Handling Motion Blur in Multi-Frame Super-Resolution

Ziyang Ma[1] Renjie Liao[2] Xin Tao[2] Li Xu[2] Jiaya Jia[2] Enhua Wu[1,3]

[1]University of Chinese Academy of Sciences &
State Key Lab. of Computer Science, Inst. of Software, CAS

[2]The Chinese University of Hong Kong

[3]FST, University of Macau, Macao China

1. Notations

In this supplementary material, we elaborate our derivations in our algorithm and analysis. Before going into the details, we would like to review our notations in the paper to make the derivations more readable.

First of all, we denote the set of input low-res images as $\Omega = \{I^L_{-N}, \cdots, I^L_0, \cdots, I^L_N\}$. Multi-frame SR aims to estimate a high-res image I corresponds to I^L_0. Here, I is a vector representing the latent high-res image. $F = \{F_{0\to -N}, \cdots, F_{0\to 0}, \cdots, F_{0\to N}\}$ is a set of warping matrices corresponding to the optical flow/homography from I to every other frame. Matrices S and $K = \{K_{-N}, \cdots, K_0, \cdots, K_N\}$ correspond to the down-sampling and filtering operations. Here, each K_i is a composition of a motion blur kernel and an anti-aliasing kernel. We define a set of latent masks $Z = \{Z_{-N}, \cdots, Z_0, \cdots, Z_N\}$ to indicate whether each pixel of each input image is useful ($Z = 1$) or useless ($Z = 0$) for SR. Each Z_i has the same size as the low-res image I^L_i. In the following derivations, we usually denote the p-th pixel of the i-th image J_i as $J_{i,p}$.

2. Derivations of the EM steps in image reconstruction

In this section, we explicitly derive the EM steps for image reconstruction. First, according to our inference algorithm, image reconstruction amounts to the following MAP problem:

$$I = \arg\max_I P(I, K, F|\Omega) = \arg\max_I P(I)P(\Omega|I, K, F). \tag{1}$$

Written as a marginalization over the latent variable Z, the above equation is equivalent to:

$$I = \arg\max_I P(I) \prod_{i=-N}^{N} \sum_{Z_i} P(I^L_i, Z_i|I, K, F). \tag{2}$$

The likelihood can be pixel-wisely decomposed as:

$$\begin{aligned} P(I^L_i, Z_i|I, K, F) &= \prod\nolimits_p P(I^L_{i,p}, Z_{i,p}|I, K, F) \\ &= \prod\nolimits_p P(I^L_{i,p}|Z_{i,p}, I, K, F)P(Z_{i,p}|I, K, F). \end{aligned} \tag{3}$$

In the paper, we have defined the pixel-wise likelihood of low-res image, $I^L_{i,p}$, as:

$$P(I^L_{i,p}|Z_{i,p}, I, K, F) = \begin{cases} \exp\{-\lambda|D_{i,p}|\}/C & \text{if } Z_{i,p} = 1 \\ 1 & \text{otherwise,} \end{cases} \tag{4}$$

Where C is a normalization constant, and the reconstruction error is denoted as $D_i = SK_iF_{0\to i}I - I^L_i$ for simplification. The prior for the latent variable $Z_{i,p}$ is:

$$P(Z_{i,p} = 1|I, K, F) = \begin{cases} \frac{\exp\{-\gamma/W_{i,p}\}}{\exp\{-\gamma/W_{i,p}\}+\exp\{-\gamma\beta\}} & \text{if } (SK_iF_{0\to i}I)_p \in [0, 1] \\ 0 & \text{otherwise.} \end{cases} \tag{5}$$

E step: Now, in the E step, under the current estimate of the high-res image I^0, we calculate the expected value of the complete-data log likelihood $\log P(\Omega, Z|I, K, F)$ with respect to the conditional distribution $P(Z|I^0, K, F, \Omega)$ as:

$$\begin{aligned} Q(I|I^0) &= E_{P(Z|I^0,K,F,\Omega)}[\log P(\Omega|Z, I, K, F) + \log P(Z|I, K, F)] \\ &= P(Z=1|I^0, K, F, \Omega) \log P(\Omega|Z=1, I, K, F) \\ &+ P(Z=0|I^0, K, F, \Omega) \log P(\Omega|Z=0, I, K, F) \\ &+ P(Z=1|I^0, K, F, \Omega) \log P(Z=1|I, K, F) \\ &+ P(Z=0|I^0, K, F, \Omega) \log P(Z=0|I, K, F). \end{aligned} \tag{6}$$

We denote $E_{P(Z|I^0,K,F,\Omega)}[Z] = P(Z=1|I^0, K, F, \Omega)$ by $E[Z]$. According to Eq. (4), we have:

$$\begin{aligned} &P(Z=1|I^0, K, F, \Omega) \log P(\Omega|Z=1, I, K, F) \\ &= E[Z] \log P(\Omega|Z=1, I, K, F) \\ &= -\sum_{i=-N}^{N} \sum_{p} E[Z_{i,p}](\lambda\, |D_{i,p}| + \log C), \end{aligned} \tag{7}$$

and

$$\begin{aligned} &P(Z=0|I^0, K, F, \Omega) \log P(\Omega|Z=0, I, K, F) \\ &= P(Z=0|I^0, K, F, \Omega) \log 1 = 0. \end{aligned} \tag{8}$$

Also, according to Eq. (5), we have:

$$\begin{aligned} &P(Z=1|I^0, K, F, \Omega) \log P(Z=1|I, K, F) \\ &= \begin{cases} \sum\limits_{i=-N}^{N} \sum\limits_{p} E[Z_{i,p}] \log \frac{\exp\{-\gamma/W_{i,p}\}}{\exp\{-\gamma/W_{i,p}\}+\exp\{-\gamma\beta\}} & \text{if } (SK_iF_{0\to i}I)_p \in [0,1] \\ \sum\limits_{i=-N}^{N} \sum\limits_{p} E[Z_{i,p}] \log 0 = -\infty & \text{otherwise,} \end{cases} \end{aligned} \tag{9}$$

and

$$\begin{aligned} &P(Z=0|I^0, K, F, \Omega) \log P(Z=0|I, K, F) \\ &= \begin{cases} \sum\limits_{i=-N}^{N} \sum\limits_{p} \{1 - E[Z_{i,p}]\} \log \frac{\exp\{-\gamma\beta\}}{\exp\{-\gamma/W_{i,p}\}+\exp\{-\gamma\beta\}} & \text{if } (SK_iF_{0\to i}I)_p \in [0,1] \\ \sum\limits_{i=-N}^{N} \sum\limits_{p} \{1 - E[Z_{i,p}]\} \log 1 = 0 & \text{otherwise.} \end{cases} \end{aligned} \tag{10}$$

From Eq. (7) to Eq. (10), we have that, up to a constant (which is independent of I):

$$Q(I|I^0) = \begin{cases} -\lambda \sum\limits_{i=-N}^{N} \sum\limits_{p} E[Z_{i,p}]\, |D_{i,p}| & \text{if } (SK_iF_{0\to i}I)_p \in [0,1] \\ -\infty & \text{otherwise.} \end{cases} \tag{11}$$

The expectation $E[Z_{i,p}] \equiv P(Z_{i,p}=1|I^0, K, F, \Omega)$ can be calculated via the Bayes' theorem:

$$P(Z_{i,p}=1|I^0, K, F, \Omega) = \frac{P(\Omega|Z_{i,p}=1, I^0, K, F) P(Z_{i,p}=1|I^0, K, F)}{P(\Omega|I^0, K, F)}, \tag{12}$$

where

$$P(\Omega|I^0, K, F) = \sum_{Z_{i,p}=0}^{1} P(\Omega|Z_{i,p}, I^0, K, F) P(Z_{i,p}|I^0, K, F). \tag{13}$$

Substituting Eq. (4) and Eq. (5) into the above two equations, we get:

$$E[Z_{i,p}] = \begin{cases} \frac{\exp\{-\lambda\left|(SK_iF_{0\to i}I^0 - I_i^L)_p\right|\}\exp\{-\gamma/W_{i,p}\}}{\exp\{-\lambda\left|(SK_iF_{0\to i}I^0 - I_i^L)_p\right|\}\exp\{-\gamma/W_{i,p}\} + C\exp\{-\gamma\beta\}} & \text{if } (SK_iF_{0\to i}I^0)_p \in [0,1] \\ 0 & \text{otherwise.} \end{cases} \tag{14}$$

M step: Given $E[Z_{i,p}]$ from the E step, this step updates the estimate I^0 to be the minimum of the negative log complete-data posterior, *i.e.*,

$$I^0 = \arg\min_I \sum_{i=-N}^{N} \lambda \left\| E[Z_i](SK_iF_{0\to i}I - I_i^L) \right\|_1 + \eta \cdot \phi_\delta(\nabla I). \tag{15}$$

where $\phi_\delta(\nabla I)$ is defined as a sum over all single-pixel penalties:

$$\phi_\delta(\nabla I) = \sum\nolimits_p \psi(\nabla I_p). \tag{16}$$

$\psi(\nabla I_p)$ is set as:

$$\psi(\nabla I_p) = \begin{cases} \|\nabla I_p\|_2/\delta & \text{if } \|\nabla I_p\|_2 < \delta \\ 1 & \text{otherwise.} \end{cases} \tag{17}$$

We first prove the following theorem:

Theorem 1. *The following equality holds:*

$$\phi_\delta(\nabla I) = \min_g \frac{1}{\delta}\|\nabla I - g\|_2 + \|g\|_0. \tag{18}$$

and the minimum is achieved when $g = \nabla I \cdot \max(\textit{sign}(\|\nabla I\|_2 - \delta), 0)$.

Proof. If $\|\nabla I\|_2 < \delta$, we consider two cases: (i). If $g = 0$, then the objective value equals to $\|\nabla I\|_2/\delta < 1$. (ii). if $g \neq 0$, the objective value is: $\|\nabla I - g\|_2/\delta + 1 \geq 1$. Hence $g = 0$ is favored.

In contrast, if $\|\nabla I\|_2 \geq \delta$, the optimal g would be ∇I, and the objective value is 1. This is because (i). if $g = 0$, the objective value will be $\|\nabla I\|_2/\delta \geq 1$. (ii). if $g \neq 0$, the objective value will also be greater than 1 due to the L_0 term.

In summary, the minimum is obtained when $g = \nabla I \cdot \max(\text{sign}(\|\nabla I\|_2 - \delta), 0)$. □

Based on the above theorem, we can rewrite Eq. (15) as:

$$\begin{aligned} I^0 &= \arg\min_I \sum_{i=-N}^{N} \lambda \left\| E[Z_i](SK_iF_{0\to i}I - I_i^L) \right\|_1 + \eta(\min_g(\frac{1}{\delta}\|\nabla I - g\|_2 + \|g\|_0)) \\ &= \arg\min_{I,g} \sum_{i=-N}^{N} \lambda \left\| E[Z_i](SK_iF_{0\to i}I - I_i^L) \right\|_1 + \eta(\frac{1}{\delta}\|\nabla I - g\|_2 + \|g\|_0). \end{aligned} \tag{19}$$

For each penalty obtained from a δ, we solve for image I via alternatively updating I and g as follows.

Fix g and estimate I: Assume $g = (g_x, g_y)$ and denote $E_i \triangleq E[Z_i]$, $\mu \triangleq \frac{\eta}{\lambda\delta}$. We perform iterative reweighted least squares (IRLS), which iterates between solving the following linear system:

$$\begin{aligned} (\sum_{i=-N}^{N} F_{0\to i}^T K_i^T S^T E_i^T W_{d_i} E_i S K_i F_{0\to i} + \mu(D_x^T W_s D_x + D_y^T W_s D_y))I = \\ \sum_{i=-N}^{N} F_{0\to i}^T K_i^T S^T E_i^T W_{d_i} E_i I_i^L + \mu(D_x^T W_s g_x + D_y^T W_s g_y). \end{aligned} \tag{20}$$

and updating the weights: $W_s \triangleq \text{diag}([(D_xI - g_x)^2 + (D_yI - g_y)^2 + \varepsilon]^{-\frac{1}{2}})$, $W_{d_i} \triangleq \text{diag}([(E_i(SK_iF_{0\to i}I - I_i^L))^2 + \varepsilon]^{-\frac{1}{2}})$, where D_x and D_y are matrix form of the x- and y- derivative filters. The linear system is solved through conjugate gradient (CG) algorithm.

Fix I and solve for g: According to Theorem 1, the optimal solution is:

$$g = \nabla I \cdot \max(\text{sign}(\|\nabla I\|_2 - \delta), 0), \tag{21}$$

which is called the shrinkage formula.

Intuitively, from Eq. (21), the gradients with small magnitude are gradually suppressed as $\delta \to 0$, but the main image structural information is preserved due to the data term in Eq. (19). In the extreme case, as $\delta \to 0$, $g \to \nabla I$, it is equivalent to solve the following L_0 regularized problem:

$$I^0 = \arg\min_I \sum_{i=-N}^{N} \lambda \left\| E[Z_i](SK_iF_{0\to i}I - I_i^L) \right\|_1 + \eta \|\nabla I\|_0. \tag{22}$$

A very simple form of the above equation has also been studied in the image filtering literature to smooth an image without deteriorating the main structures [4].

3. Derivations of the mask in the MAP point of view

In this section, we would like to show that, in the MAP point of view, solving the optimal latent variable $Z_{i,p}$ for each pixel is equivalent to minimize a temporal L_0 regularized problem. To see that, we first note, the intermediate selection of pixels can be determined by:

$$\begin{aligned} Z_{i,p} &= \arg\max_{Z_{i,p}} P(\Omega | Z_{i,p}, I^0, K, F) P(Z_{i,p} | I^0, K, F) \\ &= \arg\max_{Z_{i,p}} P(I_{i,p}^L | Z_{i,p}, I^0, K, F) P(Z_{i,p} | I^0, K, F). \end{aligned} \tag{23}$$

According to Eq. (4) and Eq. (5), we discuss two cases:
(i). If $(SK_iF_{0\to i}I^0)_p \in [0,1]$, then the right-hand-side value of Eq. (23) is proportional to:

$$\begin{cases} \exp\{-\gamma\beta\} & \text{if } Z_{i,p} = 0 \\ \exp\{-\lambda \left| (SK_iF_{0\to i}I^0 - I_i^L)_p \right|\} \exp\{-\gamma / W_{i,p}\} & \text{if } Z_{i,p} = 1. \end{cases} \tag{24}$$

The above equation tells us that the optimal $Z_{i,p}$ depends on the relationship between $\gamma\beta$ and $\lambda \left| (SK_iF_{0\to i}I^0 - I_i^L)_p \right| + \gamma / W_{i,p}$, *i.e.*,

$$Z_{i,p} = \begin{cases} 0 & \text{if } \lambda \left| (SK_iF_{0\to i}I^0 - I_i^L)_p \right| + \gamma / W_{i,p} \geq \gamma\beta \\ 1 & \text{otherwise.} \end{cases} \tag{25}$$

(ii). If $(SK_iF_{0\to i}I^0)_p \notin [0,1]$, then the right-hand-side value of Eq. (23) is:

$$\begin{cases} 0 & \text{if } Z_{i,p} = 1 \\ 1 & \text{if } Z_{i,p} = 0. \end{cases} \tag{26}$$

Hence the optimal $Z_{i,p} = 0$ in this case.

In summary, the solution to Eq. (23) is:

$$Z_{i,p} = \begin{cases} 0 & \text{if } \lambda |D_{i,p}| + \gamma / W_{i,p} \geq \gamma\beta \quad \text{or} \quad (SK_iF_{0\to i}I^0)_p \notin [0,1] \\ 1 & \text{otherwise.} \end{cases} \tag{27}$$

Where $D_i = SK_iF_{0\to i}I^0 - I_i^L$.

We next prove the relationship between Eq. (27) and the temporal L_0 regularized problem through the following theorem:

Theorem 2. *When $(SK_iF_{0\to i}I^0)_p \in [0,1]$, Eq. (27) is also the optimal solution to the following problem:*

$$Z_{i,p} = \arg\min_{Z_{i,p}\in[0,1]} \frac{\lambda}{\gamma} \left\| Z_{i,p}(SK_iF_{0\to i}I^0 - I_i^L)_p \right\|_1 + \beta \|1 - Z_{i,p}\|_0 + Z_{i,p}^2 / W_{i,p}. \tag{28}$$

Proof. The objective function value in Eq. (28) is:

$$\begin{cases} \geq \beta & \text{if } Z_{i,p} \neq 1 \\ \beta & \text{if } Z_{i,p} = 0 \\ \frac{\lambda}{\gamma}\left\|(SK_iF_{0\to i}I^0 - I_i^L)_p\right\|_1 + 1/W_{i,p} & \text{if } Z_{i,p} = 1. \end{cases} \tag{29}$$

Hence the optimal $Z_{i,p}$ will be 0 if $\lambda\left\|(SK_iF_{0\to i}I^0 - I_i^L)_p\right\|_1 + \gamma/W_{i,p} \geq \gamma\beta$, and 1 otherwise. □

4. Derivations of the Cramer-Rao bounds

Now we turn to analyzing the relationship between the estimation error of high-res signals and motion blur kernels, as well as the impact factors of motion blur estimation.

We first assume that the perfect estimation of optical flow has already been obtained. That means, for upsampling rate M, we could register back M^2 low-res images with their corresponding optical flows and concatenate them to form a high-res observation J. Therefore, under this ideal setting, the SR problem is equivalent to a deblurring problem as:

$$J = \dot{F} \otimes \dot{G} \otimes I + E. \tag{30}$$

where $\otimes$ denotes the convolution operation, J is the observed image, $\dot{F}$ is the motion blur kernel, $\dot{G}$ is the anti-aliasing kernel and E is the additive Gaussian noise with variance σ_n^2. Though the notations here are slightly abused, it would not cause difficulties in understanding since this section is relatively independent with previous ones.

To simplify the derivation, we focus on 1D latent signal and its single component in frequency domain after decomposition, *i.e.*, in the time domain, the signal takes the form $I(n) = \frac{A}{N_L}e^{\frac{i2\pi\omega_0 n}{N_H}}$. Here N_L and N_H are lengths of the observed and latent signals respectively. Due to the ideal setting aforementioned, $M = N_H/N_L = 1$. A is the complex amplitude. We then conduct discrete Fourier transform to both sides of the Eq. (30) to get:

$$\tilde{J}(\omega) = FG\tilde{I}(\omega) + \tilde{E}(\omega). \tag{31}$$

Here $\tilde{I}(\omega) = MA\delta(\omega - \omega_0)$. F, G are the DFTs of $\dot{F}$, $\dot{G}$ respectively. We assume $\dot{G}$ takes the form of Gaussian blur, *i.e.*, $G(\omega) = e^{-\omega^2\sigma_k^2/2}$, σ_k^2 is the corresponding variance. Then we can write down the negative log likelihood function for the observed signal as:

$$-\log P(\tilde{J}\,|A\,) = \frac{1}{2\sigma_n^2}\left\|\tilde{J}(\omega_0) - FGA\right\|^2. \tag{32}$$

We first analyze the relationship between estimation error of I and motion blur kernel F. The parameters of the likelihood are thus the complex amplitude of signal, $\theta = [\mathrm{Re}\{A\}, \mathrm{Im}\{A\}]$, thus we can derive the Fisher information matrix $\mathcal{I}(\theta)$. According to the definition, the (i,j)-th element in $\mathcal{I}(\theta)$ is:

$$(I(\theta))_{i,j} = E\left[\left(\frac{\partial}{\partial\theta_i}\log P(\tilde{J}\,|A\,)\right)\left(\frac{\partial}{\partial\theta_j}\log P(\tilde{J}\,|A\,)\right)|\theta\right]. \tag{33}$$

Hence, we could explicitly derive the expression of $\mathcal{I}(\theta)$ as:

$$I_\theta = \frac{F^*FG^2}{\sigma_n^2}\begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \tag{34}$$

where $*$ denotes the complex conjugate. Now we could exploit the Cramer-Rao bound [1] to give a lower bound on the variance of estimator of parameter, *i.e.*, A in our model. Specifically,

$$\mathrm{var}\left(\hat{A}\right) \geq I_\theta^{-1}(1,1) + I_\theta^{-1}(2,2) = \frac{2\sigma_n^2}{F^*F}e^{\omega^2\sigma_k^2}, \tag{35}$$

where $\hat{A}$ is the unbiased estimator of A. This bound indicates that, if a frequency component in the motion blur kernel F has small magnitude, the estimation error of the corresponding frequency component in the high-res image will be large.

If we fix A and take the motion blur kernel F as the parameter of the likelihood function, following the above procedures, we can derive a similar bound for the unbiased estimator $\hat{F}$ as:

$$\mathrm{var}\left(\hat{F}\right) \geq \frac{2\sigma_n^2}{A^*A}e^{\omega^2\sigma_k^2}. \tag{36}$$

This indicates that the error of the blur kernel estimation will increase if the image becomes more blurry, *i.e.*, σ_k increases.

5. More results and comparisons

In this section, we show more results as well as compare our algorithm with other methods (including: Single-image deblurring [5], Multi-image deblurring [7], Video upsampling [3], Multi-frame super-resolution [2] and Multi-shot imaging [6]) on some challenging real sequences in the following figures.

References

[1] Steven M Kay. *Fundamentals of statistical signal processing: estimation theory*. Prentice-Hall, Inc., 1993.

[2] Ce Liu and Deqing Sun. A bayesian approach to adaptive video super resolution. In *CVPR*, pages 209–216, 2011.

[3] Qi Shan, Zhaorong Li, Jiaya Jia, and Chi-Keung Tang. Fast image/video upsampling. *ACM Transactions on Graphics (TOG)*, 27(5):153, 2008.

[4] L. Xu, C. Lu, Y. Xu, and J. Jia. Image smoothing via l0 gradient minimization. *ACM Transactions on Graphics (TOG)*, 30(6):174, 2011.

[5] Li Xu, Shicheng Zheng, and Jiaya Jia. Unnatural l0 sparse representation for natural image deblurring. In *CVPR*, pages 1107–1114, 2013.

[6] Haichao Zhang and Lawrence Carin. Multi-shot imaging: Joint alignment, deblurring, and resolution-enhancement. In *CVPR*, pages 2925–2932, 2014.

[7] Xiang Zhu, Filip Šroubek, and Peyman Milanfar. Deconvolving psfs for a better motion deblurring using multiple images. In *ECCV*, pages 636–647, 2012.

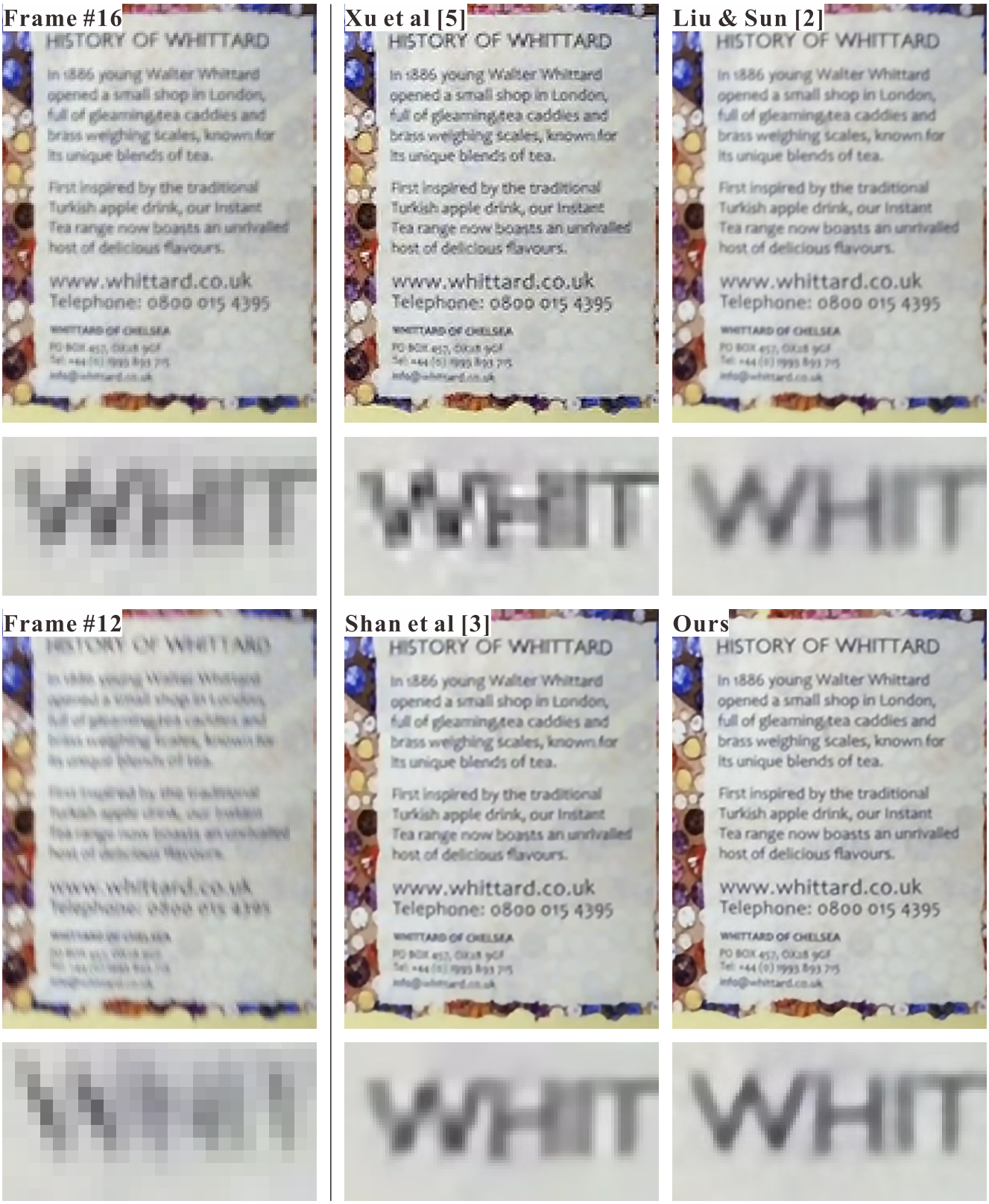


Figure 1. Comparison 1. (a) Some of the input frames (nearest $\times 4$). The image to super-resolve is Frame #16. (b) Some comparisons with close-ups.

Figure 2. Comparison 2. (a) Some of the input frames (nearest $\times 3$). The image to super-resolve is Frame #16. (b) Some comparisons with close-ups.

(a) Inputs (b) Results

Figure 3. Comparison 3. (a) Some of the input frames (nearest $\times 3$). The image to super-resolve is Frame #16. (b) Some comparisons with close-ups.

Figure 4. Comparison 4. (a) Some of the input frames (nearest $\times 4$). The image to super-resolve is Frame #16. (b) Some comparisons with close-ups.

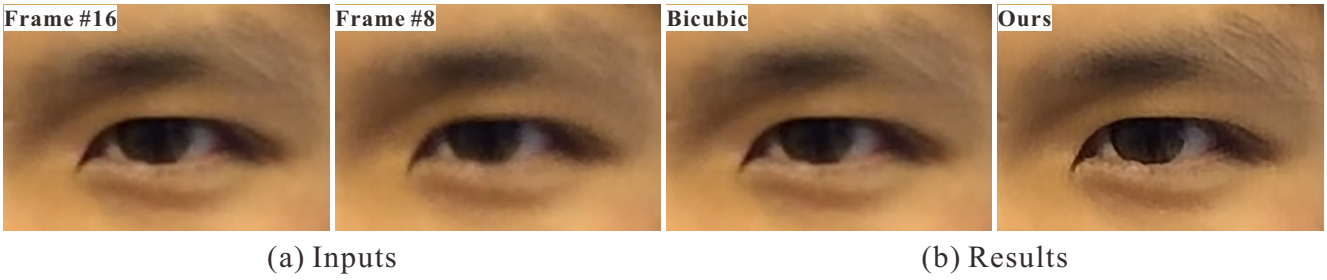
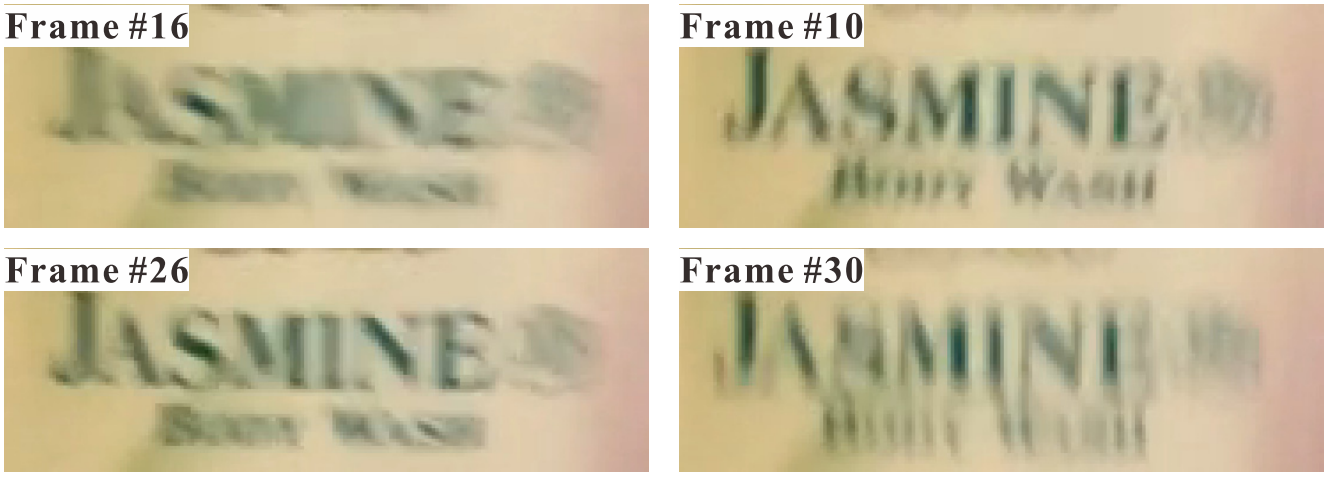
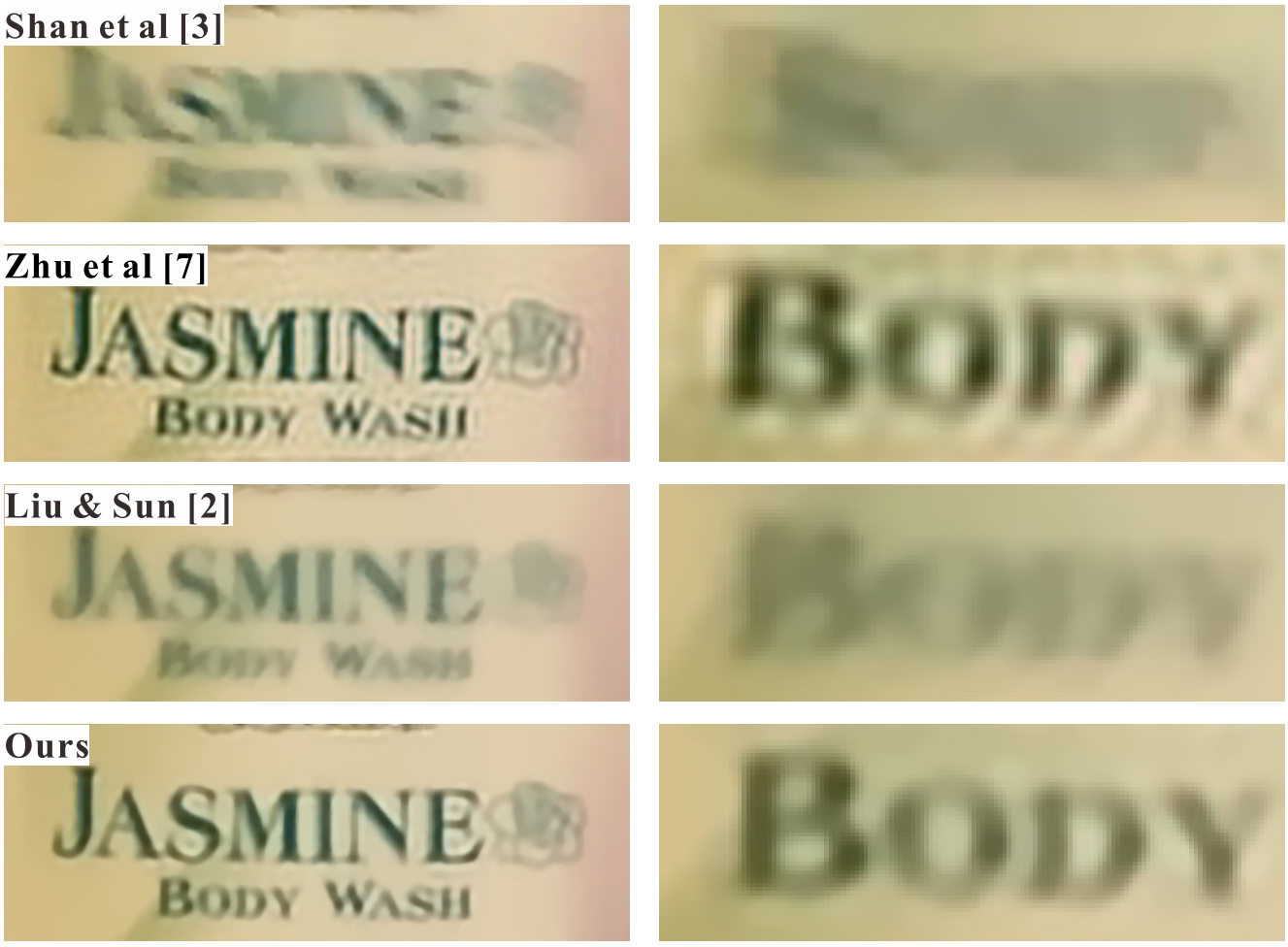


Figure 5. Comparison 5. (a) Some of the input frames (nearest $\times 3$). The image to super-resolve is Frame #16. (b) Some comparisons with close-ups.

Figure 6. Comparison 6. (a) Some of the input frames (nearest $\times 3$). The image to super-resolve is Frame #16. (b) Some comparisons with close-ups.



(a) Inputs



(b) Results

Figure 7. Comparison 7. (a) Some of the input frames (nearest $\times 3$). The image to super-resolve is Frame #16. (b) Some comparisons with close-ups.

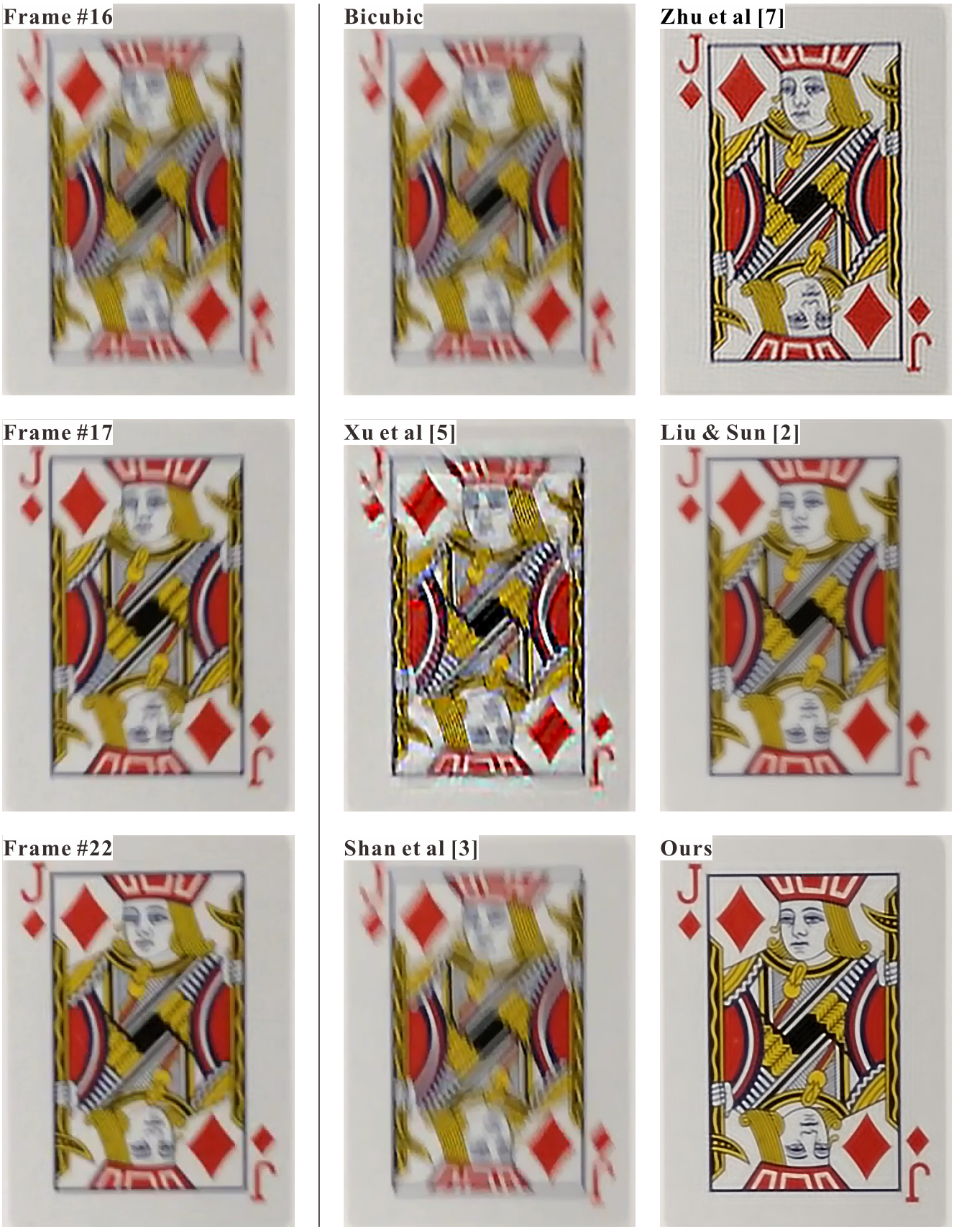
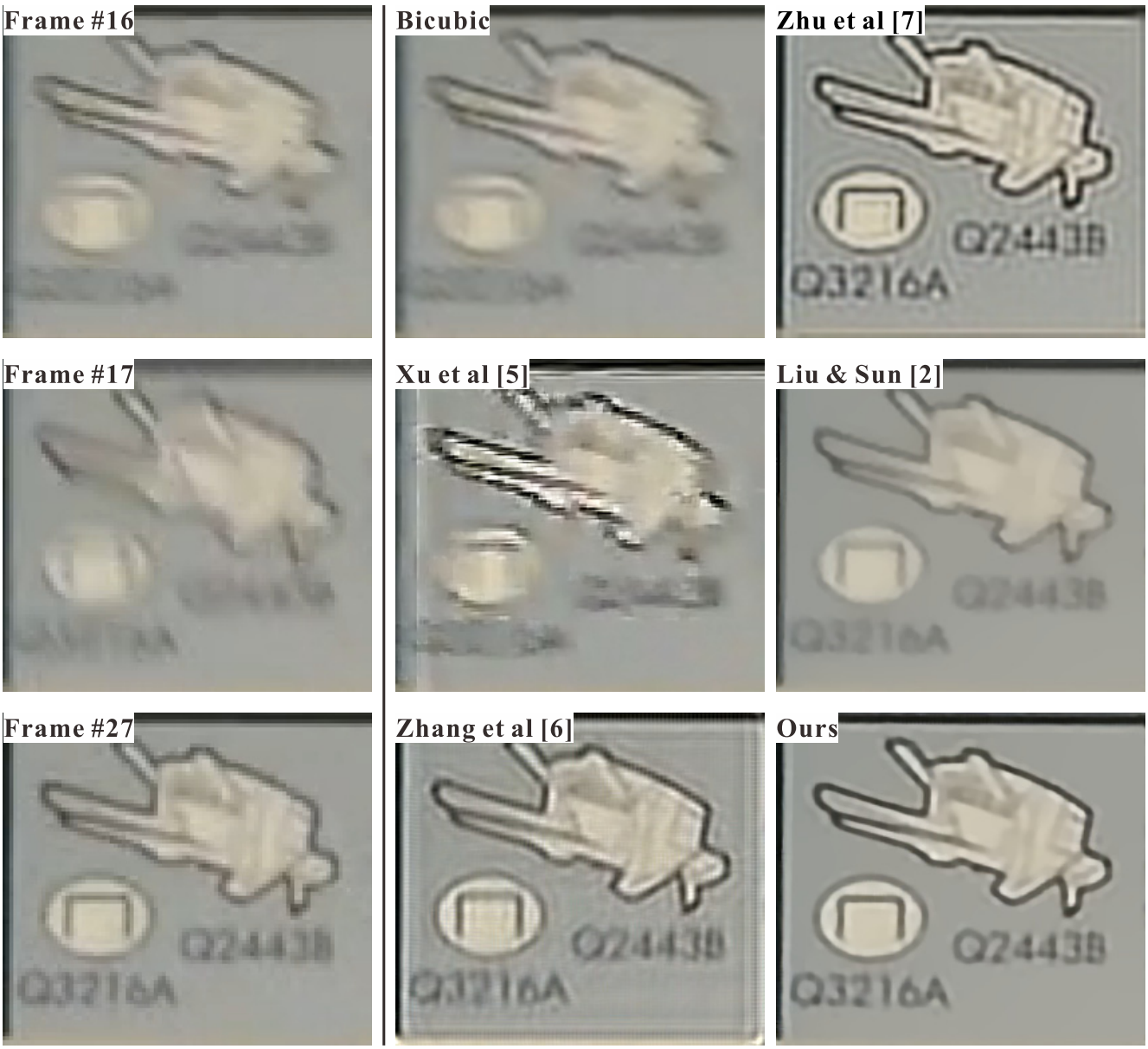


Figure 8. Comparison 8. (a) Some of the input frames (nearest $\times 2$). The image to super-resolve is Frame #16. (b) Some comparisons with close-ups.

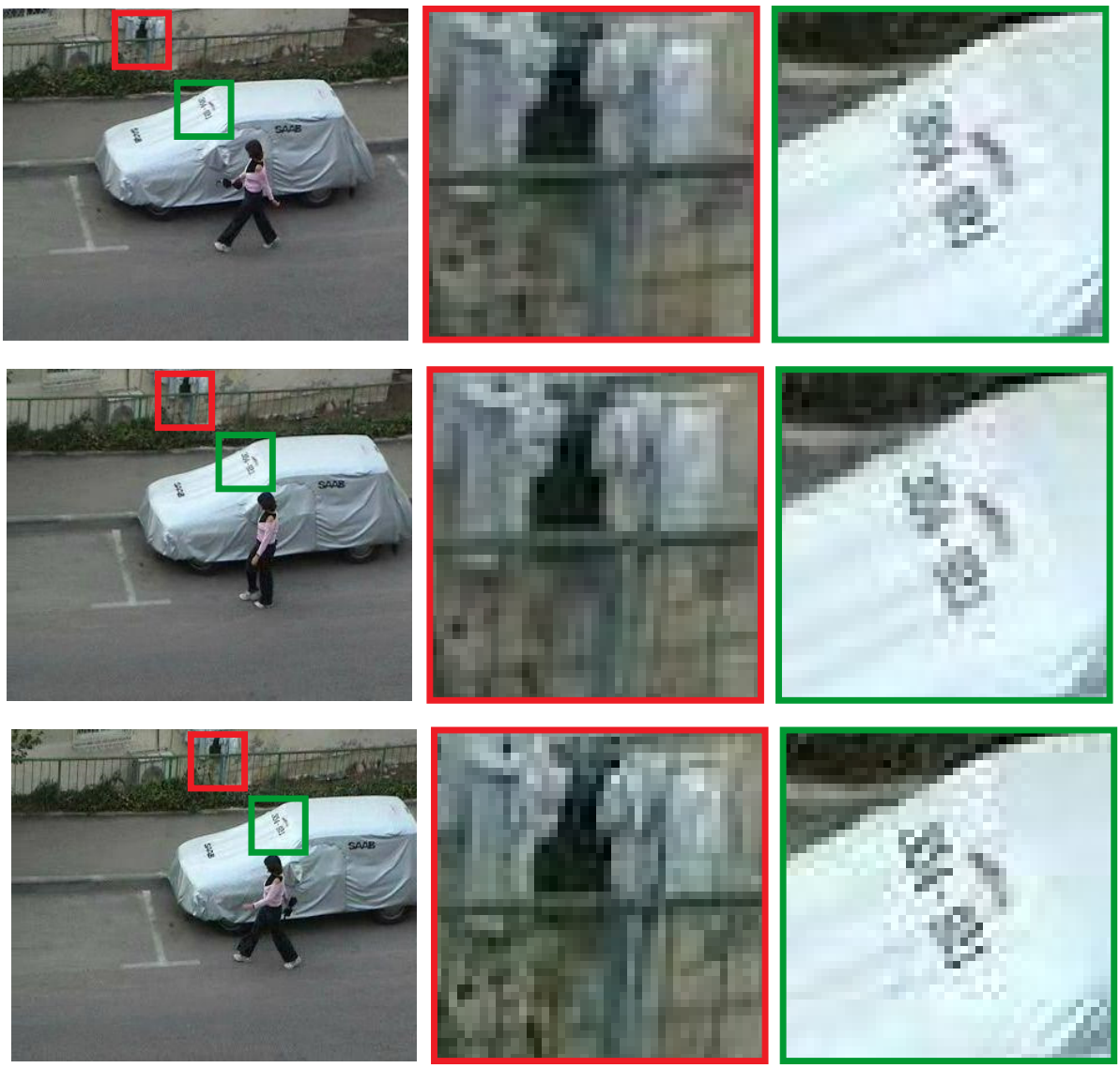


(a) Inputs (b) Results

Figure 9. Comparison 9. (a) Some of the input frames (nearest $\times 3$). The image to super-resolve is Frame #16. (b) Some comparisons with close-ups.

Figure 10. Our result on general natural sequence. (a) Some of the input frames (nearest $\times 3$). The image to super-resolve is Frame #16. (b) Some comparisons.

Figure 11. Our result on general natural sequence. (a) Some of the input frames (nearest $\times 3$). The image to super-resolve is Frame #16. (b) Some comparisons.



(a) Some input frames (with zoom-in)

(b) Our results

Figure 12. Our result on general natural sequence. (a) Some of the input frames (nearest $\times 3$). The image to super-resolve is Frame #16. (b) Our results.