Supplemental Materials

1 Proofs of Propositions

Proposition 1. *The transportation plan $\boldsymbol{\pi} \in \boldsymbol{\pi}(\boldsymbol{\mu}_0, \boldsymbol{\mu}_1)$ minimizing (10) is of the form $\boldsymbol{\pi} = \mathbf{D}_{\mathbf{v}}\mathbf{H}_t\mathbf{D}_{\mathbf{w}}$, with unique vectors $\mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ satisfying*

$$\begin{cases} \mathbf{D}_{\mathbf{v}}\mathbf{H}_t\mathbf{D}_{\mathbf{w}}\mathbf{a} = \boldsymbol{\mu}_0, \\ \mathbf{D}_{\mathbf{w}}\mathbf{H}_t\mathbf{D}_{\mathbf{v}}\mathbf{a} = \boldsymbol{\mu}_1. \end{cases} \tag{1}$$

Proof. Decompressing notation, the optimization can be written as

$$\begin{array}{rl} \min_{\boldsymbol{\pi}\in\mathbb{R}^{n\times n}} & \sum_{ij} \boldsymbol{\pi}_{ij} \ln\left[\frac{\boldsymbol{\pi}_{ij}}{e\mathbf{H}_{ij}}\right] \mathbf{a}_i\mathbf{a}_j \\ \text{s.t.} & \boldsymbol{\pi}\mathbf{a} = \boldsymbol{\mu}_0 \\ & \boldsymbol{\pi}^\top\mathbf{a} = \boldsymbol{\mu}_1. \end{array}$$

After introducing Lagrange multipliers $\boldsymbol{\lambda}_0, \boldsymbol{\lambda}_1 \in \mathbb{R}^n$, the first-order optimality conditions for this system take the form

$$-\mathbf{a}_i\mathbf{a}_j \ln\frac{\boldsymbol{\pi}_{ij}}{\mathbf{H}_{ij}} = \mathbf{a}_j\boldsymbol{\lambda}_{0i} + \mathbf{a}_i\boldsymbol{\lambda}_{1j}\ \forall i, j \in \{1, \ldots, n\}.$$

Equivalently, we can write

$$\boldsymbol{\pi}_{ij} = \mathbf{H}_{ij} \exp\left(-\frac{\boldsymbol{\lambda}_{0i}}{\mathbf{a}_i}\right) \exp\left(-\frac{\boldsymbol{\lambda}_{1j}}{\mathbf{a}_j}\right)$$

Take $\mathbf{v} \stackrel{\text{def.}}{=} \exp(-\boldsymbol{\lambda}_0 \oslash \mathbf{a})$ and $\mathbf{w} \stackrel{\text{def.}}{=} \exp(-\boldsymbol{\lambda}_1 \oslash \mathbf{a})$, where $\oslash$ denotes elementwise division. Then, this last expression shows $\boldsymbol{\pi} = \mathbf{D}_{\mathbf{v}}\mathbf{H}_t\mathbf{D}_{\mathbf{w}}$. Applying symmetry of $\mathbf{H}_t$ and substituting into the two constraints shows (1). □

Proposition 2. *The KL projection of $(\boldsymbol{\pi}_i)_{i=1}^k$ onto $\mathcal{C}_1$ satisfies $\operatorname{proj}_{\mathcal{C}_1} \boldsymbol{\pi}_i = \boldsymbol{\pi}_i\mathbf{D}_{\boldsymbol{\mu}_i \oslash \boldsymbol{\pi}_i^\top\mathbf{a}}$ for each $i \in \{1, \ldots k\}$.*

Proof. The problem decouples, and hence projection can be carried out one transportation matrix at a time. Expanding the objective for a single transportation matrix yields the following problem:

$$\begin{array}{rl} \min_{\bar{\boldsymbol{\pi}}\in\mathbb{R}^{n\times n}} & \sum_{ij} \bar{\boldsymbol{\pi}}_{ij} \ln\left[\frac{\bar{\boldsymbol{\pi}}_{ij}}{e\boldsymbol{\pi}_{ij}}\right] \mathbf{a}_i\mathbf{a}_j \\ \text{s.t.} & \bar{\boldsymbol{\pi}}^\top\mathbf{a} = \boldsymbol{\mu}, \end{array}$$

where $\bar{\boldsymbol{\pi}}$ is the projection of $\boldsymbol{\pi}$ onto $\mathcal{C}_1$. For Lagrange multiplier $\boldsymbol{\lambda} \in \mathbb{R}^n$, the first-order optimality condition for element $\bar{\boldsymbol{\pi}}_{ij}$ is

$$-\mathbf{a}_i\mathbf{a}_j \ln\frac{\bar{\boldsymbol{\pi}}_{ij}}{\boldsymbol{\pi}_{ij}} = \mathbf{a}_i\boldsymbol{\lambda}_j \implies \bar{\boldsymbol{\pi}}_{ij} = \boldsymbol{\pi}_{ij} \exp\left(-\frac{\boldsymbol{\lambda}_j}{\mathbf{a}_j}\right).$$

After taking $\mathbf{c} \stackrel{\text{def.}}{=} \exp(-\boldsymbol{\lambda} \oslash \mathbf{a})$, this expression shows $\bar{\boldsymbol{\pi}} = \boldsymbol{\pi}\mathbf{D}_{\mathbf{c}}$. Since $\bar{\boldsymbol{\pi}}^\top\mathbf{a} = \boldsymbol{\mu}$, we now can write $\mathbf{D}_{\mathbf{c}}\boldsymbol{\pi}^\top\mathbf{a} = \boldsymbol{\mu}$, showing $\mathbf{c} = \boldsymbol{\mu} \oslash \boldsymbol{\pi}^\top\mathbf{a}$, as needed. □

Proposition 3. *The KL projection of $(\boldsymbol{\pi}_i)_{i=1}^k$ onto $\mathcal{C}_2$ satisfies $\operatorname{proj}_{\mathcal{C}_2} \boldsymbol{\pi}_i = \mathbf{D}_{\boldsymbol{\mu}\oslash\mathbf{d}_i}\boldsymbol{\pi}_i$ for each $i \in \{1, \ldots k\}$, where $\mathbf{d}_i = \boldsymbol{\pi}_i\mathbf{a}$ and $\boldsymbol{\mu} = \prod_i \mathbf{d}_i^{\alpha_i/\sum_\ell \alpha_\ell}$.*

Proof. Take $(\bar{\boldsymbol{\pi}}_i)_{i=1}^k$ to be the projection onto $\mathcal{C}_2$, with unknown common marginal $\boldsymbol{\mu}$. As in [Benamou et al. 2015], expanding the optimization problem provides the form

$$\begin{array}{rl} \min_{\{\bar{\boldsymbol{\pi}}_\ell\},\boldsymbol{\mu}} & \sum_{\ell ij} \alpha_\ell\bar{\boldsymbol{\pi}}_{\ell ij} \ln\left[\frac{\bar{\boldsymbol{\pi}}_{\ell ij}}{e\boldsymbol{\pi}_{\ell ij}}\right] \mathbf{a}_i\mathbf{a}_j \\ \text{s.t.} & \bar{\boldsymbol{\pi}}_\ell\mathbf{a} = \boldsymbol{\mu}\ \forall \ell \in \{1, \ldots, k\}. \end{array}$$

The Lagrange multiplier expression for this optimization is

$$\Lambda \stackrel{\text{def.}}{=} \sum_\ell \left(\sum_{ij} \alpha_\ell\bar{\boldsymbol{\pi}}_{\ell ij} \ln\left[\frac{\bar{\boldsymbol{\pi}}_{\ell ij}}{e\boldsymbol{\pi}_{\ell ij}}\right] \mathbf{a}_i\mathbf{a}_j + \boldsymbol{\lambda}_\ell^\top(\bar{\boldsymbol{\pi}}_\ell\mathbf{a} - \boldsymbol{\mu}) \right).$$

Differentiating with respect to $\bar{\boldsymbol{\pi}}_{\ell ij}$ shows

$$0 = \frac{\partial\Lambda}{\partial\bar{\boldsymbol{\pi}}_{\ell ij}} = \alpha_\ell\mathbf{a}_i\mathbf{a}_j \ln\frac{\bar{\boldsymbol{\pi}}_{\ell ij}}{\boldsymbol{\pi}_{\ell ij}} + \boldsymbol{\lambda}_{\ell i}\mathbf{a}_j,$$

or equivalently,

$$\bar{\boldsymbol{\pi}}_{\ell ij} = \boldsymbol{\pi}_{\ell ij} \exp\left(-\frac{\boldsymbol{\lambda}_{\ell i}}{\mathbf{a}_i\alpha_\ell}\right).$$

Taking $\mathbf{c}_\ell \stackrel{\text{def.}}{=} \exp(-\boldsymbol{\lambda}_\ell \oslash \mathbf{a})$, we can write $\bar{\boldsymbol{\pi}}_\ell = \mathbf{D}_{\mathbf{c}_\ell^{1/\alpha_\ell}}\boldsymbol{\pi}_\ell$.

Differentiating Λ with respect to $\boldsymbol{\mu}$ shows

$$\mathbf{0} = \nabla_{\boldsymbol{\mu}}\Lambda = -\sum_\ell \boldsymbol{\lambda}_\ell$$

$$\implies \prod_\ell \mathbf{c}_\ell = \exp\left(-\sum_\ell \boldsymbol{\lambda}_\ell \oslash \mathbf{a}\right) = \mathbf{1}.$$

Define $\mathbf{d}_\ell \stackrel{\text{def.}}{=} \boldsymbol{\pi}_\ell\mathbf{a}$. Then, substituting our new variables into the constraint $\bar{\boldsymbol{\pi}}_\ell\mathbf{a} = \boldsymbol{\mu}$ shows

$$\begin{aligned} &\mathbf{c}_\ell^{1/\alpha_\ell} \otimes \mathbf{d}_\ell = \boldsymbol{\mu}\ \forall\ell \\ &\quad \implies \mathbf{c}_\ell = (\boldsymbol{\mu} \oslash \mathbf{d}_\ell)^{\alpha_\ell} \end{aligned}$$

Define $A \stackrel{\text{def.}}{=} \sum_\ell \alpha_\ell$. By the relationship above,

$$\begin{aligned} &\mathbf{1} = \prod_\ell \mathbf{c}_\ell = \prod_\ell (\boldsymbol{\mu} \oslash \mathbf{d}_\ell)^{\alpha_\ell} = \boldsymbol{\mu}^A \prod_\ell \mathbf{d}_\ell^{-\alpha_\ell} \\ \implies &\boldsymbol{\mu} = \prod_\ell \mathbf{d}_\ell^{\alpha_\ell/A} \end{aligned}$$

Hence, $\mathbf{c}_\ell^{1/\alpha_\ell} = \boldsymbol{\mu} \oslash \mathbf{d}_\ell$, showing $\bar{\boldsymbol{\pi}}_\ell = \mathbf{D}_{\boldsymbol{\mu}\oslash\mathbf{d}_\ell}\boldsymbol{\pi}_\ell$.

□

Proposition 4. *There exists $\beta \in \mathbb{R}$ such that the KL projection of $(\boldsymbol{\pi}_i)_{i=1}^k$ onto $\overline{\mathcal{C}}_2$ satisfies $\operatorname{proj}_{\overline{\mathcal{C}}_2} \boldsymbol{\pi}_i = \mathbf{D}_{\boldsymbol{\mu}\oslash\mathbf{d}_i}\boldsymbol{\pi}_i$ for all $i \in \{1, \ldots, k\}$, where $\mathbf{d}_i = \boldsymbol{\pi}_i\mathbf{a}$ and $\boldsymbol{\mu} = \left(\prod_i \mathbf{d}_i^{\alpha_i}\right)^\beta$.*

Proof. Similarly to the previous proposition, we write the optimization problem as follows:

$$\begin{array}{rl} \min_{\{\bar{\boldsymbol{\pi}}_\ell\},\boldsymbol{\mu}} & \sum_{\ell ij} \alpha_\ell\bar{\boldsymbol{\pi}}_{\ell ij} \ln\left[\frac{\bar{\boldsymbol{\pi}}_{\ell ij}}{e\boldsymbol{\pi}_{\ell ij}}\right] \mathbf{a}_i\mathbf{a}_j \\ \text{s.t.} & \bar{\boldsymbol{\pi}}_\ell\mathbf{a} = \boldsymbol{\mu}\ \forall \ell \in \{1, \ldots, k\} \\ & \sum_i \mathbf{a}_i\boldsymbol{\mu}_i(\ln\boldsymbol{\mu}_i - 1) \geq -H_0 - 1. \end{array}$$

When the constraint is inactive, the optimization is solved by the previous proposition. Hence, we will focus on the active case, that is, when $\sum_i \mathbf{a}_i\boldsymbol{\mu}_i(\ln\boldsymbol{\mu}_i - 1) = -H_0 - 1$.

The Lagrange multiplier expression for this optimization is

$$\begin{aligned}\Lambda \stackrel{\text{def.}}{=} \sum_{\ell} \left(\sum_{ij} \alpha_\ell \bar{\boldsymbol{\pi}}_{\ell ij} \ln \left[\frac{\bar{\boldsymbol{\pi}}_{\ell ij}}{e \boldsymbol{\pi}_{\ell ij}} \right] \mathbf{a}_i \mathbf{a}_j + \boldsymbol{\lambda}_\ell^\top (\bar{\boldsymbol{\pi}}_\ell \mathbf{a} - \boldsymbol{\mu}) \right) \\ + \gamma \left(\sum_i \mathbf{a}_i \boldsymbol{\mu}_i (\ln \boldsymbol{\mu}_i - 1) + H_0 + 1 \right)\end{aligned}$$

Differentiating with respect to $\boldsymbol{\lambda}_\ell$, γ, $\boldsymbol{\pi}$, and $\boldsymbol{\mu}$ yields the following optimality criteria:

$$\begin{aligned} \boldsymbol{\mu} &= \boldsymbol{\pi}^\ell \mathbf{a} \, \forall \ell \in \{1, \ldots, k\} \\ -H_0 - 1 &= \sum_i \mathbf{a}_i \boldsymbol{\mu}_i (\ln \boldsymbol{\mu}_i - 1) \\ 0 &= \alpha_\ell \mathbf{a}_i \mathbf{a}_j \ln \frac{\bar{\boldsymbol{\pi}}_{\ell ij}}{\boldsymbol{\pi}_{\ell ij}} + \boldsymbol{\lambda}_{\ell i} \mathbf{a}_j \, \forall i, j, \ell \\ 0 &= \gamma \mathbf{a}_i \ln \boldsymbol{\mu}_i - \sum_\ell \boldsymbol{\lambda}_{\ell i} \, \forall i \end{aligned}$$

As before, the third condition shows

$$\bar{\boldsymbol{\pi}}_{\ell ij} = \boldsymbol{\pi}_{\ell ij} \exp \left(- \frac{\boldsymbol{\lambda}_{\ell i}}{\mathbf{a}_i \alpha_\ell} \right).$$

The fourth condition shows

$$\boldsymbol{\mu}^\gamma = \exp \left(\sum_\ell \boldsymbol{\lambda}_\ell \oslash \mathbf{a} \right).$$

Take $\mathbf{c}_\ell \stackrel{\text{def.}}{=} \exp(-\boldsymbol{\lambda}_\ell \oslash \mathbf{a})$. Then, the conditions above become

$$\begin{aligned} \bar{\boldsymbol{\pi}}_{\ell ij} &= \boldsymbol{\pi}_{\ell ij} \mathbf{c}_{\ell i}^{1/\alpha_\ell} \\ \boldsymbol{\mu}_i^\gamma &= \prod_\ell \mathbf{c}_{\ell i} \end{aligned}$$

Define $\mathbf{d}_\ell \stackrel{\text{def.}}{=} \boldsymbol{\pi}_\ell \mathbf{a}$. Since $\boldsymbol{\mu} = \bar{\boldsymbol{\pi}}_\ell \mathbf{a}$, for all ℓ we can write

$$\boldsymbol{\mu}_i = \sum_j \bar{\boldsymbol{\pi}}_{\ell ij} \mathbf{a}_j = \sum_j \boldsymbol{\pi}_{\ell ij} \mathbf{c}_{\ell i}^{1/\alpha_\ell} \mathbf{a}_j = \mathbf{c}_{\ell i}^{1/\alpha_\ell} \mathbf{d}_{\ell i}.$$

Taking the log of both sides of this expression and the relationship $\boldsymbol{\mu}_i^\gamma = \prod_\ell \mathbf{c}_{\ell i}$ shows

$$\begin{aligned} \alpha_\ell \ln \boldsymbol{\mu}_i &= \ln \mathbf{c}_{\ell i} + \alpha_\ell \ln \mathbf{d}_{\ell i} \, \forall \ell \\ \gamma \ln \boldsymbol{\mu}_i &= \sum_\ell \ln \mathbf{c}_{\ell i}. \end{aligned}$$

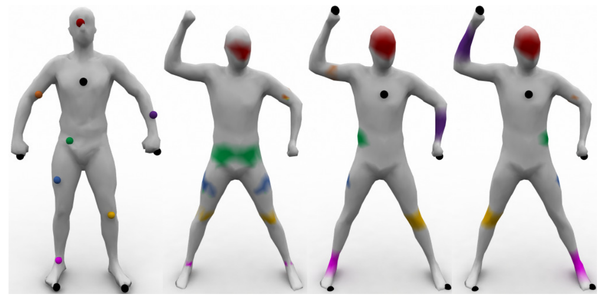
Summing the first equation over ℓ and removing the $\mathbf{c}_{\ell i}$ term by the second equation shows

$$\begin{aligned} \left(-\gamma + \sum_\ell \alpha_\ell \right) \ln \boldsymbol{\mu}_i &= \sum_\ell \alpha_\ell \ln \mathbf{d}_{\ell i} \\ \implies \boldsymbol{\mu}_i &= \prod_\ell \mathbf{d}_{\ell i}^{\alpha_\ell / (-\gamma + \sum_{\ell'} \alpha_{\ell'})} \end{aligned}$$

Identically to the previous proposition, $\bar{\boldsymbol{\pi}}_\ell = \mathbf{D}_{\boldsymbol{\mu} \oslash \mathbf{d}_\ell} \boldsymbol{\pi}_\ell$, with this new choice of $\boldsymbol{\mu}$; taking $\gamma = 0$ recovers the inactive constraint case. Defining

$$\beta \stackrel{\text{def.}}{=} \frac{1}{-\gamma + \sum_\ell \alpha_\ell}$$

provides the desired formula. $\square$



Source | $e^{-\alpha c(x,y)}$ | Result | Result (left/right flip)

Figure 3: *Additional soft map example.*

2 Proof of Formula in Algorithm 1

We simplify the convolutional distance between $\boldsymbol{\mu}_0$ and $\boldsymbol{\mu}_1$ as follows:

$$\begin{aligned} &\gamma \left[1 + \mathrm{KL}(\boldsymbol{\pi} | \mathbf{H}_t)\right] = \gamma \sum_{ij} \boldsymbol{\pi}_{ij} \ln \frac{\boldsymbol{\pi}_{ij}}{(\mathbf{H}_t)_{ij}} \mathbf{a}_i \mathbf{a}_j \\ &\quad = \gamma \sum_{ij} \boldsymbol{\pi}_{ij} \ln(\mathbf{v}_i \mathbf{w}_j) \mathbf{a}_i \mathbf{a}_j \text{ since } \mathbf{H}_t = \mathbf{D}_{\mathbf{v}} \mathbf{H}_t \mathbf{D}_{\mathbf{w}} \\ &\quad = \gamma \left[\sum_i \mathbf{a}_i (\ln \mathbf{v}_i) \sum_j \boldsymbol{\pi}_{ij} \mathbf{a}_j + \sum_j \mathbf{a}_j (\ln \mathbf{w}_j) \sum_i \boldsymbol{\pi}_{ij} \mathbf{a}_i \right] \\ &\quad = \gamma \left[\sum_i \mathbf{a}_i (\ln \mathbf{v}_i) \boldsymbol{\mu}_{0i} + \sum_j \mathbf{a}_j (\ln \mathbf{w}_j) \boldsymbol{\mu}_{1j} \right] \\ &\qquad \text{since } \boldsymbol{\pi} \mathbf{a} = \boldsymbol{\mu}_0 \text{ and } \boldsymbol{\pi}^\top \mathbf{a} = \boldsymbol{\mu}_1 \\ &\quad = \gamma \, \mathbf{a}^\top \left[(\boldsymbol{\mu}_0 \otimes \ln \mathbf{v}) + (\boldsymbol{\mu}_1 \otimes \ln \mathbf{w}) \right] \end{aligned}$$

3 Additional Examples

Figs. 1 and 2 (full page) show additional examples of color transfer on images.

Fig. 3 shows an additional example of a soft map.

References

BENAMOU, J.-D., CARLIER, G., CUTURI, M., NENNA, L., AND PEYRÉ, G. 2015. Iterative Bregman projections for regularized transportation problems. *SIAM J. Sci. Comp., to appear*.

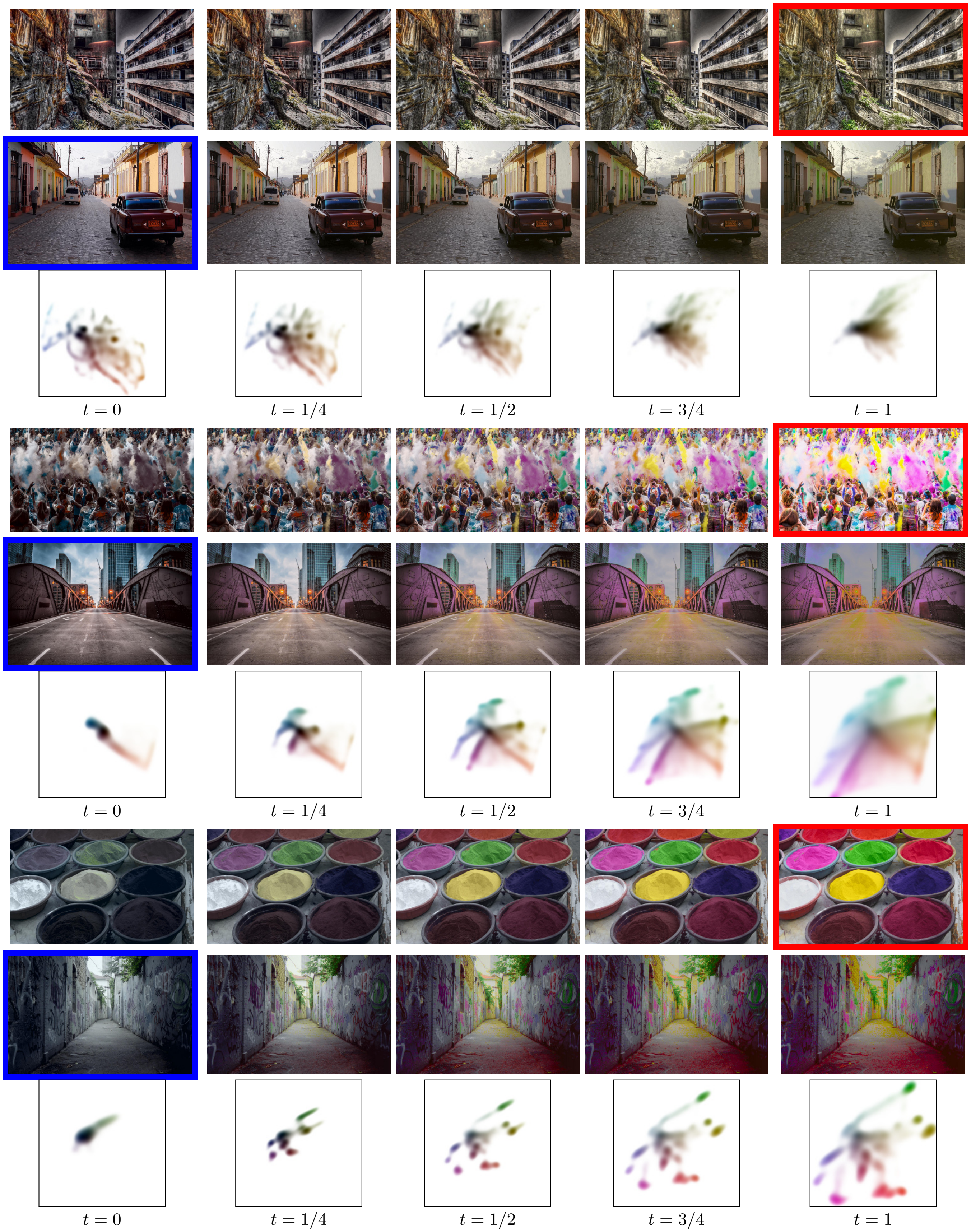


Figure 1: *Additional results: Color transfer with 2D transportation over chrominance space.*

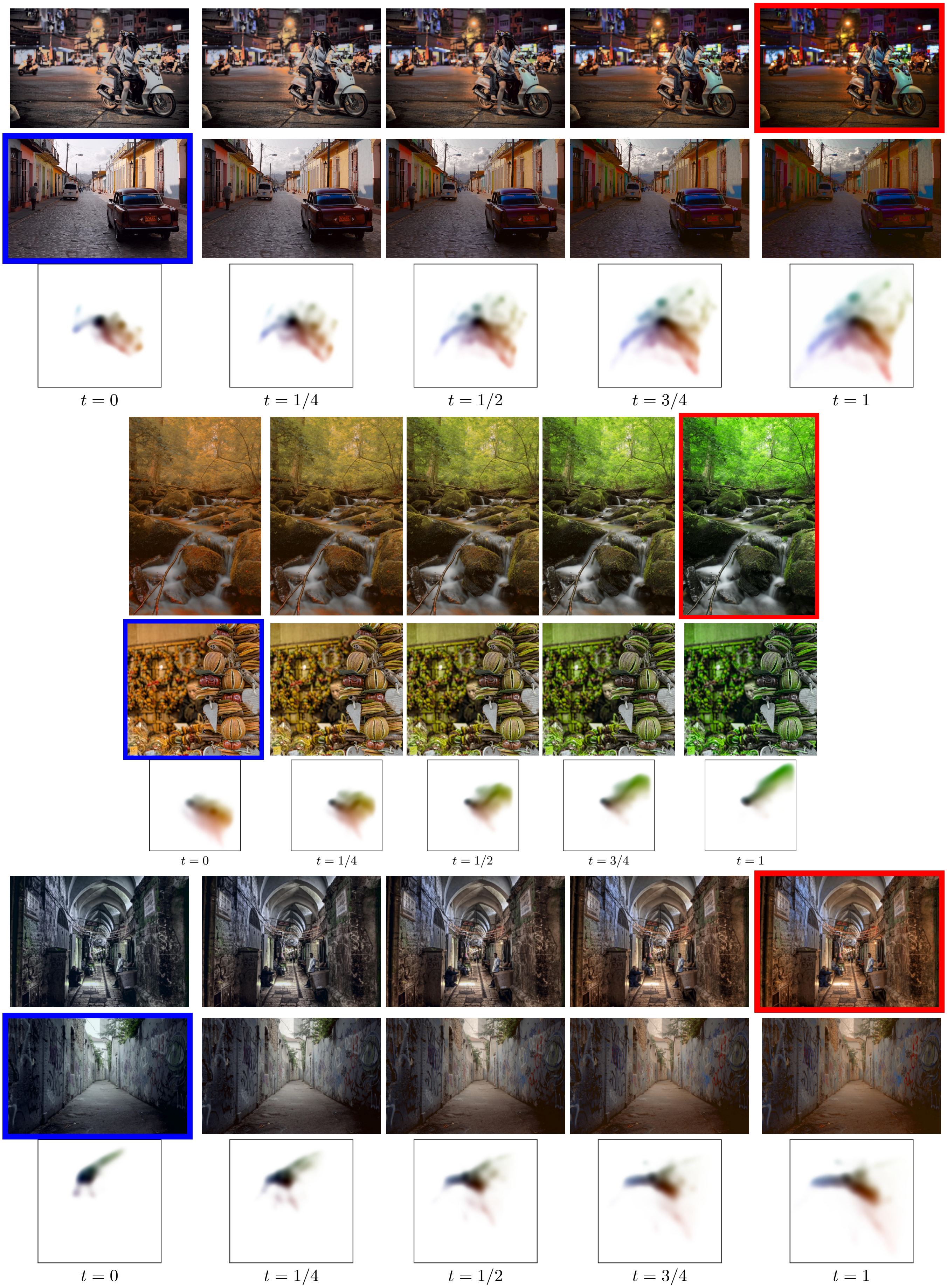


Figure 2: *Additional results: Color transfer with 2D transportation over chrominance space.*