Solving Vision Problems via Filtering
Supplementary Material

1. Experimental Results

In our main paper, we provided experimental results for a number of vision-related inverse problems. This supplement provides additional details on the formulations used, as well as more extensive visual results for the experiments.

1.1. Disparity Super-resolution

For our disparity super-resolution experiment, we use the dataset from [1], which is a subset of the Middlebury stereo dataset. We show visualizations of our $16 \times$ super-resolution disparity maps in Figure 4.

1.2. Optical Flow Estimation

In our experiments, we use the color-gradient constancy model [2] instead of the brightness-constancy one [3]. In all cases, one can express the optical flow data fidelity term as

$$
d(u) = \|H(u - u_0) + z_t\|_2^2
$$

(S1)

see (27) in our main paper. The color-constancy model gives us

$$
H = \begin{bmatrix}
Z_{ux}^R & Z_{xy}^R \\
Z_{ux}^G & Z_{xy}^G \\
Z_{ux}^B & Z_{xy}^B
\end{bmatrix}, \quad z_t = \begin{bmatrix}
z_t^R \\
z_t^G \\
z_t^B
\end{bmatrix}
$$

(S2)

in which $Z_{x,y}^{R,G,B}$ denotes the $x$- and $y$-derivatives of the target image in the $R$, $G$, and $B$ components, and $z_t^{R,G,B}$ are the difference of the reference image from the target one, in the $R$, $G$, and $B$ image components.

The gradient-constancy model on the other hand gives us the derivative data

$$
H = \begin{bmatrix}
Z_{xx} & Z_{xy} \\
Z_{yx} & Z_{yy}
\end{bmatrix}, \quad z_t = \begin{bmatrix}
z_{xt} \\
z_{yt}
\end{bmatrix}
$$

(S3)

in which $Z_{xx}$, $Z_{xy}$ and $Z_{yy}$ are the second-order derivatives of the target image, and $z_{xt}$ and $z_{yt}$ are the difference of the first-order derivatives of the reference image from the target ones. When the gradient constancy model is applied on each of the color channels, we obtain

$$
H = \begin{bmatrix}
Z_{ux}^R & Z_{xy}^R \\
Z_{ux}^G & Z_{xy}^G \\
Z_{ux}^B & Z_{xy}^B
\end{bmatrix}, \quad z_t = \begin{bmatrix}
z_t^R \\
z_t^G \\
z_t^B
\end{bmatrix}
$$

(S4)

in which we define the sub-matrices of $H$ and $z_t$ similarly to before.

Revaud et al. [4] use a weighted combination of two data terms $d(u)$ based on (S2) and (S4). This combination can be understood as forming new $H$ and $z_t$ by stacking the ones in (S2) and (S4). When the two data terms are combined using equal weights, the inverse covariance matrix $H^\dagger H$ becomes

$$
Z = \begin{bmatrix}
\sum z_{sx} z_{sx} & \sum z_{sx} z_{sy} \\
\sum z_{sx} z_{sy} & \sum z_{sy} z_{sy}
\end{bmatrix}
$$

(S5)

and the transformed signal is

$$
H^\dagger z = \begin{bmatrix}
\sum z_{sx} z_{xt} \\
\sum z_{sy} z_{yt}
\end{bmatrix}
$$

(S6)

cf. (27) in our main paper. In (S5)–(S6), the summations are over the three color channels for each of the 0th, and the 1st partial derivatives of the image. Figure 1 visualizes our flow estimates.

1.3. Image Deblurring

Figure 2 provides crops of the deblurred images from the the Kodak dataset[2], produced by different algorithms. We optimize the algorithm parameters for the different methods (Wiener, L2, and TV) via grid search. The Wiener filter uses a uniform image power spectrum model. Note the use of the bilateral filter is not optimal for de-noising as pointed out by Buades et al. [5], who demonstrate the advantages of patch-based filtering (nonlocal means denoising) over pixel-based filtering (bilateral filter). Our deblurring results are based on the bilateral filter, but one is free to use the non-local means filter (or any other filter) for the de-noising operator $\mathbf{A}$. 

Sean I. Young\(^1\) Aous T. Naman\(^2\) Bernd Girod\(^1\) David Taubman\(^2\)

\(^1\)Stanford University \(^2\)University of New South Wales
Figure 1: Optical flow (top rows) and the corresponding flow error (bottom rows) produced using the geodesic and the bilateral variants of our method. Whiter pixels correspond to smaller flow vectors.
Figure 2: Crops of images from the Kodak dataset when the B-spline blur kernel ($\tau = 8$) is used. Our method exhibits less ringing compared to the Wiener filter and the $L_2$-regularization methods, and has less staircasing artifacts than the $L_1$ (TV) method.
2. Possible Limitations

In Section 4 of our paper, we discussed that (14a) is valid only when \((C + \lambda L)^{-1}\) has a low-pass spectral response. We show this in Figure 4 (left) for the case where \(\lambda = 1\) and \(C = I\). Since \(C + \lambda L\) is Sinkhorn-normalized, it has a high-pass spectral response \(I + \lambda L\), ranging from 1 to 2. As a consequence, the inverse filter response \((I + \lambda L)^{-1}\) ranges from 1 down to 0.5. We can approximate such a filter response as a sum of low-pass and all-pass responses. In our context, an approximation of \(u^{opt} = (C + \lambda L)^{-1}Cz\) can be obtained using a convex combination of \(Cz\) and a low-pass-filtered version \(ACz\) of it. On the other hand, if \(I + \lambda L\) is a low-pass response. In this case, the inverse response (shown in Figure 4, right) is high-pass, and the solution \(u^{opt}\) cannot be approximated as a convex combination of \(Cz\) and a low-pass-filtered version of it. In practice, we can still use (14b) to solve the transformed problem.

References