Bilateral Video Magnification Filter
Supplementary Materials

Shoichiro Takeda[1] Kenta Niwa[1] Mariko Isogawa[1] Shinya Shimizu[1] Kazuki Okami[2] Yushi Aono[1]
[1]NTT Corporation [2]NTT TechnoCross Corporation
{shoichiro.takeda.us, kenta.niwa.bk, shinya.shimizu.te, yushi.aono.dy}@hco.ntt.co.jp
mariko.isogawa@ieee.org, kazuki.okami.ug@ntt-tx.co.jp

Contents

0. Introduction

This supplementary document contains the following five points related to our main paper titled bilateral video magnification filter (BVMF). We highlight reference numbers associated with our main paper in blue, and those associated with this supplementary material in red.

1. We show details of the derivations of Eqs. (6, 7), noted in subsection 4.1.
2. We show details of the setting of the length of the finite range $|\mathcal{T}|$ of BVMF, noted in subsection 4.1.
3. We evaluate the generalized BVMF that uses a temporal FIR filter instead of the LoG kernel, noted at the end of subsection 4.2.
4. In terms of excluding large motions of objects in a video, we further evaluate BVMF efficacy with different amplitudes of large motions A_3, noted in "Robustness against Large Motions (Synthesis 2)" in subsection 5.1.
5. Since the parameter tuning for each method requires different time and effort, the main paper fairly compared each method with the same parameters: the target frequency f_t and the amplification factor α. However, here we check the best results of each method via parameter tuning and validate the effectiveness of BVMF in depth.

Also, please refer to the supplementary video in the supplementary materials file, which contains the video version of the results in our main paper.

1. Derivation of Equations

1.1. Derivation of Eq. (6)

In our main paper, Eq. (6) determines the standard deviation σ_{f_t} of the Laplacian of Gaussian (LoG) kernel of Eq. (4) to maximize the passband of the LoG kernel at f_t. The LoG kernel is defined as

$$LoG(t;\sigma_{f_t}) := -\frac{t^2-\sigma_{f_t}^2}{Z\sigma_{f_t}^4}\exp\left(-\frac{t^2}{2\sigma_{f_t}^2}\right), \tag{1}$$

and we finally derived the optimal σ_{f_t} described in Eq. (6) as

$$\nabla_f \mathcal{F}[LoG(t;\sigma_{f_t})](f_t) = 0 \quad \Leftrightarrow \quad \sigma_{f_t} = \frac{\sqrt{2}}{2\pi f_t}. \tag{2}$$

For deriving Eq. (6), Eq. (2) here, we first considered a relationship between the Gaussian function $G(t;\sigma_{f_t})$ and its Fourier transform. The Gaussian function $G(t;\sigma_{f_t})$ is defined as

$$G(t;\sigma_{f_t}) := \frac{1}{\sqrt{2\pi}\sigma_{f_t}}\exp\left(-\frac{t^2}{2\sigma_{f_t}^2}\right), \tag{3}$$

and its Fourier transform is given by

$$\mathcal{F}[G(t;\sigma_{f_t})](f) = \exp\left(-2\sigma_{f_t}^2\pi^2 f^2\right), \tag{4}$$

where $\mathcal{F}[\cdot](f)$ is the 1D Fourier spectrum, namely the bandpass frequency response, of the input at a frequency f.

From Eqs. (3, 4), the second derivative of the Gaussian function $G(t;\sigma_{f_t})$ and its Fourier transform are given by

$$\frac{\partial^2}{\partial t^2}G(t;\sigma_{f_t}) = \frac{1}{\sqrt{2\pi}\sigma_{f_t}}\frac{t^2-\sigma_{f_t}^2}{\sigma_{f_t}^4}\exp\left(-\frac{t^2}{2\sigma_{f_t}^2}\right), \tag{5}$$

$$\mathcal{F}\left[\frac{\partial^2}{\partial t^2}G(t;\sigma_{f_t})\right](f) = -f^2\exp\left(-2\sigma_{f_t}^2\pi^2 f^2\right). \tag{6}$$

Note that we can get Eq. (6) via the Fourier differentiation theorem with Eq. (4). Thus, from Eqs. (5, 6), the Fourier transform of the LoG kernel $\mathcal{F}[LoG(t;\sigma_{f_t})](f)$ can be described using the Gaussian function as

$$\begin{aligned}
\mathcal{F}[LoG(t;\sigma_{f_t})](f) &= \mathcal{F}\left[-\frac{\sqrt{2\pi}\sigma_{f_t}}{Z}\frac{\partial^2}{\partial t^2}G(t;\sigma_{f_t})\right](f) \qquad (\because \text{Eqs. (1, 5)}) \\
&= -\frac{\sqrt{2\pi}\sigma_{f_t}}{Z}\mathcal{F}\left[\frac{\partial^2}{\partial t^2}G(t;\sigma_{f_t})\right](f) \\
&= \frac{\sqrt{2\pi}\sigma_{f_t}}{Z}f^2\exp\left(-2\sigma_{f_t}^2\pi^2 f^2\right) \qquad (\because \text{Eq. (6)}).
\end{aligned} \tag{7}$$

Figure 1 shows $\mathcal{F}[LoG(t;\sigma_{f_t})](f)$ with respect to f. This figure indicates that $\mathcal{F}[LoG(t;\sigma_{f_t})](f)$ has a single maximum peak at which the gradient $\nabla_f\mathcal{F}[LoG(t;\sigma_{f_t})](f)$ is 0 if $f, \sigma_{f_t} > 0$. Therefore, to set the peak of the passband of the LoG kernel at f_t, we should find the σ_{f_t} at which the gradient $\nabla_f\mathcal{F}[LoG(t;\sigma_{f_t})](f_t)$ is 0. From this fact, we derived Eq. (6), Eq. (2) here, and analytically solved it as follows.

$$\begin{aligned}
&\nabla_f\mathcal{F}[LoG(t;\sigma_{f_t})](f_t) = 0 \\
\Leftrightarrow\ &\frac{\partial}{\partial f}\left.\frac{\sqrt{2\pi}\sigma_{f_t}}{Z}f^2\exp\left(-2\sigma_{f_t}^2\pi^2 f^2\right)\right|_{f=f_t} = 0 \qquad (\because \text{Eq. (7)}) \\
\Leftrightarrow\ &\frac{\sqrt{2\pi}\sigma_{f_t}}{Z}f_t(4\sigma_{f_t}^2\pi^2 f_t^2 - 2)\exp\left(-2\sigma_{f_t}^2\pi^2 f_t^2\right) = 0 \\
\Leftrightarrow\ &(4\sigma_{f_t}^2\pi^2 f_t^2 - 2) = 0 \qquad (\because \sigma_{f_t}, Z, f_t > 0) \\
\Leftrightarrow\ &\sigma_{f_t} = \frac{\sqrt{2}}{2\pi f_t} \qquad (\because \sigma_{f_t} > 0)
\end{aligned} \tag{8}$$

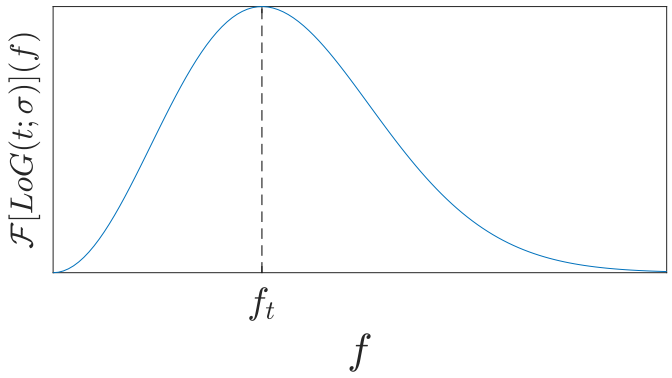

Figure 1. The Fourier transform of the LoG kernel $\mathcal{F}[LoG(t;\sigma_{f_t})](f)$ with respect to f. The imagery part of $\mathcal{F}[LoG(t;\sigma_{f_t})](f)$ is 0 and thus the vertical axis indicates the real part of $\mathcal{F}[LoG(t;\sigma_{f_t})](f)$, which means the bandpass frequency response of the LoG kernel.

1.2. Derivation of Eq. (7)

In our main paper, the purpose of Eq. (7) is to keep the original magnitude of subtle variations with f_t after applying BVMF. For this purpose, we newly formulated and solved Eq. (7) for Z so that the peak gain of the passband of $LoG(t;\sigma_{f_t})$ is unity, namely 1.0, at f_t as follows:

$$
\begin{aligned}
&|\mathcal{F}[LoG(t;\sigma_{f_t})](f_t)| = 1 \\
\Leftrightarrow\ & \frac{\sqrt{2\pi}\sigma_{f_t}}{Z} f_t^2 \exp\left(-2\sigma_{f_t}^2\pi^2 f_t^2\right) = 1 \qquad (\because \text{Eq. (7)}) \\
\Leftrightarrow\ & Z = \sqrt{2\pi}\sigma_{f_t} f_t^2 \exp\left(-2\sigma_{f_t}^2\pi^2 f_t^2\right) \\
\Leftrightarrow\ & Z = \mathcal{F}[Z \cdot LoG(t;\sigma_{f_t})](f_t) \qquad (\because \text{Eq. (7)}) \\
\Leftrightarrow\ & Z = \left|\mathcal{F}\left[-\frac{t^2-\sigma_{f_t}^2}{\sigma_{f_t}^4}\exp\left(-\frac{t^2}{2\sigma_{f_t}^2}\right)\right](f_t)\right| \qquad (\because \text{Eq. (1)}).
\end{aligned}
\tag{9}
$$

2. Setting of the Length of the Finite Range of BVMF

Considering, to extract a signal with f_t, the length of the finite range of BVMF $|\mathcal{T}|$ must have one or more full f_t Hz cycles of the input signal, we obtain this constraint as

$$
|\mathcal{T}| \geqq \frac{f_s}{f_t},
$$

where f_s is a sampling rate.

From this constraint, the minimum length of $\mathcal{T}$ can be described as $|\mathcal{T}| = f_s/f_t$, which has the lowest computational cost. However, this minimum length leads to a coarse set of frequency resolution bins defined as $\{f_{\text{bin}} = m\Delta f = mf_s/|\mathcal{T}| = mf_t \mid m = 0, \ldots, |\mathcal{T}| - 1\}$. This coarse set has no frequency resolution bins between 0 and f_t Hz. Thus, if this minimum length $|\mathcal{T}| = f_s/f_t$ is used, BVMF mis-extracts all signals with lower frequency (except for 0 Hz) than f_t as the target signal with f_t.

From this fact, we set $|\mathcal{T}|$ in our main paper as

$$
|\mathcal{T}| = \frac{2f_s}{f_t}; \tag{10}
$$

this gives a set of frequency resolution bins defined as

$$
\left\{ f_{\text{bin}} = \frac{mf_t}{2} \;\middle|\; m = 0, \ldots, |\mathcal{T}| - 1 \right\}.
$$

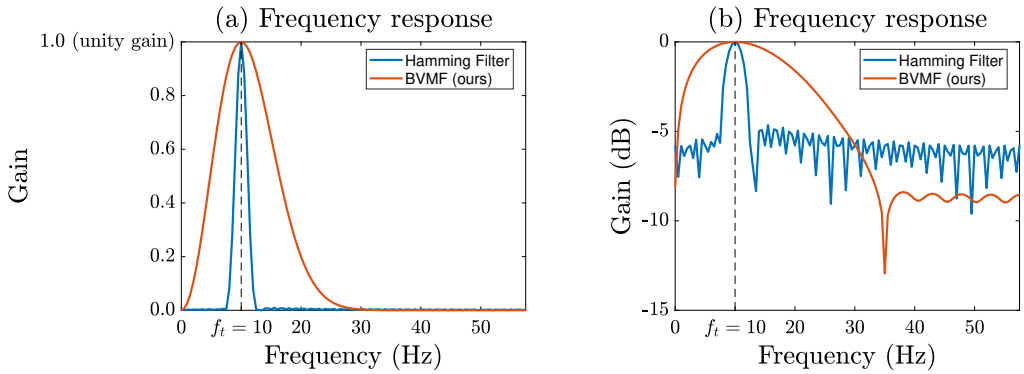

Figure 2. Bandpass frequency response of BVMF and the Hamming filter which is used in the generalized BVMF instead of the LoG kernel. The only difference between (a) and (b) is the scale of y-axis. In this comparison, the target frequency $f_t = 10$ Hz with a sampling rate $f_s = 120$ Hz. Moreover, to obtain narrower passband than the LoG kernel, we design the Hamming filter to have the four times length of the finite range of the LoG kernel.

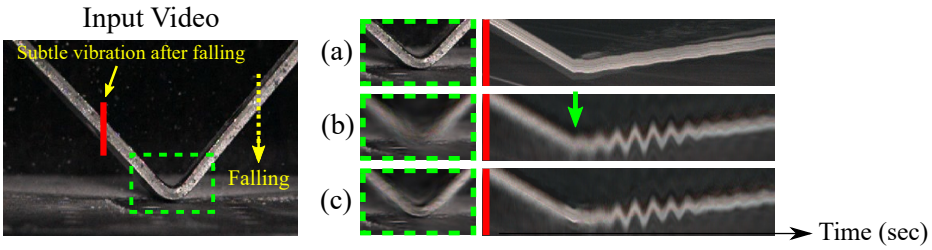

Figure 3. Motion magnification results for subtle vibrations of a corner brace after falling. The right panels show enlarged image frames in the green dot square in the input video at the moment of falling to the ground and spatiotemporal slices along the red line in the input video. (a) Original. (b) Hamming filter alone. (c) Generalized BVMF using the Hamming filter instead of the LoG kernel.

This set of frequency resolution bins has a frequency resolution bin between 0 and f_t Hz: i.e., $f_{\text{bin}} = \frac{f_t}{2}$. Therefore, using Eq. (10), the target signal with f_t can be isolated from signals with lower frequency than f_t, and thus BVMF can effectively extract the target signal with f_t.

3. Generalization of BVMF

One characteristic of the LoG kernel is to cut off some slow large motions that approximate linearly. Such linearity is often small and may be missed by the Gaussian kernel of Eq. (8) in BVMF. Thus, BVMF needs the LoG kernel to exclude large motions precisely. Meanwhile, other filters (e.g., a very narrow bandpass filter) may be good alternatives to the LoG kernel depending on the situation. As formulated in Eq. (3), LoGF kernel can be replaced with any FIR filter. Thus, we can design the generalized BVMF $\hat{\Gamma}(s(t), t)$ that can be applied to each specific situation by using an FIR filter $q(t; f_t)$ instead of the LoG kernel as

$$\begin{aligned} C_{\nu_n,\theta,f_t}(\mathbf{x}, t) &= \hat{\Gamma}(s(t), t) * S_{\nu_n,\theta}(\mathbf{x}, t), \\ \hat{\Gamma}(s(t), t) &:= G(s(t); \sigma_\varepsilon)\, q(t; f_t). \end{aligned} \tag{11}$$

For example, we used a Hamming filter for $q(t; f_t)$ as one of the narrower bandpass filters than the LoG kernel (see Fig. 2) and applied this generalized BVMF to a video, where a corner brace falls and bounces, to magnify subtle narrow-band vibrations of the object. Figure 3 shows that the generalized BVMF using a Hamming filter can magnify subtle narrow-band vibrations of the object clearly after it falls, even though various frequency vibrations occur (the Hamming filter alone cannot exclude large deceleration motion at the moment of falling to the ground, see the green dot squares). Therefore, our BVMF can be generalized to use any FIR filter instead of the LoG kernel.

To use IIR filters (e.g., a Butterworth filter) instead of the LoG kernel, we should reformulate BVMF into a recursive form. Fortunately, in the context of improving computational complexity, a recursive bilateral filter has been proposed [2], and thus

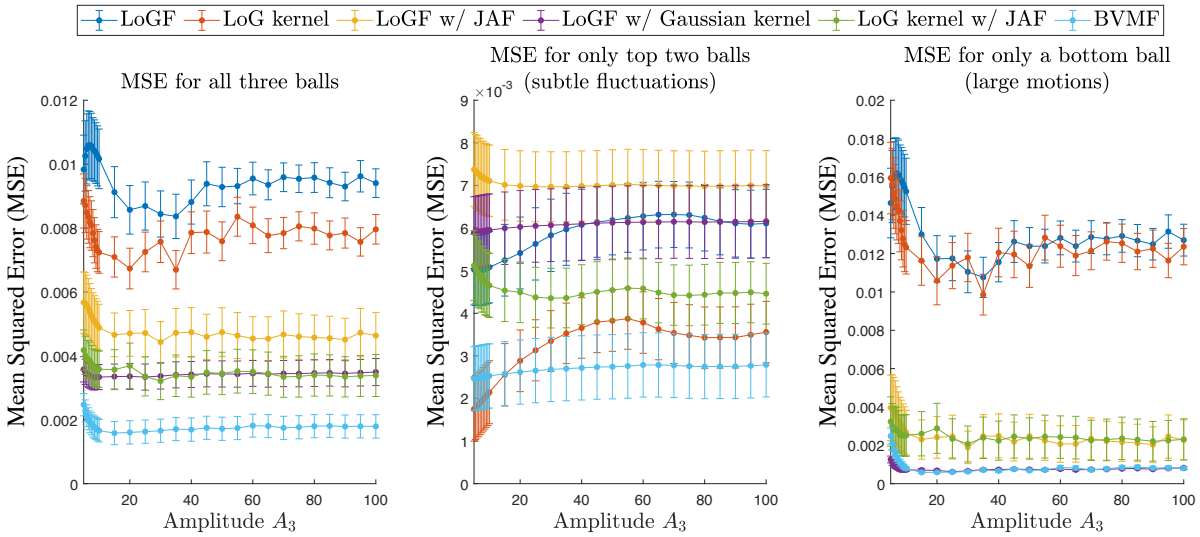

Figure 4. Mean squared error (MSE) with different amplitude $A_3 \in [5, 100]$. BVMF maintained the lowest MSE over all A_3 compared with the existing temporal filtering [3, 1] and the ablation studies (namely, the LoG kernel, LoGF w/ Gaussian kernel, and LoG kernel with JAF).

we will seek a way to incorporate this recursive technique into BVMF.

4. Further Evaluation for Effect of Large Motions of Objects

In this section, we further evaluated BVMF in terms of its robustness against large motions of objects. This experiment follows the experimental setting in "Robustness against Large Motions (Synthesis 2)" in the subsection 5.1, except for the use of different values of amplitude $A_3 \in [5, 100]$. Specifically, we created a synthetic video where the amplitude of the bottom ball motions was set as $A_3 \in [5, 100]$. In this evaluation, we applied each temporal filter to the synthetic video so as to magnifying only subtle fluctuations $d_{1,2}$ of the top two balls, but ignore large motions d_3 of the bottom ball.

Figure 4 shows mean squared error (MSE) and its scaled standard deviation over time frames at each A_3 against the ground-truth. Note that the scaled standard deviation is one-fifth of the original to fit in the plotting range of Fig. 4. Figure 4 (left) shows MSE for all three balls, the center shows MSE for only the top two balls that subtly fluctuated, while the right shows MSE for only the bottom ball that moved largely. LoGF [3] and the LoG kernel (ours) showed high MSEs over all A_3 for all three balls (see left plot) due to the collapsed shape of the bottom ball (see right plot). LoGF with JAF [1], LoGF with the Gaussian kernel, and the LoG kernel with JAF showed lower MSE than that achieved by using LoGF or the LoG kernel alone because they have a non-linear weighting filter (namely, JAF or the Gaussian kernel) that excludes large motions. However, their MSEs were not enough low because they have either the problem that large de/acceleration motions cannot be excluded by JAF or that the subtle fluctuations are magnified incorrectly by using LoGF. In contrast, BVMF yields the lowest MSEs with low standard deviation (see left, center, and right plots). This indicates that BVMF magnifies only subtle fluctuations of the top two balls (close to the ground-truth) while excluding large de/acceleration motions of the bottom ball, which maintains the shape of the bottom ball. Therefore, these results indicate that BVMF offers the best EVM performance even under the presence of large motions.

5. Magnification Results of Each Temporal Filter with Parameter Tuning

Our proposed BVMF offers rigorous physical meaning; its passband peaks at the target frequency with unity gain. In the main paper, since the parameter tuning for each method requires different time and effort, we fairly compared each method with the same parameters f_t, α at each experiment. On the other hand, in practice when a user performs the EVM algorithm, the user often cannot know the true target frequency f_t of temporal variations of interest and the desired amplification factor α in advance. Thus, the user may have to play around with these parameters to get the best results.

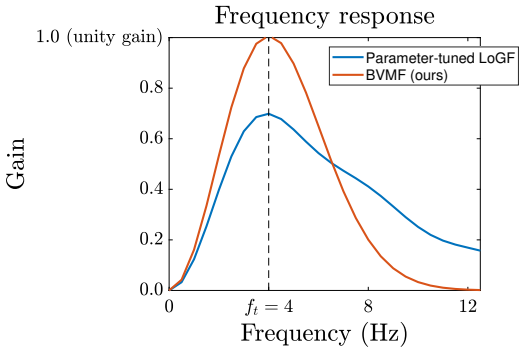

Figure 5. Bandpass frequency response of BVMF and the parameter-tuned LoGF. In this comparison, a target frequency $f_t = 4.0$ Hz with a sampling rate $f_s = 30$ Hz. Note that the parameter-tuned LoGF needs $f_t = 3$ to set its passband peak at 4 Hz.

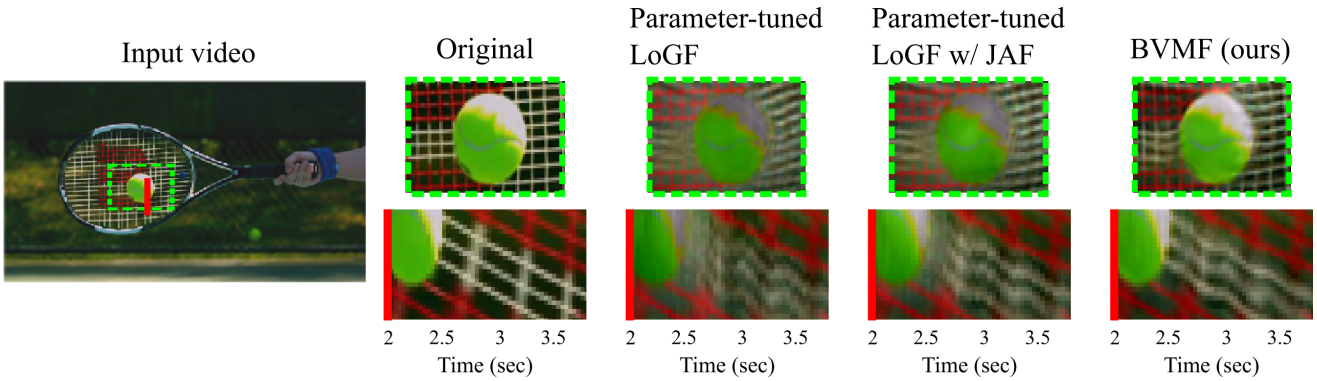

Figure 6. Motion magnification results for subtle string vibrations when a tennis racket hits a ball. The top panels show enlarged image frames in the green dot square in the input video at impact (the ball suddenly stops). The bottom panels show spatiotemporal slices along the red line in the input video.

From this point of view, we check the best result of LoGF [3] via parameter tuning. Figure 5 shows the passband of the parameter-tuned LoGF and BVMF. To set the passband peak at 4 Hz, BVMF uses $f_t = 4$ while the parameter-tuned LoGF needs $f_t = 3$. Moreover, to magnify subtle variations as much as BVMF does, we set α in the parameter-tuned LoGF as α divided by the peak of its passband (namely, α divided by 0.6989) when BVMF takes α. These parameters' settings enable the parameter-tuned LoGF to magnify subtle variations with 4.0 Hz as much as BVMF does.

Figure 6 shows motion magnification results for subtle string vibrations when a tennis racket hits a ball. Comparing Fig. 6 with Fig. 1, the parameter-tuned LoGF (with JAF) magnified the subtle string vibrations as well as BVMF did, but slightly mis-magnified non-target higher frequency string vibrations (see the bottom panels) due to its wide passband (see Fig. 5). Furthermore, the parameter-tuned LoGF (with JAF) collapsed the ball shape due to the suddenly stopping ball by hitting (see the top panels). In contrast, BVMF magnified subtle string vibrations with the target frequency while maintaining the ball shape. Therefore, BVMF is superior to LoGF even if the user plays around with these parameters (f_t, α) to get the best results.

References

[1] Shoichiro Takeda, Kazuki Okami, Dan Mikami, Megumi Isogai, and Hideaki Kimata. Jerk-aware video acceleration magnification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1769–1777, June 2018. 5

[2] Qingxiong Yang. Recursive bilateral filtering. In *The European Conference on Computer Vision (ECCV)*, pages 399–413, 2012. 4

[3] Yichao Zhang, Silvia L. Pintea, and Jan C. van Gemert. Video acceleration magnification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 502–510, 2017. https://acceleration-magnification.github.io/. 5, 6