High-Quality AR Lipstick Simulation via Image Filtering Techniques

Kanstantsin Sokal Siarhei Kazakou Igor Kibalchich Matsvei Zhdanovich
Google Research
1600 Amphitheatre Pkwy, Mountain View, 94043 CA, USA
{kanstantsin, kazakou, kibalchich, matvey}@google.com

Figure 1. Visual results of AR lipstick material simulation.

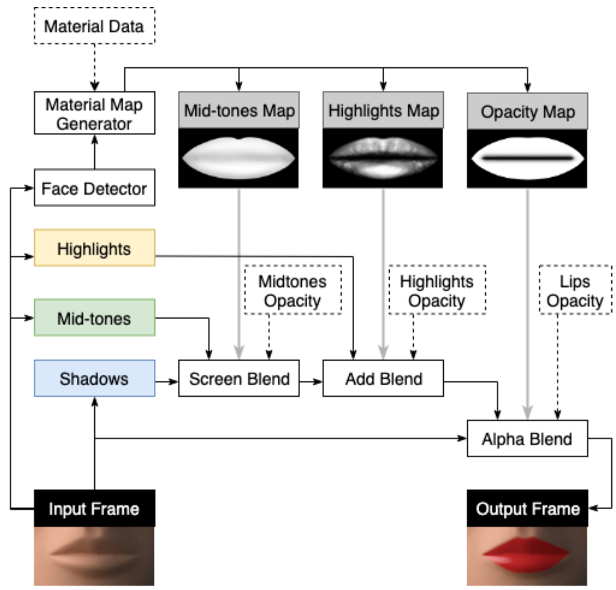

Figure 2. High-level AR lipstick material simulation pipeline. See text for details.

Abstract

This paper presents a new AR lipstick simulation technique that produces high-quality visual results. It is designed with a strong emphasis on simplicity and the use of existing image filtering techniques. Given its unique nature, the proposed technique can be implemented by both software engineers and visual designers. Our approach is robust across various skin tones and lighting conditions. It reaches real-time performance on modern mid- and high-end smartphones in 720p resolution.

1. Introduction

Recently, multiple AR makeup material simulation techniques have been proposed (including [1], [2], [4] and [5]) balancing visual result quality with performance on challenging hardware, such as mobile devices.

This paper proposes a new technique that produces high-quality visual results and is designed with a strong emphasis on simplicity and the use of existing image filtering techniques. It combines well-known fundamental image filters into a single pipeline to perform AR material property transfer (such as base color, glossiness/matteness and others) onto images of people.

Given its unique nature, the proposed technique can be implemented by both software engineers and visual designers by combining high-level blocks provided by popular image processing libraries and graphics editors. Being a composite of well-productionized high-level blocks, our approach has a clear benefit for visual designers since they can create and iterate on AR lipstick materials with tools they already have expertise with.

Our approach is robust across various of skin tones and lighting conditions (as demonstrated in Figure 1). It also maps well to the GPU programming model, which results in real-time performance on modern mid- and high-end smartphones in 720p resolution.

2. Proposed Solution

A high-level diagram of the proposed pipeline is shown in Figure 2. The dashed rectangles in this diagram (as well as in any other diagram in this paper) refer to AR material parameters, which together form the AR material definition.

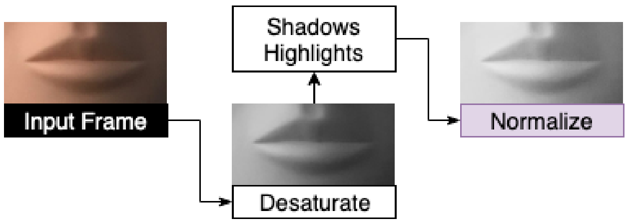

Figure 3. Intensity normalization pipeline.

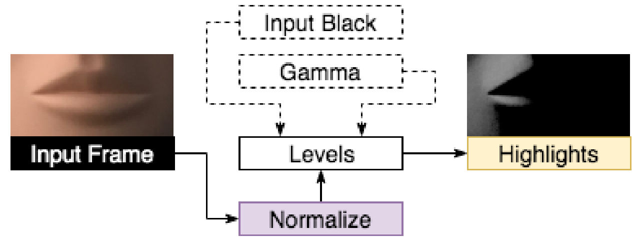

Figure 5. Highlights filtering pipeline.

Any image filter parameter, which is neither shown on a diagram nor mentioned in the text, should be assumed to have its neutral value.

Similar to previous work, our intuition is based on dividing the light spectrum into multiple ranges, filtering each of the individual ranges and then merging them before blending with the original image. After each individual range is filtered, they are combined together using various blending modes, as in [7]. Following this basic intuition, we've performed an extensive experimentation process in order to converge on the pipeline presented in this paper.

We consider three light spectrum ranges: shadows, midtones and highlights. Face detection and landmarking technology is used to identify high-fidelity facial semantic regions for lips. Those are used to generate AR material maps by warping canonical static material maps on top of the lip 3D surfaces detected on the input frame. AR material maps are used during the final blending operations for masking those areas, which should be augmented by various AR material simulation pipeline stages.

All computations are being performed in the linear RGB color space. There are global opacity values defined for each blending operation; those provide a easier way to tune the pipeline for a specific AR material without a need to modify material map generation process.

2.1. Intensity Normalization

A diagram of the intensity normalization pipeline is shown in Figure 3. In order to ensure that the transfer of the AR material base color is robust to variations of skin tone and lighting conditions, we apply a normalization step. It is invariant to above variations while maintaining local features (*e.g.* shadows, tiny cracks on lips and color gradients). Specifically for normalization, we first de-saturate the input frame and then apply the "shadows/highlights" image filter, as in [6], with carefully tuned handpicked parameters. As a result, no matter how dark or bright the input lip color was, it should approximately reach a unified well-lighted look after applying our normalization technique.

Although this technique performs good color intensity normalization, it comes the a cost of changing the color intensity distribution. This leads to a loss of visual depth as diminished gradients make the resulting picture look flat. For this reason, the normalized color intensities are only used to filter shadows and highlights; in contrast, mid-tones are based on the original color intensities in order to preserve "visual volume". Carefully balancing normalized shadows/highlights and "voluminous" mid-tones at the final blending stages helps to deliver robust recoloring while keeping the original color intensity distribution perceptually unaffected.

2.2. Shadow, Highlight and Mid-Tone Filtering

Diagrams of our shadow, highlight and mid-tone filtering pipelines are shown on Figures 4, 5 and 6 respectively.

Our main instrument for selecting and re-mapping image intensity ranges is the levels adjustment image filter, which is commonly used in graphics editors. It performs a combination of three basic intensity transformations, as described in [3]: contrast stretching from the $[B_{input}, W_{input}]$ range to the $[0, 1]$ range (B and W stand for "black" and "white"

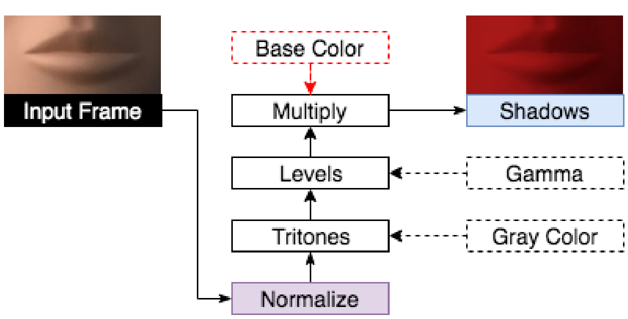

Figure 4. Shadows filtering pipeline. See text for details.

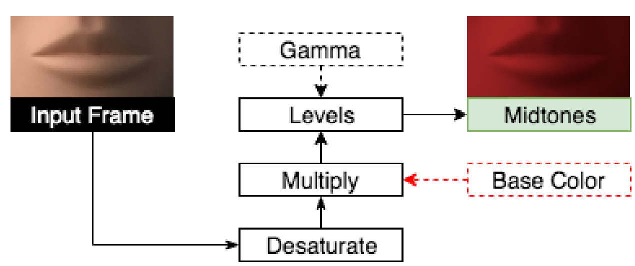

Figure 6. Mid-tones filtering pipeline.

Figure 7. Visual result quality evaluation examples. Note how our AR material (even rows) faithfully replicates real material properties (odd rows) across a variety of skin tones and lighting conditions.

respectively); gamma transformation with fixed $c = 1$ and variable γ; contrast stretching from the $[0, 1]$ range to the $[B_{output}, W_{output}]$.

As mentioned in the previous section, filtered shadows are based on normalized color intensities. After applying our normalization technique, we apply the piecewise-linear intensity transformation image filter, as in [3], to reproduce the tritones image filter, which is commonly used in graphics editors. In order to perform preliminary colorization to the normalized color intensities, we move the gray point (*i.e.* linear RGB color space midpoint) from $(0.5, 0.5, 0.5)$ to some other point defined as an AR material parameter.

For more details on each individual filtering stage, please refer to the respective diagrams.

3. Results

To evaluate the visual result quality, we performed a qualitative comparison of the material simulation results and ground truth data on various skin tones and lighting conditions (examples are shown on Figure 7). As can be seen, the approach is able to realistically and robustly augment images with AR materials.

Additionally, we conducted a user survey with over 80 participants to investigate whether people are able to distinguish between images of real lipstick and images of lips that have been altered with our AR lipstick simulation. We selected 5 real images of lipstick and 5 images augmented with our AR lipstick simulation. We then asked participants to evaluate each image and to answer whether they believe the image has been altered with the AR technique or not. Aggregated results are shown on Table 1. The results show, that people more often correctly detect real images comparing to detecting AR images as real (62.38% versus 46.43%); however, our participants struggled to correctly identify altered images in those 46.43% cases, which means that our AR simulation technology has reached high level of fidelity and seems as good as real to many people.

Actual \ Predicted	Real	AR
Real	62.38%	37.72%
AR	46.43%	53.57%

Table 1. Aggregated user survey results.

Device	Avg Time Per Frame
Google Pixel 3 (2018)	4 ms
Huawei Mate 9 (2017)	7.5 ms
iPhone 7 (2016)	8.2 ms

Table 2. Time measurement benchmark results, performance is real-time on a variety of mid- to high-end devices.

Our implementation of this pipeline runs on the two most popular mobile platforms (Android and iOS) and targets the OpenGL ES 3.0 graphics API for improved performance on GPU. Time measurement benchmark results are shown on Table 2. Time performance of the pipeline has been measured by averaging time over a large batch of input frames being processed without any intermediate CPU-GPU synchronization. The input frame size is 720x1280, which is commonly used in image processing pipelines on mobile.

4. Application

One possible application of our technology is in creating engaging AR experiences for online advertising platforms. Together with internal and external collaborators, we piloted a series of advertisement campaigns on a popular video-sharing platform that featured an AR lipstick try-on experience hosted on a popular video-sharing platform. While watching original creator content on beauty and makeup topics, viewers were given the opportunity to virtually ”try-on” AR lipstick, as demonstrated in Figure 8.

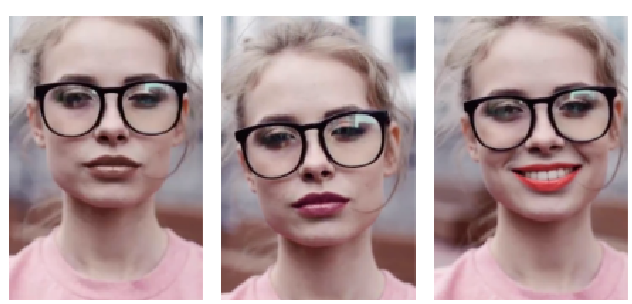

Figure 8. Example of a user virtually ”trying on” AR lipstick.

References

[1] Filipe Campos and Carlos Morimoto. Virtual makeup: Foundation, eye shadow and lipstick simulation. In *2014 XVI Symposium on Virtual and Augmented Reality*, pages 181–189, 05 2014. 1

[2] Bruno Evangelista, Houman Meshkin, Helen Kim, Anaelisa Aburto, Ben Max Rubinstein, and Andrea Ho. Realistic ar makeup over diverse skin tones on mobile. In *SIGGRAPH Asia 2018 Posters*, SA '18, pages 81:1–81:2, New York, NY, USA, 2018. ACM. 1

[3] Rafael C. Gonzalez and Richard E. Woods. *Digital Image Processing (3rd Edition)*. Prentice-Hall, Inc., Upper Saddle River, NJ, USA, 2006. 2, 3

[4] Jeong-Sik KIM and Soo-Mi CHOI. Interactive cosmetic makeup of a 3d point-based face model. *IEICE Transactions on Information and Systems*, E91.D(6):1673–1680, 2008. 1

[5] Chen Li, Kun Zhou, and Stephen Lin. Simulating makeup through physics-based manipulation of intrinsic image layers. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4621–4629, 2015. 1

[6] Ulrich Pegelow. Shadow recovery revisited, Feb 2012. 2

[7] Thomas Porter and Tom Duff. Compositing digital images. *SIGGRAPH Comput. Graph.*, 18(3):253–259, Jan. 1984. 2