A Bayesian Approach to Adaptive Video Super Resolution

Ce Liu
Microsoft Research New England

Deqing Sun
Brown University

(a) Input low-res (b) Bicubic up-sampling $\times 4$ (c) Output from our system (d) Original frame

Figure 1. Our video super resolution system is able to recover image details after $\times 4$ up-sampling.

Abstract

Although multi-frame super resolution has been extensively studied in past decades, super resolving real-world video sequences still remains challenging. In existing systems, either the motion models are oversimplified, or important factors such as blur kernel and noise level are assumed to be known. Such models cannot deal with the scene and imaging conditions that vary from one sequence to another. In this paper, we propose a Bayesian approach to adaptive video super resolution via simultaneously estimating underlying motion, blur kernel and noise level while reconstructing the original high-res frames. As a result, our system not only produces very promising super resolution results that outperform the state of the art, but also adapts to a variety of noise levels and blur kernels. Theoretical analysis of the relationship between blur kernel, noise level and frequency-wise reconstruction rate is also provided, consistent with our experimental results.

1. Introduction

Multi-frame super resolution, namely estimating the high-res frames from a low-res sequence, is one of the fundamental problems in computer vision and has been extensively studied for decades. The problem becomes particularly interesting as high-definition devices such as HDTV's dominate the market. There is a great need for converting low-res, low-quality videos into high-res, noise-free videos that can be pleasantly viewed on HDTV's.

Although a lot of progress has been made in the past 30 years, super resolving real-world video sequences still remains an open problem. Most of the previous work assumes that the underlying motion has a simple parametric form, and/or that the blur kernel and noise levels are known. But in reality, the motion of objects and cameras can be arbitrary, the video may be contaminated with noise of unknown level, and motion blur and point spread functions can lead to an unknown blur kernel.

Therefore, a practical super resolution system should simultaneously estimate optical flow [9], noise level [18] and blur kernel [12] in addition to reconstructing the high-res frames. As each of these problems has been well studied in computer vision, it is natural to combine all these components in a single framework without making oversimplified assumptions.

In this paper, we propose a Bayesian framework for adaptive video super resolution that incorporates high-res image reconstruction, optical flow, noise level and blur kernel estimation. Using a sparsity prior for the high-res image, flow fields and blur kernel, we show that super resolution computation is reduced to each component problem when other factors are known, and the MAP inference iterates between optical flow, noise estimation, blur estimation and image reconstruction. As shown in Figure 1 and later examples, our system produces promising results on challenging real-world sequences despite various noise levels and blur kernels, accurately reconstructing both major structures and fine texture details. In-depth experiments demonstrate that our system outperforms the state-of-the-art super resolution systems [1, 23, 25] on challenging real-world sequences.

We are also interested in theoretical aspects of super resolution, namely to what extent the original high-res information can be recovered under a given condition. Although previous work [3, 15] on the limits of super resolution provides important insights into the increasing difficulty of recovering the signal as a function of the up-sampling factor, most of the bounds are obtained for the entire signal with frequency perspective ignored. Intuitively, high frequency components of the original image are much harder to recover as the blur kernel, noise level and/or up-sampling factor increases. We use Wiener filtering theory to analyze the

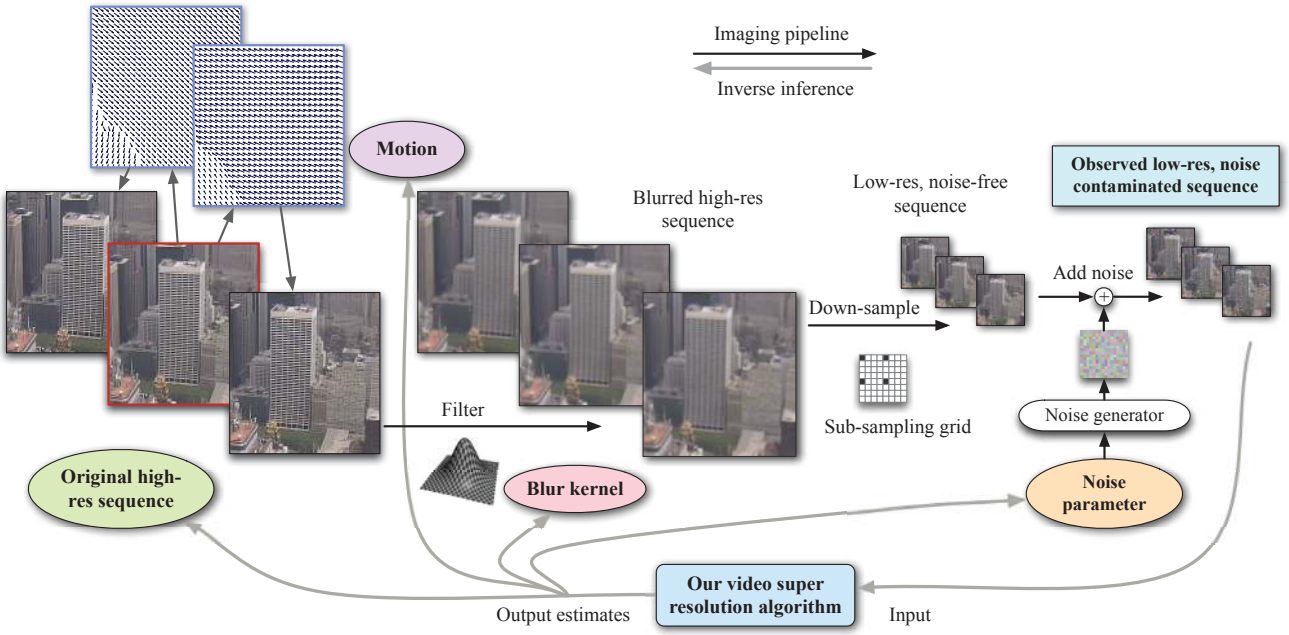


Figure 2. **Video super resolution diagram**. The original high-res video sequence is generated by warping the source frame (enclosed by a red rectangle) both forward and backward with some motion fields. The high-res sequence is then smoothed with a blur kernel, down-sampled and contaminated with noise to generate the observed sequence. Our adaptive video super resolution system not only estimates the high-res sequence, but also the underlying motion (on the lattice of original sequence), blur kernel and noise level.

frequency-wise bound, which is further verified through experiments.

2. Related Work

Since the seminal work by Tsai and Huang [26], significant progress has been made in super resolution. We refer readers to [20] for a comprehensive literature review.

Early super resolution work focused on dealing with the ill-posed nature of reconstructing a high-res image from a sequence of low-res frames [10]. The lack of constraints is often addressed by spatial priors on the high-res image [22]. Hardie *et al*. [8] jointly estimated the translational motion and the high-res image, while Bascle *et al*. [4] also considered the motion blur using an affine motion model. But these motion models are too simple to reflect the nature of real-world sequences.

To deal with the complex motion of faces, Baker and Kanade [2] proposed to use optical flow for super resolution, although in fact a parametric motion model was adopted. Fransens *et al*. [6] proposed a probabilistic formulation and jointly estimated the image, flow field and Gaussian noise statistics within an EM framework. They assumed that the blur kernel was known, and used Gaussian priors for both images and flow fields. However, Gaussian priors tend to over-smooth sharp boundaries in images and flows.

While most of these motion-based super resolution models use somewhat standard flow estimation techniques, recent advances in optical flow have resulted in much more reliable methods based on sparsity priors *e.g.* [5]. Accurate motion estimation despite strong noise has inspired Liu and Freeman [17] to develop a high quality video denoising system that removes structural noise in real video sequences. In this paper, we also want to incorporate recent advances in optical flow for more accurate super resolution.

Inspired by the successful non-local means method for video denoising, Takeda *et al*. [25] avoided explicit sub-pixel motion estimation and used 3D kernel regression to exploit the spatiotemporal neighboring relationship for video up-sampling. However, their method still needs to estimate a pixel-wise motion at regions with large motion. In addition, its data model does not include blur and so its output needs to be postprocessed by a deblurring method.

While most methods assume the blur kernel is known, some work considers estimating the blur kernel under simple settings. Nguyen *et al*. [19] used the generalized cross-correlation method to identify the blur kernel using quadratic formulations. Sroubek *et al*. [24] estimated the image and the blur kernel under translational motion models by joint MAP estimation. However, their models can barely generalize to real videos due to the oversimplified motion models.

Significant improvements on blur estimation from real images have been made in the blind deconvolution community. Levin *et al*. [14] showed that joint MAP estimation of the blur kernel and the original image favors a non-blur explanation, *i.e.* , a delta blur function and the blurred image.

Their analysis assumes no spatial prior on the blur kernel, while Joshi *et al*. [11] used a smoothness prior for the blur kernel and obtained reliable estimates. Moreover, Shan *et al*. [23] applied the recent advances in image deconvolution to super resolution and obtained promising improvement, but their method only works on a single frame and does not estimate the noise statistics.

On the theory side, there has been important work on the limit of super resolution as the up-sampling factor increases [3, 15]. Their analysis focused on the stability of linear systems while ignoring the frequency aspects of the limit. In fact, many useful tools have been developed in the signal processing community to analyze the performance of linear systems w.r.t. a particular frequency component. Though our super resolution system is nonlinear, we apply these analysis tools to a simplified linear problem and obtain how the blur kernel and noise level affect the reconstruction at each frequency.

3. A Bayesian Model for Super Resolution

Given the low-res sequence $\{J_t\}$, our goal is to recover the high-res sequence $\{I_t\}$. Due to computational issues, we aim at estimating I_t using adjacent frames $J_{t-N}, \cdots, J_{t-1}, J_t, J_{t+1}, \cdots, J_{t+N}$. To make the notations succinct, we will omit t from now on. Our problem becomes to estimate I given a series of images $\{J_{-N}, \cdots, J_N\}$. In addition, we will derive the equations using gray-scale images for simplicity although our implementation is able to handle color images.

The model of obtaining low-res sequence is illustrated in Figure 2. A full generative model that corresponds to Figure 2 is shown in Figure 3. At time $i = 0$, frame I is smoothed and down-sampled to generate J_0 with noise. At time $i = -N, \cdots, N,\ i \neq 0$, frame I is first warped according to a flow field w_i, and then smoothed and down-sampled to generate J_i with noise and outlier R_i (we need to model outliers because optical flow cannot perfectly explain the correspondence between two frames). The unknown parameters in the generative models include the smoothing kernel K, which corresponds to point spread functions in the imaging process, or smoothing filter when video is down-sampled, and parameter θ_i that controls the noise and outlier when I is warped to generate adjacent frames.

We use Bayesian MAP to find the optimal solution

$$\{I^*, \{w_i\}^*, K^*, \{\theta_i\}^*\} = \underset{I,\{w_i\},K,\{\theta_i\}}{\operatorname{argmax}}\ p(I, \{w_i\}, K, \{\theta_i\}|\{J_i\}), \tag{1}$$

where the posterior is the product of prior and likelihood:

$$p(I, \{w_i\}, K, \{\theta_i\}|\{J_i\}) \propto p(I)p(K)\prod_i p(w_i)\prod_i p(\theta_i)\cdot p(J_0|I, K, \theta_0)\prod_{i\neq 0} p(J_i|I, K, w_i, \theta_i). \tag{2}$$

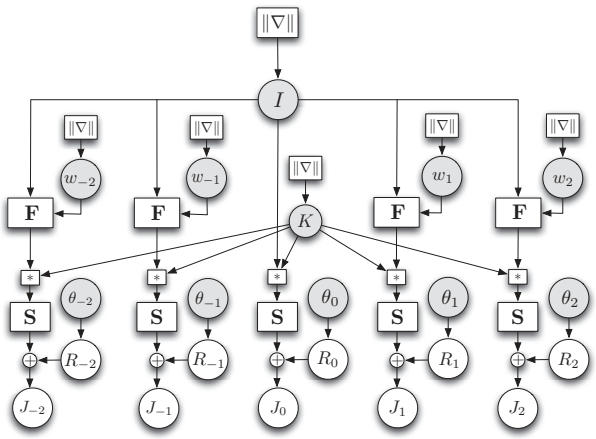


Figure 3. The graphical model of video super resolution. The circular nodes are variables (vectors), whereas the rectangular nodes are matrices (matrix multiplication). We do not put priors η, λ, ξ, α and β on I, w_i, K, and θ_i for succinctness.

Sparsity on derivative filter responses is used to model the priors of image I, optical flow field w_i and blur kernel K

$$p(I) = \frac{1}{Z_I(\eta)}\exp\left\{-\eta\left\|\nabla I\right\|\right\}, \tag{3}$$

$$p(w_i) = \frac{1}{Z_w(\lambda)}\exp\left\{-\lambda\left(\left\|\nabla u_i\right\| + \left\|\nabla v_i\right\|\right)\right\}, \tag{4}$$

$$p(K_x) = \frac{1}{Z_K(\xi)}\exp\left\{-\xi\left\|\nabla K_x\right\|\right\}, \tag{5}$$

where ∇ is the gradient operator, $\left\|\nabla I\right\| = \sum_q \left\|\nabla I(q)\right\| = \sum_q (|I_x(q)| + |I_y(q)|)$ ($I_x = \frac{\partial}{\partial x} I, I_y = \frac{\partial}{\partial y} I$) and q is the pixel index. The same notation holds for u_i and v_i, the horizontal and vertical components of the flow field w_i. For computational efficiency, we assume the kernel K is x- and y-separable: $K = K_x \otimes K_y$, where K_y has the same probability distribution as K_x. $Z_I(\eta)$, $Z_w(\lambda)$ and $Z_K(\xi)$ are normalization constants only dependant on η, λ and ξ, respectively.

To deal with outliers, we assume an exponential distribution for the likelihood

$$p(J_i|I, K, \theta_i) = \frac{1}{Z(\theta_i)}\exp\left\{-\theta_i\left\|J_i - \mathbf{SKF}_{w_i} I\right\|\right\}, \tag{6}$$

where the parameter θ_i reflects the noise level of frame i and $Z(\theta_i) = (2\theta_i)^{-\dim(I)}$. Matrices $\mathbf{S}$ and $\mathbf{K}$ correspond to down-sampling and filtering with blur kernel K, respectively. $\mathbf{F}_{w_i}$ is the warping matrix corresponding to flow w_i. Naturally, the conjugate prior for θ_i is a Gamma distribution

$$p(\theta_i; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\theta_i^{\alpha-1}\exp\{-\theta_i\beta\}. \tag{7}$$

Now that we have the probability distributions for both prior and likelihood, the Bayesian MAP inference is per-

formed using coordinate descend. Note that in this model there are only five free parameters: η, λ, ξ, α and β.

3.1. Image Reconstruction

Given the current estimates of the flow field w_i, the blur kernel K and the noise level θ_i, we estimate the high-res image by solving

$$I^* = \underset{I}{\operatorname{argmin}}\, \theta_0 \big\|\mathbf{SK}I - J_0\big\| + \eta\big\|\nabla I\big\| + \sum_{i=-N, i\neq 0}^{N} \theta_i \big\|\mathbf{SKF}_{w_i} I - J_i\big\|. \quad (8)$$

To use gradient-based methods, we replace the L1 norm with a differentiable approximation $\phi(x^2) = \sqrt{x^2+\epsilon^2}$ ($\epsilon = 0.001$), and denote the vector $\Phi(|I|^2) = [\phi(I^2(q))]$.

This objective function can be solved by the iteratively reweighted least squares (IRLS) method [16], which iteratively solves the following linear system:

$$\Big[\theta_0 \mathbf{K}^T\mathbf{S}^T\mathbf{W}_0\mathbf{SK} + \eta\Big(\mathbf{D}_x^T\mathbf{W}_s\mathbf{D}_x + \mathbf{D}_y^T\mathbf{W}_s\mathbf{D}_y\Big) + \sum_{i=-N, i\neq 0}^{N} \theta_i \mathbf{F}_{w_i}^T \mathbf{K}^T\mathbf{S}^T\mathbf{W}_i\mathbf{SKF}_{w_i}\Big] I = \theta_0\mathbf{K}^T\mathbf{S}^T\mathbf{W}_0 J_0 + \sum_{i=-N, i\neq 0}^{N} \theta_i \mathbf{F}_{w_i}^T \mathbf{K}^T\mathbf{S}^T\mathbf{W}_i J_i, \quad (9)$$

where the matrices $\mathbf{D}_x$ and $\mathbf{D}_y$ correspond to the x- and y- derivative filters. IRLS iterates between solving the above least square problem (through conjugate gradient) and estimating the diagonal weight matrices $\mathbf{W}_0 = \text{diag}(\Phi'(|\mathbf{SK}I - J_0|^2))$, $\mathbf{W}_s = \text{diag}(\Phi'(|\nabla I|^2))$, and $\mathbf{W}_i = \text{diag}(\Phi'(|\mathbf{SKF}_{w_i} I - J_i|^2))$ based on the current estimate.

3.2. Motion and Noise Estimation

Given the high-res image and the blur kernel, we jointly estimate the flow field and the noise level in a coarse-to-fine fashion on a Gaussian image pyramid. At each pyramid level, the noise level and optical flow are estimated iteratively. The Bayesian MAP estimate for the noise parameter θ_i has the following closed-form solution

$$\theta_i^* = \frac{\alpha + N_q - 1}{\beta + N_q \overline{x}},\ \overline{x} = \frac{1}{N_q}\sum_{q=1}^{N_q} \Big|\big(J_i - \mathbf{SKF}_{w_i} I\big)(q)\Big|, \quad (10)$$

where $\overline{x}$ is sufficient statistics. When noise is known, the flow field w_i is estimated as

$$w_i^* = \underset{w_i}{\operatorname{argmin}}\, \theta_i \big\|\mathbf{SKF}_{w_i} I - J_i\big\| + \lambda\big\|\nabla u_i\big\| + \lambda\big\|\nabla v_i\big\|, \quad (11)$$

where we again approximate $|x|$ by $\phi(x^2)$. Notice that this optical flow formulation is different from the standard ones: the flow is established from high-res I to low-res J_i.

By first-order Taylor expansion

$$\mathbf{F}_{w_i + dw_i} I \approx \mathbf{F}_{w_i} I + \mathbf{I}_x du_i + \mathbf{I}_y dv_i, \quad (12)$$

where $\mathbf{I}_x = \text{diag}(\mathbf{F}_{w_i} I_x)$ and $\mathbf{I}_y = \text{diag}(\mathbf{F}_{w_i} I_y)$, we can derive (following the conventions in [16])

$$\begin{bmatrix} \mathbf{I}_x^T\tilde{\mathbf{W}}_i\mathbf{I}_x + \zeta_i\mathbf{L} & \mathbf{I}_x^T\tilde{\mathbf{W}}_i\mathbf{I}_y \\ \mathbf{I}_y^T\tilde{\mathbf{W}}_i\mathbf{I}_x & \mathbf{I}_y^T\tilde{\mathbf{W}}_i\mathbf{I}_y + \zeta_i\mathbf{L} \end{bmatrix} \begin{bmatrix} du_i \\ dv_i \end{bmatrix} = -\begin{bmatrix} \zeta_i\mathbf{L}u_i \\ \zeta_i\mathbf{L}v_i \end{bmatrix} - \begin{bmatrix} \mathbf{I}_x^T \\ \mathbf{I}_y^T \end{bmatrix} \big(\tilde{\mathbf{W}}_i\mathbf{F}_{w_i} I - \mathbf{K}^T\mathbf{S}^T\mathbf{W}_i J\big) \quad (13)$$

where $\zeta_i = \frac{\lambda}{\theta_i}$, $\tilde{\mathbf{W}}_i = \mathbf{K}^T\mathbf{S}^T\mathbf{W}_i\mathbf{SK}$, and $\mathbf{L}$ is a weighted Laplacian matrix. Again, we use IRLS to solve the above equation iteratively.

One may notice that it is more expensive to solve Eqn. 13 than ordinary optical flow because in each iteration smoothing $\mathbf{K}$ and down-sampling $\mathbf{S}$ as well as the transposes $\mathbf{S}^T$, $\mathbf{K}^T$ need to be computed. We estimate optical flow from J_i to J_0 on the low-res lattice, and up-sample the estimated flow field to the high-res lattice as initialization for solving Eqn. 13.

3.3. Kernel Estimation

Without loss of generality, we only show how to estimate the x-component kernel K_x given I and J_0. Let each row of matrix $\mathbf{A}$ be the concatenation of pixels corresponding to the filter K, and define $\mathbf{M}_y : \mathbf{M}_y K_x = K_y \otimes K_x = K$. Estimating K_x leads to

$$K_x^* = \underset{K_x}{\operatorname{argmin}}\, \theta_0 \Big\|\mathbf{A}\mathbf{M}_y K_x - J\Big\| + \xi\Big\|\nabla K_x\Big\|, \quad (14)$$

which is again optimized by IRLS.

Although similar Bayesian MAP approach performed poorly for general debluring problems [14], the spatial smoothness prior on the kernel prevents kernel estimation from converging to the delta function, as shown by [11]. Experiments also show that our estimation is able to recover the underlying blur kernel.

4. Frequency-wise Performance Bounds

Intuitively, super resolution becomes more challenging when noise increases and blur kernel grows larger. Moreover, the low-frequency components of the signal should be easier to recover than high-frequency ones. In this section we will theoretically address these issues. We first show that super resolution can be reduced to deblurring under certain conditions and then use frequency-wise analysis to bound the performance of super resolution.

We assume the scene is fixed and only camera motion is present (and controllable). For up-sampling rate s, we can collect s^2 low-res images by shifting the camera one pixel at a time to cover the full high-res frame. Concatenating these s^2 low-res images we obtain an image I_{obs} with the same dimension as the original high-res I. Under this ideal setup, super resolution is reduced to a deblurring problem

$$I_{\text{obs}} = K * I + n, \quad (15)$$

where n is zero-mean, white Gaussian noise. Although this setting is very ideal with known, controllable motion and every high-res pixel fully observed, analyzing this system gives us the upper bound of super resolution. We use matrix multiplication to represent blur convolution for this linear system.

We apply the Wiener filter [7], which provides the minimum mean square error (MMSE) solution under this setting, to analyze this linear problem. We use the expectation of the ratio between the output signal by the Wiener filter and the original input signal to bound the performance of any estimator at frequency ω

$$P(\omega) = \frac{|K(\omega)|^2}{|K(\omega)|^2 + |N(\omega)|^2/|I(\omega)|^2}, \qquad (16)$$

where $K(\omega)$, $N(\omega)$ and $I(\omega)$ are the Fourier transform of the blur kernel, noise and original image, respectively. We cannot know the true $I(\omega)$ for natural images and so use the power-law spectra $|\hat{I}(\omega)|^2 \propto |\omega|^{1.64}$, which has been fitted to the Berkeley natural image database [21].

Figure 4 shows the curves for different Gaussian blur kernels. Generally, higher SNR and smaller blur kernel make super resolution easier and the high frequency components are more difficult to recover. For given blur kernel and noise level, there exists a cut-off frequency. Any frequency components above the cut-off frequency are impossible to recover, while those below it are still possible to estimate.

To test these bounds, we used the checkerboard pattern as input. We constructed the matrix for the blur kernel and performed maximum likelihood estimation of the original signal by matrix inversion. As shown in Figure 5, this method perfectly reconstructed the signal at the noise-free case. Note that the noise-free case will never happen in practice because of quantization. When we store images as integers in $[0, 255]$, the quantization error is in $[-0.5, 0.5]$. This corresponds to $\sigma_n = 0.001$ for pixel values in $[0, 1]$. For a large blur kernel ($\sigma_k = 2$), such tiny noise already makes it impossible to recover the original checkerboard pattern. While at a large noise level ($\sigma_n = 0.05$), the task is also impossible for a smaller blur kernel ($\sigma_k = 1$). All these results are consistent with the prediction by the curves in Figure 4, suggesting that the frequency-wise analysis of the linear problem by the Wiener filter provides good bounds.

5. Experimental Results

We will first examine the performance of our system under unknown blur kernel and noise level and then compare it to state-of-the-art video super resolution methods on several real-world sequences. **Please refer to the supplemental materials to view the super-resolved sequences. Please enlarge and view Figure 7, 8 and 9 on the screen for better comparison.**

Performance evaluation. We used the benchmark sequence *city* in video compression society to evaluate the

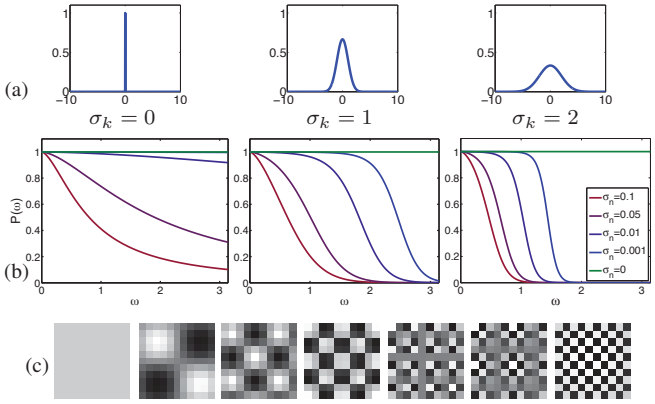
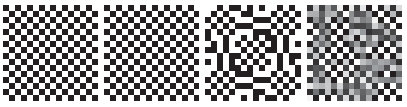


Figure 4. **Performance of the Wiener filter at different frequencies under different blur and noise conditions** (1 corresponds to perfect reconstruction). Last row from left to right: basis images corresponding to frequency $\omega = 0, 0.5, 1, \cdots, 3$; note that $\omega = 3$ corresponds to a nearly checkerboard pattern, the highest spatial frequency.



original $(2, 0)$ $(2, 0.001)$ $(1, 0.05)$

Figure 5. Reconstruction results of the checkerboard pattern. $(2, 0)$ means that $\sigma_k = 2$ and $\sigma_n = 0$, and corresponds to the right end point of green curve corresponding to $\sigma_k = 2$ in Figure 4 (see text for more details).

performance. Rich details at different scales make the *city* sequence ideal to observe how different frequency components get recovered. We simulated the imaging process by first smoothing every frame of the original video with a Gaussian filter with standard deviation σ_k. We downsample the smoothed images by a factor of 4, and add white Gaussian noise with standard deviation σ_n. As we vary the blur kernel σ_k and the noise level σ_n for evaluation, we initialize our blur kernel K_x, K_y with a standard normal distribution and initialize noise parameters θ_i using the temporal difference between frames. We use 15 forward and 15 backward adjacent frames to reconstruct a high-res image.

We first tested how our system performs under various noise levels. We fixed σ_k to be 1.6 and changed σ_n from small (0) to large (0.05). When $\sigma_n = 0$, quantization is the only source of error in the image formation process. As shown in Figure 7, our system is able to produce fine details when the noise level is low ($\sigma_n = 0.00$ or 0.01). Our system can still recover major image structure even under very heavy noise ($\sigma_n = 0.05$). These results suggest that our system is robust to unknown noise. Note that the performance drop as the noise level increases is consistent with our theoretical analysis.

Next, we tested how well our system performs under various blur kernels. We gradually increase σ_k from 1.2 to 2.4 with step size 0.4 in generating the low-res input. As shown in Figure 8, the estimated blur kernels match the ground

(a) Bicubic $\times 4$ (b) 3DKR [25] (c) Our system (d) Original

Figure 6. Closeup of Figure 9. From top to bottom: *city*, *calendar*, *foliage* and *walk*.

truth well. In general, fewer details are recovered as σ_k increase, consistent with our theoretical analysis. However, the optimal performance (in PSNR) of our system occurs for $\sigma_k = 1.6$ instead of 1.2. This seems to contradict the prediction of our theoretical analysis. In fact, small blur kernel generates strong aliasing, a fake signal that can severely degrade motion estimation and therefore prevent reconstructing the true high-frequency details.

Comparison to the state of the art. We compared our method to two recent methods [23, 25] using the public implementations downloaded from the authors' websites [1] and one state-of-the-art commercial software, "Video Enhancer" [1]. Since the 3DKR method [25] produced the best results amongst these methods, we only display the results of 3DKR in our paper.

We used three additional real-world video sequences, *calendar*, *foliage* and *walk* for comparison. The results are listed in Figures 6 and 9. Although the 3DKR method has recovered the major structures of the scene, it tends to over-smooth fine details. In contrast, our system performed consistently well across the test sequences. On the *city* sequence our system recovered the windows of the tall building while 3DKR only reconstructed some blurry outlines. On the *calendar* sequence, we can easily recognize the banner "MAREE FINE" from the output of our system, while the 3DKR method failed to recover such detail. Moreover, our system recovered the thin branches in the *foliage* sequence and revealed some facial features for the man in the *walk* sequence. The 3DKR method, however, over-smoothed these details and produced visually less appealing results.

[1]The implementation of the 3DKR method [25] does not include the last deblurring step as described in their paper. We used a state-of-the-art deconvolution method [13] to post-process its output. We used the default parameter setting of the 3DKR code to upscale the low-res video and adjusted the deconvolution method [13] to produce visually the best result for each individual sequence. The 3DKR implementation does not have valid output for pixels near the image boundaries. We filled in the gaps using gray pixels.

Table 1. **PSNR and SSIM scores**. 3DKR-b is the output of the 3DKR method before postprocessing.

PSNR	*city*	*calendar*	*foliage*	*walk*
Proposed	27.100	21.921	25.888	24.664
3DKR [25]	24.672	19.360	24.887	22.109
3DKR-b [25]	24.363	18.836	24.376	21.938
Enhancer [1]	24.619	19.115	24.476	22.303
Shan *et al*. [23]	23.828	18.539	22.858	21.018
Bicubic	23.973	18.662	24.393	22.066
SSIM				
Proposed	0.842	0.803	0.845	0.786
3DKR [25]	0.647	0.600	0.819	0.584
3DKR-b [25]	0.637	0.554	0.797	0.554
Enhancer [1]	0.677	0.587	0.803	0.604
Shan *et al*. [23]	0.615	0.544	0.747	0.554
Bicubic	0.597	0.529	0.789	0.548

We also observe failures from our system. For the fast moving pigeon in the *walk* sequence, our system produced sharp boundaries instead of preserving the original motion blur. Since motion blur has not been taken into account in our system, the sparse spatial prior favors sharp boundaries in reconstructing smooth regions such as motion blur. Furthermore, motion blur can significantly degrade motion estimation and can result in undesired artifacts.

Tables 1 summarizes the PSNR and SSIM scores[2] for these methods on the video frames in Figure 9. Our system consistently outperforms other methods across all the test sequences.

Computational performance. Our C++ implementation takes about two hours on an Intel Core i7 Q820 workstation with 16 GB RAMs when super resolving a 720×480 frame using 30 adjacent frames at an up-sampling factor of 4. The computational bottleneck is solving the optical flow equation in Eqn. 13, which takes about one minute for a pair of high-res and low-res frames. Computing flow for all adjacent frames takes more than half an hour. To compare, one IRLS iteration for image reconstruction takes about two minutes.

6. Conclusion

In this paper we demonstrated that our adaptive video super resolution system based on a Bayesian probabilistic model is able to reconstruct original high-res images with great details. Our system is robust to complex motion, unknown noise level and/or unknown blur kernel because we jointly estimate motion, noise and blur with the high-res image using sparse image/flow/kernel priors. Very promising experimental results suggest that our system outperform the state-of-the-art methods on a variety of real-world sequences. The theoretical bounds on frequency-wise reconstruction rate are consistent with our experiments, indicating that they can be good guidelines for analyzing super resolution systems.

[2]We discarded rows and columns within 20 pixels to the boundary in computing these numbers because the 3DKR method did not have valid output in these regions.

(a) $\sigma = 0.00$ (26.05db, 0.790) (b) $\sigma = 0.01$ (24.77db, 0.717) (c) $\sigma = 0.03$ (24.40db, 0.675) (d) $\sigma = 0.05$ (23.94db, 0.634)

Figure 7. **Our video super resolution system is robust to noise.** We added synthetic additive white Gaussian noise (AWGN) to the input low-res sequence, with the noise level varying from 0.00 to 0.05 (top row, left to right). The super resolution results are shown in the bottom row. The first number in the parenthesis is PSNR score and the second is SSIM score.

(a) $\sigma_k = 1.2$ (25.41db, 0.832) (b) $\sigma_k = 1.6$ (26.05db, 0.790) (c) $\sigma_k = 2.0$ (24.58db, 0.713) (d) $\sigma_k = 2.4$ (24.06db, 0.654)

Figure 8. **Our video super resolution system is able to estimate the PSF**. As we varied the standard deviation of the blur kernel (or PSF) $\sigma_k = 1.2,\ 1.6,\ 2.0,\ 2.4$, our system is able to estimate the underlying PSF. Aliasing causes performance degradation for the small blur kernel $\sigma_k = 1.2$ (see text for detail). Top: bicubic up-sampling ($\times 4$); middle: output of our system; bottom: the ground truth kernel (left) and estimated kernel (right). The first number in the parenthesis is PSNR score and the second is SSIM score.

References

[1] http://www.infognition.com/videoenhancer/, Sep. 2010. Version 1.9.5.

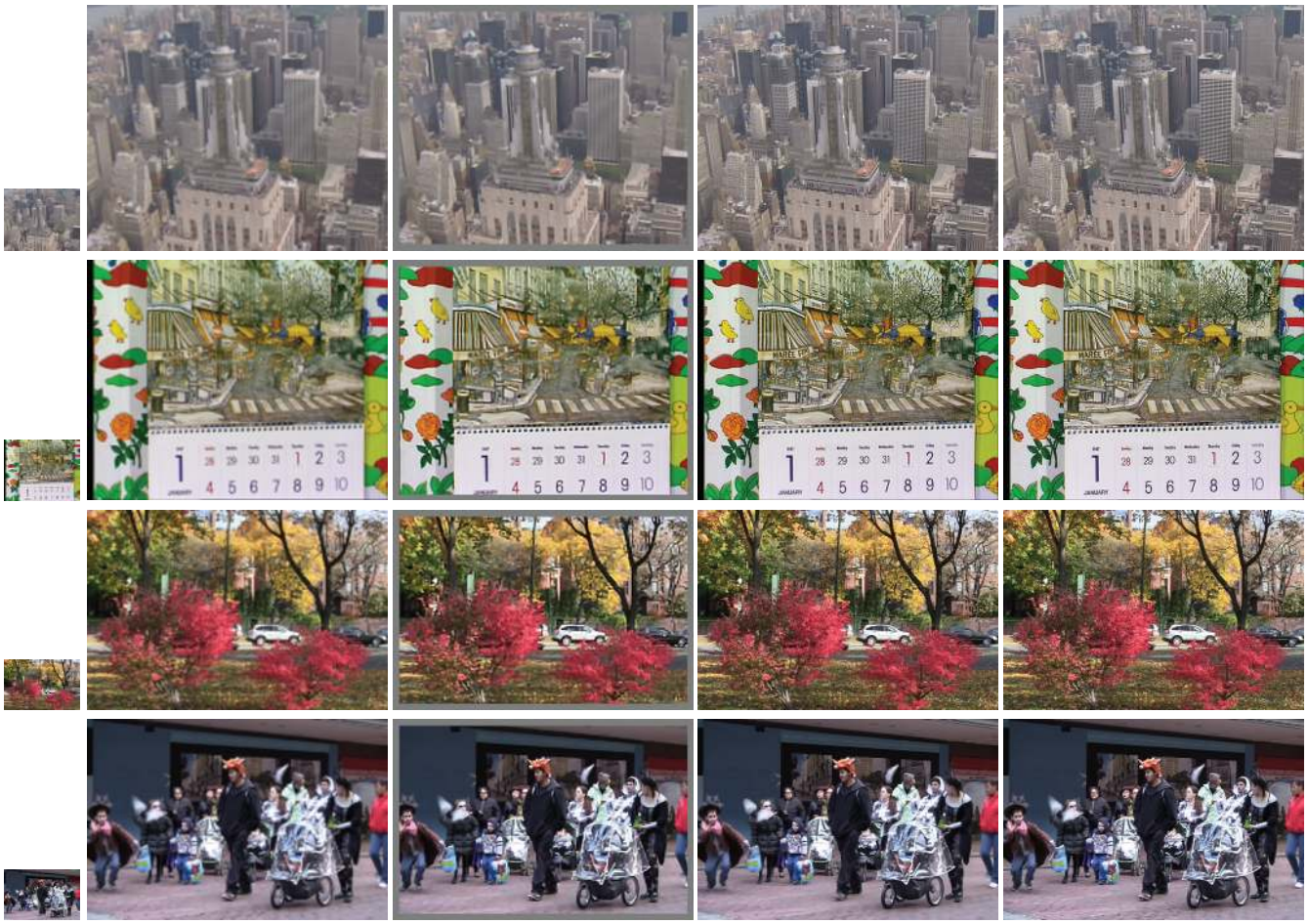
[2] S. Baker and T. Kanade. Super-resolution optical flow. Technical report, CMU, 1999.

[3] S. Baker and T. Kanade. Limits on super-resolution and how to break them. In *CVPR*, 2000.

[4] B. Bascle, A. Blake, and A. Zisserman. Motion deblurring and super-resolution from an image sequence. In *ECCV*, 1996.

[5] T. Brox, A. Bruhn, N. Papenberg, and J. Weickert. High accuracy optical flow estimation based on a theory for warping. In *ECCV*, 2004.

[6] R. Fransens, C. Strecha, and L. J. Van Gool. Optical flow based super-resolution: A probabilistic approach. *CVIU*, 106:106–115, 2007.



(a) Input (b) Bicubic ×4 (c) 3DKR [25] (d) Our system (e) Original

Figure 9. **Super resolution results.** From top to bottom are *city*, *calendar*, *foliage* and *walk* sequences. The 3DKR implementation does not have valid output for pixels near the image boundaries and we fill in the gaps using gray pixels. **Please view this figure on the screen**.

[7] R. C. Gonzalez and R. E. Woods. *Digital Image Processing*. Addison-Wesley, 2nd edition, 2001.

[8] R. Hardie, K. Barnard, and E. Armstrong. Joint MAP registration and high-resolution image estimation using a sequence of undersampled images. *IEEE TIP*, 6(12):1621 –1633, Dec. 1997.

[9] B. Horn and B. Schunck. Determining optical flow. *Artificial Intelligence*, 16:185–203, Aug. 1981.

[10] M. Irani and S. Peleg. Improving resolution by image registration. *CVGIP*, 53:231–239, 1991.

[11] N. Joshi, R. Szeliski, and D. Kriegman. PSF estimation using sharp edge prediction. In *CVPR*, 2008.

[12] D. Kundur and D. Hatzinakos. Blind image deconvolution. *Signal Proc. Magazine, IEEE*, 13(3):43 –64, 1996.

[13] A. Levin, R. Fergus, F. Durand, and W. T. Freeman. Image and depth from a conventional camera with a coded aperture. *ACM TOG*, 26, 2007.

[14] A. Levin, Y. Weiss, F. Durand, and W. Freeman. Understanding and evaluating blind deconvolution algorithms. In *CVPR*, 2009.

[15] Z. Lin and H.-Y. Shum. Fundamental limits of reconstruction based superresolution algorithms under local translation. *TPAMI*, 26:83 –97, 2004.

[16] C. Liu. *Beyond pixels: exploring new representations and applications for motion analysis.* PhD thesis, MIT, 2009.

[17] C. Liu and W. T. Freeman. A high-quality video denoising algorithm based on reliable motion estimation. In *ECCV*, 2010.

[18] C. Liu, R. Szeliski, S. B. Kang, C. L. Zitnick, and W. T. Freeman. Automatic estimation and removal of noise from a single image. *TPAMI*, 30(2):299–314, 2008.

[19] N. Nguyen, G. Golub, and P. Milanfar. Blind restoration/superresolution with generalized cross-validation using gauss-type quadrature rules. In *ICSSC*, 1999.

[20] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: a technical overview. *Signal Proc. Magazine, IEEE*, 20(3):21 – 36, 2003.

[21] S. Roth. *High-Order Markov Random Fields for Low-Level Vision*. PhD thesis, Brown University, 2007.

[22] R. Schultz and R. Stevenson. Extraction of high-resolution frames from video sequences. *IEEE TIP*, 5(6):996–1011, June 1996.

[23] Q. Shan, Z. Li, J. Jia, and C.-K. Tang. Fast image/video upsampling. *ACM TOG*, 27(5):153, 2008.

[24] F. Sroubek, G. Cristobal, and J. Flusser. Simultaneous super-resolution and blind deconvolution. *Journal of Physics: Conference Series*, 124(1), 2008.

[25] H. Takeda, P. Milanfar, M. Protter, and M. Elad. Super-resolution without explicit subpixel motion estimation. *IEEE TIP*, 18(9):1958–1975, Sep. 2009.

[26] R. Y. Tsai and T. S. Huang. Multiframe image restoration and registration. In *Advances in CVIP*, 1984.