Actionable Saliency Detection:
Independent Motion Detection Without Independent Motion Estimation
Supplementary Material

Georgios Georgiadis   Alper Ayvaci   Stefano Soatto
University of California, Los Angeles, 90095, USA
{giorgos,ayvaci,soatto}@cs.ucla.edu
http://vision.ucla.edu/

In the supplementary material, we show the calculation of $\nabla\psi(V)$ that is required to implement the gradient descent in our algorithm.

1. Calculating $\nabla\psi(V)$ for gradient descent

To calculate $\nabla\psi(V) = \left[\frac{\partial\psi(V)}{\partial V}\right]^T$ we use the following conventions. The derivative of a function $f\colon \mathbb{R}^n \to \mathbb{R}^m$ is given by an $m \times n$ matrix of partial derivatives $[Df_{ij}] = \frac{\partial f_i(x)}{\partial x_j}$. For $A \in \mathbb{R}^{n\times m}$ and $f\colon \mathbb{R}^{n\times m} \to \mathbb{R}^{p\times q}$ the derivative is given by $\frac{\partial f(A)}{\partial A} = \frac{\partial \text{vec}(f(A))}{\partial \text{vec}(A)} \in \mathbb{R}^{pq\times mn}$ where the vec operator stacks the columns of a matrix on top of each other. Using these definitions and common rules for chain and product rules we can derive $\left[\frac{\partial\psi(V)}{\partial V}\right]^T$. We can decompose $\frac{\partial\psi(V)}{\partial V}$ as follows:

$$\frac{\partial\psi(V)}{\partial V} = \frac{\partial\psi(\hat{C})}{\partial \text{vec}(\hat{C})} \frac{\partial \text{vec}(\hat{C}(\mathcal{W}C))}{\partial \text{vec}(\mathcal{W}C)} \frac{\partial \text{vec}(\mathcal{W}C)}{\partial \text{vec}(C)} \frac{\partial \text{vec}(C(V))}{\partial V} \tag{1}$$

The four terms are given by the following equations:

$$\frac{\partial\psi(\hat{C})}{\partial \text{vec}(\hat{C})} = (\mathcal{W}x)^T \otimes (\mathcal{W}x)^T \hat{C} \tag{2}$$

Using the product rule we can get $\frac{\partial \text{vec}(\hat{C}(\mathcal{W}C))}{\partial \text{vec}(\mathcal{W}C)}$. Define $f_1(\mathcal{W}C) = \mathcal{W}C$ and $g_1(\mathcal{W}C) = (C^T\mathcal{W}^T\mathcal{W}C)^{-1}C^T\mathcal{W}^T$. We then have:

$$\begin{aligned}
\frac{\partial \text{vec}(\hat{C}(\mathcal{W}C))}{\partial \text{vec}(\mathcal{W}C)} &= \frac{\partial}{\partial \text{vec}(\mathcal{W}C)}\left(I - \mathcal{W}C(C^T\mathcal{W}^T\mathcal{W}C)^{-1}C^T\mathcal{W}^T\right) && (3)\\
&= -\frac{\partial}{\partial \text{vec}(\mathcal{W}C)}\mathcal{W}C(C^T\mathcal{W}^T\mathcal{W}C)^{-1}C^T\mathcal{W}^T && (4)\\
&= -\left[\left(g_1(\mathcal{W}C)^T \otimes I_{2N}\right) f_1'(\mathcal{W}C) + \left(I_{2N} \otimes \mathcal{W}C\right) g_1'(\mathcal{W}C)\right] && (5)
\end{aligned}$$

The derivative of $f_1(\mathcal{W}C)$ with respect to $\mathcal{W}C$ is simply $f_1'(\mathcal{W}C) = I_{2N(N+3)}$. For the derivative of $g_1(\mathcal{W}C)$ we need to use the product rule. Define $f_2(\mathcal{W}C) = \left((\mathcal{W}C)^T\mathcal{W}C\right)^{-1}$ and $g_2(\mathcal{W}C) = (\mathcal{W}C)^T$. Then we can calculate the derivative of $g_1(\mathcal{W}C)$ using the following:

$$g_1'(\mathcal{W}C) = (g_2(\mathcal{W}C)^T \otimes I_{N+3}) f_2'(\mathcal{W}C) + (I_{2N} \otimes f_2(\mathcal{W}C)) g_2'(\mathcal{W}C) \tag{6}$$

The derivative of $g_2(\mathcal{W}C)$ is $g_2'(\mathcal{W}C) = T_{2N,N+3}$, where $T_{N,M} \in \mathbb{R}^{MN\times MN}$ is a permutation matrix such that

$$T_{N,M}\text{vec}(A) = \text{vec}(A^T) \tag{7}$$

To calculate the derivative of $f_2(\mathcal{W}C)$ we need to use the chain rule. Define $f_3(X) = X^{-1}$ and $g_3(\mathcal{W}C) = (\mathcal{W}C)^T\mathcal{W}C$. So, $f_3(g_3(\mathcal{W}C)) = f_2(\mathcal{W}C)$. The derivative of $f_2'(\mathcal{W}C)$ is:

$$f_2'(\mathcal{W}C) = f_3'((\mathcal{W}C)^T\mathcal{W}C)g_3'(\mathcal{W}C) \tag{8}$$

From standard results:

$$\begin{aligned} f_3'((\mathcal{W}C)^T\mathcal{W}C) &= -(((\mathcal{W}C)^T\mathcal{W}C)^{-T} \otimes ((\mathcal{W}C)^T\mathcal{W}C)^{-1}) && (9)\\ g_3'(\mathcal{W}C) &= \left(I_{(N+3)^2} + T_{(N+3),(N+3)}\right)\left(I_{N+3} \otimes (\mathcal{W}C)^T\right) && (10) \end{aligned}$$

Therefore:

$$\begin{aligned} f_2'(\mathcal{W}C) &= f_3'((\mathcal{W}C)^T\mathcal{W}C)g_3'(\mathcal{W}C) && (11)\\ &= -(((\mathcal{W}C)^T\mathcal{W}C)^{-T} \otimes ((\mathcal{W}C)^T\mathcal{W}C)^{-1}) && (12)\\ &\quad \left(I_{(N+3)^2} + T_{(N+3),(N+3)}\right)\left(I_{N+3} \otimes (\mathcal{W}C)^T\right) && (13)\\ g_1'(\mathcal{W}C) &= (g_2(\mathcal{W}C)^T \otimes I_{N+3})f_2'(\mathcal{W}C) + (I_{2N} \otimes f_2(\mathcal{W}C))g_2'(\mathcal{W}C) && (14)\\ &= -\left(\mathcal{W}C \otimes I_{N+3}\right)\left(\left((\mathcal{W}C)^T\mathcal{W}C\right)^{-T} \otimes \left((\mathcal{W}C)^T\mathcal{W}C\right)^{-1}\right) && (15)\\ &\quad \times \left(I_{(N+3)^2} + T_{(N+3),(N+3)}\right)\left(I_{N+3} \otimes (\mathcal{W}C)^T\right) && (16)\\ &\quad + \left(I_{2N} \otimes \left((\mathcal{W}C)^T(\mathcal{W}C)\right)^{-1}\right)T_{2N,N+3} && (17)\\ \frac{\partial \mathrm{vec}(\hat{C}(\mathcal{W}C))}{\partial \mathrm{vec}(\mathcal{W}C)} &= -\left[\left(g_1(\mathcal{W}C)^T \otimes I_{2N}\right)f_1'(\mathcal{W}C) + \left(I_{2N} \otimes \mathcal{W}C\right)g_1'(\mathcal{W}C)\right] && (18)\\ &= -\left[\left((C^T\mathcal{W}^T\mathcal{W}C)^{-1}C^T\mathcal{W}^T)^T \otimes I_{2N}\right)I_{2N(N+3)}\right] + (I_{2N} \otimes \mathcal{W}C) && (19)\\ &\quad \times \left(\mathcal{W}C \otimes I_{N+3}\right)\left(\left((\mathcal{W}C)^T\mathcal{W}C\right)^{-T} \otimes \left((\mathcal{W}C)^T\mathcal{W}C\right)^{-1}\right) && (20)\\ &\quad \times \left(I_{(N+3)^2} + T_{(N+3),(N+3)}\right)\left(I_{N+3} \otimes (\mathcal{W}C)^T\right) && (21)\\ &\quad + \left(I_{2N} \otimes \left((\mathcal{W}C)^T\mathcal{W}C\right)^{-1}\right)T_{2N,(N+3)} && (22)\\ & && (23) \end{aligned}$$

The derivative $\frac{\partial \mathrm{vec}(\mathcal{W}C)}{\partial \mathrm{vec}(C)}$ is given by:

$$\frac{\partial \mathrm{vec}(\mathcal{W}C)}{\partial \mathrm{vec}(C)} = I_{N+3} \otimes \mathcal{W} \tag{24}$$

In addition $\frac{\partial \mathrm{vec}(C(V))}{\partial V}$ is given by

$$\frac{\partial \mathrm{vec}(C(V))}{\partial V} = \begin{bmatrix} A_1 \\ O_{2N,3} \\ A_2 \\ O_{2N,3} \\ \vdots \\ A_N \\ O_{6N,3} \end{bmatrix} \tag{25}$$

which is the final term of the derivative of $\frac{\partial \psi(V)}{\partial V}$. $O_{M,N}$ is an $M \times N$ matrix with all elements equal to 0.