Image reconstruction by matching gradient distributions -Supplemental materials

Taeg Sang Cho, C. Lawrence Zitnick, Neel Joshi, Sing Bing Kang, Rick Szeliski, William T. Freeman

1. The KL divergence between q_E and q_R

We show that the penalty function ρ_G defined in Algorithm 1 in the paper is one way of evaluating the KL divergence between the empirical distribution q_E and the reference distribution q_R.

Recall that the KL divergence between q_E and q_R is as follows:

$$KL(q_E||q_R) = \int_z q_E(z) \ln\left(\frac{q_E(z)}{q_R(z)}\right) dz \tag{1}$$

There are different ways to represent q_E. We can parameterize q_E as follows:

$$q_E(z) = \frac{\gamma_E \lambda_E^{\left(\frac{1}{\gamma_E}\right)}}{2\Gamma(\frac{1}{\gamma_E})} \exp\left(-\lambda_E \|z\|^{\gamma_E}\right) \tag{2}$$

where the shape parameters γ_E, λ_E have been fitted to N gradient samples ∇x_i using Eq. 7 in the paper.

We can also approximate q_E as follows:

$$\tilde{q_E}(z) = \frac{1}{N}\sum_i^N \delta(z - \nabla x_i) \tag{3}$$

Therefore,

$$\begin{aligned} KL(q_E||q_R) &= \int_z q_E(z) \ln\left(\frac{q_E(z)}{q_R(z)}\right) dz \\ &\approx \int_z \tilde{q_E}(z) \ln\left(\frac{q_E(z)}{q_R(z)}\right) dz \\ &= \sum_i^N \left\{ \frac{1}{N} \ln\left(\frac{q_E(\nabla x_i)}{q_R(\nabla x_i)}\right)\right\} \\ &= \frac{1}{N}\sum_i^N \rho_G(\nabla x_i) \end{aligned} \tag{4}$$

2. Fitting samples to a generalized Gaussian distribution

Claim 1 *Suppose $x_i, i = 1...N$ are samples from an unknown distribution, and we would like to fit a parametric distribution q to the samples x_i. Let $p_E(x) = \frac{1}{N}\sum_{i=1}^N \delta(x - x_i)$ be an empirical distribution of the samples x_i, and let q be a generalized Gaussian distribution parameterized by shape parameters λ, γ. We show that a distribution q that best parameterizes the empirical distribution q_E (in the KL divergence sense) minimizes the sum of negative log-likelihood over samples x_i:*

$$\min_{\lambda,\gamma} KL(p_E||q) = \min_{\lambda,\gamma}\left\{-\sum_{i=1}^N \ln(q(x_i))\right\} \tag{5}$$

Proof: We can show that the KL divergence between p_E and q takes the following form:

$$\begin{aligned} KL(p_E||q) &= \int_x p_E(x) \ln\left(\frac{p_E(x)}{q(x)}\right) dx \\ &= \int_x \frac{1}{N}\{\sum_{i=1}^N \delta(x - x_i)\} \ln\left(\frac{\frac{1}{N}\{\sum_{i=1}^N \delta(x-x_i)\}}{q(x)}\right) dx \\ &= \frac{1}{N}\sum_{i=1}^N \ln\left(\frac{\frac{1}{N}}{q(x_i)}\right) \\ &= -\ln N - \frac{1}{N}\sum_{i=1}^N \ln\left(q(x_i)\right) \end{aligned} \tag{6}$$

□

3. Algorithm details

We derive the details of the image reconstruction procedure in the paper. We can rewrite the image reconstruction optimization function as follows:

$$\begin{aligned} &\frac{\|y - k \otimes x\|^2}{2\eta^2} + w_1 \lambda_R \|\nabla x\|^{\gamma_R} \\ &+ w_2 \left(\lambda_R \|\nabla x\|^{\gamma_R} - \lambda_E \|\nabla x\|^{\gamma_E}\right) \\ &+ w_2 \ln\left(\frac{\gamma_E {\lambda_E}^{1/\gamma_E}}{2\Gamma(1/\gamma_E)} \frac{2\Gamma(1/\gamma_R)}{\gamma_R {\lambda_R}^{1/\gamma_R}}\right) \end{aligned} \tag{7}$$

The shape parameters of the empirical distribution q_E are functions of x, but dependences are omitted to reduce clutter.

The first two rows of (7) are similar in form to the ordinary MAP estimator, therefore they can be minimized using a gradient descent technique. If we can compute the derivative of $\ln\left(\frac{\gamma_E {\lambda_E}^{1/\gamma_E}}{2\Gamma(1/\gamma_E)} \frac{2\Gamma(1/\gamma_R)}{\gamma_R {\lambda_R}^{1/\gamma_R}}\right)$ with respect to x, we can minimize the entire function in (7) using a gradient descent method. We show that it indeed is the case.

Let X be a rasterized vector of the image x. The derivative of $\ln\left(\frac{\gamma_E {\lambda_E}^{1/\gamma_E}}{2\Gamma(1/\gamma_E)} \frac{2\Gamma(1/\gamma_R)}{\gamma_R {\lambda_R}^{1/\gamma_R}}\right)$ with respect to X takes the following form:

$$
\begin{aligned}
&\frac{\partial}{\partial X} \ln\left(\frac{\gamma_E {\lambda_E}^{1/\gamma_E}}{2\Gamma(1/\gamma_E)} \frac{2\Gamma(1/\gamma_R)}{\gamma_R {\lambda_R}^{1/\gamma_R}}\right) = \\
&\alpha \frac{\partial \gamma_E}{\partial X} + \beta \frac{\partial \lambda_E}{\partial X}
\end{aligned}
\tag{8}
$$

where

$$
\begin{aligned}
\alpha &= \left(\frac{1}{\gamma_E} - \frac{\ln(\lambda_E)}{{\gamma_E}^2} + \frac{\Psi\left(\frac{1}{\gamma_E}\right)}{{\gamma_E}^2}\right) \\
\beta &= \left(\frac{1}{\gamma_E \lambda_E}\right)
\end{aligned}
\tag{9}
$$

Ψ is a digamma function. $\frac{\partial \gamma_E}{\partial X}$ and $\frac{\partial \lambda_E}{\partial X}$ can be derived as follows:

$$
\begin{aligned}
\frac{\partial \gamma_E}{\partial X} &= \frac{{\gamma_E}^2 {\lambda_E}^{(\frac{2}{\gamma_E})} \Gamma(\frac{1}{\gamma_E})}{N\Gamma(\frac{3}{\gamma_E})\left(\Psi(\frac{1}{\gamma_E}) - 3\Psi(\frac{3}{\gamma_E}) + 2\ln(\lambda_E)\right)} 2G^T GX \\
\frac{\partial \lambda_E}{\partial X} &= -\frac{\Gamma(1/\gamma_E)\gamma_E {\lambda_E}^{(1+2/\gamma_E)}}{N\Gamma(3/\gamma_E)} G^T GX
\end{aligned}
\tag{10}
$$

We show the proofs in the following subsections. Since $\ln\left(\frac{\gamma_E {\lambda_E}^{1/\gamma_E}}{2\Gamma(1/\gamma_E)} \frac{2\Gamma(1/\gamma_R)}{\gamma_R {\lambda_R}^{1/\gamma_R}}\right)$ is differentiable, we can optimize (7) using a gradient descent technique. Furthermore, at fixed $[\gamma_E, \lambda_E]$, (8) is linear in X, suggesting that an iterative reweighted least squares (IRLS) method can minimize (7).

Let Y be a rasterized vector of the observed image y, and K be the convolution matrix of the blur kernel k. We take the derivative of the optimization function (7) with respect to X:

$$
\begin{aligned}
&-\frac{K^T(Y - KX)}{\eta^2} + 2w_1 \lambda_R \gamma_R G^T \|GX\|^{\gamma_R - 1} \\
&+ w_2 \Big(2\lambda_R \gamma_R G^T \|GX\|^{\gamma_R - 1} - 2\lambda_E \gamma_E G^T \|GX\|^{\gamma_E - 1} \\
&+ (\alpha - \lambda_E |GX|^{\gamma_E} \ln(|GX|)) \circ \frac{\partial \gamma_E}{\partial X} \\
&+ (\beta - |GX|^{\gamma_E}) \circ \frac{\partial \lambda_E}{\partial X} \Big)
\end{aligned}
\tag{11}
$$

where G is a gradient operator, and $\circ$ is a Hadamard element-wise matrix multiplication operator.

IRLS algorithm approximates the solution of a non-linear equation (11) by iteratively solving a linear equation that approximates (11). We approximate $\gamma G^T \|GX\|^{\gamma-1}$ as follows:

$$
\gamma G^T \|GX\|^{\gamma - 1} = \gamma G^T W G X \tag{12}
$$

where W is a reweighting matrix. We update W iteratively such that minimizing $\gamma G^T \|GX\|^{\gamma-1}$ matches minimizing $\gamma G^T W G X$.

We handle the non-linearity due to $\lambda_E |GX|^{\gamma_E} \ln(|GX|)$ and $|GX|^{\gamma_E}$ by evaluating them once with the image reconstructed from the previous iteration, and *fixing these coefficients* during the actual minimization with respect to X. We iterate this process until convergence. We use a minimum residual method to solve the *linear* system in (11).

We can easily modify this algorithm to derive the IDR algorithm details.

3.1. The derivative of $\ln\left(\frac{\gamma_E {\lambda_E}^{1/\gamma_E}}{2\Gamma(1/\gamma_E)} \frac{2\Gamma(1/\gamma_R)}{\gamma_R {\lambda_R}^{1/\gamma_R}}\right)$

We note that γ_R, λ_R are independent of X, so we can focus on taking the derivative of γ_E, λ_E. We can rewrite $\ln\left(\frac{\gamma_E {\lambda_E}^{1/\gamma_E}}{2\Gamma(1/\gamma_E)}\right)$ as follows:

$$
\ln\left(\frac{\gamma_E {\lambda_E}^{1/\gamma_E}}{2\Gamma(1/\gamma_E)}\right) = \ln(\gamma_E) + \frac{1}{\gamma_E}\ln(\lambda_E) - \ln\left(2\Gamma(\frac{1}{\gamma_E})\right) \tag{13}
$$

There exists a relationship between the Gamma function Γ and the digamma function Ψ:

$$
\frac{d\Gamma(z)}{dz} = \Gamma(z)\Psi(z) \tag{14}
$$

We can use that relationship to show that

$$
\begin{aligned}
&\frac{\partial}{\partial X} \ln\left(\frac{\gamma_E {\lambda_E}^{1/\gamma_E}}{2\Gamma(1/\gamma_E)} \frac{2\Gamma(1/\gamma_R)}{\gamma_R {\lambda_R}^{1/\gamma_R}}\right) \\
&\quad = \frac{1}{\gamma_E}\frac{\partial \gamma_E}{\partial X} + \frac{1}{\gamma_E \lambda_E}\frac{\partial \lambda_E}{\partial X} - \frac{1}{{\gamma_E}^2}\ln(\lambda_E)\frac{\partial \gamma_E}{\partial X} \\
&\qquad + \frac{1}{{\gamma_E}^2}\Psi(\frac{1}{\gamma_E})\frac{\partial \gamma_E}{\partial X} \\
&\quad = \alpha\frac{\partial \gamma_E}{\partial X} + \beta\frac{\partial \lambda_E}{\partial X}
\end{aligned}
\tag{15}
$$

3.2. The derivative of λ_E with respect to X

We show that

$$
\frac{\partial \lambda_E}{\partial X} = -\frac{\Gamma(1/\gamma_E)\gamma_E {\lambda_E}^{(1+2/\gamma_E)}}{N\Gamma(3/\gamma_E)} G^T GX \tag{16}
$$

where N is the total number of samples.

We can compute the second moment m_2 of gradient samples of X as follows:

$$m_2 = \frac{1}{N} X^T G^T G X \tag{17}$$

where G is a gradient operator, and we assume that the mean of gradients GX is zero.

The second moment m_2 is related to generalized Gaussian shape parameters γ_E, λ_E as follows:

$$m_2 = \frac{\Gamma(3/\gamma_E)}{\lambda_E^{\frac{2}{\gamma_E}} \Gamma(1/\gamma_E)} \tag{18}$$

We take the derivative of m_2 with respect to X. From (17),

$$\frac{\partial m_2}{\partial X} = \frac{2}{N} G^T G X \tag{19}$$

For tractability, we assume that γ_E is independent of X. From (18),

$$\frac{\partial m_2}{\partial X} = \frac{\Gamma(3/\gamma_E)}{\Gamma(1/\gamma_E)} \frac{2}{\gamma_E} \lambda_E^{-\frac{2}{\gamma_E}-1} \frac{\partial \lambda_E}{\partial X} \tag{20}$$

From (19) and (20), we can show that

$$\frac{\partial \lambda_E}{\partial X} = -\frac{\Gamma(1/\gamma_E) \gamma_E \lambda_E^{(1+2/\gamma_E)}}{N \Gamma(3/\gamma_E)} G^T G X \tag{21}$$

3.3. The derivative of γ_E with respect to X

We show that

$$\frac{\partial \gamma_E}{\partial X} = \frac{\gamma_E^2 \lambda_E^{(\frac{2}{\gamma_E})} \Gamma(\frac{1}{\gamma_E})}{N \Gamma(\frac{3}{\gamma_E}) \left(\Psi(\frac{1}{\gamma_E}) - 3\Psi(\frac{3}{\gamma_E}) + 2\ln(\lambda_E) \right)} 2 G^T G X \tag{22}$$

where N is the total number of samples.

Again, we use the relationship:

$$m_2 = \frac{\Gamma(3/\gamma_E)}{\lambda_E^{\frac{2}{\gamma_E}} \Gamma(1/\gamma_E)} \tag{23}$$

We take the derivative of m_2 with respect to X assuming that m_2 is independent of λ_E.

$$\begin{aligned} \frac{\partial m_2}{\partial X} = & \frac{1}{\left(\gamma_E^{\left(\frac{2}{\gamma_E}\right)} \Gamma\left(\frac{1}{\gamma_E}\right) \right)^2} \times \\ & \left\{ \Gamma(\frac{3}{\gamma_E}) \Psi(\frac{3}{\gamma_E}) (-\frac{3}{\gamma_E^2}) \lambda_E^{\frac{2}{\gamma_E}} \Gamma(\frac{1}{\gamma_E}) \right. \\ & \left. -\Gamma(\frac{3}{\gamma_E}) \left(\frac{\partial}{\partial \gamma_E} \left(\lambda_E^{\frac{2}{\gamma_E}} \Gamma(\frac{1}{\gamma_E}) \right) \right) \right\} \end{aligned} \tag{24}$$

We can show that

$$\begin{aligned} & \left(\frac{\partial}{\partial \gamma_E} \left(\lambda^{\frac{2}{\gamma_E}} \Gamma(\frac{1}{\gamma_E}) \right) \right) \\ & = -\lambda_E^{\frac{2}{\gamma_E}} \Gamma(\frac{1}{\gamma_E}) \left(\frac{1}{\gamma_E^2} \right) \left(\Psi(\frac{1}{\gamma_E}) + 2\ln(\lambda_E) \right) \end{aligned} \tag{25}$$

Using above relationships and the derivative of m_2 with respect to X ((19)), we can show that

$$\frac{\partial \gamma_E}{\partial X} = \frac{\gamma_E^2 \lambda_E^{(\frac{2}{\gamma_E})} \Gamma(\frac{1}{\gamma_E})}{N \Gamma(\frac{3}{\gamma_E}) \left(\Psi(\frac{1}{\gamma_E}) - 3\Psi(\frac{3}{\gamma_E}) + 2\ln(\lambda_E) \right)} 2 G^T G X \tag{26}$$

4. User comments

In our user study, we asked users to comment on their selection of the visually pleasing image. We present a subset of comments from the users.

4.1. Comments from users that favored the image reconstructed using the IDR algorithm

- road/gravel is clearer
- I like the picture on the right more because certain spots of the picture have more detail than the picture on the left and the yellow in the train seems to POP more
- Detail looks more realistic.
- bushes are clearer
- the picture on the left is more clear
- Detailing looks more realistic. The second one looks like a painting.
- not sure. just more appealing
- The fur on the mother bear was more visible and real like than the first picture. Detailing was shown slightly more in the cubs land and water as well.
- Mother bear's fur is more realistic
- Can see gravel more clearly
- Leaves on trees in background look more distinct
- Better resolution
- theres more detail and not as blury
- the image is sharper
- the color is more vivid and you can see the true color of the bush rather than the blur.
- Focused mainly on the clarity of the tree in front. Branches seemed more defined than the other tree. Building looked nearly the same though.
- a bit crisper imagery

- I like how the trees/bushes look more detailed more real
- You can see the individual hairs
- In many places on the selected image the hair looks more realistic (grainier fluffier and less of a blob).
- prefer the look of the grassy bank in this one as it looks clearer - the other just looks like a smudge
- The 2nd image is a bit more focused than the other.
- seems a slightly more focused picture so it's clamer on the eye
- A tiny bit more detail can be seen on the path.
- The trees in the selected image are much more in focus. Overall the image is less blurry but I can make out individual details about the path and the trees.
- the path in the foreground seems more natural
- seems a little more in focus- looking at the grass as the rest of the pic seems equally as unfocused
- the other image looks like some one poured water on it

4.2. Comments from users that favored MAP estimates

- neat and clean
- I picked the one that looked a little abstract like a painting
- The train in the selected image looks much clearer and the building in the background seems less blurry after staring at the two for a while.
- better clarity
- less specks
- less blocky
- The leaf in the road is easier to see.
- Sharper focus
- yellow vehicle more defined
- Less pixely than the other one
- it looks like an artist's rendering
- just a bit crisper
- clearer image w/ less specks
- this is little more clear
- clearer images

4.3. Comments from users that selected "*There is no difference.*" option

- Both are too bright.
- I don't see a difference.
- both images are looking in every aspect same to me.
- Leaves of tree seem to be better focused - slightly
- seems a little more focused - looking at the path leading to the little conifer
- Each photo had attributes that was more appealing in presentation than the other. Picture one seemed to be slightly clearer with the larger tree in front view. While picture two the side of the building and sky was clearer.
- I focused on the telephone pole
- look the same