Abstract—This paper proposes a new colour transfer method with Optimal transport to transfer the colour of a source image to match the colour of a target image of the same scene. We propose to formulate the problem in higher dimensional spaces (than colour spaces) by encoding overlapping neighborhoods of pixels containing colour information as well as spatial information. Since several recoloured candidates are now generated for each pixel in the source image, we define an original procedure to efficiently merge these candidates which allows denoising and artifact removal as well as colour transfer. Experiments show quantitative and qualitative improvements over previous methods. Our method can be applied to different contexts of colour transfer such as transferring colour between different camera models, camera settings, illumination conditions and colour retouch styles for photographs.

Index Terms—optimal transport, colour transfer, image enhancement, JPEG compression blocks.

I. INTRODUCTION

In the digital media post-production industry, colour transfer methods have been developed to transform a source colour image into a specified target colour image to match colour statistics or eliminate colour variations between different photographs. Colour variations between photographs often happen due to illumination changes, using different cameras, different in-camera settings or due to tonal adjustments of the users. Applications in image processing problems range from generating colour consistent image mosaicing and stitching [1] to colour enhancement and style manipulation [2].

When computing the transfer function, considering colour information only does not take into account the fact that coherent colours should be transferred to neighboring pixels, which can create results with blocky artifacts emphasizing JPEG compression blocks, or increase noise. In this paper we propose to extend the colour problem to higher dimensional spaces by encoding overlapping neighborhoods of pixels, taking into account both their colour and spatial information. We solve the high dimensional problem in 1D space using an iterative projection approach. Our method can be applied to different contexts of colour transfer such as transferring colour between different camera models, camera settings, illumination conditions and colour retouch styles for photographs.

II. RELATED WORKS

Optimal transport (OT) applications in imaging have been widely studied in the last decade. The OT problem consists of estimating the minimum cost (referred to as the Wasserstein Distance [3], [4] or as the Earth Movers Distance [5]) of transferring a source distribution to a target distribution. As a byproduct of OT distance estimation, the mapping μ itself between the two distributions is also provided.

Monge’s formulation of OT [4] defines the deterministic coupling y∗ = φ(x∗) between random vectors x∗ ∼ f(x∗) and y∗ ∼ g(y∗) that capture the colour information of the source and target images respectively, and its solution minimizes the total transportation cost:

$$\arg\min_\phi \int \|x^\phi - \phi(x^\phi)\|^2 f(x^\phi) \, dx^\phi$$

such that: $$f(x^\phi) = g(\phi(x^\phi)) \cdot |\det \nabla \phi(x^\phi)|$$

with f the probability density function (pdf) of x∗ and g the pdf of y∗. The solution for φ can be found using existing algorithms such as linear programming, and the Hungarian and Auction algorithms [6]. However, in practice it is difficult to find a solution for colour images when \(\dim(x^\phi) = \dim(y^\phi) = d > 1\) as the computational complexity of these solvers increases in multidimensional spaces [3]. But for d = 1, with x∗, y∗ ∈ ℝ, a solution for φ is straightforward and can be defined using the increasing rearrangement [4]:

$$\phi = G^{-1} \circ F$$

where F and G are the cumulative distributions of the colour values in the source and target images respectively. The 1D solution in (2) has been used to tackle problems in multidimensional colour spaces and of particular interest is the Iterative Distribution Transfer (IDT) algorithm for colour transfer proposed by Pitié et al. [7]. They proposed to iteratively project colour values \(\{x_i\}_{i=1}^n\) and \(\{y_j\}_{j=1}^m\) originally in \(\mathbb{R}^d\) to a 1D subspace and solve the OT using (2) in this 1D subspace and then propagate the solution back to \(\mathbb{R}^d\) space. This operation is repeated with different directions in 1D space until convergence. This strategy was inspired by the idea of the Radon Transform [8] which states the following proposition: if the target and source colour points are aligned in all possible 1D projective spaces, then matching is also achieved in \(\mathbb{R}^d\).
space. Note that the implementation of IDT approximates $F$ and $G$ using cumulative histograms\(^1\).

The Sliced Wasserstein Distance (SWD) algorithm was later proposed, similarly using the iterative projection approach of IDT, and was applied to texture mixing applications \([9]\) and colour transfer \([10]\). The SWD algorithm uses quantile matching instead of cumulative histogram matching to solve (2). More specifically, SWD sorts the $n$ 1D projections of the source and target images respectively to define quantiles with regular increments of size $\frac{1}{n}$ between 0 and 1 for both source and target distributions. IDT can also be considered as an approximation of a quantile matching approach but with irregular increments derived from the cumulative histograms of the source and target images - as source and target quantiles do not match exactly, interpolation can be used to compute solution (2) \([8]\). The SWD algorithm can be computed in $O(n \log(n))$ operations using a fast sorting algorithm \([9]\). When a small number of observations are available, using SWD is best but with a large number of observations, histogram matching with IDT is more efficient.

The warping map $\phi$ derived using the closed form (2) has a number of important characteristics: it is parameter free, no strong hypotheses are made about the distributions (i.e. $f$ and $g$), and by design it is an increasing function. More importantly, it provides a tractable iterative solution for imaging applications formulated in higher dimensional spaces \([10]\). However, undesirable visual artifacts can be produced in the output image and a post processing step is added to ensure that the gradient field of the recoloured source image is as close as possible to the original source image \([8]\). Similarly, Bonneel et al. \([10]\) use an iterative post-processing technique \([11]\) to regularize the transportation map.

### III. Patch-based colour transfer

We propose to solve colour transfer by encoding overlapping neighborhoods of pixels, accounting for both colour information as well as pixel location, and extending the problem to higher dimensional space. We solve the transfer problem in 1D space using an iterative projection approach. In our context, we found that using the quantile matching approach that employs sorting operations is better than the histogram technique. The problem with building histograms is that the regular bin size used for every projection must be chosen, and so to alleviate the need to optimize the bin size parameter we choose the sorting technique \([10]\). Due to overlapping regions in patches, several recoloured candidates are now generated for each pixel in the source image. Therefore, we also define an original procedure to efficiently merge these candidates. Our method consists of the following steps:

1. **Combine colour and spatial information.** Zhixin et al. \([12]\) have proposed to incorporate spatial information for portrait relighting that transfers shading with awareness of the face geometry in source and target images. Similarly we propose to incorporate pixel positions with colour information to take into account spatial information when recolouring same content images. The pixel’s colour $x_c$ and its pixel position $x^p$ are concatenated into a vector $x = (x^c, x^p)^T$ such that $\dim(x) = \dim(x^c) + \dim(x^p)$. Adding this spatial information extends Monge’s formulation in (1.a) such that $\phi(x)$ minimizes the following cost:

\[
\int(\|x^c - \phi_c(x)\|^2 + \|x^p - \phi_p(x)\|^2) \, f(x) \, dx \tag{3.a}
\]

such that $f(x) = g(\phi(x)) \mid \det \nabla \phi(x)} \tag{3.b}$

where $\phi$ denotes the pdf of $x$ and $g$ the pdf of $y$ in the high dimensional space ($\mathbb{R}^d$, where $d = \dim(x)$). $\phi_c$ and $\phi_p$ are the components of $\phi$ that return the transformed colour and spatial features respectively.

2. **Change the coordinate system.** In the case when we combine colour and spatial information, the colours have integer values from 0 to 255, and the spatial values can be anything depending on the size of the image. In order to produce consistent results regardless of the size of the image, we normalize all the colour and position coordinates to lie in $[0, 1]$. We then stretch that space in the direction of the spatial coordinates by a factor $w = 2.5$ to make it harder to move the pixels in the spatial domain than in the colour domain, because since we are focusing on transferring colour between images of a same scene, we know that the scenes are overlapped.

3. **Create patch vectors.** We encode overlapping neighborhoods of pixels to preserve local topology information. Starting from the origin of the coordinate system of the images (upper left corner), we use a sliding window operation of window size $k \times k$ to extract overlapping patches. From each individual patch we create a high dimensional vector in $\mathbb{R}^d \times k \times k$ by concatenating the pixels in the patch with

---

\(^1\)https://github.com/frcs/colour-transfer
their \( d \) dimensional features (colour values only such that \( x = (x^c)^T \) Fig. 1 (A), or colour and spatial features such that \( x = (x^c, x^p)^T \) Fig. 1 (B)). We apply this process to the source and target images to create patch vector sets \( \{x_i\} \) and \( \{y_j\} \) for each respectively. Note, the formulation that includes spatial information favors local mapping between patches such that it penalizes mappings between patches that are spatially distant from each other even if their colours are similar.

4. Iterative projection transfer. The sorting operation from 4. Iterative projection transfer.

\[
\mathbf{R}^{d \times k \times k}
\]

at iteration \( t \). At each iteration, we generate a random orthogonal basis for the \( R^{d \times k \times k} \) space and project the source and target points onto these axes. Along each axis, \( \phi \) is estimated and then applied to the source points to create an intermediate value \( \hat{x}(t+1) \). In SWD the final modified source points at the \( t+1 \) iteration are obtained by taking a partial step with \( \alpha=0.2 \) such that \( x(t+1) = (1-\alpha) x(t) + \alpha \hat{x}(t+1) \). We have chosen here \( \alpha=1 \) to get an updating step similar to Pitié et al [7], hence not enforcing any pulling constraint on the transfer function towards the identity as Bonneel et al. [10]. After convergence, we retain only the transformed colour information (the \( \phi_c(x) \) component of \( \phi \)) and discard any changes in positions (\( \phi_p(x) \)).

5. Merge recoloured candidates. Because of the overlapping regions between patches, several recoloured candidates are now generated for each pixel in the source image. We compute the average colour value from the candidates and use it in conjunction with the original position to create the output recoloured image.

IV. EXPERIMENTAL ASSESSMENT

We provide quantitative and qualitative evaluations to validate both of our optimal transport solutions - using colour patches only, annotated in the results as \( \text{Our}_c \), and using colour patches with spatial information, annotated as \( \text{Our}_cp \). We compare our methods to different state of the art colour transfer methods [2], [8], [13], [14]. In these evaluations we use image pairs with similar content from an existing dataset provided by Hwang et al.\(^2\) [2]. The dataset includes registered pairs of images (source and target) taken with different cameras and settings, and different illuminations and recolouring styles.

A. Colour space and parameters settings

We use the YCbCr colour space because of its ability to correlate colour channels. We transform the luminance (Y) and chrominance (CbCr) components separately and recombine the resulting recoloured sources (Y and CbCr) to create the final recoloured source image. For \( \text{Our}_cp \), when processing chrominance, each pixel is represented by its 2D chrominance (CbCr) value and its 2D spatial position. When processing luminance, each pixel is represented by its 1D luminance (Y) values and its 2D spatial position. The spatial information corresponds to the position coordinates of the pixel in the image, with the origin of the coordinate system in the upper left corner (Fig. 1). Our patches with combined colour and spatial features create a vector in 75 dimensions (\( 3 \times 5 \times 5 \)) for the luminance and position component, and 100 dimensions (\( 4 \times 5 \times 5 \)) for the chrominance and position component. For \( \text{Our}_c \), pixel position is not accounted for, and only chrominance and luminance are used. When creating patch vectors, we experimented with different patch sizes and found that a size of \( 5 \times 5 \) captures enough of a pixel’s neighbourhood and does not increase computational complexity too much. The code we used to implement our method is available online\(^3\).

B. Evaluation metrics

To quantitativingly assess the recolouring results, four metrics are used: peak signal to noise ratio (PSNR) [15], structural similarity index (SSIM) [16], colour image difference (CID) [17] and feature similarity index (FSIMc) [18]. These metrics are often used when considering source and target images of the same content [2], [14], [19], [20]. Although the registration errors may affect the evaluation metrics, the measures for different methods were computed using the same registration errors. Note that the results using PMLS were provided by the authors [2]. We also compared PMLS with two other recent techniques [21], [22] that also incorporate correspondences into their framework, but PMLS has been shown to perform better than these two [13] and so is the one reported here.

C. Experimental Results

Table I presents the quantitative results. Our method with colour and spatial information (\( \text{Our}_cp \)) outperforms other state of the art methods in most cases as measured by CID and FSIMc and performs similarly to other top methods PMLS and L2 as measured by the PSNR and SSIM metrics.

Figure 2 provides qualitative results. For clarity, the results are presented in image mosaics, created by switching between the target image and the transformed source image column wise (Figure 2, top row). If the colour transfer is accurate, the resulting mosaic should look like a single image (ignoring the small motion displacement between source and target images), otherwise column differences appear. As can be noted, our approach \( \text{Our}_cp \) is visually the best at removing the column differences.

While PMLS provides equivalent results to our methods in terms of metrics, it introduces visual artifacts if the input images are not registered correctly (Figure 3), while our method is robust to registration errors. Note that although the accuracy of the PSNR, SSIM, CID and FSIMc metrics relies on the fact that the input images are registered correctly; if this is not the case, these metrics may not accurately capture all artifacts (Figures 3 and 2). In addition, due to the merging step of our algorithm (cf. step 5 Sec. III), our approach allows us to create a smooth colour transfer result, and can also alleviate JPEG compression artifacts and noise (cf. Figure 3 for comparison).

\(^2\)https://sites.google.com/site/unimono/pmls

\(^3\)https://github.com/leeshp/PCT_OT
Our method $\text{Our}_{\text{cp}}$ can also correctly transfer colours between images that contain moving objects, as can be seen in Figure 3 with examples ‘illum’ and ‘mart’. However, $\text{Our}_{\text{cp}}$ can create shadow artifacts when there are large changes between target and source images (Figure 3 in example ‘building’). In this case $\text{Our}_{\text{c}}$ does not suffer from these shadow artifacts and creates good colour transfer results. In the future, we would like to create a hybrid framework that integrates both $\text{Our}_{\text{c}}$ and $\text{Our}_{\text{cp}}$ and switches between the two using the parameter $w$. A motion estimation technique can be used to determine which pixels in the source image have moved in the target image. With pixels affected by large motion, the position’s weight would be set to $w = 0$, making the algorithm account for $\text{Our}_{\text{c}}$ only. When pixels move less, the position’s weight would be set to $w = 2.5$, making the algorithm account for $\text{Our}_{\text{cp}}$.

V. CONCLUSION

Several contributions to colour transfer within OT have been made in this paper, showing quantitative and qualitative improvements over state of the art methods. In particular, first, neighborhoods of pixels (patches) are used with OT algorithm in high dimensional space, and second, spatial information as well as colour content of pixels are both encoded in the high dimensional feature vectors. This original construction implies a new reconstruction step since each recoloured pixel benefits from the contribution of several estimated candidates using a
simple averaging solution which allows denoising and artifact removal as well as colour transfer.

REFERENCES


