In-camera, Photorealistic Style Transfer for On-set Automatic Grading

Itziar Zabaleta

> Universitat Pompeu Fabra

Marcelo Bertalmío

> Universitat Pompeu Fabra

Written for presentation at the

SMPTE 2018 Annual Technical Conference & Exhibition

Abstract. In professional cinema, the intended artistic look of the movie informs the creation of a static 3D LUT that is applied on set, where further manual modifications to the image appearance are registered as 10-parameter transforms in a color decision list (CDL). The original RAW footage and its corresponding LUT and CDL are passed-on to the post-production stage where the fine-tuning of the final look is performed during color grading.

In many cases, the director wants to emulate the style and look present in a reference image, e.g. a still from an existing movie, or a photograph, or a painting, or even a frame from a previously shot sequence in the current movie. The manual creation of a LUT and CDL for this purpose may require a significant amount of work from very skilled artists and technicians, while the state of the art in the academic literature offers promising but partial solutions to the photorealistic style transfer problem, with limitations regarding artifacts, speed and manual interaction.

In this paper, we propose a method that automatically transfers the style, in terms of luminance, color palette and contrast, from a reference image to the source raw footage. It consists of three separable operations: global luminance matching, global color transfer and local contrast matching. As it just takes into account the statistics of source and reference images, no training is required. The total transform is not static but adapts to the changes in the source footage. The computational complexity of the procedure is extremely low and allows for real-time implementation in-camera, for on-set monitoring. While the method is proposed as a substitute for the need to specify a LUT and a CDL, it's compatible with further refinements performed via LUTs, CDLs and grading, both on-set and in post-production. The results are free from artifacts and provide an excellent approximation to the intended look, bringing savings in pre-production, shooting and post-production time.

Keywords. Style transfer, color transfer, color grading, video.

Introduction

Color is an important part of film as it plays a critical role in how we perceive the film and the characters. It can build harmony or tension within a scene, or bring attention to a key theme. If it is chosen carefully, a well-placed movie color palette evokes mood and sets the tone for the film. Ideally, the intended color interaction would be conceived before-hand, then recorded faithfully, and enhanced carefully by the colorist under strict supervision. In the past color grading was a photochemical process performed at a photographic laboratory, but nowadays is generally performed digitally in a color suite. Usually, in professional cinema, 3D LUTs are created by the director of photography (DoP) to set a look as well as to endow images with specific looks. Later on, these 3D LUTs along with the RAW footage are passed to the post-production stage, where the LUTs are used as a starting point for the final color grading of the movie. This whole process is very costly in terms of budget and time and it may require a significant amount of work from very skilled artists and technicians.

We propose a method that can ease the workload of cinematographers and photographers during post-production. In many cases, the director wants to emulate the style and look present in a reference image, e.g. a still from an existing movie, a photograph, or even a previously shot sequence in the current movie. Given a reference image, our approach automatically transfers the style, in terms of tone, color palette and contrast to the source RAW footage. It is a low computational cost process that can be implemented in-camera, so light and color can be adjusted on-set while seeing the resulting image on the screen. While the method is proposed as a substitute for some of the post-production tasks, it is compatible with further refinements, both on-set and in post-production. The method is applied directly on the RAW images and it generates a display-ready result that matches the style of the reference image.

This paper makes the following contributions:

- We introduce a simple but effective method for style transfer from a still image to a video. It consists of different low-computational cost operations: an exponential transformation, PCA analysis, a matrix multiplication and a convolution. It produces time coherent results and is free of flickering artifacts.
- We understand style as a combination of different attributes of images: luminance, color and contrast. Our approach is divided into three separable and independent steps for each of the mentioned attributes. Our method offers control over the modifications made to the images.

Related work

Color transfer is the process of editing an image so that it matches the colors of a reference. Reinhard and his colleagues [Reinhard] introduced a pioneering method for color transfer. Their method takes advantage of the decorrelation property of the $L\alpha\beta$ color space to transfer the mean and standard deviation between each channel of the two images. The $L\alpha\beta$ color space is constructed so that it decorrelates natural, foliage-dominated hyperspectral images, but it cannot be guaranteed to successfully decorrelate each scene type. To overcome this limitation, principal component analysis (PCA) can be used to find a decorrelated color space for each image [Kotera], [Xiao]. These methods use statistical properties to transfer colors between images. The transformations are applied to each channel of the image separately; therefore

they are simple and computationally low cost methods. However, sometimes, all the subtleties in the color distribution of images are not captured by these per-channel approaches.

By treating the 3D color distribution as a whole the local information is better preserved; some methods [Pitié] consist of reshaping the 3D distribution so it matches the reference. However, this translating process is not straightforward and some strategies have to be used to simplify this complex problem, such as an optimization-based approach or dividing the problem into fewer dimensions [Pitié].

When the goal is to match the colors between images that share some content, correspondences between images can be found in order to ensure that common features end up with the same colors. The method proposed by HaCohen *et. al.* [HaCohen] is designed to match the colors between images of the same scene under different illumination or taken with different camera conditions. First, dense correspondences are found using the "PatchMatch" method, then a global color transformation is applied, consisting in per-channel non-linear tone curves along with a saturation matrix for cross-channel adjustments. In [Vazquez] Vazquez and Bertalmío propose a method to match the colors of images of the same scene under different illuminations or camera settings. Their method consists of three steps: set values for the gamma of each image, find pixel correspondences between images and finally compute a transformation matrix that minimizes the distance between corresponding pixels. These methods produce good results when input images share content, but in the case of very dissimilar pictures, a good performance is not guaranteed.

Recently, there is a line of research that increases in popularity: style transfer using convolutional neural networks [Gatys] [Luan]. These methods were initially designed to transfer the style of an artwork to a photography [Gatys], and they are able to produce good non-realistic results from images with very different content and style, however, they fail in the case of producing photorealistic results free of painting-like distortions. These methods have been adapted to transfer the style between photographs by using a semantic segmentation of the source and reference images [Luan]. Although the results look visually more satisfying compared to the previously mentioned method, they are not completely free of painting-like artifacts.

Bonneel et al. [Bonneel] extended the concept of style transfer to video. They propose a method that estimates a per-frame statistical-based color transform of previously segmented input images and then applies an interpolation to these transformations to prevent artifacts in the video such as bleeding and flickering. Their method works for a wide range of situations but requires some human intervention for segmentation and for parameter adjustment. Another approach that handles image sequences has been proposed by Frigo et al. [Frigo] based on their previous work of style transfer for still images. This method is designed for artistic style transfer, that is for the reference being a painting. It consists of two stages, one for texture transfer using their method called "split and match" and then a color and contrast transfer step.

Style transfer for still images

We start by proposing a style transfer method for still images, that extends our recent work "Photorealistic style transfer for cinema shoots" [Zabaleta]. It consists of a three-step algorithm: global luminance transfer, global color transfer and local contrast transfer.

Luminance transfer. In this step, the demosaiced white balanced RAW image, which is linear RGB data captured by camera sensors, is non-linearly transformed to match the luminance of the reference image.

We use the tone-mapping algorithm proposed by Cyriac *et. al.* [Cyriac], that calculates a variable exponent γ such that when applied to the original RAW image I, it performs a constrained histogram equalization creating a non-linear image I_{eq},

$$I_{eq} = TM(I) = I^{\gamma(I(x))} \tag{1}$$

Let gc be the gamma correction formula explained in [Poynton],

$$gc(x) = \begin{cases} 4.5\,x & 0 \leq x < 0.018 \\ 1.099\,x^{0.45} - 0.099 & 0.018 \leq x \leq 1 \end{cases} \tag{2}$$

where x denotes each channel R, G, and B and $gc(x)$ denotes each gamma corrected channel $gc(R)$, $gc(G)$ and $gc(B)$.
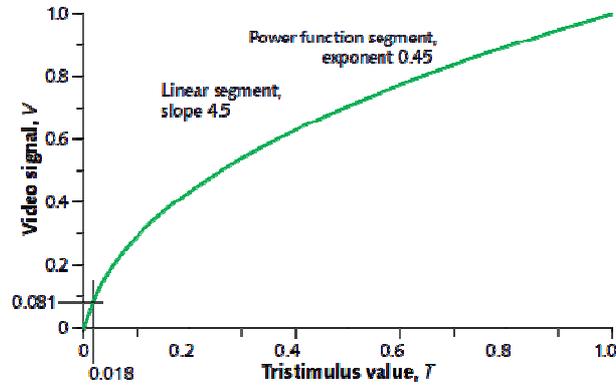

Figure 1. Gamma correction function, linear between 0 and 0.081 intensity values, and exponential between 0.081 and 1. Image from [Poynton].

To recover RGB values proportional to scene tristimulus values, invert the previous equation

$$gc^{-1}(y) = \begin{cases} \dfrac{y}{4.5} & 0 \leq y < 0.081 \\ \left(\dfrac{y+0.099}{1.099}\right)^{\frac{1}{0.45}} & 0.081 \leq y \leq 1 \end{cases} \tag{3}$$

In [Poynton] it is explained that a pure power function with an exponent of 0.45 suffices for gamma correction, however, in order to minimize noise in dark regions, a limitation in the slope of the function near black is necessary.

Let S_0 be the original source (demosaiced and white balanced) RAW image and R the reference image, which is a non-linear image (e.g. a JPEG image). If we assume that the reference image R has been created with the gamma correction in Eq. 2, and taking into account that the tone-mapping method mentioned above performs an approximate histogram equalization we can assume that

$$TM_S(S_0) \simeq TM_R(\,gc^{-1}(R)) \tag{4}$$

so the resulting image S_1 from this luminance matching step will be

$$S_1 = gc\left(TM_R^{-1}\left(TM_S(S_0)\right)\right) \tag{5}$$

Figure 2. From left to right: Demosaiced white balanced source RAW image S_0, luminance matched source image S_1, and reference image R. A non-linear transformation is applied to the source image so its luminance histogram approximates the reference luminance histogram. Video from RED Digital Cinema and image from pexels.com.

Color transfer. We seek to transfer the colors of the reference image R to the resulting non-linear image S_1 from the previous step. Statistical properties are often used to define a global mapping between two input images, especially when correspondences between images are not available. Following the same line as in [Kotera] [Reinhard] [Xiao], we transfer the statistics (mean and standard deviation) along each channel separately in a decorrelated color space but with some key modifications outlined below. Principal Component Analysis (PCA) is applied to the RGB source and reference images to find decorrelated spaces for each of them. We have done experiments of PCA in other color spaces, such as IPT [Ebner], but results show that color transfer by PCA performance is better in RGB.

The color transfer method can be summarized as a traslation for matching the mean between images, a rotation for axes alignment and a scaling for matching the standard deviation along each axis. The matching matrix M_C for color transfer is

$$M_C = T_R \cdot R_R \cdot V \cdot R_S^{-1} \cdot T_S \tag{6}$$

where T_R and T_S are the traslation matrices with the mean information of the reference and source images respectively, R_R and R_S are the rotation matrices composed of the eigenvectors of the PCA covariance matrices for reference and source images respectively, and V is the scaling matrix

$$V = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \sqrt{\lambda_2^R/\lambda_2^S} & 0 \\ 0 & 0 & \sqrt{\lambda_3^R/\lambda_3^S} \end{pmatrix} \tag{7}$$

where λ_i^R and λ_i^S are the eigenvalues of the PCA covariance matrices along the i axis, $i \in \{2,3\}$, for reference and source images respectively.

The resulting source image S_2 from this color transfer step will be

$$S_2(x) = M_C \cdot S_1(x) \tag{8}$$

where M_C is the matching matrix explained above, x is a pixel and S_1 is the resulting image from the previous luminance matching step.

Figure 3. From left to right: Luminance matched source image S_1 from the previous step, color-matched image S_2 and reference image R.

This type of technique has been used in some previous works [Kotera] [Xiao], but we add some modifications to it. As it can be observed in the scaling matrix V, the standard deviation along the first axis of the PCA decomposition is not modified. As it is mentioned in [Brown] [Mudrova], we can assume that the first axis of PCA corresponds to the luminance channel of the image. Luminance modifications are performed in the other steps of our method, in the color transfer step only chromaticity channels are modified, which correspond to the second and third axes in PCA. The second difference with respect to classic color transfer methods based on PCA decomposition is a constraint added to the rotation matrices to prevent large color shifts. By doing PCA analysis and simply assigning the source axes to the reference axes, the geometry of the resulting mapping might not be as it was intended. The resulting image could have the color proportions as expected, but locally the colors would be swapped. To avoid that, the method finds the rotation where the source axes are assigned to the closest reference axes.

Algorithm 1 Constraint to rotation matrices

Input: R_S eigenvectors matrix for source image, R_R eigenvectors matrix for reference image

Output: R'_R new reference rotation matrix

1:　　**for** vector v_i^S ($i = 1,2,3$) from source image **do**

2:　　　　calculate the angle between v_i^S and each vector in reference image v_i^R and the
3:　　　　opposite vectors $-v_i^R$ and select the closest to v_i^S

4:　　　　add a new column to R'_R with the selected vector from the previous step

5:　　**end for**

Local contrast transfer. We make an adaptation of the local contrast normalization formula used in [Cyriac]:

$$I'(x) = \mu(x) + (I(x) - \mu(x)) \cdot \frac{k}{\sigma} \tag{9}$$

where x is a pixel, $I(x)$ is the value at pixel x, μ is the local mean of I, k is a constant, σ is the standard deviation of σ, and $I'(x)$ is the final output value for pixel x. The value μ is obtained by convolving I with a kernel W which is a linear combination of two Gaussian kernels.

As it has been mentioned above, the first component of the PCA keeps the luminance and contrast information of the images, so for modifying the contrast, we will work on this first axis of PCA.

The resulting image S_3 will be obtained by a local contrast transformation applied to S_2, the image obtained from the color transfer step:

$$S_3 = LC(S_2) \tag{10}$$

We seek to transfer the local contrast from the reference to the source image. If we define local contrast of an image as the difference between light intensity and its local mean value, $I(x) - \mu$, our approach transfers the standard deviation of the local contrast from the reference to the source image through the following formula:

$$S_3(x) = \mu(x) + \left(S_2(x) - \mu(x)\right) \cdot \frac{\sigma_R}{\sigma_S} \tag{11}$$

where x is a pixel, $S_2(x)$ is the value at pixel x, μ is the local mean of S_2, σ_S is the standard deviation of the local contrast of the source image S_2 and σ_R is the standard deviation of the local contrast of the reference image R.

This modification is applied only to the luminance axis which is the first axis in PCA of S_2, so the second and third axes in PCA of S_2 and S_3 remain equal.

Figure 4. From left to right: Color-matched image S_2 from the previous step, local contrast transferred image S_3 (final result) and reference image R.

Video style transfer

We extend now the problem of transferring the style from a reference image to a source image to the problem of transferring the style from a reference image to all the frames in a video. Applying an independent style transfer to each frame of a video sequence might result in strong texture flickering and temporal incoherence, even if neighbor frames are similar in content and color. It is also computationally expensive to calculate the parameters for each frame without making profit from the common content between frames. In order for our method to have a low-

computational cost and to impose coherence between frames, we propose to apply the same style transfer transformation to all the frames in the video sequence.

First, the luminance, color transfer and contrast transfer transformations (explained in the previous section) are calculated for the first frame of the video, then those transformations are applied to each frame in the sequence.

Let f be the luminance transformation explained in the previous section,

$$S_1 = f(S_0) \tag{12}$$

where S_0 is the first frame of the video. Let g be the color transfer transformation explained in the previous section,

$$S_2 = g(S_1) \tag{13}$$

where S_1 is the resulting image from previous equation and S_2 is the color matched image for S_1. Let k be the quotient σ_R / σ_S obtained when applying the local contrast transformation to S_2. Then, the same transformation f calculated above is applied to each frame of the video, followed by g, and finally, the local contrast of each frame is adjusted multiplying it by the constant k, calculated before.

Temporal coherence is guaranteed by applying the same transformation to all the frames in the video. This approach has very low computational cost and produces temporal-coherent and flickering-free results.

Experiments and results

In this section, we provide some results obtained by our style transfer method for image sequences. We demonstrate style transfer results on a wide range of source and reference sequences (see Fig. 5), where the luminosity and color range vary from one to another.

The method has been tested for videos with resolution of 1920×960 pixels and some of the resulting frames can be seen in Fig.6. The full videos can be found in the following link:

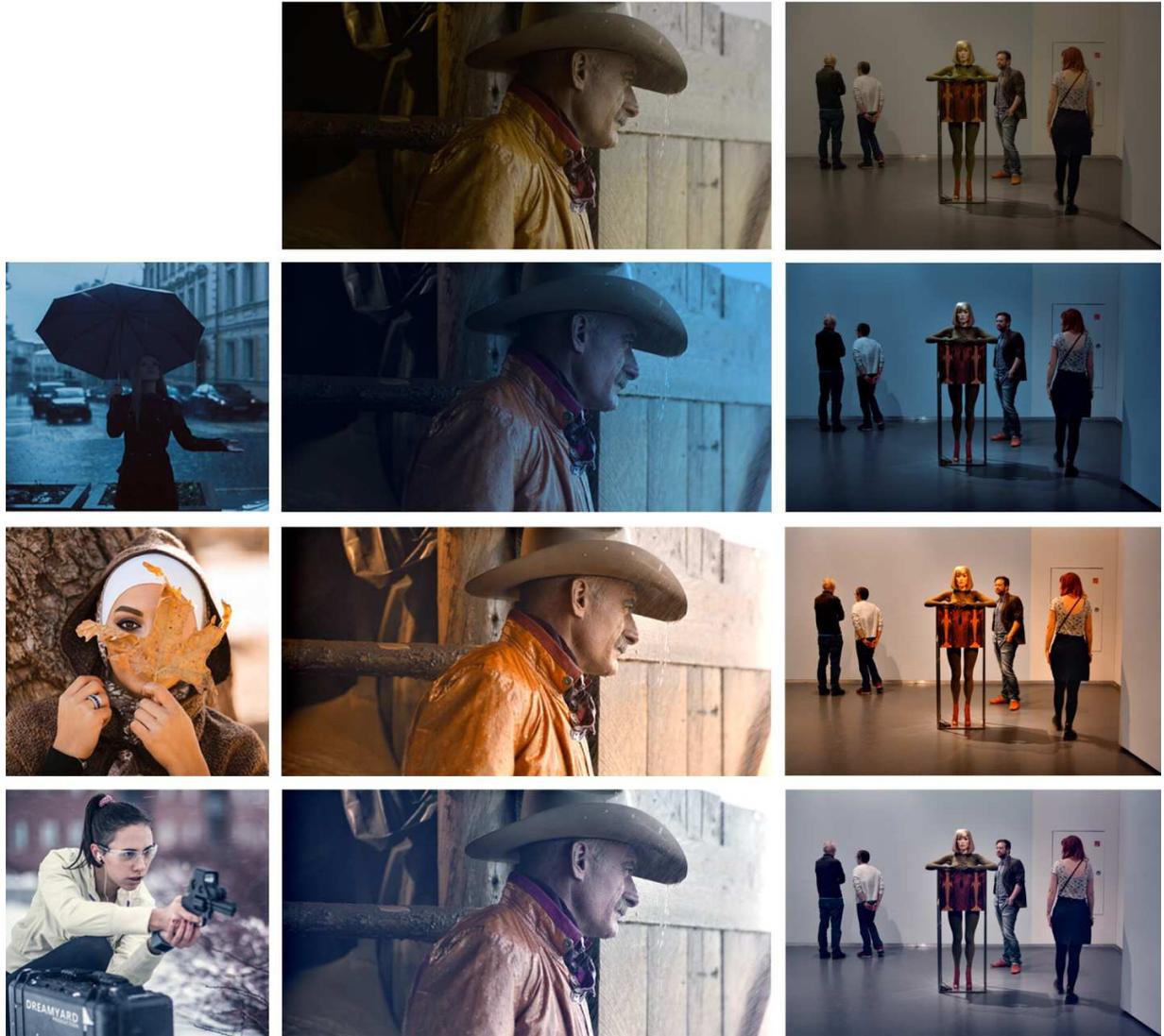http://ip4ec.upf.edu/VideoStyleTransfer

Figure 5. Original images on top, reference images on the left and resulting images. Images from the authors, from pexels.com and from RED Digital Cinema.
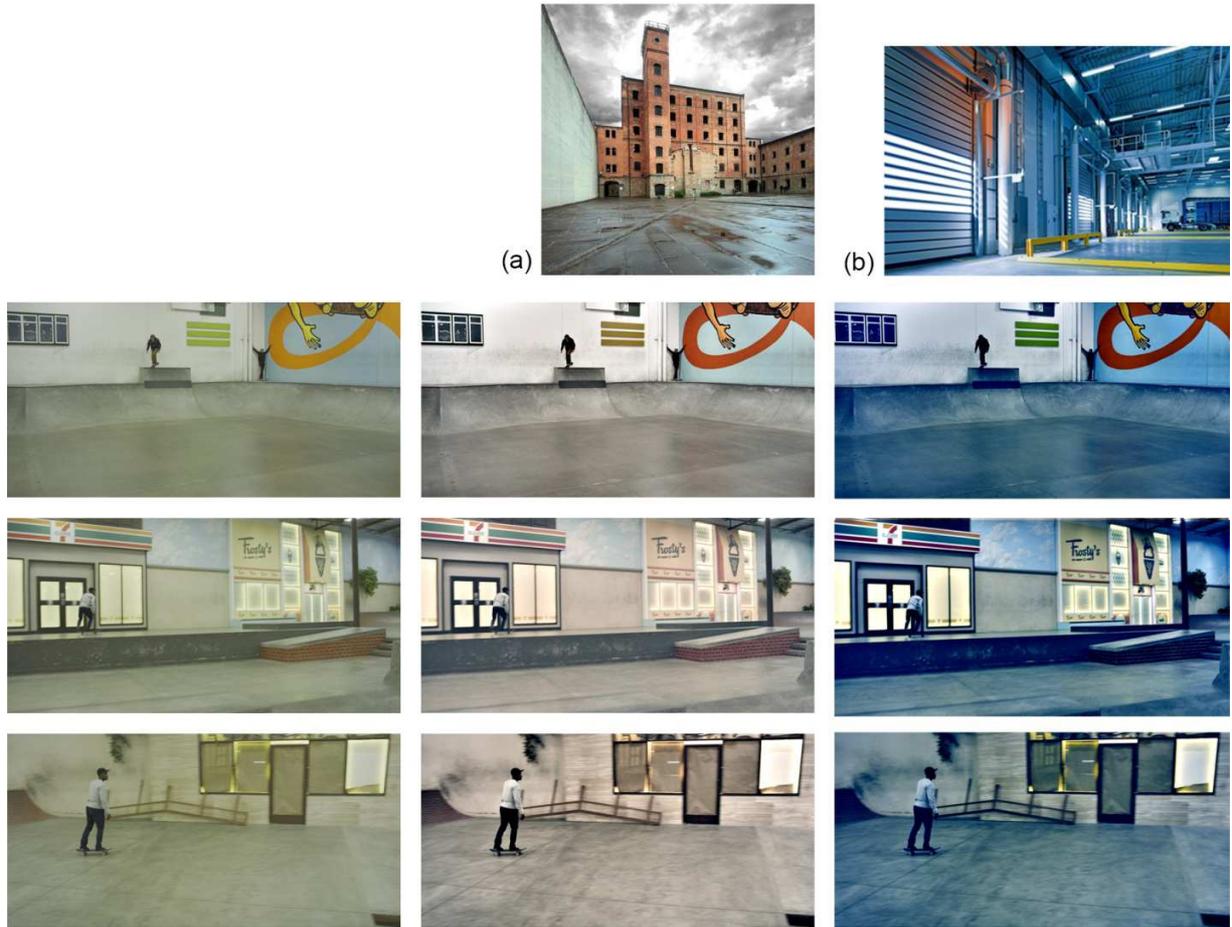
Figure 6. Left: Selected frames from the original RAW video (gamma corrected for visualization). Center: Resulting frames for reference (a). Right: Resulting frames for reference (b). Video from RED Digital Cinema and images from pexels.com.

We must say that quantitative evaluation of stylization is very difficult since there is no ground truth and different artistic choices can be rated differently by users. However, the results are in general visually pleasing, and free of artifacts or flickering effects and show temporal and spatial consistency.

The selection of the reference image is very important for the method. As we can see in Fig.7, when the source and reference images are very different in content and composition, the results are not as good as expected. Generally, statistical color-mapping methods suffer from the same limitation and some of them use segmentation or user interaction to solve this issue [Kotera] [Xiao].

Figure 7. From left to right: Original image (gamma corrected for visualization), style transferred image and reference image. The dissimilarity between the source and the reference result in an incoherent style transferred image, with unreal skin color. Images Video from RED Digital Cinema and image from pexels.com.

Conclusion

We have presented a method to transfer the style from one reference image to a video sequence extending our previous work [Zabaleta]. It consists of three separable and independent stages that could be implemented as a whole or separately: a luminance transfer step, a color transfer step and a local contrast transfer step. The style transfer transformations are calculated for the first frame of the video and then applied to all the frames in the sequence. Each operation has very low computational cost making the method suitable for in-camera implementation.

We are working in adding some constraints to our method to improve the results in the case of color transfer for skin tones. When the source and the reference image are very dissimilar the results might be incoherent, and it is especially noticeable in the case of skin. Another limitation of our method is the selection of the frame used to calculate the transformations which will be applied to each frame of the video sequence: If the first frame of the video is not representative of the rest of the frames in the sequence, the results could be sub-optimal.

Acknowledgments

References

[Bonneel] N. Bonneel, K. Sunkavalli, S. Paris, H. Pfister, "Example-based video color grading", ACM Transactions on Graphics (Proceedings of SIGGRAPH 2013), 2013.

[Brown] M. Brown, S. Susstrunk, "Multi-spectral SIFT for Scene Category Recognition", International Conference on Computer Vision and Pattern Recognition, 2011.

[Cyriac] P. Cyriac, D. Kane and M. Bertalmío, "Optimized tone curve for in-camera image processing", IS&T Electronic Imaging Conference, 2016.

[Ebner] F. Ebner, "Derivation and modelling of HUE uniformity and development of the IPT color space", Ph.D. degree dissertation, 1998.

[Frigo] O. Frigo, N. Sabater, J. Delon, P. Hellier, "Video Style Transfer by Consistent Adaptive Patch Sampling", The Visual Computer, 2018.

[Gatys] L. A. Gatys, A. S. Ecker, M. Bethge, "Image style transfer using convolutional neural networks", IEEE Computer Vision and Pattern Recognition, 2016.

[HaCohen] Y. HaCohen, E. Shechtman, D. B. Goldman, D. Lischinski, "Non-rigid dense correspondence with applications for image enhancement", ACM Transactions on Graphics (Proceedings of SIGGRAPH 2011), 2011.

[Kotera] H. Kotera, "A Scene-referred color transfer for pleasant imaging on display", IEEE International Conference on Image Processing, 2005.

[Luan] F. Luan, S. Paris, E. Shechtman, K. Bala, "Deep photo style transfer", IEEE Computer Vision and Pattern Recognition, 2017.

[Mudrova] M. Mudrová, A. Procházka, "Principal Component Analysis in image processing", Proceedings of MATLAB Technical Computing Conference, 2005.

[Pitié] F. Pitié, A. Kokaram, R. Dahyot, "Automated colour grading using colour distribution transfer", Computer Vision and Image Understanding, 2007.

[Poynton] C. Poynton, "Gamma", in Digital Video and HD: Algorithms and Interfaces, Morgan Kaufmann, 2003.

[Reinhard] E. Reinhard, M. Ashikhmin, B. Gooch and P. Shirley, "Color transfer between images", IEEE Computer Graphics and Applications, 2001.

[Xiao] X. Xiao, L. Ma, "Color transfer in correlated color space", Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications (VRCIA), 2006.

[Vazquez] J. Vázquez-Corral, M. Bertalmío, "Color stabilization along time and across shots of the same scene, for one or several cameras of unknown specifications", IEEE Transactions on Image Processing, 2014.

[Zabaleta] I. Zabaleta, M. Bertalmio, "Photorealistic style transfer for cinema shoots", to appear Proceedings in Colour and Visual Computing Symposium, 2018.