.cz/home/sykorad/Futschik21-EG.pdf>

[4] Fišer, J et al., Color Me Noisy: Example-based Rendering of Handcolored Animations with Temporal Noise Control. *(Computer Graphics Forum 33, 4 2014)*.

[5] Fišer, J et al., Example-Based Synthesis of Stylized Facial Animations (*ACM Transactions on Graphics 2017*). <https://dl.acm.org/doi/10.1145/3072959.3073660> <https://github.com/jamriska/ebsynth>

[6] Fišer, J et al., StyLit: Illumination-Guided Example-Based Stylization of 3D Renderings. *(ACM Transactions on Graphics 35, 4 2016)*

[7] Texler, A et al. FaceBlit: Instant Real-time Example-based Style Transfer to Facial Videos (*Proceedings of the ACM on CGIT April 2021*). <https://dl.acm.org/doi/10.1145/3451270> <https://github.com/AnetaTexler/FaceBlit>