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

Ubiquitous motion blur, which is caused by camera shaking or fast object moving, easily fails most previous methods in multi-frame super-resolution (MFSR). Moreover, in high-res videos, regions of interest (ROI) are still small areas, where even small motion blur may cause severe degradations. As shown in Fig. 1. Although common in captured videos, motion blur has not been sufficiently studied for the MFSR task in literatures.

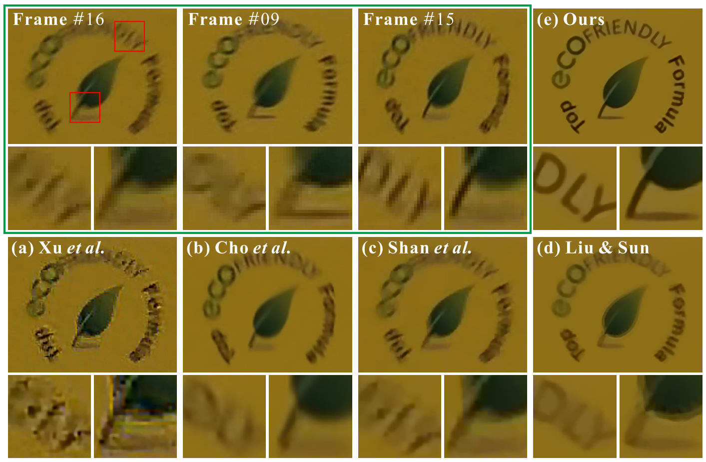

Figure 1: MFSR results on a real video sequence. Green box: Input frames (150×120) *directly cropped* from an iPhone video. Three clearest ones are selected. Motion blur and compression artifacts are present. (a) Result of single image deblurring [4]. (b) Result of video deblurring [1]. (c) Result of video upsampling [3]. (d) MFSR result [2]. (e) Our result ($\times 3$).

We proposed an integrated framework to estimate motion blur and high-res image with quality feedback and control. Formally, we want to estimate a high-res image I corresponds to I_0^L from a set of low-res images $\Omega = \{I_{-N}^L, \cdots, I_0^L, \cdots, I_N^L\}$. The imaging model is:

$$I_i^L = SK_iF_{0\to i}I + n \quad \text{where } i = -N, \cdots, N. \tag{1}$$

Here, I is the latent high-res image. $F = \{F_{0\to -N}, \cdots, F_{0\to 0}, \cdots, F_{0\to N}\}$ is a set of matrices corresponding to the optical flow from I to each frame. S and K_i correspond to down-sampling and filtering matrices. $K_i = K_aK_{b_i}$, where K_a is the anti-aliasing kernel, K_{b_i} is the motion blur kernel. n is the noise. Given this model, I, K and F are estimated through MAP:

$$\{I, K, F\} = \arg\max_{I,K,F} P(I)P(K)P(F)P(\Omega|I,K,F). \tag{2}$$

To robustly handle degenerated low-res inputs in the presence of motion blur, the key of our approach is to introduce a binary latent variable $Z = \{Z_{-N}, \cdots, Z_0, \cdots, Z_N\}$ to classify each pixel from each input image as either useful ($Z = 1$) or useless ($Z = 0$). We want to exclude pixels that are largely blurred compared to other temporal correspondences. This is because we observed that those blurred edges could easily mislead kernel estimation in MFSR, while a careful temporal selection of clear pixels can compose an image with sharp structures which is beneficial to kernel estimation. We write Eq.(2) as:

$$\{I, K, F\} = \arg\max_{I,K,F} P(I) \prod_{i=-N}^{N} [P(K_i)P(F_{0\to i}) \sum_{Z_i} P(I_i^L, Z_i|I,K,F)]. \tag{3}$$

We adopt classical regularizer for $P(K)$, $P(F)$, and pixel-wise decompose:

$$P(I_i^L, Z_i|I,K,F) = \prod_p P(I_{i,p}^L|Z_{i,p}, I, K, F)P(Z_{i,p}|I,K,F). \tag{4}$$

where p index each pixel. The likelihood term is set as:

$$P(I_{i,p}^L|Z_{i,p}, I, K, F) \propto \begin{cases} \exp\{-\lambda\,|D_{i,p}|\} & \text{if } Z_{i,p} = 1 \\ 1 & \text{otherwise} \end{cases} \tag{5}$$

where the error $D_i = SK_iF_{0\to i}I - I_i^L$. We propose to define the prior term based on temporal relative sharpness of the pixels from the conditional Bernoulli distribution:

$$P(Z_{i,p} = 1|I,K,F) \propto \begin{cases} \exp\{-\gamma/W_{i,p}\} & \text{if } SK_iF_{0\to i}I \in [0,1] \\ 0 & \text{otherwise} \end{cases} \tag{6}$$

where $W_{i,p} = W_p(V_{i,p})$. $W_p(x)$ ($x \in [0,1]$) is the cumulative distribution function of $V_{i,p}$ for $i \in -N, \cdots, N$, which measures the temporal relative sharpness of pixel p across all frames:

$$V_{i,p} = \frac{\sum_{q\in\mathcal{N}(p)} \left\|\nabla J_{i,q}\right\|_1}{\sum_{j=-N}^{N} \sum_{q\in\mathcal{N}(p)} \left\|\nabla J_{j,q}\right\|_1 + \varepsilon}. \tag{7}$$

Here, J_i is the i-th registered image, and $J_{i,p}$ denotes its p-th pixel. To keep useful salient structures while suppressing noise, we employ a family of sparse image priors. We show that the binary latent variable and the high-res image/flow/kernel can be iteratively estimated in an EM framework.

Our method produces satisfying results on challenging real-world low-res noisy and blurred sequences. In addition to ROI and motion blur, it also works for natural images corrupted with compression blur, some examples are shown in Fig. 2, more can be found in our paper and supplementary file.

(a) Selected input frames (with zoom-in) (b) Our results

Figure 2: More results. (a) Some input frames from *real-world* infrared/surveillance video sequence. (b) Our MFSR results.

This is an extended abstract. The full paper is available at the Computer Vision Foundationwebpage.

[1] Sunghyun Cho, Jue Wang, and Seungyong Lee. Video deblurring for hand-held cameras using patch-based synthesis. *ACM Transactions on Graphics (TOG)*, 31(4):64, 2012.

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